JOHANNES KEPLER
UNIVERSITÄT LINZ

Eingereicht von
Svetoslav Nakov, MSc.

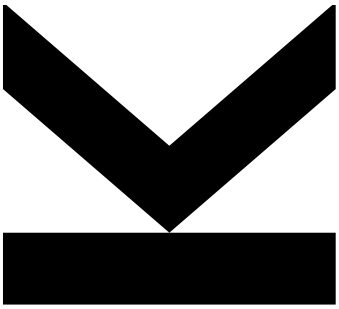
Angefertigt am
Johann Radon Institute for Computational and Applied Mathematics

Betreuer und Erstbeurteiler
Prof. Dr. Johannes Kraus

Zweitbeurteiler
Prof. Dr. Dirk Praetorius

June 2019

The Poisson-Boltzmann Equation: Analysis, A Posteriori Error Estimates and Applications



Dissertation

zur Erlangung des akademischen Grades

Doktor der Technischen Wissenschaften

im Doktoratsstudium der

Technischen Wissenschaften

JOHANNES KEPLER
UNIVERSITÄT LINZ
Altenbergerstraße 69
4040 Linz, Österreich
www.jku.at
DVR 0093696

Abstract

The Poisson-Boltzmann equation (PBE) gives a mean field description of the electrostatic potential in a system of molecules in ionic solution. It is a commonly accepted and widely used approach to the modelling of the electrostatic fields in and around biological macromolecules such as proteins, RNA or DNA. The PBE is a semilinear elliptic equation with a nonlinearity of exponential type, a measure right hand-side, and jump discontinuities of its coefficients across complex surfaces that represent the molecular structures under study. These features of the PBE pose a number of challenges to its rigorous analysis and numerical solution.

This thesis is devoted to the existence and uniqueness analysis of the PBE and the derivation of a posteriori error estimates for the distance between its exact solution and any admissible approximation of it, measured either in global energy norms or in terms of a specific goal quantity represented in terms of a linear functional. These error estimates allow for the construction of adaptive finite element methods for the fully reliable and computationally efficient solution of the PBE in large systems with complicated molecular geometries and distribution of charges.

One of the main focuses of this work is the rigorous analysis of the Poisson-Boltzmann equation and its linearized version, the LPBE. The starting point is to give a weak formulation which is appropriate for elliptic equations with measure data, such as the delta distributions due to fixed point charges in the molecular regions. For this weak formulation we are able to show existence of a solution by means of 2-term and 3-term splittings, where the full potential is decomposed into a singular Coulomb potential and a more regular part, a particular representative of which can be defined by a weak formulation involving H^1 Sobolev spaces. In the case of the LPBE we are also able to show the uniqueness of the full electrostatic potential.

Another main goal of this thesis is the derivation of a posteriori error estimates for the linearized and fully nonlinear Poisson-Boltzmann equation. More precisely, we derive two types of a posteriori error estimates: global estimates for the error in the electrostatic potential, measured in the so-called energy norm, and goal-oriented error estimates for the electrostatic interaction between molecules. We apply the first type of error estimates to the study of the electrostatic potential in and around the insulin protein with PDB ID 1RWE, the Alexa 488 and 594 dyes, as well as the membrane protein-conducting channel SecYEG. In all these applications we obtain guaranteed and fully computable bounds on the relative errors in global energy norms. Moreover, we are able to establish a near best approximation result for the regular part of the electrostatic potential which is the basis for deriving a priori error estimates in energy norm for the finite element method. The second type of error estimates, also called goal-oriented a posteriori error estimates, are employed in the computation of the electrostatic interaction between the dyes Alexa 488 and Alexa 594 being either in their ground state or transition state. The latter configuration is related to the calculation of the efficiency of the Fröster resonance energy transfer (FRET) between the two dyes.

Abstract

Die Poisson-Boltzmann-Gleichung (PBE) dient der Beschreibung des gemittelten elektrostatischen Potentials in einem System von Molekülen in ionischer Lösung. Sie stellt einen allgemein anerkannten und weit verbreiteten Ansatz zur Modellierung der elektrostatischen Felder innerhalb und in der Umgebung von biologischen Makromolekülen, wie Proteinen, RNA oder DNA, dar. Die PBE ist eine semilineare elliptische partielle Differentialgleichung mit einer Nichtlinearität des exponentiellen Typs, einem über Delta-Distributionen definierten Quellterm und einer Koeffizientenfunktion, die im Allgemeinen an der die molekulare Struktur begrenzenden Fläche Diskontinuitäten aufweist. Diese Merkmale machen die mathematische Analyse und die numerische Lösung der PBE zu anspruchsvollen Aufgaben.

Die vorliegende Dissertation befasst sich mit der Analyse der Existenz- und Eindeutigkeit der Lösung des PBE und der a posteriori Fehlerabschätzung für deren zulässige Approximationen, gemessen entweder in globalen Energienormen oder bezüglich bestimmter linearer Zielfunktionale. Die gewonnenen Fehlerschätzer bilden die Basis für die Konstruktion adaptiver Finite-Elemente-Methoden und für die absolut zuverlässige und rechnerisch effiziente Lösung der PBE im Falle großer biologischer Makromoleküle mit komplizierten Geometrien und Ladungsverteilungen.

Ein Schwerpunkt dieser Arbeit ist die sorgfältige Analyse der Poisson-Boltzmann-Gleichung und ihrer linearisierten Version, der LPBE. Ausgangspunkt dafür ist eine spezielle variationelle Formulierung, die sich für elliptische Gleichungen mit Daten (Quelltermen) in distributioneller Form eignet, wie dies zum Beispiel bei der Modellierung von Punktladungen in Molekülen durch Delta-Distributionen der Fall ist. Für diese schwache Formulierung wird dann die Existenz einer Lösung mit Hilfe von 2- und 3-Term-Zerlegungen nachgewiesen, bei denen das Gesamtpotential in ein singuläres Coulomb-Potential und in eine reguläre Komponente zerlegt wird. Letztere kann als Lösung ein spezielles Variationsproblems in H^1 Sobolev-Räumen definiert und gefunden werden kann. Im Falle der LPBE wird überdies die Eindeutigkeit des elektrostatischen (Gesamt-) Potentials nachgewiesen.

Ein weiteres Hauptziel dieser Arbeit ist die Herleitung von a posteriori Fehlerabschätzungen für die linearisierte sowie für die nichtlineare Poisson-Boltzmann-Gleichung. Genauer gesagt werden zwei Arten von a posteriori Fehlerabschätzungen vorgestellt: einerseits globale Schätzer für den Fehler im elektrostatischen Potential, gemessen in der sogenannten Energienorm und andererseits zielorientierte Fehlerschätzer für die elektrostatische Wechselwirkung zwischen Molekülen. Die erste Art von Fehlerabschätzungen wird im Zuge der Untersuchung des elektrostatischen Potentials des Insulinproteins mit PDB ID 1RWE, der Chromophore Alexa 488 und 594, sowie des Membranproteins SecYEG angewandt. In all diesen Fällen erhalten wir garantierte und vollständig berechenbare Schranken für den relativen Fehler in der globalen Energienorm. Darüber hinaus wird für die reguläre Komponente des elektrostatischen Potentials ein "Near-Best-Approximationsergebnis" gezeigt, das die Grundlage für die Herleitung

von a priori Fehlerabschätzungen in der Energienorm von Finite-Elemente-Näherungen bildet. Der zweite Typ von Fehlerschätzern, auch zielorientierte a posteriori Fehlerschätzer genannt, wird bei der Berechnung der elektrostatischen Wechselwirkung zwischen den Chromophoren Alexa 488 und Alexa 594 verwendet, die sich dabei entweder im Grundzustand oder im Übergangszustand befinden. Letztere Konfiguration wird auch zur Berechnung (bzw. Simulation) der Effizienz des sogenannten Förster-Resonanz-Energie-Transfers (FRET) zwischen den beiden Farbträgern herangezogen.

Acknowledgments

First and foremost, I would like to thank my supervisor Prof. Johannes Kraus for his continuous guidance and understanding during my doctoral study, as well as for letting me follow my own ideas and time schedule. His valuable scientific advice along with support in dealing with various everyday and administrative matters smoothed my way through the doctoral study. I would also like to thank Prof. Kraus for encouraging me to attend different conferences, seminars, workshops and also to complete a six months stay abroad. At the same time, I want to express my gratitude to Prof. Sergey Repin for introducing me to the topic of a posteriori error estimation and for his patience in our collaboration. The time and effort of Prof. Dirk Praetorius put into reviewing my thesis is also greatly appreciated.

I want to gratefully acknowledge the financial support received from the Doctorate College program "Nano-Analytics of Cellular Systems (NanoCell)" of the Austrian Science Fund (FWF) (grant number: W 1250), without which this work would not be possible. The financial support from the projects LIT-JKU-2017-4-SEE-004 and FWF P31074 is also highly appreciated. While the results in this thesis were mainly achieved in the framework of the Doctorate College NanoCell, I am also thankful for the convenient and friendly working environment provided to me by the Radon Institute for Computational and Applied Mathematics (RICAM) of the Austrian Academy of Sciences. Here I want to thank Prof. Ulrich Langer and Prof. Walter Zulehner for the positive atmosphere created at our doctoral seminar, and Wolfgang Forsthuber and Florian Tischler for their continuous technical support.

Next I want to thank my colleagues Dr. Roman Andreev, Dr. Kamran Sadiq, Dr. José A. Iglesias, Dr. Sergio Rodrigues, Dr. Hannes Meinlschmidt, and Dr. Behzad Azmi for their help and many fruitful discussions that we had. Along with them, I also want to thank my colleague Ekaterina Sobakinskaya from the Doctorate College NanoCell, who helped me gain a better understanding of the biophysical systems that we studied. Furthermore, I want to thank Dr. Tihomir Ivanov for the time he put into reading parts of my thesis and to Stefan Vasilev for the diverse conversations that we had.

Last, but not least, I would like to thank my family for supporting me throughout my whole life and to my friends for their understanding during my doctoral study.

Svetoslav Nakov
Linz, June 2019

Contents

Chapter 1

Introduction

Biomolecular electrostatic models play an important role in the quantitative analysis of biological macromolecules such as proteins, RNA or DNA in solution [124,126,176]. A commonly accepted and widely used approach is based on solving the nonlinear Poisson-Boltzmann equation (PBE) which was described independently by Gouy [93] and Chapman [46] and later generalized by Debye and Hückel [60]. The PBE gives a mean field description of the electrostatic potential in a system of biomolecules. The molecules are modeled as fixed partial charges in a low dielectric cavity and ions are treated as continuous fluid-like particles moving independently in the high dielectric region, outside the molecular domain, under the influence of a mean electric potential. Applications include computations of the electrostatic potential at the solvent-accessible molecular surface, pKa values of biomolecules, the encounter rates between molecules in solution, or the free energy of association in conjunction with its salt dependence (see, e.g. [81]). Biomolecular association (e.g. the association of ligand and proteins) depends on the shape of the molecules and their electrostatic field. Therefore, adequate mathematical models must properly account for the effects induced by both geometrical properties and by the distribution of charges.

State of the art

Existence and uniqueness of a solution to the PBE

The Poisson-Boltzmann equation is a semilinear interface elliptic problem with exponential nonlinearity and a measure right-hand side. These features make it very challenging to analyze and approximate numerically the PBE. Despite its extensive use in the biophysics community, there are very few works discussing the existence and uniqueness analysis of this equation. Efforts on providing a solution theory for the PBE are made for example in [48, 101, 103, 123, 188] and for the modified PBE with finite size ions in [47].

Numerical solution of the PBE

Simple-shape molecular models, e.g., electrostatic models for globular proteins as used in [113], had been replaced in the early 1980s by models based on more complex geometries. This development was driven by the progress of finite element (FE), boundary element (BE), and finite difference (FD) methods for solving nonlinear partial differential equations (PDE), see e.g. [133]. Numerous software packages for the simulation of biomolecular electrostatic effects that are presently available, such as APBS, CHARMM, DelPhi, UHBD, MEAD, and mFES, reflect the popularity and success of the PBE model.

Finite difference methods

The most popular method for solving the PBE has been the FD method based on a regular 3D lattice. Its popularity in the PBE community is thanks to its simpler implementation compared to the FE and BE methods. This is reflected by the large number of solvers based on this method, such as DelPhi [114], GRASP [140], MEAD [11], UHBD [134], CHARMM [37], the PMG solver in the APBS software suite [102].

The FD method was first adapted to biological macromolecules in the early 1980s in [185] and then utilized and modified in numerous works (see, e.g. [91,92,114,141,163,168,175,185]). In the FD method, the PBE is discretized on a cubic lattice (grid) by approximating the differential equation with a difference equation at every grid point. Here, the derivatives at a grid point are approximated as finite differences between function values sampled at surrounding lattice points. The resulting set of simultaneous linear/nonlinear equations, with unknowns the electrostatic potential at the grid points, is usually solved by means of some iterative method.

Regular lattices allow for a simple discretization of the differential operators and the use of highly efficient multigrid solvers for the resulting system of algebraic equations. However, the lack of adaptivity results in a noticeable trade-off between the grid spacing (resolution) and the accuracy of the boundary condition. More precisely, the number of grid points in a regular cubic lattice is n^3, where n is the number of grid points in each coordinate direction. By taking into account that the usual grid spacing in electrostatic computations ranges between $0.2\,\mathring{A}$ and $1\,\mathring{A}$, it is clear that the problem size can easily approach 1000^3 unknowns for relatively small systems of several hundred atoms. This is highly prohibitive in terms of memory and computational costs. To speed up the convergence and reduce the memory requirements, a so-called focusing technique was proposed in [92,114]. This involves solving the equation on a coarse grid, covering a large region, followed by a solution on a finer grid, covering a smaller region, with a boundary condition, interpolated from the values of the coarse grid solution. This approach was further improved in [7] by implementing it in parallel

and then applied to the electrostatics of microtubules and ribosoms. Due to the noticeably improved performance, focusing schemes are implemented in most of the FD based solvers for the PBE in which regular cubic lattices are used.

Unlike in FE and BE methods, in FD methods, the molecular surface is implicitly defined which causes difficulties handling sharp interfaces between regions with a relatively high jump in the dielectric coefficient. More precisely, for each integer or half-integer grid point the dielectric coefficient is defined by means of some averaging of its values at the neighboring points which have assigned values based on which region they belong to (see [92, 141]). The inverse Debye-Hückel parameter is defined in a similar fashion. Due to this averaging there is a region around the true interface in which these coefficients vary between their respective values on both sides of the interface. The thickness of this region depends on the grid spacing h and this is where the discretization error is typically much higher compared to the FE and BE methods.

Finite difference methods also exhibit problems caused by nonsmooth source terms, such as the delta distributions, introduced to model fixed point charges. A redistribution of the charges around their closest grid points is necessary in order to minimize the errors occurring in the short range potential, i.e., the errors close to the charges (see [92, 141] for different approaches to the redistribution). However, redistribution techniques in general lead to a so-called grid artifact (see [12, 164]). The latter can be easily avoided by using Galerkin type FE approximations and combining them with proper splitting techniques for the full potential. By applying such decompositions of the full potential, the delta distributions describing the discrete charge density due to the point charges in the solute are transformed to a more regular distribution, which is interpreted as a certain surface charge density concentrated on the molecular surface. We note that these splitting techniques can also be used in conjunction with FD methods (see, e.g. [141, 195]). However, the surface charge density, supported on the interface between the molecular region and the solution region, remains hard to handle with FD methods.

Another disadvantage of the FD method is that the errors, caused by the specific way of handling interfaces and partial charge distribution in the solute, strongly depend on the particular position and orientation of the molecular system relative to the used FD grid. One approach to cope with this problem is to use finer grids, which naturally leads to a larger number of grid points. Another approach is to use a rotational averaging scheme. The idea is to make a number of calculations with a slightly different relative position and orientation between the molecular system and the FD grid, and finally take some average (see [92, 141]). A third approach that could cope better with curved interfaces and regions where the solution is sharply varying (mainly around the fixed partial charges and across interfaces) is to involve some kind of adaptivity in the FD method in conjunction with splitting techniques for the full potential. Finite difference methods, based on adaptive Cartesian grids (ACG), have

been developed for solving nonlinear equations in domains with curved boundaries (see, e.g [83, 136, 142]). Here, the grid representation is based on quadtrees in two dimensions and octrees in three dimensions. The ACG is recursively adapted by identifying mesh cells, which intersect the molecular surface, and test them whether they satisfy one of the following criteria: finest grid-spacing is reached or intersected cell lies more than a prescribed distance from the nearest atomic charge site. The cell, which does not meet either of these criteria is uniformly subdivided into four (in 2D) or eight (in 3D) smaller cells. Using a coarser mesh away from the surface reduces the total number of grid points, which for ACG often is several orders of magnitudes less than that required in a conventional lattice grid. However, the underlying mesh still does not conform to the solute surface and thus the accurate evaluation of the electrostatic potential near the interface remains problematic. A method that can reduce the error in the potential near the interface for FD methods based on ACG is proposed in [97]. Recently, an FD method based on ACG, was implemented in the finite difference solver CPB, particularly aimed at the approximation of the PBE for complex biomolecular structures [27]. This solver also utilizes a splitting technique for the full potential to eliminate the singular behavior at charge sites.

Boundary element methods

Another quite popular approach to the numerical solution of the linearized and nonlinear PBE is the boundary element method (BEM). It was first adapted to the computation of the electrostatic potential of macromolecules in [193] and later different algorithmic improvements were proposed in [154, 194]. Here, the PBE is reformulated into boundary integral equations, living on the boundary of the solute domain, by employing Green's theorem. Therefore, only the two-dimensional molecular surface needs to be discretized. The resulting boundary integral equations are solved by using collocation, Galerkin, or least squares methods. The first one is the most commonly used by the PBE community since it is the simplest to apply in practice.

The BE method has some advantages over other numerical methods like the FE and FD methods, which involve volume-domain discretization. Since only the molecular surface is discretized, the number of unknowns is greatly reduced. Moreover, boundary conditions at infinity are exactly treated as opposed to the FE and FD methods. In contrast to the FDM, the jump in the normal component of the electrostatic field across the molecular surface (interface) is explicitly accounted for, which results in a more accurate solution near the solute surface.

However, the BE method also has some disadvantages that make it not so flexible compared to the FE method. Although the number of unknowns is considerably smaller than in FE or FD methods, the arising linear systems from the discretization of the boundary integral equations have dense matrices which results in high memory and solution costs in direct solution approaches. Furthermore, when dealing with the nonlinear PBE, volume integrals

appear in the boundary integral equations, which additionally reduces the efficiency of this method. The computation of the coefficient matrices of the discretized boundary integral equations involves integration of singular functions which causes problems related to accuracy and/or stability. Different modifications, aimed at improving the efficiency of the BEM for the numerical solution of the linearized or nonlinear PBE, have been considered for example in [80,130–132,183]. Another approach to enhance the efficiency of the BEM is to use adaptive mesh refinement based on reliable and efficient error indicators. General convergence theory for BEM, not necessarily tailored towards the PBE, is presented for example in [74–77,79,88]. A comprehensive review of convergence theory for adaptive BEM can be found in [73].

Finite element methods

The FE methods are the most successful for elliptic PDE, since they combine geometrical flexibility and satisfactory convergence analysis with the ability to handle nonlinear problems involving interface jump conditions and nonsmooth source terms. Moreover, they enjoy a wide range of efficient iterative solvers for the resulting sparse linear systems. One of the first applications of the FE method to continuum electrostatics of biomolecules in solution appears in [147]. Since then, the popularity of this approach has increased synchronically with the increase of computational power and advancement of FE mesh generation techniques [8,54,101,102,109]. As representatives of the solvers, implementing the FE method for the numerical approximation of the PBE and the linearized PBE, we mention APBS [102] (utilizing adaptive mesh refinement based on a residual type error indicator) and mFES [164], respectively.

The FE method is based on the so-called weak formulation of the problem, which is the natural setting for the theoretical analysis and numerical solution of problems involving discontinuous coefficients and right hand sides which are not classical functions, but rather functionals. This is the case with the PBE, where the dielectric coefficient undergoes a jump discontinuity across the interface between the solute and solvent regions and the right-hand side is a linear combination of delta functions representing the fixed partial charges in the solute region. In FE methods, the approximate solution u_h is sought in a finite dimensional space V_h associated with a partition $\mathscr{T}_h$ into finite elements of the computational domain Ω, where the subscript h refers to the maximum size of the elements. In the so called conforming finite element method, the finite dimensional space V_h in which the approximate solution is defined is a subspace of the infinite dimensional functional space V to which the exact solution belongs. Moreover, as $h \to 0$ the resulting finite dimensional subspaces V_h approximate better and better the infinite dimensional space V and u_h converges to the exact solution u. The most popular FE spaces V_h used for the approximation of the PBE consist of continuous piecewise polynomial functions associated with a partition of Ω into triangles in 2D and tetrahedrons in 3D.

In contrast to the FD methods, based on uniform cubic grids, the FE methods allow for local refinement of the underlying mesh and a better resolution of extremely fine features in the protein geometry, while solving the PBE on large computational domains, spanning millions of angstroms. At the same time, the number of mesh points can be orders of magnitude smaller than the one for comparable in size FD uniform cubic lattices, even for moderate lattice spacing h. This is achieved by using extremely graded meshes towards the boundary of the computational domain. Clearly, this feature of the FE method eliminates the exceedingly restrictive trade-off, typical for FD methods with uniform cubic lattices, between fines of the grid in regions of interest and accuracy of the boundary condition.

Another advantage of the FE methods is that the underlying mesh can be constructed in such a way that it conforms to the solute-solvent interface. Therefore, unlike FD methods, based on either uniform or adaptive Cartesian grids, the complex solute-solvent interface can be explicitly defined. This allows for accurate approximation of the electrostatic potential near the interface even in the presence of large jump in the dielectric coefficient. However, the generation of molecular surface meshes which are also suitable for FE and BE calculations is not a trivial task. Among the more popular surface mesh generators are MSMS [165], LSMS [43], TDTSurf [191], TMS [49], NanoShaper [61], GAMer [192] (also performs volumetric mesh generation). Once, the surface mesh is generated, a volume tetrahedral mesh can be obtained, for example, by using Netgen [166], TetGen [170], Gmsh [89].

Major advances in the quality of the numerical solution of the PBE regarding accuracy and efficiency are due to proper mesh adaptation techniques, see, e.g., [48, 103]. Adaptive FE methods exploit error indicators, which must be reliable and efficient in that upon multiplication by constants of the same order they provide bounds for the actual error from above and below. Efficient error indicators can be constructed by different methods closely related to different approaches to the a posteriori error estimation problem. In this context, we mention goal-oriented methods for the error measured in terms of a specific goal functional, residual based methods and functional type methods for the error in global energy norms.

The goal oriented error estimates are useful when one is interested in a specific quantity that depends on the solution (goal quantity) rather than the solution itself over the whole computational domain. Such error estimates for linear and nonlinear problems are considered in numerous publications, see e.g. [13, 78, 104, 105, 115, 138, 144]. Nonetheless, it seems difficult to find works which consider applications to the PBE equation.

In residual based approaches to the a posteriori error estimation of the global error in energy norm, the error is bounded from above and below by multiples of the residual norm. Depending on the way the residual norm is estimated one has explicit or implicit residual methods, see, e.g. [3, 4, 6, 45, 181] and [69, 152, 180, 182] for nonlinear problems. A residual

based error indicator is used to drive the adaptive mesh refinement in the FE solution of the nonlinear PBE in [48, 103], where also the convergence of the method is proven. A more general convergence theory for adaptive FE methods is developed, for example, in [44, 87, 171]. Some popular softwares for adaptive mesh refinement in 3D are, for example, TetGen [170], Netgen [166], mmg3d [62].

Functional type error estimates have been developed in the framework of duality theory for convex variational problems [65, 139, 155–157]. They provide estimates that generate guaranteed bounds on the distance to the exact solution valid for the whole class of energy admissible functions. In contrast to the residual based, these functional type estimates contain neither mesh dependent constants nor do they rely on any special conditions or assumptions on the exact solution (e.g., higher regularity) or its approximation (e.g., Galerkin orthogonality), which means that they are fully computable. However, this type of error estimates has not been applied to continuum electrostatics computations until recently in [116, 117].

On this work

This thesis is concerned with the solution theory and a posteriori error estimation for the Poisson-Boltzmann equation, as well as with the numerical solution of a series of practical problems which demonstrate and substantiate the developments in the presented work. More precisely, existence and uniqueness results are proved by using different decomposition techniques, along with a priori L^∞ estimates for the regular components of this solution. It is shown that a particular representative of these regular components can be defined by a weak formulation involving H^1 spaces and that they can be approximated numerically in a standard way by using, for example, the finite element method. Further, a posteriori error estimates, leading to guaranteed and fully computable bounds on the global energy norm of the error, are derived for these regular components of the solution to the PBE. In addition to this type of estimates, also goal oriented error estimates are presented for the electrostatic interaction between molecules, where the potential in the system is governed by the linearized PBE.

Structure of the thesis

The thesis is organized as follows. In Chapter 2, we fix the notation and recall some well known results on Lebesgue and Sobolev spaces, convex analysis and regularity theory for linear elliptic interface problems.

In Chapter 3, we consider the solution theory for the linearized and fully nonlinear Poisson-Boltzmann equation.

- In Section 3.1, we introduce the Poisson-Boltzmann equation, describing the physical

problem of continuum electrostatics, and all relevant geometrical regions, involved in the description of a general system of molecules immersed in solution.

- In Section 3.2, we start with the solution theory for the LPBE, the right hand side of which is a measure, given by a linear combination of delta functions. The first step here is to give a meaningful notion of a solution to this equation, which also ensures uniqueness. This is done by following [21, 63, 86]. Once the weak formulation is defined, we prove existence by means of either 2- or 3-term splitting of the solution. The uniqueness of the solution is proved by involving a duality argument together with a regularity result for linear elliptic interface problems [66]. This section ends with a discussion of the regularity properties of the regular components of the solution in the 2- and 3-term splittings.

- Section 3.3 deals with the analysis of the nonlinear PBE and has a similar structure to Section 3.2. First, the weak formulation for the LPBE is naturally extended to the nonlinear case and then existence of a solution is shown by utilizing the 2- and 3-term splittings mentioned above. We show that a particular representative of the regular components of the solution in both 2- and 3-term splittings can be defined by means of weak formulations involving H^1 spaces. The existence theorems based on the 2- and 3-term splittings are stated in Section 3.3.1 and Section 3.3.2, respectively. Moreover, in Section 3.3.4 we summarize some regularity results for the well behaved component of the solution in these splittings and give the respective conditions under which they hold.

- Section 3.4 concludes this chapter with an L^∞ a priori estimate for the solution of a more general semilinear elliptic problem with neither sign nor growth conditions on the nonlinearity.

In Chapter 4, we derive functional type a posteriori error estimates which provide guaranteed and fully computable bounds on the global energy norm of the error for the nonlinear PBE. In addition, a series of numerical examples are presented.

- In Section 4.1, we consider a more general semilinear elliptic interface problem with the same type of nonlinearity as in the PBE. First we present an abstract framework, based on duality theory for convex variational problems, for the derivation of functional type a posteriori error estimates by following [65, 139, 155]. We establish an abstract error identity for which an explicit form is given for the considered problem. First we consider the case of homogeneous Dirichlet boundary condition. In this situation, we present in detail the computation of all terms in the abstract error identity; obtain a near best approximation result for conforming, not necessarily finite element, spaces; show how to explicitly compute the respective error terms for the case of a more general nonlinearity; discuss the effect of data oscillation related to the approximation of the equilibration condition on the dual variable; and present academic numerical examples

in the case of homogeneous interface jump condition. Next, we show that all the results obtained in the case of homogeneous Dirichlet boundary condition remain valid also in the case of nonhomogeneous one.

- In Section 4.2 we describe a procedure, based on the patchwise equilibrated flux reconstruction in [30], to obtain a good conforming approximation of the dual variable. This method allows for an efficient evaluation of the error indicator in the adaptive algorithm and can be easily realized in parallel.

- In Section 4.3, we apply the obtained results for the more general semilinear elliptic interface problem to the a posteriori error estimation of the nonlinear PBE. To be more precise, we derive error estimates for the regular component of the solution in the 2- and 3-term splittings.

  In certain situations it might be beneficial to make one more additional splitting of the regular component in both 2- and 3-term splittings. Section 4.3.1 and Section 4.3.2 focus on the derivation of error majorants and minorants for the individual components of the solution that appear in the different splittings. Since each subproblem that defines a particular component of the solution depends on the the solution of the previous one, this causes a certain perturbation error additionally to the approximation error. Thus, in Section 4.3.1 and Section 4.3.2, we also derive overall estimates for the error, which take into account both the perturbation and approximation error.

- Section 4.3.3 is devoted to the application of the described in this chapter methodology to realistic systems consisting of macromolecules immersed in ionic solution. We solve adaptively the PBE to find the electrostatic potential in and around the insulin protein with PDB ID 1RWE, the Alexa 488 and 594 dyes, as well as the membrane channel SecYEG.

In Chapter 5, we present goal oriented error estimates for the electrostatic interaction between molecules in ionic solution.

- Section 5.1 introduces the physical problem and introduces the goal functional describing the electrostatic interaction between molecules. Next, in Section 5.2, we recall some known results on goal oriented a posteriori error estimates, based on the dual weighted residual (DWR) method, for elliptic problems with L^2 right-hand side and a regular goal functional, defined by an L^2 function.

- Since the goal functional defining the electrostatic interaction is composed of point evaluations and the exact solution of the primal problem is a harmonic function at these point evaluations, we can apply an equivalent goal functional defined by averaging over balls centered at the points of interest. In Section 5.3, we derive four different representations for the error that involve the solution of an adjoint problem. In two of

the estimates, we use the equivalent goal functional, defined by averaging over balls, and in the other two we employ the original goal functional composed of delta functions. We end this section with a summary of all four error estimates and the respective error indicators.

- In Section 5.4, we present numerous tests performed on problems with analytically known solutions and demonstrate the efficiency of the adaptive FE solvers, based on the proposed goal oriented error estimates. In addition, we also make a comparison of our adaptive solvers with the results obtained with the software package MEAD *version 2.2.8a* [11].

- Section 5.5 ends this chapter with two practical biophysical applications related to the computation of Fröster resonance energy transfer (FRET) and the electrostatic interaction between chromophores in their ground state. Again, a thorough comparison with MEAD is carried out.

Finally, in Chapter 6 we draw some conclusions and discuss possible future developments.

Main achievements

The main achievements of this work can be organized into three groups, corresponding to Chapter 3, Chapter 4, and Chapter 5.

- **Existence and uniqueness analysis**

  - Posing appropriate weak formulation for the LPBE and PBE (Definition 3.2 and Definition 3.12);

  - Existence and uniqueness for the LPBE with both 2- and 3-term splittings (Theorem 3.4 and Remark 3.8);

  - Existence and uniqueness for the regular component u in both 2- and 3-term splittings (Section 3.3.3, Theorem 3.21);

  - Existence for the PBE with both 2- and 3-term splittings (Theorem 3.13 and Theorem 3.15);

  - A priori L^∞ estimates for the regular component in both 2- and 3-term splittings (Proposition 3.18 and the discussion on p. 53);

  - A priori L^∞ estimate for a more general semilinear elliptic problem (Theorem 3.29).

• **Functional a posteriori error estimates**

  – Guaranteed and fully computable a posteriori error estimates for a more general semilinear elliptic problem with a nonhomogeneous interface jump condition on the normal flux (Proposition 4.5, equation (4.51));

  – Near best approximation result without assumptions on the mesh regularity (Proposition 4.12 and Proposition 4.16);

  – Effect of data oscillation related to the approximation of the equilibration condition on the dual variable (estimate (4.81));

  – Equilibrated flux reconstruction adapted to nonhomogeneous interface problems and implemented in parallel in 2D and 3D using lowest order Raviart-Thomas space (Section 4.2);

  – Overall error estimation for the PBE with the 2- and 3-term splittings (Proposition 4.18, Proposition 4.23, Proposition 4.27);

  – Implementation in FreeFem++ [98] of an adaptive FE solver based on the derived functional a posteriori error estimates and application to realistic systems, including insulin protein with PDB code 1IRW and the protein-conducting channel SecYEG (pages 154–176).

• **Goal oriented error estimates**

  – Four error representations in terms of the solution of an adjoint problem with both bounded and unbounded in H^1 goal functionals (Proposition 5.6, Proposition 5.10, Proposition 5.12);

  – Implementation in FreeFem++ of adaptive FE solvers based on the four error representations;

  – Verification of the implemented solvers on hundreds of configurations with analytically known solutions (Section 5.4);

  – Application of the implemented solvers to the calculation of the electrostatic interaction between molecules in ionic solution for hundreds of molecular dynamics frames and comparison with MEAD (Section 5.5).

Some parts of this thesis have already been accepted for publication in a peer-reviewed journal or have been made available online in the repository arXiv. More precisely, parts of Chapter 3 and Chapter 4 are presented in

- [117] J. Kraus, S. Nakov, and S. Repin. Reliable numerical solution of a class of nonlinear elliptic problems generated by the Poisson-Boltzmann equation. *Computational Methods in Applied Mathematics*, forthcoming.

- [116] J. Kraus, S. Nakov, and S. Repin. Reliable computer simulation methods for electrostatic biomolecular models based on the Poisson-Boltzmann equation. Preprint on arXiv:1805.11441, 2018.

Chapter 2

Preliminaries

In this chapter we fix the notation and give some relevant definitions and theorems that will be used throughout the thesis. We start with a brief discussion of some classes of continuous functions. Next we recall the spaces of Lebesgue measurable functions with summability index p. We continue with the definition and some relevant properties of the Sobolev spaces $W^{m,p}(\Omega)$. Finally, we recall some facts from functional analysis that will be of use in our considerations.

2.1 Classes of continuous functions

Let Ω be a domain in $\mathbb{R}^d$, $d \in \mathbb{N}$, i.e., an open subset of $\mathbb{R}^d$. Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d) \in \mathbb{N}_0$ be a multiindex of order $|\alpha| = \alpha_1 + \ldots \alpha_d$. For any nonnegative integer m, let $C^m(\Omega)$ be the space of continuous functions v in Ω for which all their partial derivatives

$$D^\alpha v = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}} v$$

of orders $|\alpha| \leq m$ are continuous in Ω. For $m = 0$, we denote $C^0(\Omega) \equiv C(\Omega)$ the space of continuous functions in Ω and by $C^\infty(\Omega)$ the space of infinitely differentiable functions in Ω. The space $C_0^\infty(\Omega)$ consists of those functions v in $C^\infty(\Omega)$ that have compact support in Ω. Further, the space $C^m(\overline{\Omega})$ consists of all functions v in $C^m(\Omega)$ for which all their derivatives $D^\alpha v$ of orders $0 \leq |\alpha| \leq m$ are uniformly continuous and bounded in Ω. As such, every function v in $C^m(\overline{\Omega})$ and all its derivatives $D^\alpha v$ with $|\alpha| \leq m$ possess unique, bounded, continuous extensions up to the boundary $\partial\Omega$ of Ω. By $C^\infty(\overline{\Omega})$ we denote the space defined by $C^\infty(\overline{\Omega}) := \bigcap_{m=0}^{\infty} C^m(\overline{\Omega})$. In particular the space of restrictions of functions in $C_0^\infty(\mathbb{R}^n)$ to the domain Ω is a subspace of $C^\infty(\overline{\Omega})$.

If $0 < \lambda \leq 1$, the space $C^{m,\lambda}(\overline{\Omega})$ is defined as the subspace of $C^m(\overline{\Omega})$ for which every function v and its partial derivatives $D^\alpha v$ of orders $|\alpha| \leq m$ satisfy a Hölder condition with exponent

λ, i.e., there exists a constant K such that

$$|D^\alpha v(x) - D^\alpha v(y)| \leq K\,|x-y|^\lambda \quad \text{for all } x, y \in \Omega.$$

We say that $f : \Omega \to \mathbb{R}$ is Lipschitz continuous in Ω if f is Hölder continuous with exponent $\lambda = 1$ and we write $f \in C^{0,1}(\Omega)$.

2.2 Lebesgue and Sobolev spaces

Let $A \subset \mathbb{R}^d$. By χ_A we denote the characteristic function of A and it is given by

$$\chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases} \tag{2.1}$$

If the set A is measurable, we denote its Lebesgue measure by $|\Omega|$. For a function $f : \mathbb{R}^d \to \mathbb{R}$, by $f_{\restriction_A}$ we denote its restriction to the set A.

Lebesgue integral

We will need two important properties of the Lebesgue integral, namely Fatou's Lemma and the dominated convergence theorem, which we formulate here.

Theorem 2.1 (Fatou's Lemma, see, e.g., [2, 71, 177])
Let $\Omega \subset \mathbb{R}^d$ be a measurable set and let $f_n : \Omega \to [0, \infty]$, $n = 1, 2, \ldots$ be a sequence of measurable functions. Then

$$\int_\Omega \liminf_{n\to\infty} f_n dx \leq \liminf_{n\to\infty} \int_\Omega f_n dx.$$

Note that the integrals in Theorem 2.1 may be finite or infinite.

Theorem 2.2 (Dominated convergence theorem (DCT), see, e.g., [2, 71, 177])
Let $\Omega \subset \mathbb{R}^d$ be a measurable set and let f, f_n, $n = 1, 2, \ldots$ be measurable functions. Suppose that

$$f_n(x) \to f(x) \;\; a.e.\; x \in \Omega$$

and that there exists a nonnegative function $g \in L^1(\Omega)$, i.e., $\int_\Omega g dx < \infty$ such that

$$|f_n(x)| \leq g(x) \;\; a.e.\; x \in \Omega.$$

Then

$$\lim_{n\to\infty} \int_\Omega |f_n - f|\, dx = 0.$$

L^p spaces

Let Ω be a domain in $\mathbb{R}^d$ and let $1 \leq p < \infty$. The vector space $L^p(\Omega)$ consists of all Lebesgue measurable functions $u : \Omega \to \mathbb{R}$ such that

$$\|u\|^p_{L^p(\Omega)} := \int_\Omega |u|^p \, dx < \infty.$$

When $p = \infty$, the space $L^\infty(\Omega)$ is defined as the space consisting of all measurable functions u which are essentially bounded, i.e.,

$$\|u\|_{L^\infty(\Omega)} := \inf \{C > 0 \text{ such that } |u(x)| \leq C \text{ for almost each } x \in \Omega\}.$$

Further, $L^p_{loc}(\Omega)$ denotes the space of all measurable functions which are in $L^p(K)$ for every compact set K in Ω. We identify in $L^p(\Omega)$, $1 \leq p \leq \infty$ all functions that agree almost everywhere with respect to the Lebesgue measure. Thus the elements of $L^p(\Omega)$ are equivalence classes of measurable functions u for which the quantity $\|u\|_{L^p(\Omega)}$ is finite. With this in mind, it is easy to verify that the functional $\|\cdot\|_{L^p(\Omega)}$ is a norm in $L^p(\Omega)$ for all $1 \leq p \leq \infty$. Moreover, with this norm, the space $L^p(\Omega)$ is a complete vector space, and thus a Banach space. In the special case $p = 2$, the space $L^2(\Omega)$ becomes a Hilbert space for the inner product $(\cdot, \cdot)_\Omega$ defined by

$$(u, v)_\Omega := \int_\Omega uv dx \text{ for all } u, v \in L^2(\Omega).$$

If there is no ambiguity, we will denote the inner product in $L^2(\Omega)$ just by $(\cdot, \cdot)$ and will skip the subscript $''\Omega''$. If A is a subset of Ω we will denote the inner product in $L^2(A)$ by $(\cdot, \cdot)_A$ to distinguish it from $(\cdot, \cdot)$. Similarly to the scalar case, we can introduce the space $[L^p(\Omega)]^d$ of vector valued functions $\boldsymbol{u} = (u_1, u_2, \ldots, u_d) : \Omega \to \mathbb{R}^d$ with the norm

$$\|\boldsymbol{u}\|^p_{[L^p(\Omega)]^d} := \int_\Omega |\boldsymbol{u}|^p_p \, dx,$$

where $|\boldsymbol{u}|_p$ denotes the Euclidean norm in $\mathbb{R}^d$ given by $|\boldsymbol{u}|^p_p := \sum_{i=1}^d |u_i|^p$. When $p = 2$, this space is also a Hilbert space. We will denote the inner product in the same way as for the scalar case

$$(\boldsymbol{u}, \boldsymbol{v})_\Omega := \int_\Omega \boldsymbol{u} \cdot \boldsymbol{v} dx = \int_\Omega (u_1 v_1 + \ldots u_d v_d) \, dx \text{ for all } \boldsymbol{u}, \boldsymbol{v} \in \left[L^2(\Omega)\right]^d,$$

where we will again skip the subscript $''\Omega''$ if there is no ambiguity. For a subset A of Ω we will keep the subscript $''A''$ to distinguish the inner product in $\left[L^2(A)\right]^d$ from that in $\left[L^2(\Omega)\right]^d$. If $p = 2$, we will skip the index in the notation for the Euclidean norm in $\mathbb{R}^n$, $n \in \mathbb{N}$, i.e., we will write $|\xi|$ instead of $|\xi|_2$.

Remark 2.3
Everywhere in this work we denote scalar functions and position vectors (points in $\mathbb{R}^d$) in standard italic font, e.g., $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, $u(x)$, $u : \Omega \subset \mathbb{R}^d \to \mathbb{R}$. On the other hand, vector or matrix valued functions we denote by letters in bold italic font, e.g., $\boldsymbol{f}(x) = (f_1(x), f_2(x), \ldots, f_n(x))$, $\boldsymbol{f} : \mathbb{R}^d \to \mathbb{R}^n$.

Theorem 2.4 (Hölder's inequality)
Let $1 \leq p \leq \infty$ and let q denote the Hölder conjugate of p defined by $\frac{1}{p} + \frac{1}{q} = 1$. If $u \in L^p(\Omega)$, $v \in L^q(\Omega)$, then $uv \in L^1(\Omega)$ and it is satisfied

$$\|uv\|_{L^1(\Omega)} \leq \|u\|_{L^p(\Omega)} \|v\|_{L^q(\Omega)}.$$

Theorem 2.5 (Minkowski's inequality, i.e., triangle inequality)
Let $1 \leq p \leq \infty$. Then

$$\|u + v\|_{L^p(\Omega)} \leq \|u\|_{L^p(\Omega)} + \|v\|_{L^p(\Omega)}.$$

Theorem 2.6 (Lemma 3.31 in [2])
Let $u \in L^1_{loc}(\Omega)$ satisfy $\int_\Omega u\varphi dx = 0$ for all $\varphi \in C_0^\infty(\Omega)$. Then $u(x) = 0$ a.e. in Ω.

Theorem 2.7 (Approximation by compactly supported smooth functions)
$C_0^\infty(\Omega)$ is dense in $L^p(\Omega)$ for all $1 \leq p < \infty$.

Weak derivatives

We can generalize the notion of a classical derivative for functions in $L^1_{loc}(\Omega)$. Let $u, v \in L^1_{loc}(\Omega)$ and let α be a multiindex. We say that v is the α^{th} weak partial derivative of u and write $D^\alpha u = v$ provided that

$$\int_\Omega u D^\alpha \varphi dx = (-1)^{|\alpha|} \int_\Omega v\varphi dx$$

for all test functions $\varphi \in C_0^\infty(\Omega)$.

Sobolev $W^{m,p}$ spaces

Now, for an open set Ω in $\mathbb{R}^d$, we can define the Sobolev space $W^{m,p}(\Omega)$ for all $1 \leq p \leq \infty$ and for all integers $m \geq 0$ by

$$W^{m,p}(\Omega) := \{u \in L^p(\Omega) : D^\alpha u \in L^p(\Omega) \text{ for all multiindices } \alpha \text{ with } |\alpha| \leq m\},$$

where $D^\alpha u$ denotes the α^{th} weak partial derivative of u and we set $W^{0,p}(\Omega) := L^p(\Omega)$. We can introduce a norm in $W^{m,p}(\Omega)$ by the function $\|\cdot\|_{W^{m,p}(\Omega)}$ given by

$$\|u\|^p_{W^{m,p}(\Omega)} := \sum_{0 \leq |\alpha| \leq m} \|D^\alpha u\|^p_{L^p(\Omega)}, \text{ if } p < \infty,$$

$$\|u\|_{W^{m,\infty}(\Omega)} := \max_{0 \leq |\alpha| \leq m} \|D^\alpha u\|_{L^\infty(\Omega)}.$$

With this norm, $W^{m,p}(\Omega)$ is a complete space and thus a Banach space. In the special case $p = 2$, $W^{m,2}(\Omega)$ becomes a Hilbert space and we denote it by $H^m(\Omega)$. Further, by $W_0^{m,p}(\Omega)$ we denote the closure of $C_0^\infty(\Omega)$ in $W^{m,p}(\Omega)$ and by such it is also a Banach space. When $p = 2$, we write $H_0^m(\Omega) = W_0^{m,2}(\Omega)$.

Theorem 2.8 (see Proposition 9.4 in [96])
Let $\Omega \subset \mathbb{R}^d$ be an open set, $u, v \in W^{1,p}(\Omega) \cap L^\infty(\Omega)$ with $1 \leq p \leq \infty$. Then $uv \in W^{1,p}(\Omega) \cap L^\infty(\Omega)$ and

$$\frac{\partial}{\partial x_i}(uv) = \frac{\partial u}{\partial x_i}v + u\frac{\partial v}{\partial x_i},\ i = 1, 2, \ldots, d.$$

Theorem 2.9 (see Theorem 1 in Section 5.2.3 in [70])
Let $\Omega \subset \mathbb{R}^d$ be an open set, $\xi \in C_0^\infty(\Omega)$, and $u \in W^{m,p}(\Omega)$. Then $\xi u \in W^{m,p}(\Omega)$ and Leibniz' formula applies

$$D^\alpha(\xi u) = \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} D^\beta \xi D^{\alpha-\beta} u,$$

where $\binom{\alpha}{\beta} = \frac{\alpha!}{\beta!(\alpha-\beta)!}$.

Theorem 2.10 (see Theorem 2.2.4 in [111] or Lemma 9.5 in [96])
Let $\Omega \subset \mathbb{R}^d$ be an open set, $u \in W^{1,p}(\Omega)$ with $1 \leq p < \infty$, and assume that the support of u is a compact subset of Ω. Then $u \in W_0^{1,p}(\Omega)$.

In this thesis, we will work only with bounded sets in $\mathbb{R}^d$. Thus, let Ω be a bounded and open set in $\mathbb{R}^d$. We say that Ω has Lipschitz boundary $\partial\Omega$ if each point x on $\partial\Omega$ has a neighborhood U_x whose intersection with $\partial\Omega$ is the graph of a Lipschitz continuous function (see [2]). We will also say that Ω is a bounded Lipschitz domain.

Theorem 2.11 (see Theorem 3.22 in [2])
Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. Then the set of restrictions to Ω of functions in $C_0^\infty(\mathbb{R}^d)$ is dense in $W^{m,p}(\Omega)$ for $1 \leq p < \infty$.

Theorem 2.12 (Poincaré's inequality, see, e.g., Theorem 12.17 in [122], Theorem 1.4.3.4 in [95], Corollary 9.19 in [96])
Suppose that $1 \leq p < \infty$ and that Ω is a bounded domain in $\mathbb{R}^d$. Then there exists a constant $C = C(p, \Omega)$ such that

$$\|u\|_{L^p(\Omega)} \leq C\|\nabla u\|_{L^p(\Omega)} \text{ for all } u \in W_0^{1,p}(\Omega). \tag{2.2}$$

Traces in $W^{1,p}(\Omega)$

In general, functions in Sobolev spaces do not have well defined values on sets with measure zero. However, for bounded Lipschitz domains one can extend the notion of restriction of a continuous function to the boundary of Ω by introducing the trace operator. We have the following theorem.

Theorem 2.13 (Trace Theorem, see Theorems 15.10 and 15.23 in [122])
Let $\Omega \subset \mathbb{R}^d$, $d \geq 2$ be a bounded domain with Lipschitz boundary $\partial\Omega$ and let $1 \leq p < \infty$. Then there exists a continuous linear operator

$$\gamma_p : W^{1,p}(\Omega) \to L^p(\partial\Omega)$$

such that
(i) $\gamma_p(u) = u$ on $\partial\Omega$ for all $u \in W^{1,p}(\Omega) \cap C(\overline{\Omega})$,
(ii) for all $\varphi \in C_0^1(\mathbb{R}^n)$, $u \in W^{1,p}(\Omega)$, and $i = 1, 2, \ldots, d$, the following integration by parts formula holds

$$\int_\Omega u \frac{\partial \varphi}{\partial x_i} dx = -\int_\Omega \varphi \frac{\partial u}{\partial x_i} dx + \int_{\partial\Omega} \varphi \gamma_p(u) n_i ds, \tag{2.3}$$

where n_i is the i-th component of the outward unit normal vector to the boundary $\partial\Omega$.

As a consequence of the integration by parts formula (2.3) we can obtain the following form of the Gauss-Ostrogradsky Theorem.

Theorem 2.14
Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary $\partial\Omega$ and a unit outward normal vector $\boldsymbol{n}_{\partial\Omega}$. Let $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_d) \in \left[W^{1,p}(\Omega)\right]^d$, $\varphi \in C_0^1(\mathbb{R}^d)$, and $1 \leq p < \infty$. Then the following integration by parts formula holds

$$\int_\Omega \boldsymbol{\psi} \cdot \nabla \varphi dx = -\int_\Omega \varphi \operatorname{div} \boldsymbol{\psi} dx + \int_{\partial\Omega} \gamma_p(\boldsymbol{\psi}) \cdot \boldsymbol{n}_{\partial\Omega} \varphi ds, \tag{2.4}$$

where $\operatorname{div} \boldsymbol{\psi} = \frac{\partial \psi_1}{\partial x_1} + \ldots + \frac{\partial \psi_d}{\partial x_d}$ *and* $\gamma_p(\boldsymbol{\psi}) := (\gamma_p(\psi_1), \ldots, \gamma_p(\psi_d))$.

By a standard density argument, i.e., by applying Theorem 2.11 and the trace theorem, Theorem 2.13, we obtain a more general form of (2.4) which is valid for functions $\varphi \in W^{1,q}(\Omega)$ with $\frac{1}{p} + \frac{1}{q} = 1$.

Theorem 2.15
Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary $\partial\Omega$ and a unit outward normal vector $\boldsymbol{n}_{\partial\Omega}$. Let $1 < p < \infty$ and let q be its Hölder conjugate defined by $\frac{1}{p} + \frac{1}{q} = 1$. If $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_d) \in \left[W^{1,p}(\Omega)\right]^d$ and $v \in W^{1,q}(\Omega)$, then the following integration by parts formula holds

$$\int_\Omega \boldsymbol{\psi} \cdot \nabla v dx = -\int_\Omega v \operatorname{div} \boldsymbol{\psi} dx + \int_{\partial\Omega} \gamma_p(\boldsymbol{\psi}) \cdot \boldsymbol{n}_{\partial\Omega} \gamma_q(v) ds. \tag{2.5}$$

The next theorem characterizes the space $W_0^{1,p}(\Omega)$ as the subspace of functions in $W^{1,p}(\Omega)$ with zero trace.

Theorem 2.16 (Theorem 15.29 in [122])
Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain whose boundary is $\partial\Omega$, let $1 \leq p < \infty$, and let $u \in W^{1,p}(\Omega)$. Then $\gamma_p(u) = 0$ if and only if $u \in W_0^{1,p}(\Omega)$.

The next theorem is a refined version of Theorem 2.16 and it is due to Gagliardo [84].

Theorem 2.17 (Refined version of Theorem 2.13, see, e.g., Theorem 1.5.1.3 in [95])
Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain whose boundary is $\partial\Omega$ and let $1 < p < \infty$. Then the mapping $u \mapsto \gamma_p(u)$ which is defined for $u \in C^{0,1}(\overline{\Omega})$, has a unique continuous extension as an operator from $W^{1,p}(\Omega)$ onto $W^{1-1/p,p}(\partial\Omega)$. This operator has a right continuous inverse independent of p.

We will use this theorem for $p = 2$. In particular, there exists a constant $C_{tr} = C_{tr}(p, d, \Omega) > 0$ such that $\gamma_2(u) \in H^{1/2}(\partial\Omega)$ for all $u \in W^{1,2}(\Omega) \equiv H^1(\Omega)$ and

$$\|\gamma_2(u)\|_{H^{1/2}(\partial\Omega)} \leq C_{tr}\|u\|_{H^1(\Omega)}.$$

Conversely, there exists a constant $C_{inv} = C_{inv}(p, d, \Omega) > 0$ such that for any given function $w \in H^{1/2}(\partial\Omega)$ there is an extension $u \in H^1(\Omega)$ such that

$$\|u\|_{H^1(\Omega)} \leq C_{inv}\|w\|_{H^{1/2}(\partial\Omega)}.$$

The dual of $H^{1/2}(\partial\Omega)$ is denoted by $H^{-1/2}(\partial\Omega)$ and the duality product in $H^{-1/2}(\partial\Omega) \times H^{1/2}(\partial\Omega)$ is denoted by $\langle\cdot,\cdot\rangle_{H^{-1/2}(\partial\Omega)\times H^{1/2}(\partial\Omega)}$.

For $s \in (0, 1)$, the space $W^{s,p}(\partial\Omega)$ above consists of all functions $g \in L^p(\partial\Omega)$ such that

$$\|g\|_{W^{s,p}(\partial\Omega)} := \left(\|g\|^p_{L^p(\partial\Omega)} + \int_{\partial\Omega}\int_{\partial\Omega} \frac{|g(x) - g(y)|^p}{|x - y|^{d-1+sp}} ds(x)ds(y) \right)^{\frac{1}{p}} < \infty$$

The space $W^{s,p}(\partial\Omega)$ equipped with the norm $\|\cdot\|_{W^{s,p}(\partial\Omega)}$ is a Banach space. For more information on Sobolev spaces of fractional order, we refer to [2, 95, 122, 125]. Next, for a bounded Lipschitz domain Ω, a function $g \in C^{0,1}(\partial\Omega)$, and $1 < p < \infty$, by $W_g^{1,p}(\Omega)$ we denote the functional space defined by

$$W_g^{1,p}(\Omega) := \{v \in W^{1,p}(\Omega), \text{ such that } \gamma_p(v) = g \text{ on } \partial\Omega\}.$$

The dual of $W_0^{1,p}(\Omega)$

The space of bounded linear functionals over $W_0^{1,p}(\Omega)$ is denoted by $W^{-1,p'}(\Omega)$ where $1 \leq p < \infty$ and p' is the Hölder conjugate of p. By $H^{-1}(\Omega)$ we denote the dual of $H_0^1(\Omega)$ and the duality pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$ we denote by $\langle\cdot,\cdot\rangle$.

In the proofs of uniqueness and a priori L^∞ estimates for semilinear elliptic equations without any growth conditions on the nonlinearity, we will need the following result due to H. Brézis and F. Browder, 1978.

Theorem 2.18 (A property of Sobolev spaces, H. Brézis and F. Browder, 1978, [35])
Let Ω be an open set in $\mathbb{R}^d$, $T \in H^{-1}(\Omega) \cap L^1_{loc}(\Omega)$, and $v \in H^1_0(\Omega)$. If there exists a function $f \in L^1(\Omega)$ such that $T(x)v(x) \geq f(x)$, a.e in Ω, then $Tv \in L^1(\Omega)$ and the duality product $\langle T, v\rangle$ in $H^{-1}(\Omega) \times H^1_0(\Omega)$ coincides with $\int_\Omega Tvdx$.

Remark 2.19
In other words, we have the following situation: a locally summable function $b \in L^1_{loc}(\Omega)$ defines a bounded linear functional T_b over the dense subspace $D(\Omega) \equiv C^\infty_0(\Omega)$ of $H^1_0(\Omega)$ through the integral formula $\langle T_b, \varphi\rangle = \int_\Omega b\varphi dx$. It is clear that the functional T_b is uniquely extendable by continuity to a bounded linear functional $\overline{T}_b$ over the whole space $H^1_0(\Omega)$. Now the question is whether this extension is still representable by the same integral formula for any $v \in H^1_0(\Omega)$ (if the integral makes sense at all). If the function $v \in H^1_0(\Omega)$ is fixed, then Theorem 2.18 gives a sufficient condition for bv to be summable and for the extension $\overline{T}_b$ evaluated at v to be representable with the same integral formula as above, i.e $\langle \overline{T}_b, v\rangle = \int_\Omega bvdx$.

The space $H(\mathrm{div};\Omega)$

Let Ω be a domain in $\mathbb{R}^d$ and let $\boldsymbol{\psi} \in \left[L^1_{loc}(\Omega)\right]^d$. We say that $\boldsymbol{\psi}$ possesses a weak divergence if there exists a locally integrable function $w \in L^1_{loc}(\Omega)$ such that

$$\int_\Omega \boldsymbol{\psi} \cdot \nabla\varphi dx = -\int_\Omega w\varphi dx \text{ for all } v \in C^\infty_0(\Omega).$$

The function w is called a weak divergence of $\boldsymbol{\psi}$, it is obviously unique, and we right $\mathrm{div}\,\boldsymbol{\psi} = w$.

Now, let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary $\partial\Omega$ whose outward unit normal vector is denoted by $\boldsymbol{n}_{\partial\Omega}$. The space of vector functions with square integrable weak divergence is denoted by $H(\mathrm{div};\Omega)$ and is defined by

$$H(\mathrm{div};\Omega) := \left\{\boldsymbol{\psi} \in \left[L^2(\Omega)\right]^d : \mathrm{div}\,\boldsymbol{\psi} \in L^2(\Omega)\right\}$$

with the graph norm

$$\|\boldsymbol{\psi}\|_{H(\mathrm{div};\Omega)} := \left(\|\boldsymbol{\psi}\|^2_{L^2(\Omega)} + \|\mathrm{div}\,\boldsymbol{\psi}\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}}.$$

With the obvious inner product, $H(\mathrm{div};\Omega)$ is a Hilbert space. An alternative characterization of $H(\mathrm{div};\Omega)$ is as the closure of $\left[C^\infty(\overline{\Omega})\right]^d$ in the norm $\|.\|_{H(\mathrm{div};\Omega)}$, see, e.g. [59, 137]. For a function $\boldsymbol{\psi} \in \left[C^\infty(\overline{\Omega})\right]^d$, the normal trace operator γ_n is defined almost everywhere on $\partial\Omega$ by

$$\gamma_n(\boldsymbol{\psi}) := \text{ restriction of } \boldsymbol{\psi} \cdot \boldsymbol{n}_{\partial\Omega} \text{ to } \partial\Omega. \tag{2.6}$$

Theorem 2.20 (Trace theorem for $H(\mathrm{div};\Omega)$, see, e.g., Theorem 2 in Section 1.3 in [59], Theorem 3.24 in [137])
Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary $\partial\Omega$ and a unit outward normal vector $\boldsymbol{n}_{\partial\Omega}$. Then the mapping γ_n, defined by (2.6), can be extended by continuity to a continuous linear operator from $H(\mathrm{div};\Omega)$ onto $H^{-1/2}(\partial\Omega)$. Moreover, the following divergence formula holds for all functions $\boldsymbol{\psi} \in H(\mathrm{div};\Omega)$ and $v \in H^1(\Omega)$:

$$\int_\Omega \boldsymbol{\psi} \cdot \nabla v dx = -\int_\Omega v \,\mathrm{div}\, \boldsymbol{\psi} dx + \langle \gamma_n(\boldsymbol{\psi}), \gamma_2(v) \rangle_{H^{-1/2}(\partial\Omega)\times H^{1/2}(\partial\Omega)} \tag{2.7}$$

Now we give a particular version of the Sobolev embeddings, summarized in the following theorem. For a full version of the Sobolev embedding theorem, we refer to Theorem 4.12 in [2].

Theorem 2.21 (Sobolev embedding Theorem)
Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. Let $j \geq 0$ and $m \geq 1$ be integers and let $1 \leq p < \infty$.
Case A *If $mp > d > (m-1)p$, then*

$$W^{j+m,p}(\Omega) \hookrightarrow C^{j,\lambda}(\overline{\Omega}), \ \ \textit{for } 0 < \lambda \leq m - \frac{d}{p},$$

and if $d = (m-1)p$, then

$$W^{j+m,p}(\Omega) \hookrightarrow C^{j,\lambda}(\overline{\Omega}), \ \ \textit{for } 0 < \lambda < 1.$$

In particular

$$W^{m,p}(\Omega) \hookrightarrow L^q(\Omega), \ \ \textit{for } p \leq q \leq \infty.$$

Case B *If $mp = d$, then*

$$W^{j+m,p}(\Omega) \hookrightarrow W^{j,q}(\Omega), \ \ \textit{for } p \leq q < \infty,$$

and in particular

$$W^{m,p}(\Omega) \hookrightarrow L^q(\Omega), \ \ \textit{for } p \leq q < \infty.$$

Case C *If $mp < d$, then*

$$W^{j+m,p}(\Omega) \hookrightarrow W^{j,q}(\Omega), \ \ \textit{for } p \leq q \leq p^* = \frac{dp}{d-mp}.$$

In particular,

$$W^{m,p}(\Omega) \hookrightarrow L^q(\Omega), \ \ \textit{for } p \leq q \leq p^* = \frac{dp}{d-mp}.$$

The embedding constants in the embeddings above depend only on d, m, p, q, j, and the domain Ω.

The next theorem specifies under what conditions the above embeddings are compact.

Theorem 2.22 (Rellich-Kondrachov Theorem, see Theorem 6.3 in [2])
Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. Let $j \geq 0$ and $m \geq 1$ be integers, and let $1 \leq p < \infty$.
Case A *The following embeddings are compact:*

$$W^{j+m,p}(\Omega) \hookrightarrow C^j(\overline{\Omega}), \quad \text{if } mp > m,$$
$$W^{j+m,p}(\Omega) \hookrightarrow C^{j,\lambda}(\overline{\Omega}), \quad \text{if } mp > d \geq (m-1)p, \quad \text{and } 0 < \lambda < m - \frac{d}{p}.$$

Case B *If $mp = d$, then the following embeddings are compact:*

$$W^{j+m,p}(\Omega) \hookrightarrow W^{j,q}(\Omega), \quad \text{if } 1 \leq q < \infty.$$

Case C *If $mp < d$, then the following embeddings are compact:*

$$W^{j+m,p}(\Omega) \hookrightarrow W^{j,q}(\Omega), \quad \text{if } 1 \leq q < \frac{dp}{d-mp}.$$

2.3 Variational problems and convex analysis

Theorem 2.23 (Lax-Milgram Theorem, see, e.g., Theorem 2.7.7 in [32])
Let H be a Hilbert space with the norm $\|\cdot\|_H$ and let the bilinear form $a(\cdot,\cdot) : H \times H \to \mathbb{R}$ satisfies the following two conditions:

(i) Boundedness: $\exists\, \overline{c} > 0$ such that $a(u,v) \leq \overline{c}\|u\|_H\|v\|_H, \quad \forall u, v \in H,$

(ii) Coercivity: $\exists\, \underline{c} > 0$ such that $a(u,u) \geq \underline{c}\|u\|_H^2, \quad \forall u \in H.$

Then, for any $F \in H^*$, the variational problem

$$\text{Find } u \in H \text{ such that } a(u,v) = \langle F, v\rangle, \forall v \in H$$

has a unique solution u which satisfies the a priori bound

$$\|u\|_H \leq \frac{1}{\underline{c}}\|F\|_{H^*}.$$

Strong and weak convergence

Let V be a Banach space with a norm $\|\cdot\|_V$ and let $\{u_n\}_{n=1}^{\infty} \subset V$. We say that u_n converges strongly or in norm to an element $u \in V$ if $\|u_n - u\|_V \to 0$ as $n \to \infty$. We say that u_n converges weakly to u and we write $u_n \rightharpoonup u$ if $\langle f, u_n\rangle \to \langle f, u\rangle$ for all $f \in V^*$, the dual of V.

Convex analysis

We continue by recalling some facts from convex analysis. Let V be a linear vector space and let $A \subset V$ is a convex set, we say that $J : A \to (-\infty, +\infty]$ is convex if

$$J\left(tx_1 + (1-t)x_2\right) \leq tJ(x_1) + (1-t)J(x_2), \quad \forall x_1, x_2 \in A, \quad \forall t \in [0,1]$$

and J is strictly convex if

$$J\left(tx_1 + (1-t)x_2\right) < tJ(x_1) + (1-t)J(x_2), \quad \forall x_1, x_2 \in A, \quad \forall t \in (0,1).$$

We say that a convex function of V in $\overline{\mathbb{R}} := [-\infty, \infty]$ is proper if it nowhere takes the value $-\infty$ and is not identically equal to $+\infty$.

Now, we recall the notion of lower semi-continuity (l.s.c) and weak lower semi-continuity (w.l.s.c.) in the setting that is relevant to our further considerations and for more information we refer to [65, 82, 96]. Let V be a Banach space and $J : V \to (-\infty, +\infty]$. We say that J is sequentially l.s.c. if for every sequence $\{u_n\}_{n=1}^{\infty} \subset V$ such that $u_n \to u \in V$ it is satisfied $J(u) \leq \liminf_{n\to\infty} J(u_n)$. Similarly, we say that J is sequentially w.l.s.c. if for every sequence $u_n \rightharpoonup u \in V$ it is satisfied $J(u) \leq \liminf_{n\to\infty} J(u_n)$. It is easy to see that if J_1 and J_2 are sequentially l.s.c., respectively, sequentially w.l.s.c., then $J_1 + J_2$ is also sequentially l.s.c., respectively, sequentially w.l.s.c.. When it comes to proving existence results for convex minimization problems, we will need the following two facts. If $A \subset V$ is norm closed and convex, then it is also weakly closed. From this it follows that if J is a convex l.s.c. functional, then J is also sequentially w.l.s.c., see, e.g. Corollary 3.9 in [96] or Corollary 2.2 in [65].

Definition 2.24 (Fenchel conjugate, see, e.g., [65, 96, 162])
Let V be a normed vector space and let $J : V \to (-\infty, +\infty]$ be a functional such that $J \not\equiv +\infty$. The functional $J^* : V^* \to (-\infty, +\infty]$ defined by

$$J^*(v^*) = \sup_{v \in V} \{\langle v^*, v\rangle - J(v)\}$$

is called the polar (Fenchel conjugate) functional to J.

Obviously, J^* is l.s.c. and convex as the supremum of the family of continuous affine functionals $\langle \cdot, v\rangle - J(v)$. In a similar fashion we can define the bipolar (biconjugate) J^{**} to J.

Definition 2.25 (Fenchel biconjugate, see, e.g., [65, 96])
Let V be a Banach space and let $J : V \to (-\infty, +\infty]$ be a functional such that $J \not\equiv +\infty$. The functional $J^{**} : V \to \overline{\mathbb{R}}$ defined by

$$J^{**}(v) = \sup_{v^* \in V^*} \{\langle v^*, v\rangle - J^*(v^*)\}$$

is called the bipolar (Fenchel biconjugate) functional to J.

The following theorem will be used in Chapter 4and it is of particular importance to the derivation of functional a posteriori error estimates which are based on the duality theory.

Theorem 2.26 (Fenchel-Moreau, see, e.g., [65, 96])
Let V be a normed vector space. Assume that $J : V \to (-\infty, +\infty]$ is a convex, l.s.c. proper functional. Then $J^{**} = J$.

Definition 2.27 (see, e.g., [65])
Let $J : V \to \overline{\mathbb{R}}$. We call the limit as $\lambda \to 0^+$, if it exists, of

$$\frac{J(u + \lambda v) - J(u)}{\lambda}$$

the directional derivative of J at u in the direction v and denote it by $J'(u; v)$. If there exists $u^* \in V^*$ such that:

$$\forall v \in V, \qquad J'(u; v) = \langle u^*, v \rangle$$

we say that J is Gateaux-differentiable at u, call u^* the Gateaux-differential at u of J, and denote it by $J'(u)$.

Definition 2.28
Let $\mathscr{C}$ be a subset in the normed vector space V with norm $\|\cdot\|$. We say that $J : V \to \overline{\mathbb{R}}$ is coercive in $\mathscr{C}$ if

$$\lim J(v) = +\infty, \qquad \text{for } v \in \mathscr{C}, \quad \|v\| \to \infty. \tag{2.8}$$

For coercive functionals we have the following useful property.

Proposition 2.29
If $J : \mathscr{C} \to \overline{\mathbb{R}}$ is coercive, then the sets $\mathscr{C}_\alpha := \{v \in \mathscr{C} : J(v) \leq \alpha\}$ are bounded.

Proof. Assume that $\mathscr{C}_\alpha$ is unbounded. Then, for any $n \in \mathbb{N}$ there is $v_n \in \mathscr{C}_\alpha$ such that $\|v_n\| \geq n$. From the coercivity condition (2.8) it follows that $J(v_n) \to +\infty$ which is a contradiction with the fact that $J(v_n) \leq \alpha$. □

Versions of the next theorem on existence of a minimizer can be found, for example, in Proposition 1.2 in [65], Theorem 5.4.1 in [139], Theorem 2.11 in [9], Theorem 7.3.7 in [118].

Theorem 2.30 (Existence of a minimizer)
Let V be a reflexive Banach space with norm $\|\cdot\|$ and let $\mathscr{C}$ be a non-empty closed convex subset of V. Let $J : \mathscr{C} \to (-\infty, +\infty]$ be a convex proper sequentially lower semi-continuous functional. Let us assume that either $\mathscr{C}$ is bounded or that J is coercive over $\mathscr{C}$. Then the problem

$$\text{Find } u \in \mathscr{C} \text{ such that } J(u) = \inf_{v \in \mathscr{C}} J(v) \tag{2.9}$$

has at least one solution. It has a unique solution if J is strictly convex.

Proof. Let $\{u_n\}_{n=1}^{\infty} \subset \mathscr{C}$ be a minimizing sequence, i.e.,

$$\lim_{n\to\infty} J(u_n) = \inf_{v\in\mathscr{C}} J(v) = \beta,$$

such that $+\infty > J(u_1) \geq J(u_2) \geq \ldots \geq J(u_n) \geq \ldots$. Note that a priori $\beta \in [-\infty, +\infty)$. We observe that the sequence u_n is bounded in V. Indeed, if $\mathscr{C}$ is bounded, this is obvious, and if J is coercive, this follows from Proposition 2.29 with $\alpha = J(u_1)$. The set $\mathscr{C}$ is convex and norm closed, thus it is weakly closed. Since, V is a reflexive Banach space and $\mathscr{C}$ is weakly closed, we can extract a weakly convergent subsequence $\{u_{n_k}\}_{k=1}^{\infty}$ which converges to an element $u \in \mathscr{C}$, i.e., $u_{n_k} \rightharpoonup u$. The functional J is convex and sequentially l.s.c. and thus it is sequentially w.l.s.c.. Hence,

$$J(u) \leq \liminf_{n\to\infty} J(u_{n_k}) = \beta,$$

u is a solution to (2.9), and $\beta > -\infty$. □

Remark 2.31
In Theorem 2.30 we allow the functional J to take the value $+\infty$. In this case, J being convex over $\mathscr{C}$ is equivalent to $dom(J) := \{v \in \mathscr{C} : J(v) < \infty\}$ being a convex set and J being convex over $dom(J)$ (see, e.g., [65]).

2.4 Regularity of linear elliptic interface problems

Here we list two theorems concerning the regularity of linear elliptic (interface) problems that we will use often in this work.

Theorem 2.32 (Boundedness of weak solutions, see, e.g., [174] or Theorem B.2 in [112])
Let $a_{ij}(x) \in L^{\infty}(\Omega)$ satisfy

$$\underline{c}\,|\xi|^2 \leq a_{ij}(x)\xi_i\xi_j \quad \text{for all } \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d, \text{ a.e. } x \in \Omega,$$

where $\underline{c} > 0$. Let $f_0, f_1, \ldots, f_d \in L^s(\Omega)$ for $s > d$ and let

$$u \in H_0^1(\Omega) : \int_{\Omega} \sum_{i,j=1}^{d} a_{ij} \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_i} dx = \int_{\Omega} \left(f_0 v + f_1 \frac{\partial v}{\partial x_1} + \ldots + f_d \frac{\partial v}{\partial x_d} \right) dx, \; \forall v \in H_0^1(\Omega).$$

Then

$$\|u\|_{L^{\infty}(\Omega)} \leq \frac{K}{\underline{c}} \sum_{i=0}^{d} \|f_i\|_{L^s(\Omega)} |\Omega|^{\frac{1}{d}-\frac{1}{s}}, \tag{2.10}$$

where K is a constant independent of $\underline{c}$.

A key element in the proof of Theorem 2.32 is the following Lemma, which we will also use in our work.

Lemma 2.33 (Lemma B.1. in [112])
Let $\varphi(t)$ denote a function which is nonnegative and nonincreasing for $k_0 \leq t < \infty$. Further, let

$$\varphi(h) \leq C\frac{\varphi(k)^\beta}{(h-k)^\alpha}, \forall h > k > k_0, \tag{2.11}$$

where C, α are positive constants and $\beta > 1$. If $e \in \mathbb{R}$ is defined by $e^\alpha := C\varphi(k_0)^{\beta-1}2^{\frac{\alpha\beta}{\beta-1}}$, then $\varphi(k_0 + e) = 0$.

Theorem 2.34 (Optimal regularity of elliptic interface problems, Theorem 1.1 in [66])
Assume that $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain and let $\Omega_0 \subset \Omega$ be another domain with a C^1 boundary, which does not touch the boundary of Ω. Let $\boldsymbol{\mu}$ be a function on Ω with values in the set of real, symmetric $d \times d$ matrices which is uniformly continuous on both Ω_0 and $\Omega \setminus \overline{\Omega_0}$. Additionally, $\boldsymbol{\mu}$ is supposed to satisfy the usual ellipticity condition

$$\exists\, \underline{c} \text{ such that } \underline{c}\,|\xi|^2 \leq \boldsymbol{\mu}(x)\xi \cdot \xi \quad \text{for all } \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d, \text{ a.e. } x \in \Omega.$$

Then there is a $p > 3$ such that for every $\lambda \geq 0$,

$$-\nabla \cdot \boldsymbol{\mu}\nabla + \lambda : W_0^{1,q}(\Omega) \to W^{-1,q}(\Omega)$$

is a topological isomorphism for all $q \in (p', p)$ with p' being the Hölder conjugate of p. If Ω itself is also a C^1 domain, then p may be taken as $+\infty$.

Chapter 3

Existence and uniqueness analysis

This chapter is devoted to the solution theory of the Poisson-Boltzmann equation (PBE) and the linearized Poisson-Boltzmann equation (LPBE). We start with the introduction of the PBE and the LPBE, as well as all relevant geometrical regions that describe the system under study. The main difficulties in the analysis of the PBE are related to the exponential nonlinearity and the measure data (represented by a linear combination of delta functions) on the right-hand side of the equation.

The first step in our analysis is to give a meaningful notion of a solution to the LPBE which also ensures uniqueness. This is done in Section 3.2 by following the ideas in [21, 63, 85, 153]. In Section 3.2.1 we prove existence and uniqueness of a solution to the LPBE by means of a 2-term splitting (see, e.g. [48, 141, 195]), where the full potential ϕ is decomposed into a singular Coulomb potential G and a more regular reaction field part u. The existence of a solution is showed by finding a particular representative of the reaction field potential u which satisfies a standard weak formulation involving H^1 spaces. The uniqueness of the potential ϕ is proven by employing an adjoint problem with a more regular right hand side and by making use of an appropriate regularity result for linear interface problems. In some situations, it can be inappropriate to use the 2-term splitting in practice. For this reason, in Section 3.2.2, we also consider a 3-term splitting of the potential ϕ, introduced in [51], and again show that a particular solution ϕ can be obtained by considering a standard weak formulation, involving H^1 spaces, for the regular component in this decomposition. We finish the analysis of the LPBE with Section 3.2.3 where we discuss some regularity properties of the reaction field component of the potential in the 2- and 3-term splittings.

In Section 3.3, we continue with the analysis of the (nonlinear) PBE. This section has a similar structure to Section 3.2. First, we extend in a natural way the definition of a weak solution from the linear case and again show the existence of a solution ϕ by utilizing the 2- and 3-term splittings. A particular representative of the regular component u for both 2- and 3-term splittings can be defined as the solution of a weak formulation that involves H^1

spaces. This time, the problem contains a nonlinearity of exponential type, and therefore showing existence and uniqueness even for the weak formulation involving H^1 spaces is not a trivial task. We analyze this weak formulation in detail in Section 3.3.3, where we also prove a priori L^∞ bounds on the regular part u of the potential. An important tool in the proofs of existence, uniqueness, and boundedness of the component u is a property of the Sobolev spaces proved in [35]. The existence theorems based on the 2- and 3-term splittings are stated in Section 3.3.1 and Section 3.3.2, respectively. However, in the case of the PBE, we are unable to prove the uniqueness of the full potential ϕ: the proof of uniqueness of the potential ϕ for the LPBE employs a duality argument which is not available in the nonlinear case.

Similarly to the case of the LPBE, in Section 3.3.4, we summarize some regularity results for the well behaved component u in the 2- and 3-term splittings, and give the respective conditions under which they hold.

As a generalization of the a priori L^∞ estimate for the regular component u in the 2- and 3-term splittings for the PBE, in Section 3.4, we prove such estimate for the solution of a more general semilinear elliptic equation with neither sign nor growth conditions on the nonlinearity.

3.1 Problem formulation

3.1.1 Poisson-Boltzmann equation

Let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ be a bounded domain with Lipschitz boundary $\partial\Omega$ whose outward unit normal vector is denoted by $\boldsymbol{n}_{\partial\Omega}$. The domain Ω contains two Lipschitz subdomains, Ω_m and Ω_s, which denote the molecular region and the solvent region, respectively. It is assumed that $\overline{\Omega}_m \subset \Omega$, i.e., the molecular region is strictly contained in Ω. Each of the two subdomains Ω_m and Ω_s is allowed to be a disconnected set, which can be represented as the union of Lipschitz domains. The boundary of Ω_m is denoted by Γ, the interface between the molecular region and the solvent region, and its outward unit normal vector is denoted by $\boldsymbol{n}_\Gamma$. Finally, we can write $\Omega = \Omega_m \cup \Gamma \cup \Omega_s$. There are several definitions of the molecular surface that have been used in practice, the most common of which is the solvent excluded surface (SES). The SES is formed by the contact points of the Van der Waals surface and a solvent probe sphere that is rolled over it (see [67, 94, 158, 165]). In 1983 Connolly gave an analytic description of the solvent excluded surface and therefore it is also known as the Connolly surface (see [53] for the piecewise analytic definition of this surface).

To model the electrostatic potential in a system of biomolecules with the presence of moving ions the so-called ion exclusion layer (IEL) is introduced. This is a layer around the bio-molecules in which no ions can penetrate and it is defined as the difference between the union of the inflated Van der Waals spheres of the atoms by a counterion radius R_{ion} and

the molecular region defined by the SES. Alternatively, it is defined as the difference between the region, enclosed by the Connolly surface of the molecule with inflated Van der Waals spheres of the atoms by a counterion radius R_{ion}, and the usual molecular region defined by the Connolly surface. We denote this region by Ω_{IEL}. The part of Ω_s without the ion exclusion layer Ω_{IEL} is accessible for ions and we denote it by Ω_{ions}. With this notation, it holds $\Omega_s = \left(\overline{\Omega_{IEL}} \setminus \Gamma\right) \cup \Omega_{ions}$ (see Figure 3.1).

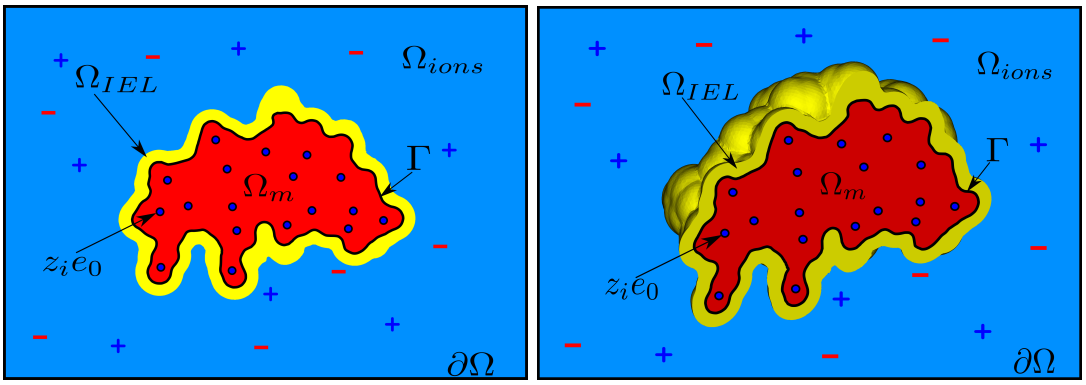


Figure 3.1: Computational domain Ω with molecular domain Ω_m and solution domain $\Omega_s = \overline{\Omega_{IEL}} \setminus \Gamma \cup \Omega_{ions}$.

The electrostatic potential φ is governed by the Poisson equation which is derived from Gauss's law of electrostatics. In CGS (centimeter-gram-second) units, the Poisson equation reads

$$-\nabla \cdot (\epsilon \nabla \varphi) = 4\pi\rho, \tag{3.1}$$

where $\rho(x)$ is the charge density at point x and ϵ is the dielectric coefficient, which is assumed to be constant in the molecule region Ω_m and Lipschitz continuous in the solvent region Ω_s with a possible jump discontinuity across the interface Γ, i.e.,

$$\epsilon(x) = \begin{cases} \epsilon_m, & x \in \Omega_m, \\ \epsilon_s(x), & x \in \Omega_s. \end{cases} \tag{3.2}$$

In the molecular region Ω_m, there are only fixed partial charges and therefore the charge density is

$$\rho_m = \sum_{i=1}^{N_m} z_i e_0 \delta_{x_i},$$

where N_m is the number of fixed partial charges, z_i is the valency of the i-th partial charge, x_i its position, and $e_0 = 4.8032424 \times 10^{-10}$ esu ($= 1.60217662 \times 10^{-19}$ Coulombs) is the elementary charge. For all electrostatics units and physical constants that we will be using see Table 3.1 and Table 3.2. The Poisson equation for the potential φ in Ω_m reads

$$-\nabla \cdot (\epsilon_m \nabla \varphi) = 4\pi\rho_m.$$

In the region Ω_{IEL} there are no fixed partial charges, nor moving ions and therefore the charge density there is $\rho_{IEL} = 0$. The Poisson equation for φ in Ω_{IEL} reads

$$-\nabla \cdot (\epsilon_s \nabla \varphi) = \rho_{IEL} = 0.$$

In the region Ω_{ions}, there are moving ions whose charge density is assumed to follow Boltzmann distribution and is given by

$$\rho_{ions} = \sum_{j=1}^{N_{ions}} M_j \xi_j e_0 \mathrm{e}^{-\frac{\xi_j e_0 \varphi}{k_B T}},$$

where N_{ions} is the number of different ion species in the solvent, ξ_j is the valency of the j-th ion species, $M_j = \frac{\#\text{ions}}{\text{cm}^3}$ is its average concentration in Ω_{ions}, $k_B = 1.38064852 \times 10^{-16}\text{erg K}^{-1}$ is the Boltzmann constant, and T is the absolute temperature. Therefore, the Poisson equation for the potential φ in Ω_{ions} reads

$$-\nabla \cdot (\epsilon_s \nabla \varphi) = 4\pi \rho_{ions}.$$

If we denote the total charge density in Ω by ρ, it holds $\rho = \rho_m + \rho_{IEL} + \rho_{ions}$ and we can write one equation in the whole computational domain Ω

$$\begin{aligned}
-\nabla \cdot (\epsilon \nabla \varphi) - \chi_{\Omega_{ions}} 4\pi \sum_{j=1}^{N_{ions}} M_j \xi_j e_0 \mathrm{e}^{-\frac{\xi_j e_0 \varphi}{k_B T}} &= 4\pi e_0 \sum_{i=1}^{N_m} z_i \delta_{x_i} =: \mathscr{F} \quad \text{in } \Omega, && \text{(3.3a)} \\
[\varphi]_\Gamma &= 0, && \text{(3.3b)} \\
[\epsilon \nabla \varphi \cdot \boldsymbol{n}_\Gamma]_\Gamma &= 0, && \text{(3.3c)} \\
\varphi &= g \quad \text{on } \partial\Omega, && \text{(3.3d)}
\end{aligned}$$

where $[\cdot]_\Gamma$ denotes the jump across the interface Γ of the enclosed quantity and we have taken into account the continuity condition on the potential and the normal component of the displacement field $\epsilon \nabla \varphi$ across the interface Γ. We notice that in fact the physical problem prescribes a vanishing potential at infinite distance from the boundary of Ω_m, i.e., $\lim_{|x| \to \infty} = 0$. In practice, one uses a bounded computational domain and imposes the boundary condition (3.35d) instead, where the function $g \in C^{0,1}(\partial\Omega)$ can usually be calculated accurately enough by solving a simpler problem, possibly with a known analytical solution.

Under the assumption that there are only two ion species in the solution with the same concentration $M_1 = M_2 = M$, which are univalent but with opposite charge, i.e $\xi_j = (-1)^j$, $j = 1, 2$, we obtain the equation

$$-\nabla \cdot (\epsilon \nabla \varphi) + \chi_{\Omega_{ions}} 8\pi M e_0 \sinh\left(\frac{e_0 \varphi}{k_B T}\right) = 4\pi e_0 \sum_{i=1}^{N_m} z_i \delta_{x_i} \quad \text{in } \Omega. \tag{3.4}$$

By introducing the new functions $\phi = \frac{e_0\varphi}{k_BT}$ and $g = \frac{e_0 g}{k_BT}$ in (3.4) we arrive at the equation for the dimensionless potential ϕ

$$\begin{aligned} -\nabla\cdot(\epsilon\nabla\phi) + \overline{k}^2\sinh(\phi) &= \frac{4\pi e_0^2}{k_BT}\sum_{i=1}^{N_m} z_i\delta_{x_i} =: \mathcal{F} \quad \text{in } \Omega, && (3.5a)\\ [\phi]_\Gamma &= 0, && (3.5b)\\ [\epsilon\nabla\phi\cdot\boldsymbol{n}_\Gamma]_\Gamma &= 0, && (3.5c)\\ \phi &= g \quad \text{on } \partial\Omega. && (3.5d) \end{aligned}$$

The coefficient $\overline{k}$ is defined by

$$\overline{k}^2(x) = \begin{cases} 0, & x \in \Omega_m \cup \Omega_{IEL},\\ \overline{k}^2_{ions} = \dfrac{8\pi N_A e_0^2 I_s}{1000 k_BT}, & x \in \Omega_{ions}, \end{cases} \tag{3.6}$$

where $N_A = 6.022140857\times 10^{23}$ is Avogadro's number and the ionic strength I_s, measured in moles per liter (molar), is given by

$$I_s = \frac{1}{2}\sum_{j=1}^{2} c_i\xi_j^2 = \frac{1000M}{N_A}$$

with $c_1 = c_2 = \frac{1000M}{N_A}$, the average molar concentration of each ion (see [12, 101, 141]). Equation (3.5) is often referred to as the Poisson-Boltzmann equation [103, 143, 168]. When there are no ions present, $M_j = 0$, $j = 1, 2, \ldots, N_{ions}$, $I_s = 0$, $\overline{k}^2 = 0$ in Ω, $\Omega_{IEL} = \emptyset$, $\Omega_{ions} \equiv \Omega_s$, and equation (3.5a) becomes the linear Poisson equation of electrostatics.

$$\begin{aligned} -\nabla\cdot(\epsilon\nabla\phi) &= \mathcal{F} \quad \text{in } \Omega_m \cup \Omega_s, && (3.7a)\\ [\phi]_\Gamma &= 0, && (3.7b)\\ [\epsilon\nabla\phi\cdot\boldsymbol{n}_\Gamma]_\Gamma &= 0, && (3.7c)\\ \phi &= g \quad \text{on } \partial\Omega. && (3.7d) \end{aligned}$$

The Poisson-Boltzmann equation (3.5a) can be linearized by expanding sinh in Maclaurin series. We obtain the linearized Poisson-Boltzmann equation for the electrostatic potential ϕ

$$\begin{aligned} -\nabla\cdot(\epsilon\nabla\phi) + \overline{k}^2\phi &= \mathcal{F} \quad \text{in } \Omega, && (3.8a)\\ [\phi]_\Gamma &= 0, && (3.8b)\\ [\epsilon\nabla\phi\cdot\boldsymbol{n}_\Gamma]_\Gamma &= 0, && (3.8c)\\ \phi &= g \quad \text{on } \partial\Omega. && (3.8d) \end{aligned}$$

Remark 3.1
We note that the LPBE (3.8) *and the Poisson problem* (3.7) *are often given for the electrostatic*

potential φ, and not for the dimensionless potential ϕ in order to avoid the scaling with the factor $\frac{e_0}{k_B T}$. The LPBE for the potential φ with dimension $[\varphi] = \left[\frac{charge}{length}\right]$ reads

$$\begin{aligned} -\nabla \cdot (\epsilon \nabla \varphi) + \overline{k}^2 \varphi &= \mathscr{F} \quad \text{in } \Omega, && (3.9a)\\ [\varphi]_\Gamma &= 0, && (3.9b)\\ [\epsilon \nabla \varphi \cdot \boldsymbol{n}_\Gamma]_\Gamma &= 0, && (3.9c)\\ \varphi &= g \quad \text{on } \partial\Omega. && (3.9d) \end{aligned}$$

We will use this form of the LPBE in Chapter 5.

For the subsequent analysis of the PBE and the LPBE we will need the function G given by

$$G = \sum_{i=1}^{N_m} G_i = -\frac{2e_0^2}{\epsilon_m k_B T} \sum_{i=1}^{N_m} z_i \ln|x - x_i|, \text{ if } d = 2, \tag{3.10}$$

$$G = \sum_{i=1}^{N_m} G_i = \frac{e_0^2}{\epsilon_m k_B T} \sum_{i=1}^{N_m} \frac{z_i}{|x - x_i|}, \text{ if } d = 3. \tag{3.11}$$

The function G describes the Coulomb part of the potential due to the partial charges $\{z_i e_0\}_{i=1}^{N_m}$ in a uniform dielectric medium with a dielectric constant ϵ_m. It is well known that G is the distributional solution of the problem

$$-\nabla \cdot (\epsilon_m \nabla G) = \frac{4\pi e_0^2}{k_B T} \sum_{i=1}^{N_m} z_i \delta_{x_i} = \mathcal{F} \quad \text{in } \mathbb{R}^d,\ d \in \{2, 3\}. \tag{3.12}$$

What is meant by (3.12) is that (see, e.g. p.106 in [173])

$$-\int_{\mathbb{R}^d} \epsilon_m G \Delta v dx = \langle \mathcal{F}, v \rangle \quad \text{for all } v \in C_0^\infty(\mathbb{R}^d). \tag{3.13}$$

In particular, (3.13) is valid for all $v \in C_0^\infty(\Omega)$. The function G is weakly differentiable with a weak derivative equal almost everywhere to its classical derivative. Moreover, G and ∇G are in $L^p(\Omega)$ for all $p < \frac{d}{d-1}$ and thus $G \in \bigcap_{p < \frac{d}{d-1}} W^{1,p}(\Omega)$ (alternatively, we can use the AC characterization of $W^{1,p}(\Omega)$ functions - see Theorem 10.35 in [122]). Therefore, by applying the integration by parts formula (see Theorem 2.15), for any $v \in C_0^\infty(\Omega)$, we obtain

$$\int_\Omega \epsilon_m G \Delta v dx = \underbrace{\int_{\partial\Omega} \epsilon_m G \nabla v \cdot \boldsymbol{n}_{\partial\Omega} ds}_{=0} - \int_\Omega \epsilon_m \nabla G \cdot \nabla v dx$$

$$\int_\Omega \epsilon_m \nabla G \cdot \nabla v dx = \langle \mathcal{F}, v \rangle \quad \text{for all } v \in C_0^\infty(\Omega). \tag{3.14}$$

For a fixed $q > d$, owing to the Sobolev embedding $W_0^{1,q}(\Omega) \hookrightarrow C^{0,\lambda}(\overline{\Omega})$, $0 < \lambda \leq 1 - d/q$ (see Theorem 2.21), $\mathcal{F}$ is bounded in $W_0^{1,q}(\Omega)$:

$$|\langle \mathcal{F}, v\rangle| = \left| \frac{4\pi e_0^2}{k_B T} \sum_{i=1}^{N_m} z_i v(x_i) \right| \leq \frac{4\pi e_0^2}{k_B T} \sum_{i=1}^{N_m} |z_i| \|v\|_{L^\infty(\Omega)} \leq C_E \frac{4\pi e_0^2}{k_B T} \sum_{i=1}^{N_m} |z_i| \|v\|_{W^{1,q}(\Omega)},$$

where C_E is the constant in the inequality $\|v\|_{L^\infty(\Omega)} \leq C_E \|v\|_{W^{1,q}(\Omega)}$. Therefore, by the density of $C_0^\infty(\Omega)$ in $W_0^{1,q}(\Omega)$ we see that (3.14) is valid for all $v \in W_0^{1,q}(\Omega)$ and consequently for all $v \in \bigcup_{q>d} W_0^{1,q}(\Omega)$. Therefore, $G \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$ satisfies the weak formulation of (3.12), i.e.,

$$\int_\Omega \epsilon_m \nabla G \cdot \nabla v dx = \langle \mathcal{F}, v\rangle \quad \text{for all } v \in \bigcup_{q>d} W_0^{1,q}(\Omega). \tag{3.15}$$

Table 3.1: Some units expressed in Centimetre-Gram-Second (CGS) system of units. Here mol denotes the amount of chemical substance that contains exactly $6.02214076 \times 10^{23}$ (Avogadro's number) constitutive particles.

abbreviation	unit	represents	expression in CGS
g	gram	mass	g
cm	centimetre	length	cm
s	second	time	s
Å	angstrom	length	10^{-8} cm
l	liter	volume	1000 cm^3
M (molar)	moles per liter	concentration	mol/l
esu (statcoulomb)	electrostatic unit	electric charge	$\text{cm}^{3/2}\text{g}^{1/2}\text{s}^{-1}$
erg	erg	energy	g cm^2s^{-2}

Table 3.2: Some physical constants. Here K denotes Kelvin, a unit for temperature.

abbreviation	name	value in CGS derived units
N_A	Avogadro's number	$6.022140857 \times 10^{23}$
e_0	elementary charge	$4.8032424 \times 10^{-10}$ esu
k_B	Boltmann's constant	$1.38064852 \times 10^{-16}$erg K^{-1}

3.2 Linearized Poisson Boltzmann equation

Now, our goal is to give a meaningful notion of a solution to the linear problem (3.8) which ultimately will ensure uniqueness and can be also extended to the case of the nonlinear PBE. There are different ways to define a solution to a linear problem of the form

$$\begin{aligned} -\operatorname{div}(\boldsymbol{A}(x)\nabla u) &= f \ \text{ in } \Omega, \\ u &= 0 \ \text{ on } \partial\Omega, \end{aligned} \tag{3.16}$$

where Ω is a bounded Lipschitz domain, f is a bounded Radon measure, and the coefficient matrix $\boldsymbol{A}$ is such that $\boldsymbol{A} \in [L^\infty(\Omega)]^{d\times d}$ and there is a constant $\mu > 0$ for which

$$\boldsymbol{A}(x)\xi \cdot \xi \geq \mu \left|\xi\right|^2 \text{ for a.e. } x \in \Omega \text{ and all } \xi \in \mathbb{R}^d. \tag{3.17}$$

Here we mention two approaches. The first one is due to Stampacchia [174], where he introduced a notion of a solution to (3.16) defined by duality, and the second one is due to Boccardo and Gallouët [21], where they defined a distributional solution as the limit of solutions of (3.16) obtained for more regular data f.

The solution u defined by duality is unique and it can be shown that it satisfies the weak formulation

$$\begin{aligned} &u \in W_0^{1,q}(\Omega) \text{ for every } q < \frac{d}{d-1}, \\ &\int_\Omega \boldsymbol{A}\nabla u \cdot \nabla v dx = \int_\Omega v df, \ \forall v \in C_0^\infty(\Omega). \end{aligned} \tag{3.18}$$

Unfortunately, to show existence of a solution by the framework of Stampacchia is, in general, not possible for nonlinear equations since the notion of duality is unavailable.

On the other hand, solutions that are obtained by approximation in the approach of Boccardo and Gallouët also verify the weak formulation (3.18). Moreover, this approach can be extended for more general nonlinear elliptic problems. However, in dimension $d \geq 3$ and for a general diffusion coefficient matrix $\boldsymbol{A}$ which is in $L^\infty(\Omega)$ and which satisfies the uniform ellipticity condition (3.17), the weak formulation (3.18) does not ensure uniqueness as it is shown by a counterexample due to Serrin [153, 167]. Here we note that for $d = 2$, (3.18) always has a unique solution due to a regularity result of Meyers (see Theorem 2 in [85] and Theorem 4.1, Theorem 4.2 in [17]). The question of existence and uniqueness for more general linear and nonlinear elliptic problems involving measure data is studied for example in [10, 15, 22, 24, 33, 36, 56, 57, 63, 146, 150].

However, under some assumptions on the regularity of the coefficient matrix $\boldsymbol{A}$, one can still show the uniqueness of a weak solution to (3.16) by employing an adjoint problem with a more regular right-hand side. In such cases, the approach involving duality techniques and the one involving approximation techniques lead to one and the same solution and thus both approaches are equivalent. In particular, uniqueness can be shown if $\boldsymbol{A}^*$, the transposed of $\boldsymbol{A}$, satisfies the assumptions in Theorem 2.34. We will apply this technique in the proof of uniqueness in Theorem 3.4 in the next section.

The considerations above lead us to the following definition of a weak solution to the linear problem (3.8).

Definition 3.2
We call the measurable function ϕ a weak solution of (3.8) *if it satisfies*

$$\phi \in \bigcap_{p<\frac{d}{d-1}} W_g^{1,p}(\Omega), \quad \int_\Omega \epsilon \nabla\phi \cdot \nabla v dx + \int_\Omega \overline{k}^2 \phi v dx = \langle \mathcal{F}, v\rangle, \; \forall v \in \bigcup_{q>d} W_0^{1,q}(\Omega). \tag{3.19}$$

Remark 3.3
Note that we can define the weak formulation in Definition 3.2 with test functions in the much smaller space $C_0^\infty(\Omega)$ as in the weak formulation (3.18)*. Such a weak formulation is equivalent for linear problems to the one given above by applying a standard density argument.*

3.2.1 2-term splitting $\phi = G + u$

A commonly used technique (see, e.g. [48, 141, 195]) to solve and analyze the PBE is to split the solution according to $\phi = G + u$, where u is a well behaved regular component and G is defined by (3.10) in 2D or (3.11) in 3D. The function u describes the so-called reaction field potential. The reaction field is defined as the field that includes all forces acting on a biomolecule due to the presence of the solvent (see [141, 163]). By substituting the expression $\phi = G + u$ in (3.19) and by taking into account (3.15) we obtain the weak formulation that u has to satisfy:

$$\begin{aligned} &\text{Find } u \in \bigcap_{p<\frac{d}{d-1}} W_{g-G}^{1,p}(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla u \cdot \nabla v dx + \int_\Omega \overline{k}^2 uv dx = \int_{\Omega_s} (\epsilon_m - \epsilon_s)\nabla G \cdot \nabla v dx - \int_\Omega \overline{k}^2 G v dx \\ &=: \langle \mathcal{G}_2, v\rangle - \int_\Omega \overline{k}^2 G v dx \;\text{ for all } v \in \bigcup_{q>d} W_0^{1,q}(\Omega). \end{aligned} \tag{3.20}$$

Since the space $W^{1,2}(\Omega)$ lies between the spaces $\bigcup_{q>d} W^{1,q}(\Omega)$ and $\bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$ we can find a particular function u that satisfies (3.20) by posing a standard weak formulation for u that involves H^1 spaces. Once we have a u that solves (3.20) we need to show that it is indeed the only solution. This quick sketch of existence and uniqueness for (3.20) is summarized in the following theorem.

Theorem 3.4
There exists a weak solution ϕ of equation (3.8) *satisfying* (3.19)*. A particular ϕ satisfying* (3.19) *can be given in the form $\phi = G + u$, where $u \in H^1_{g-G}(\Omega) \cap H^2_{loc}(\Omega_m) \cap H^2_{loc}(\Omega_s)$ is the unique solution of the equation*

$$\int_\Omega \epsilon \nabla u \cdot \nabla v dx + \int_\Omega \overline{k}^2 uv dx = \langle \mathcal{G}_2, v\rangle - \int_\Omega \overline{k}^2 G v dx \;\; \textit{for all } v \in H_0^1(\Omega). \tag{3.21}$$

If we assume in addition that $\Gamma \in C^1$, then ϕ is unique (for example, when Γ is the Connolly surface it is often a C^1 surface).

Remark 3.5
The homogenized version of (3.21) *is given by:*

$$\begin{aligned}&\textit{Find } u_0 \in H_0^1(\Omega) \textit{ such that}\\ &\int_\Omega \epsilon\nabla u_0\cdot\nabla v dx + \int_\Omega \overline{k}^2 u_0 v dx = \langle \mathcal{G}_2, v\rangle - \int_\Omega \overline{k}^2 G v dx - \int_\Omega \epsilon\nabla u_{g-G}\cdot\nabla v dx - \int_\Omega \overline{k}^2 u_{g-G} v dx,\end{aligned} \tag{3.22}$$

where $u_{g-G}:\Omega\to\mathbb{R}$ *is in* $H^1(\Omega)$ *with* $\gamma_2(u_{g-G}) = g - G$ *on* $\partial\Omega$ *and* $u = u_{g-G} + u_0$. *The existence of* u_{g-G} *follows from the trace theorem (Theorem 2.17).*

Notice that $\overline{k} = 0$ *in* $\Omega_m \cup \Omega_{IEL}$ *and that the function* G *is smooth in the solvent region* $\overline{\Omega}_s$. *As a consequence, the integrals* $\int_{\Omega_s}(\epsilon_m - \epsilon_s)\nabla G\cdot\nabla v dx$ *and* $-\int_\Omega \overline{k}^2 G v dx$ *are well defined and each of them defines a bounded linear functional over* $H_0^1(\Omega)$. *The expression* $-\int_\Omega \epsilon\nabla u_{g-G}\cdot\nabla v dx - \int_\Omega \overline{k}^2 u_{g-G} v dx$ *also defines a functional in* $H^{-1}(\Omega)$ *and hence the whole right-hand side of* (3.22) *defines an element from* $H^{-1}(\Omega)$.

Remark 3.6
Notice that by applying the integration by parts formula, the term $\langle\mathcal{G}_2, v\rangle$ *on the right-hand side of* (3.21) *can be rewritten in the form*

$$\begin{aligned}\int_{\Omega_s}(\epsilon_m-\epsilon_s)\nabla G\cdot\nabla v dx &= -\int_\Gamma (\epsilon_m-\epsilon_s)\nabla G\cdot\boldsymbol{n}_\Gamma v ds + \int_{\partial\Omega}(\epsilon_m-\epsilon_s)\nabla G\cdot\boldsymbol{n}_{\partial\Omega} v ds\\ &\quad - \int_{\Omega_s}\left(\nabla(-\epsilon_s)\cdot\nabla G + (\epsilon_m-\epsilon_s)\Delta G\right) v dx\\ &= -\int_\Gamma (\epsilon_m-\epsilon_s)\nabla G\cdot\boldsymbol{n}_\Gamma v ds + \int_{\Omega_s}\nabla\epsilon_s\cdot\nabla G v dx\end{aligned} \tag{3.23}$$

where we have used the facts that ϵ_m *is constant in* Ω_m, $\epsilon_s \in C^{0,1}(\overline{\Omega}_s)$ *(and therefore* $\epsilon_s \in W^{1,\infty}(\Omega_s)$*), that* G *is harmonic in a neighborhood of* Ω_s, *and that* $\Gamma, \partial\Omega \in C^{0,1}$. *Now, it is seen that* (3.21) *is the weak formulation of a linear interface elliptic problem with a jump condition on the normal flux* $[\epsilon\nabla u\cdot\boldsymbol{n}_\Gamma]_\Gamma = -(\epsilon_m-\epsilon_s)\nabla G\cdot\boldsymbol{n}_\Gamma = -[\epsilon\nabla G\cdot\boldsymbol{n}_\Gamma]_\Gamma$.

Remark 3.7
To see that (3.23) *indeed holds, we first observe that since* $G \in C^\infty(\overline{\Omega}_s)$, *it holds for smooth functions* $\epsilon_s \in C^\infty(\overline{\Omega}_s)$ *and* $v \in C_0^\infty(\Omega)$ *by the classical divergence theorem. Next, if* $\epsilon_s \in C^{0,1}(\overline{\Omega}_s)$, *it follows that* $\epsilon_s \in W^{1,\infty}(\Omega_s)$ *(see, e.g. Exercise 11.46 in [122]). Since* Ω_s *is a bounded Lipschitz domain, by Theorem 2.11, for any* $1 \le p < \infty$ *there exists a sequence* $\{\epsilon_s^n\}_{n=1}^\infty \subset C^\infty(\overline{\Omega}_s)$ *such that* $\epsilon_s^n \to \epsilon_s$ *in* $W^{1,p}(\Omega_s)$. *Equation* (3.23) *is satisfied for each* ϵ_s^n. *By letting* n *to infinity, and by using Hölder's inequality together with the trace theorem*

in $W^{1,p}(\Omega_s)$ *for* ϵ_s, ϵ_s^n, *we see that* (3.23) *is also satisfied for* $\epsilon_s \in W^{1,\infty}(\Omega_s)$. *Finally, if* $v \in H_0^1(\Omega)$ *is an arbitrary test function and* $\{v_n\}_{n=1}^{\infty} \subset C_0^{\infty}(\Omega)$ *such that* $v_n \to v \in H_0^1(\Omega)$, *then for all* $n \in \mathbb{N}$

$$\int_{\Omega_s} (\epsilon_m - \epsilon_s)\nabla G \cdot \nabla v_n dx = -\int_{\Gamma} (\epsilon_m - \epsilon_s)\nabla G \cdot \boldsymbol{n}_{\Gamma} v_n ds + \int_{\Omega_s} \nabla \epsilon_s \cdot \nabla G v_n dx.$$

Again, by letting n *to infinity and using Hölder's inequality together with the trace theorem in* $H^1(\Omega_s)$ *for* v, v_n, *we see that* (3.23) *is satisfied for any* $\epsilon_s \in C^{0,1}(\overline{\Omega}_s)$ *and* $v \in H_0^1(\Omega)$.

Proof of Theorem 3.4.
Existence: Let $u \in H^1_{g-G}(\Omega)$ be the unique weak solution of problem (3.21), given by the Lax-Milgram Theorem applied to the homogeneous problem (3.22) (see Remark 3.5). It is clear that u is also in $\bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$ since $p < \frac{d}{d-1} \leq 2$ and $W^{1,2}(\Omega) \equiv H^1(\Omega)$. Thus, for each $p < \frac{d}{d-1}$, u is in $W^{1,p}(\Omega)$ with a trace on $\partial\Omega$ equal to $g - G$. We conclude that $G + u \in \bigcap_{p<\frac{d}{d-1}} W_g^{1,p}(\Omega)$. Since $\nabla G \in L^{\infty}(\Omega_s)$ and for each $q > d$ we have $H_0^1(\Omega) \supset W_0^{1,q}(\Omega)$, (3.21) is also valid for all $v \in \bigcup_{q>d} W_0^{1,q}(\Omega)$. By adding together (3.15) and (3.21) we conclude that ϕ satisfies the weak formulation (3.19). Finally, by testing (3.21) with $v \in C_0^{\infty}(\Omega_m)$, $v \in C_0^{\infty}(\Omega_s)$ and taking into account (3.23) and the facts that $G \in C^{\infty}(\overline{\Omega}_s)$, $\epsilon_s \in C^{0,1}(\overline{\Omega}_s)$, we see that $u \in H^2_{\text{loc}}(\Omega_m) \cap H^2_{\text{loc}}(\Omega_s)$. Furthermore, testing with $v \in C_0^{\infty}(\Omega_m)$ shows that $u \in H^t_{\text{loc}}(\Omega_m)$ for all $t \geq 2$ (see p. 309 in [70] for interior regularity of elliptic problems).

If in addition $\epsilon_s \in C^{\infty}(\overline{\Omega}_s)$, then by testing with $v \in C_0^{\infty}(\Omega_{ions})$ and taking into account (3.23), we see that $u \in H^t_{\text{loc}}(\Omega_m) \cap H^t_{\text{loc}}(\Omega_{IEL}) \cap H^t_{\text{loc}}(\Omega_{ions})$ for all integer $t \geq 2$.

Uniqueness: The idea to prove the uniqueness of ϕ is from [63] where the authors show uniqueness of a linear problem with constant diffusion coefficient ϵ. The difference is that the type of regularity result for the adjoint problem used in [63] is not applicable in the case of discontinuous ϵ. Instead we apply the regularity result from Theorem 2.34. It is enough to show that if ϕ satisfies the homogeneous problem (3.19) with $\mathcal{F} = 0$ then $\phi = 0$. For a fixed $\theta \in L^{\infty}(\Omega)$, we consider the auxiliary problem

$$\begin{aligned} &\text{Find } w \in H_0^1(\Omega) \text{ such that} \\ &\int_{\Omega} \epsilon \nabla w \cdot \nabla v dx + \int_{\Omega} \overline{k}^2 w v dx = \int_{\Omega} \theta v dx, \ \forall v \in H_0^1(\Omega). \end{aligned} \tag{3.24}$$

By the Lax-Milgram Theorem, this problem has a unique solution $w \in H_0^1(\Omega)$. In view of the Sobolev embedding theorem, for $d = 3$, $H^1(\Omega) \hookrightarrow L^6(\Omega)$ and for $d = 2$, $H^1(\Omega) \hookrightarrow L^r(\Omega_m)$, $\forall r : 1 \leq r < \infty$. Therefore, $w \in L^6(\Omega)$ and consequently $(-\overline{k}^2 w + \theta) \in L^6(\Omega)$. Since $\int_{\Omega} \left(-\overline{k}^2 w + \theta\right) v dx$ defines a bounded linear functional in $W^{-1,p'}$ for all $\frac{6}{5} \leq p < \frac{d}{d-1}$ and

since $\Gamma \in C^1$, by applying Theorem 2.34, we see that $w \in W_0^{1,q_0}(\Omega)$ for some $q_0 \in (d, 6]$. By a density argument we see that (3.24) holds for all test functions $v \in W_0^{1,q_0'}(\Omega)$ with $1/q_0 + 1/q_0' = 1$. Thus, we can use w as a test function in (3.19) ($\mathcal{F} = 0$) and ϕ as a test function in (3.24). In this way we obtain

$$0 = \int_\Omega \epsilon \nabla w \cdot \nabla \phi dx + \int_\Omega \overline{k}^2 w \phi dx = \int_\Omega \theta \phi dx. \tag{3.25}$$

Since θ was an arbitrary function in $L^\infty(\Omega)$, it follows that $\phi = 0$ a.e. in Ω. □

From Theorem 3.4 it is seen that if the potential ϕ is needed, one can avoid the numerical approximation of the full potential ϕ which has singularities at the fixed partial charges $\{z_i e_0\}_{i=1}^{N_m}$. Instead, it is enough to find an approximation u_h only to the much better behaved reaction field potential u by numerically solving (3.21). This idea sounds very appealing in theory, but it is not always appropriate to use in practice. The reason is that if the component u happens to have almost the same magnitude as G but opposite sign, then adding up both components gives a number which is much smaller in magnitude compared to u, i.e., we have $|\phi| = |G + u| \ll |u|$ (see, e.g. [103]). This typically happens in the solvent region Ω_s and under the conditions that the ratio $\frac{\epsilon_m}{\epsilon_s}$ is much smaller than 1 and that the ionic strength I_s is nonzero. In this case a small relative error in u is already a substantial relative error in $\phi = G + u$. Here, we summarize the cases, in which the 2-term splitting can be recommended:

- only the reaction field u is needed, for example, when calculating the solvation energy difference

$$\Delta G^{\text{vac}\to\text{solv}} = G^{\text{vac}}(\epsilon_{\text{vac}}, \epsilon_m) - G^{\text{solv}}(\epsilon_s, \epsilon_m) = \frac{1}{2}\sum_{i=1}^{N} q_i(u_{\text{vac}}(x_i) - u_{\text{solv}}(x_i)),$$

  where $q_i = z_i e_0$, u_{vac} and u_{solv} are the reaction field potentials for the molecule in vacuum and in solvent, respectively;

- the ratio $\frac{\epsilon_m}{\epsilon_s}$ is close to one and I_s is close to zero ($\overline{k}$ is close to zero);

3.2.2 3-term splitting $\phi = G + u^H + u$

In order to overcome the difficulty connected with the 2-term splitting, one may use a splitting of ϕ into 3 components, two of which add up to zero in Ω_s. Such a splitting is given by

$$\phi = G + u^H + u, \tag{3.26}$$

where $\phi = u$ in Ω_s, i.e. $u^H = -G$ in Ω_s, and has been also used in [51, 103]. By substituting the expression $\phi = G + u^H + u$ in (3.19) and using (3.15) and the fact that $u^H = -G$ in Ω_s,

we obtain the weak formulation that u has to satisfy:

$$\begin{aligned}&\text{Find } u \in \bigcap_{p<\frac{d}{d-1}} W_g^{1,p}(\Omega) \text{ such that}\\ &\int_\Omega \epsilon\nabla u\cdot\nabla v dx + \int_\Omega \overline{k}^2 uv dx = -\int_{\Omega_m}\epsilon_m\nabla u^H\cdot\nabla v dx + \int_{\Omega_s}\epsilon_m\nabla G\cdot\nabla v dx =: \langle\mathcal{G}_3, v\rangle \qquad (3.27)\\ &\text{for all } v\in\bigcup_{q>d} W_0^{1,q}(\Omega).\end{aligned}$$

Notice that if $u \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$, then since $\phi \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$, it follows that u^H is also in the space $\bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$ and therefore the integral $\int_{\Omega_m}\epsilon_m\nabla u^H\cdot\nabla v dx$ makes sense. Also notice that we still have not defined u^H in Ω_m ($u^H = -G$ in Ω_s). The only condition that has to be satisfied so far is that $u^H + u \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}_{g-G}(\Omega)$. Thus, if we define u^H in Ω_m such that $u^H \in H^1(\Omega)$ it will hold $u^H + u \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}_{g-G}(\Omega)$. Again, since we have

$$\bigcup_{q>d} W^{1,q}(\Omega) \subset W^{1,2}(\Omega) \subset \bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$$

we will find a particular solution u of problem (3.27) if we consider the standard weak formulation

$$\begin{aligned}&\text{Find } u\in H^1_g(\Omega) \text{ such that}\\ &\int_\Omega \epsilon\nabla u\cdot\nabla v dx + \int_\Omega \overline{k}^2 uv dx = \langle\mathcal{G}_3, v\rangle \text{ for all } v\in H^1_0(\Omega),\end{aligned} \qquad (3.28)$$

where $\mathcal{G}_3$ is a well defined functional in $H^{-1}(\Omega)$ since we have chosen u^H to be in $H^1(\Omega)$. From the Lax-Milgram Theorem (see Remark 3.5) it follows that problem (3.28) has a unique solution $u \in H^1_g(\Omega)$ for any fixed $u^H \in H^1(\Omega)$. Also notice that if we test (3.28) with $v \in H^1_0(\Omega_m)$ we obtain

$$\int_{\Omega_m}\epsilon_m\nabla u\cdot\nabla v dx = -\int_{\Omega_m}\epsilon_m\nabla u^H\cdot\nabla v dx \text{ for all } v\in H^1_0(\Omega_m).$$

It is convenient for the a posteriori analysis in Section 4.3 to define u^H to be harmonic in Ω_m, i.e.,

$$u^H \in H^1_{-G}(\Omega_m) \quad \text{and} \quad \int_{\Omega_m}\nabla u^H\cdot\nabla v dx = 0 \text{ for all } v\in H^1_0(\Omega_m), \qquad (3.29)$$

where the Dirichlet boundary condition $u^H = -G$ on $\partial\Omega$ ensures that u^H has the same trace on Γ from both sides and therefore $u^H \in H^1(\Omega)$.

Remark 3.8
By means of the 3-term splitting we have obtained one particular solution

$$\phi = G + u^H + u \in \bigcap_{p<\frac{d}{d-1}} W_g^{1,p}(\Omega)$$

of (3.19)*. By Theorem 3.4 it is the unique solution of* (3.19)*.*

Remark 3.9
Notice that the right-hand side of equation (3.28) *depends on the solution of* (3.29)*. Therefore, in practice we first have to find an approximation* $\tilde{u}^H$ *to* u^H *in* Ω_m *and then numerically solve* (3.28) *with* $\tilde{u}^H$ *in it. We will discuss this in detail in Chapter 4. Also, note that* $\mathcal{G}_3$ *represents a jump condition on the normal component of* $\epsilon\nabla u$*. Indeed, by applying the divergence theorem (Theorem 2.20) we obtain*

$$\langle \mathcal{G}_3, v\rangle = -\int_{\Omega_m} \epsilon_m \nabla u^H \cdot \nabla v dx + \int_{\Omega_s} \epsilon_m \nabla G \cdot \nabla v dx \tag{3.30}$$

$$= -\langle \gamma_{\boldsymbol{n}_\Gamma,\Omega_m}\left(\epsilon_m \nabla u^H\right), \gamma_{2,\Gamma}(v)\rangle_{H^{-1/2}(\Gamma)\times H^{1/2}(\Gamma)} + \langle \gamma_{\boldsymbol{n}_\Gamma,\Omega_s}\left(\epsilon_m \nabla G\right), \gamma_{2,\Gamma}(v)\rangle_{H^{-1/2}(\Gamma)\times H^{1/2}(\Gamma)},$$

where we have used the facts that ϵ_m *is constant,* $\nabla u^H \in H(\mathrm{div};\Omega_m)$ *(see* (3.29)*), and that* G *is harmonic in a neighborhood of* Ω_s*. In* (3.30)*,* $\gamma_{\boldsymbol{n}_\Gamma,\Omega_m}$ *and* $\gamma_{\boldsymbol{n}_\Gamma,\Omega_s}$ *are the normal trace operators in* $H(\mathrm{div};\Omega_m)$ *and* $H(\mathrm{div};\Omega_s)$*, respectively, and* $\gamma_{2,\Gamma}(v)$ *is the trace of* v *on* Γ*. If* ∇u^H *is more regular, then by using the integration by parts formula (Theorem 2.15)* (3.30) *can be rewritten in terms of surface integrals over* Γ*, i.e.,*

$$\langle \mathcal{G}_3, v\rangle = \int_\Gamma -\epsilon_m \nabla\left(u^H + G\right)\cdot \boldsymbol{n}_\Gamma v ds,\ \forall v \in H_0^1(\Omega). \tag{3.31}$$

This means that if the function u *is smooth in each subdomain, it should satisfy the jump condition* $[\epsilon\nabla u \cdot \boldsymbol{n}_\Gamma]_\Gamma = -\epsilon_m \nabla\left(u^H + G\right)\cdot \boldsymbol{n}_\Gamma$*.*

3.2.3 Regularity of the component u in the 2-term and 3-term splittings

First, using Theorem 2.32, we will show the boundedness of the regular component u in both splittings without the assumption that $\Gamma \in C^1$. Then, using Theorem 2.34 together with the assumption that $\Gamma \in C^1$, we will show that u is actually in $W^{1,q}(\Omega)$ for some $q > d$, which implies that u is also Hölder continuous. The fact that u is Hölder continuous also follows from the regularity results of De Giorgi-Nash-Moser. In the case of the 2-term splitting no additional assumption on the smoothness of the interface Γ is needed, and in the case of the 3-term splitting we have to assume only $\Gamma \in C^{0,1}$.

2-term splitting
In the case of the 2-term splitting, where $\phi = G + u$, the regular component $u = u_{g-G} + u_0 \in H^1_{g-G}(\Omega)$ satisfies problem 3.21 and $u_0 \in H_0^1(\Omega)$ satisfies the homogenized problem (3.22).

Before we continue, we note that u_{g-G} can be chosen in $H^2(\Omega)$ if the function g, prescribing the Dirichlet boundary condition, is given as the trace of a function $\tilde{g} \in H^2(\Omega)$. Indeed, we can construct $\tilde{G} \in H^2(\Omega)$ such that $\gamma_2(\tilde{G}) = G$ on $\partial\Omega$: just take $\tilde{G} := (\psi G)_{|\overline{\Omega}} \in C^\infty(\overline{\Omega})$, where $\psi \in C_0^\infty(\mathbb{R}^d)$ is such that it is equal to 1 in a neighborhood of $\partial\Omega$ and with support in $\mathbb{R}^d \setminus \overline{\Omega}_m$. For the function ψ one can mollify the characteristic function of the set $(\partial\Omega)^{+\delta} := \{x \in \mathbb{R}^d : \text{dist}(x, \partial\Omega) < \delta\}$ with a mollifier η_ρ for $\rho < \delta/2$, where $\delta < \frac{1}{2}\text{dist}(\Gamma, \partial\Omega)$. Consequently, $\tilde{g} - \tilde{G} \in H^2(\Omega)$, $\gamma_2(\tilde{g} - \tilde{G}) = g - G$ and we define $u_{g-G} := \tilde{g} - \tilde{G}$. By using the fact that $u = u_{g-G} + u_0$, equation (3.22) can be rewritten in the form

$$\int_\Omega \epsilon \nabla u_0 \cdot \nabla v dx = -\int_\Omega \overline{k}^2 (G + u) v dx + \int_\Omega \chi_{\Omega_s}(\epsilon_m - \epsilon_s) \nabla G \cdot \nabla v dx - \int_\Omega \epsilon \nabla u_{g-G} \cdot \nabla v dx, \tag{3.32}$$

Now, since $u \in H^1(\Omega)$, by the Sobolev emebedding theorem it follows that $u \in L^6(\Omega)$ for $d = 2, 3$. Since $\overline{k}^2 \in L^\infty(\Omega)$, $\overline{k} = 0$ in $\Omega_m \cup \Omega_{IEL}$, G is smooth in $\overline{\Omega}_s$ and $u \in L^6(\Omega)$ it follows that $\overline{k}^2(G + u) \in L^6(\Omega)$. We also have that $\chi_{\Omega_s}(\epsilon_m - \epsilon_s)\nabla G \in [L^\infty(\Omega)]^d$ and that $\epsilon \nabla u_{g-G} \in \left[L^6(\Omega)\right]^d$. Now, from Theorem 2.32 it follows that $u_0 \in L^\infty(\Omega)$. Since $u_{g-G} \in H^2(\Omega) \subset L^\infty(\Omega)$ it follows $u \in L^\infty(\Omega)$.

To show that $u \in W^{1,q}(\Omega)$ for some $q > d$, observe that the right-hand side of (3.32) defines a bounded linear functional over $W_0^{1,p}(\Omega)$ for all $\frac{6}{5} \leq p < \frac{d}{d-1}$. If we assume that $\Gamma \in C^1$, then from Theorem 2.34 it follows that $u_0 \in W^{1,\overline{q}}(\Omega)$ for some $d < \overline{q} \leq 6$ ($\overline{q}$ is at most 6, since the right-hand side of (3.32) defines a bounded linear functional over $W_0^{1,p}(\Omega)$ for p which is at least $\frac{6}{5}$). Finally, since $u_{g-G} \in W^{1,\overline{q}}(\Omega)$ (by the Sobolev embedding Theorem 2.21), it follows that $u \in W^{1,\overline{q}}(\Omega)$ and hence u is Hölder continuous (again by Theorem 2.21). The Hölder continuity of u_0 and u also follows from the regularity results of De Giorgi-Nash-Moser which hold for any bounded and measurable coefficient ϵ (see, e.g., Theorem 2.12 in [107], p. 65, Theorem 3.5 in [50]). Thus, for the Hölder continuity of u_0 and u the assumption that Γ is C^1 is not needed.

3-term splitting

In the case of the 3-term splitting, where $\phi = G + u^H + u$, the component $u \in H_g^1(\Omega)$ satisfies problem (3.28). Similarly to the case of the 2-term splitting, we can rewrite it in a homogenized form, i.e.,

Find $u_0 \in H_0^1(\Omega)$ such that

$$\begin{aligned} &\int_\Omega \epsilon \nabla u_0 \cdot \nabla v dx + \int_\Omega \overline{k}^2 u_0 v dx \\ &= -\int_{\Omega_m} \epsilon_m \nabla u^H \cdot \nabla v dx + \int_{\Omega_s} \epsilon_m \nabla G \cdot \nabla v dx - \int_\Omega \epsilon \nabla u_g \cdot \nabla v dx - \int_\Omega \overline{k}^2 u_g v dx \text{ for all } v \in H_0^1(\Omega), \end{aligned} \tag{3.33}$$

where $u_g \in H^1(\Omega)$ with $\gamma_2(u_g) = g$ on $\partial\Omega$ and $u = u_g + u_0$. If g is given as the trace of a function $\tilde{g} \in H^2(\Omega)$, then we define $u_g := \tilde{g}$ and in this case, by the Sobolev embedding Theorem 2.21, $\nabla u_g \in L^6(\Omega)$ for $d = 2, 3$. In order to apply Theorem 2.32 to (3.33), we need to ensure that $\nabla u^H \in [L^s(\Omega_m]^d$ for some $s > d$. This follows by applying Theorem 2.34 to the homogenized version of (3.29) (here it is enough if Γ is only Lipschitz continuous):

$$\begin{aligned} &\text{Find } u_0^H \in H_0^1(\Omega_m) \text{ such that} \\ &\int_{\Omega_m} \nabla u_0^H \cdot \nabla v dx = -\int_{\Omega_m} \nabla u_{-G}^H \cdot \nabla v dx \text{ for all } v \in H_0^1(\Omega_m), \end{aligned} \tag{3.34}$$

where $u_{-G}^H \in H^1(\Omega_m)$ and $\gamma_2\left(u_{-G}^H\right) = -G$ on Γ. Again, we can choose u_{-G}^H to be in $H^2(\Omega_m)$. To see this, let $r > 0$ be so small that all balls $B(x_i, r)$ centered at x_i, $i = 1, \ldots, N_m$ and with radius r are strictly contained in Ω_m. Then, we define the function $u_{-G}^H := (\psi G)_{|\overline{\Omega}_m} \in C^\infty(\overline{\Omega_m})$, where $\psi \in C_0^\infty(\mathbb{R}^d)$ is such that it is equal to 1 in a neighborhood of Γ and with support in $\mathbb{R}^d \setminus \bigcup_{i=1}^{N_m} B(x_i, r)$. It follows that the right-hand side of (3.34) defines a bounded linear functional over $W_0^{1,p}(\Omega_m)$ for all $1 \leq p < \infty$ and by Theorem 2.34 it follows that $u_0^H \in W^{1,\overline{q}}(\Omega_m)$ for some $\overline{q} > d$. Now, $u^H = u_{-G}^H + u_0^H \in W^{1,\overline{q}}(\Omega_m)$. By recalling that $u_0 \in H^1(\Omega) \subset L^6(\Omega)$, $\nabla u_g \in \left[L^6(\Omega)\right]^d$, $u_g \in L^\infty(\Omega)$, $\nabla G \in [L^\infty(\Omega)]^d$, we can apply Theorem 2.32 to (3.33) to obtain that $u_0 \in L^\infty(\Omega)$. Since $u_g \in H^2(\Omega) \subset L^\infty(\Omega)$ it follows $u \in L^\infty(\Omega)$.

If we additionally assume that $\Gamma \in C^1$, by recalling that $u^H \in W^{1,\overline{q}}(\Omega_m)$ and applying Theorem 2.34 to (3.33), we obtain that $u_0 \in W_0^{1,\overline{\overline{q}}}(\Omega)$ for some $\overline{\overline{q}} \in (d, \min\{\overline{q}, 6\}]$. As a consequence, $u = u_g + u_0$ is also in $W_0^{1,\overline{\overline{q}}}(\Omega)$ and thus Hölder continuous. As in the 2-term splitting, the Hölder continuity of u_0 and u also follows from the regularity results of De Giorgi-Nash-Moser under the assumption that $\nabla u^H \in [L^s(\Omega_m)]^d$ for some $s > d$. As we showed above, the latter is ensured if $\Gamma \in C^{0,1}$. We can summarize the above results in the following theorem.

Theorem 3.10
Assume that the function g prescribing the Dirichlet boundary condition on $\partial\Omega$ is given as the trace of some function $\tilde{g} \in H^2(\Omega)$. The following statements hold true:

(i) the unique $u \in H^1(\Omega)$ in the 2-term splitting defined by the standard weak formulation (3.21) is Hölder continuous in $\overline{\Omega}$ and thus belongs to $L^\infty(\Omega)$;

(ii) if $\Gamma \in C^{0,1}$, then the unique $u \in H^1(\Omega)$ in the 3-term splitting, defined by the standard weak formulation (3.28) is also Hölder continuous in $\overline{\Omega}$ and thus belongs to $L^\infty(\Omega)$;

(iii) if we assume additionally that $\Gamma \in C^1$, then for both 2-term and 3-term splittings u is in $W^{1,q}(\Omega)$ for some $q > d$ and hence it is also Hölder continuous;

Remark 3.11
Note that when Γ *is not* C^1 *the uniqueness of the potential* ϕ *is not clear. Therefore, in Theorem 3.10, when* Γ *is not assumed* C^1*, we specify that we are talking about the particular* u *given by* (3.21) *or* (3.28)*. However, when* $\Gamma \in C^1$*, then* u *is unique for all weak formulations: for* (3.20) *and* (3.21)*, and for* (3.27) *and* (3.28) *if* u^H *is fixed by* (3.29)*.*

3.3 Poisson-Boltzmann equation

In this section we will consider the general problem where N_{ions} different ion species are present in the solvent, each one with valency ξ_j and average number density M_j, $j = 1, 2, \dots N_{ions}$. From (3.3) it follows that the nonlinear interface problem for the dimensionless potential ϕ, related to the unknown electrostatic potential φ via $\varphi = \frac{e_0\phi}{k_BT}$, reads

$$-\nabla \cdot (\epsilon \nabla \phi) - \chi_{\Omega_{ions}} \frac{4\pi e_0^2}{k_B T} \sum_{j=1}^{N_{ions}} M_j \xi_j \mathrm{e}^{-\xi_j \phi} = \frac{4\pi e_0^2}{k_B T} \sum_{i=1}^{N_m} z_i \delta_{x_i} = \mathcal{F} \quad \text{in } \Omega, \tag{3.35a}$$

$$[\phi]_\Gamma = 0, \tag{3.35b}$$

$$[\epsilon \nabla \phi \cdot \boldsymbol{n}_\Gamma]_\Gamma = 0, \tag{3.35c}$$

$$\phi = g \quad \text{on } \partial\Omega, \tag{3.35d}$$

We will refer to (3.35) as the general PBE. For convenience, we will denote $\chi_{\Omega_{ions}} M_j$ by $\overline{M}_j(x)$, where obviously $\overline{M}_j(x) = 0$ if $x \in \Omega_m \cup \Omega_{IEL}$ and $\overline{M}_j(x) = M_j$ if $x \in \Omega_{ions}$. Note that if the condition

$$\sum_{j=1}^{N_{ions}} M_j \xi_j = 0 \tag{3.36}$$

holds, the solvent is electroneutral and we refer to this as the charge neutrality condition. We further denote by $a(\cdot, \cdot)$ the bilinear form defined by $a(u, v) = \int_\Omega \epsilon \nabla u \cdot \nabla v dx$ for all $u, v \in H^1(\Omega)$, by $b(x, \cdot)$ the nonlinearity in (3.35), i.e.,

$$b(x, s) := -\frac{4\pi e_0^2}{k_B T} \sum_{j=1}^{N_{ions}} \overline{M}_j(x) \xi_j \mathrm{e}^{-\xi_j s}, \ \forall x \in \Omega, \ \forall s \in \mathbb{R} \tag{3.37}$$

and by $B(x, \cdot)$ an antiderivative of it given by

$$B(x, s) := \frac{4\pi e_0^2}{k_B T} \sum_{j=1}^{N_{ions}} \overline{M}_j(x) \mathrm{e}^{-\xi_j s} \geq 0, \ \forall x \in \Omega, \ \forall s \in \mathbb{R}. \tag{3.38}$$

Since $\frac{d}{ds} b(x, s) \geq 0$ for every $x \in \Omega$ it follows that the nonlinearity $b(x, \cdot)$ is monotone increasing. This in particular implies that

$$(b(x, s_1) - b(x, s_2))(s_1 - s_2) \geq 0, \ \forall s_1, s_2 \in \mathbb{R}, \ \forall x \in \Omega. \tag{3.39}$$

Note that $b(x,0) = -\frac{4\pi e_0^2}{k_B T}\sum_{j=1}^{N_{ions}} \overline{M}_j(x)\xi_j$ and therefore when (3.36) is satisfied we have $b(x,0) = 0, \forall x \in \Omega$. Also note that in the special case when $M_1 = M_2 = M$ and $\xi_j = (-1)^j$, $j = 1, 2$, equation (3.35) becomes the Poisson-Boltzmann equation (3.5). In this case, $b(x,s) = \overline{k}^2(x)\sinh(s)$ and $B(x,s) = \overline{k}^2(x)\cosh(s)$.

A natural way to extend the weak formulation (3.19) that we defined for the LPBE and for which we proved existence and uniqueness, is as follows.

Definition 3.12
We call ϕ a weak solution of problem (3.35) if $\phi \in \bigcap_{p<\frac{d}{d-1}} W_g^{1,p}(\Omega)$ is such that $b(x,\phi)v \in L^1(\Omega)$ for all $v \in \bigcup_{q>d} W_0^{1,q}(\Omega)$ and

$$\int_\Omega \epsilon\nabla\phi\cdot\nabla v dx + \int_\Omega b(x,\phi)v dx = \langle \mathcal{F}, v\rangle,\ \forall v \in \bigcup_{q>d} W_0^{1,q}(\Omega) \tag{3.40}$$

The question of existence and nonexistence of a solution to nonlinear elliptic equations with measure data has been studied for example in [15, 21, 23, 33, 36]. In [33] it is shown that even the simple equation $-\Delta u + |u|^{p-1}u = \delta_a$ with $u = 0$ on $\partial\Omega$ and $a \in \Omega$ does not have a solution in $L^p_{loc}(\Omega)$ for any $p \geq \frac{d}{d-2}$ when $d \geq 3$. We will show the existence of a solution to (3.40) by finding a ϕ which has a particular form. More precisely, we will find a particular solution by means of the 2-term and 3-term splittings of ϕ as we did in the case of the LPBE. If ϕ is unique, then the particular ϕ found through 2-term and 3-term splitting is the same. The reason for the existence of a solution of (3.40), despite the fact that the nonlinearity is of exponential type and much worse than $|u|^{p-1}$, is that $b(x,s) = 0$, $\forall x \in \Omega_m$, which is where all the delta functions in $\mathcal{F}$ are positioned. However, unlike in the case of the LPBE, we are not able to show uniqueness of ϕ.

Now we give a somehow detailed overview of the concepts and specific features of our problem that will be used in this section. To show existence of a solution we employ the 2-term or the 3-term splitting. Similarly to the case of the LPBE, it will be seen that a particular representative of the component u, in the case of both 2-term and 3-term splittings, can be found by solving a well posed weak formulation involving H^1 solution space and a test space V. We will see that for both 2-term and 3-term splittings, the weak formulation defining a particular u can be written in the general form

$$\begin{aligned} &\text{Find } u \in H^1_{\overline{g}}(\Omega) \text{ such that } b(x,u+w)v \in L^1(\Omega) \text{ for all } v \in V \text{ and} \\ &a(u,v) + \int_\Omega b(x,u+w)v dx = \int_\Omega \boldsymbol{f}\cdot\nabla v dx \text{ for all } v \in V, \end{aligned} \tag{3.41}$$

where $w \in L^\infty(\Omega_{ions})$, $\boldsymbol{f} = (f_1, f_2, \ldots, f_d) \in [L^s(\Omega)]^d$ with $s > d$, and $\overline{g}$ specifies a Dirichlet

boundary condition on $\partial\Omega$. In the case of the 2-term splitting we have

$$w = G, \quad \boldsymbol{f} = \chi_{\Omega_s}(\epsilon_m - \epsilon_s)\nabla G, \quad \text{and} \quad \overline{g} = g - G \text{ on } \partial\Omega \tag{3.42}$$

whereas in the case of the 3-term splitting we have

$$w = 0, \quad \boldsymbol{f} = -\chi_{\Omega_m}\epsilon_m \nabla u^H + \chi_{\Omega_s}\epsilon_m \nabla G, \quad \text{and} \quad \overline{g} = g \text{ on } \partial\Omega. \tag{3.43}$$

In (3.41), one can choose the test space V to be $H_0^1(\Omega)$, or $H_0^1(\Omega) \cap L^\infty(\Omega)$, or even $C_0^\infty(\Omega)$. A priori, all the weak formulations for the three choices of the test space V are reasonable to consider in the sense that if the function u is regular enough, for example $u \in H^1(\Omega) \cap C^0(\overline{\Omega}) \cap C^1(\overline{\Omega}_m) \cap C^1(\overline{\Omega}_s) \cap C^2(\Omega_m) \cap C^2(\Omega_s)$, then it can be shown that u also satisfies the strong form of the elliptic interface problem (3.41). Such weak formulations, with test space which is only a dense subspace of $H_0^1(\Omega)$, for very general nonlinear elliptic equations but with a sign and/or growth condition on the nonlinearity with respect to u are studied for example in [14, 16, 19, 20, 25, 26, 34, 38–41, 72, 99, 120, 121, 151, 178, 186]. However, the growth conditions with respect to the second argument of b, as well as the sign condition $b(x, s)s \geq 0$, $\forall s \in \mathbb{R}$, are not fulfilled in the case of the general PBE where the charge neutrality condition (3.36) does not hold.

Before we comment on the different choices of a test space V, we note that in dimension $d \leq 2$ it holds that $\mathrm{e}^u \in L^2(\Omega)$, $\forall u \in H_0^1(\Omega)$ (see [110, 179]) and thus for $d \leq 2$, $b(x, u + w) \in L^2(\Omega)$, $\forall u \in H_0^1(\Omega)$ and a standard weak formulation is available. On the other hand, in dimension $d \geq 3$, there are functions $u \in H_0^1(\Omega)$ such that exponents of them are not even summable. For example $u = \ln \frac{1}{|x|^d} \in H_0^1(B(0,1))$ but $\mathrm{e}^u \notin L^1(B(0,1))$ where $B(0,1)$ is the unit ball in $\mathbb{R}^d$. Therefore, the condition $b(x, u + w)v \in L^1(\Omega)$, $\forall v \in V$ in the weak formulation (3.41) is necessary.

Choosing $V = H_0^1(\Omega)$ in (3.41) allows for a straight forward proof of uniqueness of a solution: if u_1 and u_2 are two solutions of (3.41), then by testing the difference of the equations that u_1 and u_2 satisfy with $v := u_1 - u_2 \in H_0^1(\Omega)$ and by using the monotonicity of $b(x, \cdot)$ and the coercivity of $a(\cdot, \cdot)$, it follows that $u_1 - u_2 = 0$. However, showing existence of a solution is not as conventional as it might seem at a first glance. On the other hand, by choosing a smaller test space V showing existence of a solution u gets easier while showing the uniqueness becomes harder to prove. In all cases, the uniqueness of a solution is possible to show because of the fact that the nonlinearity $b(x, \cdot)$ is monotone increasing. Semilinear elliptic equations with multiple or even infinitely many solutions are considered for example in [127,148,189,190].

Uniqueness

Particularly interesting is the formulation where $V = C_0^\infty(\Omega)$ because it potentially has the largest set of solutions as being the most general formulation when one searches for a solution in $H^1(\Omega)$. When $V = H_0^1(\Omega) \cap L^\infty(\Omega)$ or $V = C_0^\infty(\Omega)$ showing uniqueness of the

weak solution u is not a trivial task since the difference of $u_1 - u_2$ of two solutions is not necessarily in V. In the case when $V = H_0^1(\Omega) \cap L^\infty(\Omega)$ we can test with the (truncated) test functions $T_k(u_1 - u_2) \in H_0^1(\Omega) \cap L^\infty(\Omega)$, $k \geq 0$, where $T_k(s) := \max\{-k, \min\{k, s\}\}$ and use the monotonicity of $b(x, \cdot)$ and the coercivity of $a(\cdot, \cdot)$ to obtain $u_1 - u_2 = 0$. This method and the method that we mentioned for the case $V = H_0^1(\Omega)$ do not work when $V = C_0^\infty(\Omega)$ because the difference $u_1 - u_2$ of two weak solutions and their truncations $T_k(u_1 - u_2)$ are not necessarily in $C_0^\infty(\Omega)$. We overcome this difficulty by applying Theorem 2.18 due to H. Brezis and F. Browder. This theorem gives a sufficient condition to be able to evaluate the extension of an integral bounded linear functional $T_b \in L^1_{loc}(\Omega)$, defined on the dense subspace $C_0^\infty(\Omega)$ of $H_0^1(\Omega)$, at a given element $w \in H_0^1(\Omega)$ through the same defining integral. In particular, if u_1 and u_2 are two solutions, then

$$a(u_1 - u_2, v) + \int_\Omega (b(x, u_1 + w) - b(x, u_2 + w))\, v dx = 0,\ \forall v \in C_0^\infty(\Omega). \tag{3.44}$$

Since $a(u_1 - u_2, \cdot)$ defines a bounded linear functional over $H_0^1(\Omega)$, the functional T_b defined by the formula $\langle T_b, v\rangle := \int_\Omega (b(x, u_1 + w) - b(x, u_2 + w))\, v dx$, $\forall v \in C_0^\infty(\Omega)$ satisfies the condition $T_b \in H^{-1}(\Omega) \cap L^1_{loc}(\Omega)$ in Theorem 2.18. By using the monotonicity of $b(x, \cdot)$ we see that $(b(x, u_1 + w) - b(x, u_2 + w))\,(u_1 - u_2) \geq 0 =: f(x) \in L^1(\Omega)$. Therefore by Theorem 2.18 it follows that $(b(x, u_1 + w) - b(x, u_2 + w))\,(u_1 - u_2) \in L^1(\Omega)$ and the duality product $\langle T_b, u_1 - u_2\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)}$ coincides with $\int_\Omega (b(x, u_1 + w) - b(x, u_2 + w))\,(u_1 - u_2) dx$. This means that

$$a(u_1 - u_2, u_1 - u_2) + \int_\Omega (b(x, u_1 + w) - b(x, u_2 + w))\,(u_1 - u_2) dx = 0, \tag{3.45}$$

which implies $u_1 - u_2 = 0$. Of course, this approach can also be applied to show uniqueness when $V = H_0^1(\Omega) \cap L^\infty(\Omega)$ instead of using the truncations $T_k(u_1 - u_2)$. The uniqueness of a solution to all three formulations is now clear.

Note that if we have an a priori L^∞ bound on the solution u of (3.41), then the term $b(x, u + w) \in L^\infty(\Omega)$. As a consequence, one can use a density argument (density of $C_0^\infty(\Omega)$ in $H_0^1(\Omega)$ or density of $H_0^1(\Omega) \cap L^\infty(\Omega)$ in $H_0^1(\Omega)$) to show that the weak formulations with $V = C_0^\infty(\Omega)$ and $V = H_0^1(\Omega) \cap L^\infty(\Omega)$ are equivalent to the weak formulation with $V = H_0^1(\Omega)$. Since the latter one possesses at most one solution u, which as we have seen is easy to prove, the same will be true for the other two weak formulations. Moreover, if we show the existence of one of these three problems, then the existence for the others will also follow. In the case of linear problems, all three weak formulations are obviously equivalent without the need of any a priori information on the solutions rather than the fact that they are in $H^1(\Omega)$.

A priori L^∞ bound

Showing an a priori L^∞ estimates on u for the three choices of V features similar difficulties

to the ones encountered in the proofs of uniqueness. The case when $V = H_0^1(\Omega)$ is again the easiest: if u is a solution of (3.41) and $u_{\overline{g}} \in H^1(\Omega)$ is such that $\gamma_2(u_{\overline{g}}) = \overline{g}$, then one can test with the functions $G_k(u - u_{\overline{g}}) \in H_0^1(\Omega)$, where $G_k(s) := \operatorname{sgn}(u) \max\{|u| - s, 0\}$ and again use the monotonicity of $b(x, \cdot)$ together with the additional assumption that $u_{\overline{g}} \in L^\infty(\Omega)$ and $\nabla u_{\overline{g}} \in [L^s(\Omega)]^d$ for some $s > d$. We note that G_k is Lipschitz continuous with $G_k(0) = 0$ and hence by Stampacchia's theorem $G_k(u - u_{\overline{g}}) \in H_0^1(\Omega)$. Similar test functions G_k have been used by G. Stampacchia and D. Kinderlehrer in the proof of Theorem 2.32 in the case of linear elliptic problems. When $V = H_0^1(\Omega) \cap L^\infty(\Omega)$ or $V = C_0^\infty(\Omega)$ the test functions $G_k(u - u_{\overline{g}})$ are not necessarily in V. In the case $V = H_0^1(\Omega) \cap L^\infty(\Omega)$ one can test with the functions $T_{2k}(G_k(u - u_{\overline{g}})) \in H_0^1(\Omega) \cap L^\infty(\Omega)$ and use a refined version of Lemma 2.33. However, this approach is quite technical and it is not applicable when $V = C_0^\infty(\Omega)$. To overcome these difficulties, once again we can utilize Theorem 2.18 to show that we can test in (3.41) with the functions $G_k(u - u_{\overline{g}})$. More precisely, from (3.41) it follows that the functional T_b defined by $\langle T_b, v\rangle := \int_\Omega b(x, u + w)v dx$, $\forall v \in C_0^\infty(\Omega)$ satisfies the condition $T_b \in H^{-1}(\Omega) \cap L^1_{loc}(\Omega)$. Moreover, one can show that $b(x, u + w)G_k(u - u_{\overline{g}}) \geq f_k(x)$ for certain functions $f_k \in L^1(\Omega)$, and therefore by Theorem 2.18 it follows that $b(x, u + w)G_k(u - u_{\overline{g}}) \in L^1(\Omega)$ and that $\langle T_b, G_k(u - u_{\overline{g}})\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} = \int_\Omega b(x, u + w)G_k(u - u_{\overline{g}})dx$. Now, one can use the monotonicity of $b(x, \cdot)$ and apply techniques similar to the ones used by D. Kinderlehrer and G. Stampacchia in [112] for linear problems to show the boundedness of u. We will present this kind of techniques in detail in the proof of Theorem 3.29 in Section 3.4 where we will show optimal (in terms of the summability exponents of the data) a priori L^∞ estimate on the solution of a more general semilinear elliptic interface problem. For (3.41) we will take another approach in which we can make use of Theorem 2.32 and which involves the additional splitting of u into functions u^N and u^L.

Existence

To show the existence of a weak solution u to (3.41), we cannot apply for example the theorem of Browder-Minty for monotone operators because the nonlinearity $b(x, \cdot + w)$ does not induce a bounded mapping T_b from $H_0^1(\Omega)$ to its dual $H^{-1}(\Omega)$ through the formula $\langle T_b(z), v\rangle = \int_\Omega b(x, z + w)v dx$, $\forall v \in H_0^1(\Omega)$. Actually, the form $\int_\Omega b(x, \cdot + w)v dx$ is not well defined for all $z \in H_0^1(\Omega)$ because, as we already mentioned, $b(x, z + w)$ might not even be in $L^1_{loc}(\Omega)$. The reason for this is that we do not have a polynomial growth condition on $b(x, \cdot)$ which would suffice for the appropriate summability of $b(x, z + w)$ such that the bound $|\int_\Omega b(x, z + w)v dx| \leq C\|v\|_{H^1(\Omega)}$, $\forall v \in H_0^1(\Omega)$ holds. Here we note that the existence results in [34, 187] could be applied to (3.41) to show the existence of a solution u under the assumption that the charge neutrality condition (3.36) holds and in the case of a homogeneous Dirichlet boundary condition. Since $\overline{g}$ is in general not identically zero on $\partial\Omega$ and the charge neutrality condition (3.36) does not necessarily hold in the case of the general PBE, these results cannot be applied directly.

To show existence and a priori L^∞ estimate we choose an approach that involves the additional splitting of u into u^N and u^L and we note that the existence and a priori L^∞ estimate can be achieved without this additional splitting (see p. 61). In both 2-term and 3-term splittings, the component u^L satisfies the linear nonhomogeneous elliptic interface problem

$$\text{Find } u^L \in H^1_{\bar{g}}(\Omega) \text{ such that} \quad a(u^L, v) = \int_\Omega \boldsymbol{f} \cdot \nabla v dx \text{ for all } v \in V. \tag{3.46}$$

Then, the component u^N has to satisfy

$$\begin{aligned} &\text{Find } u^N \in H^1_0(\Omega) \text{ such that } b(x, u^N + u^L + w)v \in L^1(\Omega) \text{ for all } v \in V \text{ and} \\ &a(u^N, v) + \int_\Omega b(x, u^N + u^L + w)v dx = 0 \text{ for all } v \in V. \end{aligned} \tag{3.47}$$

Problem (3.46) is linear and hence all weak formulations for the three choices of V are obviously equivalent. From the Lax-Milgram Theorem it is clear that there exists a unique $u^L \in H^1_{\bar{g}}(\Omega)$ satisfying (3.46). Moreover, from Theorem 2.32 it follows that $u^L \in L^\infty(\Omega)$.

To show existence of u^N satisfying (3.47), we introduce the strictly convex energy functional $J^N : H^1_0(\Omega) \to \mathbb{R} \cup \{+\infty\}$, defined by

$$J^N(v) = \frac{1}{2} a(v, v) + \int_\Omega B(x, v + u^L + w) dx \tag{3.48}$$

whose minimizer over $H^1_0(\Omega)$ is shown to satisfy (3.47) for $V = C^\infty_0(\Omega)$ and $V = H^1_0(\Omega) \cap L^\infty(\Omega)$. It should be noted that J^N is not Gateaux differentiable at each point in $H^1_0(\Omega)$ and thus we cannot conclude straightforwardly that the minimizer of J^N over $H^1_{\bar{g}}(\Omega)$ is a solution to the weak formulation (3.47) (see Remark 3.25). Instead, by applying the Lebesgue DCT and using the fact that the nonlinearity $B(x, u^N_{min} + u^L + w)$ is in $L^1(\Omega)$ at the minimizer u^N_{min} we can see that the unique minimizer u^N_{min} of J^N is a solution of (3.47) with $V = C^\infty_0(\Omega)$ and $V = H^1_0(\Omega) \cap L^\infty(\Omega)$. As we explained above, in order to show that this minimizer is also a solution to (3.47) with $V = H^1_0(\Omega)$, we prove an a priori $L^\infty(\Omega)$ estimate on the weak solution u^N of (3.47) with either $V = C^\infty_0(\Omega)$ or $V = H^1_0(\Omega) \cap L^\infty(\Omega)$. Then, by a density argument we see that this u^N is also a solution to (3.47) with $V = H^1_0(\Omega)$. Since all three weak formulations have at most one solution (for example proved by using Theorem 2.18 and testing with the difference $u^N_1 - u^N_2$ of two solutions), then this u^N is their only solution.

3.3.1 2-term splitting

As in the case of the LPBE, we split ϕ into the singular Coulomb potential G and a reaction field potential u. By substituting the expression $\phi = G + u$ in (3.40) and by taking into

account (3.15) we obtain the weak formulation that u has to satisfy:

$$\text{Find } u \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}_{g-G}(\Omega) \text{ such that } b(x,u+G)v \in L^1(\Omega) \text{ for all } v \in \bigcup_{q>d} W^{1,q}_0(\Omega) \text{ and}$$

$$\int_\Omega \epsilon \nabla u \cdot \nabla v dx + \int_\Omega b(x,u+G)v dx = \int_{\Omega_s} (\epsilon_m - \epsilon_s)\nabla G \cdot \nabla v dx = \langle \mathcal{G}_2, v\rangle \tag{3.49}$$

$$\text{for all } v \in \bigcup_{q>d} W^{1,q}_0(\Omega).$$

Since the space $W^{1,2}(\Omega)$ lies between the spaces $\bigcup_{q>d} W^{1,q}(\Omega)$ and $\bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$, similarly to the case of the LPBE, we can find a particular function u that satisfies (3.49) if we can pose a standard weak formulation for u that involves H^1 spaces, i.e.,

$$u \in H^1_{g-G}(\Omega) \text{ such that } a(u,v) + \int_\Omega b(x,u+G)v dx = \langle \mathcal{G}_2, v\rangle \text{ for all } v \in H^1_0(\Omega).$$

However, as we explained in the beginning of this section, for $d = 3$ there are functions $u \in H^1(\Omega)$ such that $e^u \notin L^1_{loc}(\Omega)$ and therefore $b(x, u+G)$ might not even be in $L^1_{loc}(\Omega)$. For this reason, in the case $d = 3$ it makes sense to consider the following 3 weak formulations and reveal the relations between them.

$$\begin{aligned} &\text{Find } u \in H^1_{g-G}(\Omega) \text{ such that } b(x,u+G)v \in L^1(\Omega) \text{ for all } v \in V \text{ and} \\ &a(u,v) + \int_\Omega b(x,u+G)v dx = \langle \mathcal{G}_2, v\rangle \text{ for all } v \in V, \end{aligned} \tag{3.50}$$

where $V = C^\infty_0(\Omega)$, $V = H^1_0(\Omega) \cap L^\infty(\Omega)$, or $V = H^1_0(\Omega)$. We know that in dimension $d = 3$, the Sobolev space $H^1(\Omega)$ is continuously embedded in $L^6(\Omega)$. We also know that for $v \in L^6(\Omega)$, from Hölder's inequality it follows that a sufficient condition for $zv \in L^1(\Omega)$ is $z \in L^{\frac{6}{5}}(\Omega)$ (see Remark 3.26). Since $b(x, u+G)$ is not a priori known to lie in $L^{\frac{6}{5}}(\Omega)$ we can not apply a density argument to conclude that all three formulations are equivalent. We will prove this fact by showing an a priori L^∞ estimate on the solution u for all three weak formulations. In this case, since $G \in L^\infty(\Omega_{ions})$, $b(x, u+G)$ will be in $L^\infty(\Omega)$ and hence a density argument will work and all three weak formulations will be equivalent.

We will show existence, uniqueness and boundedness to (3.50) for all three choices of the test space V in Section 3.3.3.

Before we continue with the 3-term splitting, we formulate the existence result for (3.40), which is a consequence of the existence of a solution to (3.50) with $V = H^1_0(\Omega)$ or $V = H^1_0(\Omega) \cap L^\infty(\Omega)$. Indeed, if there is a $u \in H^1_{g-G}(\Omega)$ which satisfies (3.50) with $V = H^1_0(\Omega)$ or $V = H^1_0(\Omega) \cap L^\infty(\Omega)$, then since $H^1_{g-G}(\Omega) \subset \bigcap_{p<\frac{d}{d-1}} W^{1,p}_{g-G}(\Omega)$ and $\bigcup_{q>d} W^{1,q}_0(\Omega) \subset V$, it follows that this u satisfies (3.49). Thus, $\phi = G + u$ satisfies (3.40).

Theorem 3.13
There exists a weak solution ϕ of equation (3.35) *satisfying* (3.40). *A particular ϕ satisfying* (3.40) *can be given in the form $\phi = G + u$, where $u \in H^1_{g-G}(\Omega) \cap L^\infty(\Omega)$ is the unique solution of* (3.50) *with $V = H^1_0(\Omega)$.*

As we mentioned earlier, the existence and uniqueness of such u solving (3.50) with $V = H^1_0(\Omega)$ will become clear in Section 3.3.3.

Remark 3.14
In particular Theorem 3.13 provides us with an explicit way to construct a solution of (3.35). *Moreover, the unknown component u satisfies a standard weak formulation involving H^1 spaces and can be approximated numerically by standard finite element methods. Note also that despite the fact that u is unique, we do not claim uniqueness of the full potential ϕ for the nonlinear case of the general PBE. It seems that the proof we gave in the linear case cannot be adapted so easily to the case of a semilinear equation.*

3.3.2 3-term splitting

In the 3-term splitting, we have $\phi = G + u^H + u$ where u^H is such that $G + u^H = 0$ in Ω_s. If we require that $u \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$, since $G \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$ it means that u^H should also be in the space $\bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega)$. By substituting the expression $\phi = G + u^H + u$ in (3.40) and by taking into account (3.15) we obtain the weak formulation that the component u has to satisfy:

$$\begin{aligned}&\text{Find } u \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}_g(\Omega) \text{ such that } b(x,u)v \in L^1(\Omega) \text{ for all } v \in \bigcup_{q>d} W^{1,q}_0(\Omega) \text{ and} \\ &\int_\Omega \epsilon \nabla u \cdot \nabla v dx + \int_\Omega b(x,u) v dx = -\int_{\Omega_m} \epsilon_m \nabla u^H \cdot \nabla v dx + \int_{\Omega_s} \epsilon_m \nabla G \cdot \nabla v dx = \langle \mathcal{G}_3, v \rangle \quad (3.51) \\ &\text{for all } v \in \bigcup_{q>d} W^{1,q}_0(\Omega).\end{aligned}$$

If we define u^H in Ω_m such that $u^H \in H^1(\Omega)$ it will hold $u^H + u \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}_{g-G}(\Omega)$. The argument here is similar to the argument in the 3-term splitting for the LPBE: since we have

$$\bigcup_{q>d} W^{1,q}(\Omega) \subset W^{1,2}(\Omega) \subset \bigcap_{p<\frac{d}{d-1}} W^{1,p}(\Omega),$$

we will find a particular solution u of problem (3.51) if we can pose a standard weak formulation for u that involves H^1 spaces. Like in the case of the 2-term splitting, we consider the following 3 weak formulations with test spaces $V = C^\infty_0(\Omega)$, $V = H^1_0(\Omega) \cap L^\infty(\Omega)$, and

$V = H_0^1(\Omega)$:

$$\begin{aligned} &\text{Find } u \in H_g^1(\Omega) \text{ such that } b(x,u)v \in L^1(\Omega) \text{ for all } v \in V \text{ and} \\ &\int_\Omega \epsilon \nabla u \cdot \nabla v dx + \int_\Omega b(x,u) v dx = \langle \mathcal{G}_3, v \rangle \text{ for all } v \in V, \end{aligned} \tag{3.52}$$

where $\mathcal{G}_3$ is a well defined functional in $H^{-1}(\Omega)$ since we have chosen u^H to be in $H^1(\Omega)$. For the a posteriori analysis in Section 4.3 it is convenient to define u^H to be harmonic in Ω_m, i.e.,

$$u^H \in H^1_{-G}(\Omega_m) \quad \text{and} \quad \int_{\Omega_m} \nabla u^H \cdot \nabla v dx = 0 \text{ for all } v \in H_0^1(\Omega_m). \tag{3.29}$$

The right hand-side $\langle \mathcal{G}_3, v \rangle$ of (3.51) represents a jump condition on the normal component of $\epsilon \nabla u$. More precisely, if the function u is smooth in each subdomain, it should satisfy the jump condition $[\epsilon \nabla u \cdot \boldsymbol{n}_\Gamma]_\Gamma = -\epsilon_m \nabla \left(u^H + G\right) \cdot \boldsymbol{n}_\Gamma$ (see Remark 3.9).

We will show existence, uniqueness and boundedness to (3.52) for all three choices of the test space V in Section 3.3.3.

We can formulate another existence result for (3.40), which is a consequence of the existence of a solution to (3.52) with $V = H_0^1(\Omega)$ or $V = H_0^1(\Omega) \cap L^\infty(\Omega)$: if there is a $u \in H_g^1(\Omega)$ which satisfies (3.52) with $V = H_0^1(\Omega)$ or $V = H_0^1(\Omega) \cap L^\infty(\Omega)$, then since $H_g^1(\Omega) \subset \bigcap_{p < \frac{d}{d-1}} W^{1,p}_{g-G}(\Omega)$ and $\bigcup_{q>d} W_0^{1,q}(\Omega) \subset V$, it follows that this u satisfies (3.51). Thus, $\phi = G + u^H + u$ satisfies (3.40).

Theorem 3.15
There exists a weak solution ϕ of equation (3.35) satisfying (3.40). A particular ϕ satisfying (3.40) can be given in the form $\phi = G + u^H + u$, where $u \in H_g^1(\Omega) \cap L^\infty(\Omega)$ is the unique solution of (3.52) with $V = H_0^1(\Omega)$.

Remark 3.16
Note that since we have not proven uniqueness for the problem (3.40), the particular solutions given by means of the 2-term and 3-term splittings in Theorem 3.13 and Theorem 3.15, respectively, might be different.

3.3.3 Existence, uniqueness, and boundedness of the component u in the 2-term and 3-term splittings

In this section, we prove existence, uniqueness, and a priori L^∞ estimates on the solutions of (3.50) and (3.52) for all three choices of the test space V. First we observe that these weak

formulations for both 2-term and 3-term splittings can be written in the common form

$$\begin{aligned} &\text{Find } u \in H^1_{\overline{g}}(\Omega) \text{ such that } b(x, u+w)v \in L^1(\Omega) \text{ for all } v \in V \text{ and} \\ &a(u,v) + \int_\Omega b(x, u+w)v dx = \int_\Omega \boldsymbol{f} \cdot \nabla v dx \text{ for all } v \in V. \end{aligned} \tag{3.41}$$

In the case of the 2-term splitting we have

$$w = G, \quad \boldsymbol{f} = \chi_{\Omega_s}(\epsilon_m - \epsilon_s)\nabla G, \quad \text{and} \quad \overline{g} = g - G \text{ on } \partial\Omega \tag{3.42}$$

whereas in the case of the 3-term splitting we have

$$w = 0, \quad \boldsymbol{f} = -\chi_{\Omega_m}\epsilon_m \nabla u^H + \chi_{\Omega_s}\epsilon_m \nabla G, \quad \text{and} \quad \overline{g} = g \text{ on } \partial\Omega. \tag{3.43}$$

We assume that the function $\overline{g}$ specifying the Dirichlet boundary condition on $\partial\Omega$ is given as the trace of some $H^2(\Omega)$ function (see the discussion in Section 3.2.3). Note that $w \in L^\infty(\Omega_{ions})$ and $\boldsymbol{f} = (f_1, f_2, \ldots, f_d) \in [L^s(\Omega)]^d$ with $s > d$, since G is smooth in $\overline{\Omega}_s$, $\epsilon_s \in C^{0,1}(\overline{\Omega}_s)$, and $\nabla u^H \in [L^s(\Omega_m)]^d$, $s > d$ if Γ is $C^{0,1}$ (see the discussion in Section 3.2.3).

Uniqueness

First, we prove uniqueness of a solution to (3.41) for all three choices of the test space V. Suppose that u_1 and u_2 are two solutions of (3.41). Then, we have

$$a(u_1 - u_2, v) + \int_\Omega \left(b(x, u_1 + w) - b(x, u_2 + w)\right) v dx = 0, \, \forall v \in V. \tag{3.53}$$

In the case of $V = H^1_0(\Omega)$, $u_1 - u_2 \in V$ and thus we can test (3.53) with $v := u_1 - u_2$ to obtain

$$a(u_1 - u_2, u_1 - u_2) + \int_\Omega \left(b(x, u_1 + w) - b(x, u_2 + w)\right)(u_1 - u_2) dx = 0.$$

Since $a(\cdot,\cdot)$ is coercive and $b(x,\cdot)$ is monotone increasing, we obtain $u_1 - u_2 = 0$. When $V = H^1_0(\Omega) \cap L^\infty(\Omega)$ and $V = C^\infty_0(\Omega)$, the difference $u_1 - u_2$ is not necessarily in V and we cannot test with it in (3.53). In this case, V is a dense subspace of $H^1_0(\Omega)$ which contains $C^\infty_0(\Omega)$. Since $a(u,\cdot)$ defines a bounded linear functional over $H^1_0(\Omega)$, from (3.53) it follows that the linear functional T_c defined by

$$\langle T_c, v\rangle := \left(b(x, u_1 + w) - b(x, u_2 + w), v\right), \, \forall v \in V$$

is bounded over V. Moreover, $c \in L^1_{loc}(\Omega)$ and $c(u_1 - u_2) \geq 0 \in L^1(\Omega)$ a.e. $x \in \Omega$. This means that we can apply Theorem 2.18 to the functional T_c and the test function $v = u_1 - u_2 \in H^1_0(\Omega)$ (see also Remark 2.19 after Theorem 2.18). Again, using the coercivity of $a(\cdot,\cdot)$ and the monotonicity of $b(x,\cdot)$ (see (3.39)), we conclude that $u_1 - u_2 = 0$.

Existence

We continue with the proof of existence of a solution to (3.41) simultaneously for the test spaces $V = H_0^1(\Omega) \cap L^\infty(\Omega)$ and $V = C_0^\infty(\Omega)$. The existence of a solution to (3.41) with $V = H_0^1(\Omega)$ will follow from the existence for the first two choices of V once we prove the a priori L^∞ bound on their common solution.

It is convenient to split u into u^L and u^N, i.e., $u = u^L + u^N$, where u^L satisfies the linear nonhomogeneous interface elliptic problem

$$\text{Find } u^L \in H_{\overline{g}}^1(\Omega) \text{ such that } a(u^L, v) = \int_\Omega \boldsymbol{f} \cdot \nabla v dx \text{ for all } v \in V, \tag{3.46}$$

and the component u^N has to satisfy

$$\begin{aligned} &\text{Find } u^N \in H_0^1(\Omega) \text{ such that } b(x, u^N + u^L + w)v \in L^1(\Omega) \text{ for all } v \in V \text{ and} \\ &a(u^N, v) + \int_\Omega b(x, u^N + u^L + w)v dx = 0 \text{ for all } v \in V. \end{aligned} \tag{3.47}$$

Problem (3.46) is linear and hence all weak formulations for the three choices of V are obviously equivalent. From the Lax-Milgram Theorem it is clear that there exists a unique $u^L \in H_{\overline{g}}^1(\Omega)$ satisfying (3.46). Moreover, from Theorem 2.32 it follows that $u^L \in L^\infty(\Omega)$. Indeed, let $u_{\overline{g}}^L \in H^2(\Omega)$ be such that $\gamma_2(u_{\overline{g}}^L) = \overline{g}$ on $\partial\Omega$. Then we consider the homogenized version of (3.46) given by

$$a(u_0^L, v) = \int_\Omega \boldsymbol{f} \cdot \nabla v dx - \int_\Omega \epsilon \nabla u_{\overline{g}}^L \cdot \nabla v dx \text{ for all } v \in V, \tag{3.54}$$

where $u_0^L \in H_0^1(\Omega)$ and $u^L := u_{\overline{g}}^L + u_0^L$. Since $\nabla u_{\overline{g}}^L \in H^1(\Omega)$ it follows by the Sobolev embedding theorem (Theorem 2.21) that $\nabla u_{\overline{g}}^L \in \left[L^6(\Omega)\right]^d$ for $d = 2, 3$. Therefore, by recalling that $\boldsymbol{f} \in [L^s(\Omega)]^d$, $s > d$ we can apply Theorem 2.32 to obtain that $u_0^L \in L^\infty(\Omega)$ and hence $u^L \in L^\infty(\Omega)$.

To show existence of u^N satisfying (3.47), we introduce the energy functional $J^N : H_0^1(\Omega) \to \mathbb{R} \cup \{+\infty\}$, defined by

$$J^N(v) = \frac{1}{2} a(v, v) + \int_\Omega B(x, v + u^L + w) dx, \tag{3.55}$$

where it is understood that $J^N(v) = +\infty$ whenever $B(x, v + u^L + w) \notin L^1(\Omega)$, i.e.,

$$J^N(v) := \begin{cases} \dfrac{1}{2} a(v, v) + \displaystyle\int_\Omega B(x, v + u^L + w) dx, \text{ if } B(x, v + u^L + w) \in L^1(\Omega), \\ +\infty, \text{ if } B(x, v + u^L + w) \notin L^1(\Omega). \end{cases} \tag{3.56}$$

We consider the variational problem

$$\text{Find } u^N_{\min} \in H^1_0(\Omega) \text{ such that } J^N(u^N_{\min}) = \min_{v \in H^1_0(\Omega)} J^N(v). \tag{3.57}$$

Notice that for $d \leq 2$ it holds $\mathrm{e}^v \in L^2(\Omega)$ for all $v \in H^1_0(\Omega)$ (e.g., see [110,179]) and therefore $dom(J^N) = \left\{v \in H^1_0(\Omega) \text{ such that } J^N(v) < \infty\right\}$ is a linear space coinciding with $H^1_0(\Omega)$. However, in dimension $d = 3$, $dom(J^N)$ is only a convex set and not a linear space (see Remark 3.23). In fact, $\mathrm{dom}(J^N)$ is not even a closed subspace of $H^1_0(\Omega)$. Indeed, $\mathrm{dom}(J^N)$ contains $C^\infty_0(\Omega)$ which is dense in $H^1_0(\Omega)$. If $\mathrm{dom}(J^N)$ were closed, it would coincide with $H^1_0(\Omega)$ and we know that this is not true in dimension $d \geq 3$ (see the examples in Remark 3.23).

Since $dom(J^N)$ is convex and obviously J^N is convex over $dom(J^N)$ it follows that J^N is convex over $H^1_0(\Omega)$ (see Remark 2.31 and [65]). To show existence of a minimizer of J^N over the reflexive Banach space H^1_0 one has to verify the following assertions:

(1) J^N is proper, i.e., J^N is not identically equal to $+\infty$ and does not take the value $-\infty$;

(2) J^N is sequentially weakly lower semicontinuous (s.w.l.s.c.), i.e., if $\{v_n\}_{n=1}^\infty \subset H^1_0(\Omega)$ and $v_n \rightharpoonup v$ (weakly in $H^1_0(\Omega)$) then $J^N(v) \leq \liminf_{n\to\infty} J^N(v_n)$;

(3) J^N is coercive, i.e., $\lim_{n\to\infty} J^N(v_n) = +\infty$ whenever $\|v_n\|_{H^1(\Omega)} \to \infty$.

Assertion (1) is obvious since $\int_\Omega B(x, u + u^L + w)dx \geq 0$ and $J^N(0)$ is finite. To see that (2) is fulfilled, notice that J^N is the sum of the functionals $A(v) := \frac{1}{2}a(v, v)$ and $\int_\Omega B(x, v + u^L + w)dx$. The former is convex and Gateaux differentiable, and therefore s.w.l.s.c. (for the proof of this implication, see, e.g. Corollary 2.4 in [169]). Indeed, by using the symmetry of $a(\cdot, \cdot)$, for any $z, v \in H^1_0(\Omega)$ and any $\lambda \in [0, 1]$, for the convexity of $\frac{1}{2}a(v, v)$ we obtain

$$\begin{aligned}
&A(\lambda z + (1-\lambda)v) \\
=&\frac{1}{2}a(\lambda z + (1-\lambda)v, \lambda z + (1-\lambda)v) = \frac{1}{2}\left(\lambda^2 a(z,z) + 2\lambda(1-\lambda)a(z,v) + (1-\lambda)^2 a(v,v)\right) \\
\leq&\frac{1}{2}\left(\lambda^2 a(z,z) + \lambda(1-\lambda)\left(a(z,z) + a(v,v)\right) + (1-\lambda)^2 a(v,v)\right) \\
=&\lambda A(z) + (1-\lambda)A(v),
\end{aligned} \tag{3.58}$$

where we have used the inequality $a(z, v) \leq \sqrt{a(z, z)}\sqrt{a(v, v)} \leq \frac{1}{2}\left(a(z, z) + a(v, v)\right)$. To see that $A(\cdot)$ is Gateaux differentiable, we use the symmetry of $a(\cdot, \cdot)$ together with its linearity

in each argument. For any $z \in H_0^1(\Omega)$ and any direction $v \in H_0^1(\Omega)$ we have

$$\begin{aligned} &\lim_{\lambda\to 0^+} \frac{A(z+\lambda v)-A(z)}{\lambda} = \lim_{\lambda\to 0^+} \frac{1}{2}\frac{a(z+\lambda v, z+\lambda v)-a(z,z)}{\lambda} \\ &= \lim_{\lambda\to 0^+} \frac{1}{2}\frac{a(z,z)+\lambda^2 a(v,v)+2\lambda a(z,v)-a(z,z)}{\lambda} \\ &= \lim_{\lambda\to 0^+} \left(\frac{1}{2}\lambda a(v,v)+a(z,v)\right) = a(z,v). \end{aligned} \tag{3.59}$$

Since $a(z,\cdot)$ is a bounded linear functional over $H_0^1(\Omega)$, by the definition of Gateaux differentiability (see Definition 2.27), it follows that A is Gateaux differntiable at z with a Gateaux differential at z equal to the functional $A'(z)$, defined by $\langle A'(z), v\rangle := a(z,v)$, $\forall v \in H_0^1(\Omega)$. However, for $d=3$, the functional $\int_\Omega B(x, v+u^L+w)dx$ is not Gateaux differentiable (see Remark 3.25). Nevertheless, one can still show that the functional $\int_\Omega B(x, v+u^L+w)dx$ is s.w.l.s.c. using Fatou's lemma and the compact embedding of $H_0^1(\Omega)$ into $L^2(\Omega)$ (see Theorem 2.22) as follows. Let $\{v_n\}_{n=1}^\infty \subset H_0^1(\Omega)$ be a sequence which converges weakly in $H_0^1(\Omega)$ to an element $v \in H_0^1(\Omega)$, i.e., $v_n \rightharpoonup v$. Since the embedding $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$ is compact it follows that $v_n \to v$ (strongly) in $L^2(\Omega)$, and therefore we can extract a pointwise almost everywhere convergent subsequence $v_{n_m}(x) \to v(x)$ (see Theorem 4.9 in [96]). Since $B(x,\cdot)$ is a continuous function for any $x \in \Omega$ and $x \mapsto B(x,s)$ is measurable for any $s \in \mathbb{R}$ it means that B is a Carathéodory function and as a consequence the function $x \mapsto B\left(x, v_{n_m}(x)+u^L(x)+w(x)\right)$ is measurable for all $k \in \mathbb{N}$ (see Proposition 3.7 in [55]). By noting that $B\left(x, z(x)+u^L(x)+w(x)\right) \geq 0$ for all $z \in H_0^1(\Omega)$ and using the fact that $B(x,\cdot)$ is a continuous function for any $x \in \Omega$, from Fatou's lemma we obtain

$$\begin{aligned} \liminf_{m\to\infty} \int_\Omega B\left(x, v_{n_m}(x)+u^L(x)+w(x)\right)dx &\geq \int_\Omega \liminf_{m\to\infty} B\left(x, v_{n_m}(x)+u^L(x)+w(x)\right)dx \\ &= \int_\Omega B\left(x, v(x)+u^L(x)+w(x)\right)dx. \end{aligned} \tag{3.60}$$

Now it is clear that if $\{v_{n_m}\}_{m=1}^\infty$ is an arbitrary subsequence of $\{v_n\}_{n=1}^\infty$, then there exists a further subsequence $\{v_{n_{m_s}}\}_{s=1}^\infty$ for which (3.60) is satisfied. This means that (3.60) is also satisfied for the whole sequence $\{v_n\}_{n=1}^\infty$, and hence $\int_\Omega B\left(x, v+u^L+w\right)dx$ is s.l.w.s.c. (see Remark 3.22).

The coercivity of J^N follows by observing that

$$J^N(v) = \frac{1}{2}a(v,v) + \int_\Omega B\left(x, v+u^L+w\right)dx \geq \epsilon_{\min}\|\nabla v\|^2_{L^2(\Omega)} \geq \frac{\epsilon_{\min}}{1+C_P^2}\|v\|_{H^1(\Omega)},$$

where C_P is the constant in the inequality $\|v\|_{L^2(\Omega)} \leq C_P\|\nabla v\|_{L^2(\Omega)}$, $\forall v \in H_0^1(\Omega)$. Now, the existence of a minimizer $u^N_{\min}$ of J^N over the reflexive Banach space $H_0^1(\Omega)$ follows by

Theorem 2.30. Moreover, since $a(v,v)$ is a strictly convex functional it follows that J^N is also strictly convex, and therefore this minimizer is unique.

Theorem 3.17
There exists a unique $u^N_{\min} \in H^1_0(\Omega)$ *such that* $J^N(u^N_{\min}) = \min_{v \in H^1_0(\Omega)} J^N(v)$.

We will show that the minimizer $u^N_{\min}$ of J^N over $H^1_0(\Omega)$ satisfies (3.47) for $V = C^\infty_0(\Omega)$ and $V = H^1_0(\Omega) \cap L^\infty(\Omega)$. Notice that J^N is not Gateaux differentiable at each point in $H^1_0(\Omega)$ and thus we cannot conclude straightforwardly that the minimizer $u^N_{\min}$ is a solution to the weak formulation (3.47) (see Remark 3.25).

Now by using the Lebesgue dominated convergence theorem and the fact that at the unique minimizer $u^N_{\min}$ of J^N we have $B\left(x, u^N_{\min} + u^L + w\right) \in L^1(\Omega)$, we will show that $u^N_{\min}$ is also a solution to (3.47). We have that $J^N\left(u^N_{\min} + \lambda v\right) - J^N\left(u^N_{\min}\right) \geq 0$ for all $v \in H^1_0(\Omega)$ and all $\lambda \geq 0$, i.e.,

$$\begin{aligned}&\frac{1}{2}a\left(u^N_{\min} + \lambda v, u^N_{\min} + \lambda v\right) + \int_\Omega B\left(x, u^N_{\min} + \lambda v + u^L + w\right) dx \\ &-\frac{1}{2}a\left(u^N_{\min}, u^N_{\min}\right) - \int_\Omega B\left(x, u^N_{\min} + u^L + w\right) dx \geq 0,\end{aligned}$$

which, by using the symmetry of $a(\cdot,\cdot)$, is equivalent to

$$\lambda a\left(u^N_{\min}, v\right) + \frac{\lambda}{2}a(v,v) + \int_\Omega \left(B\left(x, u^N_{\min} + \lambda v + u^L + w\right) - B\left(x, u^N_{\min} + u^L + w\right)\right) dx \geq 0. \tag{3.61}$$

Divide both sides of (3.61) by $\lambda > 0$ and let $\lambda \to 0^+$ to obtain

$$a\left(u^N_{\min}, v\right) + \lim_{\lambda \to 0^+} \int_\Omega \frac{B\left(x, u^N_{\min} + \lambda v + u^L + w\right) - B\left(x, u^N_{\min} + u^L + w\right)}{\lambda} dx \geq 0. \tag{3.62}$$

To compute the limit in the second term of (3.62), we will apply the Lebesgue dominated convergence theorem. We have

$$\begin{aligned}f_\lambda(x) &:= \frac{B\left(x, u^N_{\min}(x) + u^L(x) + w(x) + \lambda v(x)\right) - B\left(x, u^N_{\min}(x) + u^L(x) + w(x)\right)}{\lambda} \\ &\xrightarrow{\lambda \to 0^+} b\left(x, u^N_{\min}(x) + u^L(x) + w(x)\right) v(x) \quad \text{for a.e} \quad x \in \Omega\end{aligned} \tag{3.63}$$

By the mean value theorem we have

$$f_\lambda(x) = b\left(x, u^N_{\min} + u^L(x) + w(x) + \Theta(x)\lambda v(x)\right) v(x), \quad \text{where} \quad \Theta(x) \in (0,1), \forall x \in \Omega$$

and hence, if $v \in L^\infty(\Omega)$, we can obtain the following bound on f_λ:

$$
\begin{aligned}
|f_\lambda(x)| &= \left| -\frac{4\pi e_0^2}{k_B T} v(x) \sum_{j=1}^{N_{ions}} \overline{M}_j(x) \xi_j \mathrm{e}^{-\xi_j \left(u^N_{\min}(x) + u^L(x) + w(x) + \Theta(x)\lambda v(x) \right)} \right| \\
&\leq \max_j |\xi_j| \|v\|_{L^\infty(\Omega)} \frac{4\pi e_0^2}{k_B T} \sum_{j=1}^{N_{ions}} \overline{M}_j(x) \mathrm{e}^{-\xi_j \left(u^N_{\min}(x) + u^L(x) + w(x) \right) - \xi_j \Theta(x) \lambda v(x)} \\
&\leq \max_j |\xi_j| \max_j \mathrm{e}^{|\xi_j| \|v\|_{L^\infty(\Omega)}} \|v\|_{L^\infty(\Omega)} \frac{4\pi e_0^2}{k_B T} \sum_{j=1}^{N_{ions}} \overline{M}_j(x) \mathrm{e}^{-\xi_j \left(u^N_{\min}(x) + u^L(x) + w(x) \right)} \\
&= \max_j |\xi_j| \max_j \mathrm{e}^{|\xi_j| \|v\|_{L^\infty(\Omega)}} \|v\|_{L^\infty(\Omega)} B\left(x, u^N_{\min}(x) + u^L(x) + w(x) \right) \in L^1(\Omega),\ \forall \lambda \leq 1.
\end{aligned}
\tag{3.64}
$$

From the Lebesgue dominated convergence theorem, by using (3.63) and (3.64), it follows that the limit in (3.62) is equal to $\int_\Omega b\left(x, u^N_{\min} + u^L + w\right) v dx$, and therefore we obtain

$$
a\left(u^N_{\min}, v\right) + \int_\Omega b\left(x, u^N_{\min} + u^L + w\right) v dx \geq 0,\ \forall v \in H^1_0(\Omega) \cap L^\infty(\Omega). \tag{3.65}
$$

This means that $u^N = u^N_{\min}$ is a solution to the weak formulation (3.47) for $V = H^1_0(\Omega) \cap L^\infty(\Omega)$ and $V = C^\infty_0(\Omega)$. The uniqueness of the solution u^N of (3.47) is done in the same way as the uniqueness of (3.41). Now, it is clear that $u = u^N + u^L \in H^1_g$ is the unique solution to (3.41) for the test spaces $V = H^1_0(\Omega) \cap L^\infty(\Omega)$ and $V = C^\infty_0(\Omega)$.

To show that $u^N = u^N_{\min}$ is also a solution to (3.47) with $V = H^1_0(\Omega)$ and that $u = u^N + u^L$ is a solution to (3.41) with $V = H^1_0(\Omega)$ we will prove the a priori L^∞ bound on the solution of (3.47) with $V = H^1_0(\Omega) \cap L^\infty(\Omega)$ and $V = C^\infty_0(\Omega)$.

A priori L^∞ bound on the component u^N

Next, we show that the solution to Problem (3.47) is essentially bounded regardless of which test space V we have.

Proposition 3.18

The unique weak solution u^N to Problem (3.47) *belongs to $L^\infty(\Omega)$. Moreover, there is a positive constant $\overline{e} > 0$, depending only on d, Ω, $\epsilon_{\min} := \min_{x \in \Omega} \epsilon(x)$, such that $\|u^N\|_{L^\infty(\Omega)} \leq \|u^L + w\|_{L^\infty(\Omega_{ions})} + \overline{e}$. If the charge neutrality condition* (3.36) *holds, then $\overline{e} = 0$.*

Proof. To prove that u^N is bounded we apply Theorem 2.18 once again. The first step is to show that (3.47) holds for the test functions

$$
v = G_s(u^N) := \operatorname{sgn}(u^N) \max\{|u^N| - s, 0\} = \begin{cases} u^N - s & \text{for } x \in \left\{u^N > s\right\}, \\ 0 & \text{for } x \in \left\{u^N \in [-s, s]\right\}, \\ u^N + s & \text{for } x \in \left\{u^N < -s\right\}. \end{cases} \tag{3.66}
$$

with $s \geq \|u^L + w\|_{L^\infty(\Omega_{ions})}$. We notice that similar test functions $G_s(u^N)$ have been used in [112, Theorem B.2] in the context of linear elliptic problems.

It is easy to see that $G_s(0) = 0$ and that $G_s(\cdot)$ is Lipschitz continuous for any s. Therefore, by Stampacchia's theorem (e.g., see [90,112]) it follows that $G_s(u^N) \in H_0^1(\Omega)$ and that the weak partial derivatives of $G_s(u^N)$ are given by

$$\frac{\partial G_s(u^N)}{\partial x_i} = \begin{cases} \frac{\partial u^N}{\partial x_i} & \text{for } x \in \left\{u^N > s\right\}, \\ 0 & \text{for } x \in \left\{u^N \in [-s, s]\right\}, \\ \frac{\partial u^N}{\partial x_i} & \text{for } x \in \left\{u^N < -s\right\}. \end{cases} \tag{3.67}$$

Next, the functional T_b defined by $\langle T_b, v\rangle := \int\limits_\Omega b(x, u^N + u^L + w)v dx$, $\forall v \in C_0^\infty(\Omega)$ is bounded and linear over the dense subspace $C_0^\infty(\Omega)$ of $H_0^1(\Omega)$ and $b(x, u^N + u^L + w) \in L^1_{loc}(\Omega)$. This fact follows from (3.47) and the fact that the functional $a(u^N, \cdot)$ belongs to $H^{-1}(\Omega)$. Then, in view of Theorem 2.18, to show that $\langle T_b, G_s(u^N)\rangle = \int\limits_\Omega b(x, u^N + u^L + w)G_s(u^N)dx$ it suffices to verify that

$$b(x, u^N + u^L + w)G_s(u^N) \geq f \text{ a.e. for some } f \in L^1(\Omega). \tag{3.68}$$

By choosing $s \geq \|u^L + w\|_{L^\infty(\Omega_{ions})}$, using the monotonicity of $b(x, \cdot)$, and recalling that $b(x, s) = 0$ for all $x \in \Omega_m \cup \Omega_{IEL}$, we obtain

- for $x \in \Omega_{ions} \cap \left\{u^N > s\right\}$:
  $b(x, u^N + u^L + w)G_s(u^N) = b(x, u^N + u^L + w)(u^N - s) \geq b(x, 0)(u^N - s)$;
- for $x \in \Omega_{ions} \cap \left\{u^N \in [-s, s]\right\}$:
  $b(x, u^N + u^L + w)G_s(u^N) = 0$;
- for $x \in \Omega_{ions} \cap \left\{u^N < -s\right\}$:
  $b(x, u^N + u^L + w)G_s(u^N) = b(x, u^N + u^L + w)(u^N + s) \geq b(x, 0)(u^N + s)$.

(3.69)

By taking into account the equality

$$b(x, 0) = -\frac{4\pi}{k_B T} \sum_{j=1}^{N_{ions}} M_j \xi_j = const \text{ for } x \in \Omega_{ions}, \tag{3.70}$$

we see that $b(x, 0) \in L^\infty(\Omega)$ and hence $b(x, 0)(u^N - s) \in L^1\left(\Omega_{ions} \cap \left\{u^N > s\right\}\right)$ and $b(x, 0)(u^N + s) \in L^1\left(\Omega_{ions} \cap \left\{u^N < -s\right\}\right)$. Therefore, from (3.69) it follows that (3.68) holds for the summable function f defined by

$$f(x) = \begin{cases} 0 & \text{for } x \in \Omega_m \cup \Omega_{IEL} \cup \left(\Omega_{ions} \cap \left\{u^N \in [-s, s]\right\}\right), \\ b(x, 0)(u^N - s) & \text{for } x \in \Omega_{ions} \cap \left\{u^N > s\right\}, \\ b(x, 0)(u^N + s) & \text{for } x \in \Omega_{ions} \cap \left\{u^N < -s\right\}. \end{cases} \tag{3.71}$$

Now we are ready to prove that $u^N \in L^\infty(\Omega)$. First, we consider the case when the charge neutrality condition (3.36) holds. From (3.68), (3.70), and (3.71), it follows that

$$\int_\Omega b(x, u^N + u^L + w) G_s(u^N) dx \geq \int_\Omega b(x, 0) G_s(u^N) dx = 0. \tag{3.72}$$

Moreover, by using the definition of $a(\cdot, \cdot)$ and the expression (3.67) for the weak partial derivatives of $G_s(u^N)$, we obtain

$$\begin{aligned} a(u^N, G_s(u^N)) &= \int_\Omega \epsilon \nabla u^N \cdot \nabla G_s(u^N) = \int_\Omega \epsilon \nabla G_s(u^N) \cdot \nabla G_s(u^N) dx \\ &\geq \epsilon_{\min} \|\nabla G_s(u^N)\|^2_{L^2(\Omega)} \geq \frac{\epsilon_{\min}}{C_P^2} \|G_s(u^N)\|^2_{L^2(\Omega)}, \end{aligned} \tag{3.73}$$

where C_P is the constant in the inequality $\|v\|_{L^2(\Omega)} \leq C_P \|\nabla v\|_{L^2(\Omega)}$ that holds for all $v \in H^1_0(\Omega)$. Finally, using (3.47), (3.72), and (3.73) we obtain

$$\|G_s(u^N)\|^2_{L^2(\Omega)} \leq 0 \text{ for all } s \geq \|u^L + w\|_{L^\infty(\Omega_{ions})}.$$

Consequently $\left|u^N\right| \leq s$ almost everywhere for all $s \geq \|u^L + w\|_{L^\infty(\Omega_{ions})}$.

In the case where the charge neutrality condition (3.36) does not hold, we have

$$\int_\Omega b(x, u^N + u^L + w) G_s(u^N) dx \geq \int_\Omega b(x, 0) G_s(u^N) dx. \tag{3.74}$$

We further estimate $a(u^N, G_s(u^N))$ from below and $-\int_\Omega b(x, 0) G_s(u^N) dx$ from above using the Sobolev embedding $H^1(\Omega) \hookrightarrow L^q(\Omega)$ where $q < \infty$ for $d = 2$, and $q = \frac{2d}{d-2}$ for $d \geq 3$. Let q^* denote the Hölder conjugate to q. Then, $q^* = \frac{q}{q-1} > 1$ for $d = 2$, and $q^* = \frac{2d}{d+2}$ for $d \geq 3$. In order to treat both cases in which we are interested simultaneously, namely $d = 2, 3$, we can take $q = 6$ and $q^* = 6/5$. By C_E we denote the embedding constant in the inequality $\|v\|_{L^6(\Omega)} \leq C_E \|v\|_{H^1(\Omega)}$, $\forall v \in H^1(\Omega)$, which depends only on the domain Ω and d. For $a(u^N, G_s(u^N))$, we have

$$a(u^N, G_s(u^N)) = \int_\Omega \epsilon \nabla G_s(u^N) \cdot \nabla G_s(u^N) dx \geq \frac{\epsilon_{\min}}{1 + C_P^2} \|G_s(u^N)\|^2_{H^1(\Omega)} \tag{3.75}$$

and for $-\int_\Omega b(x, 0) G_s(u^N) dx$ we obtain

$$\begin{aligned} -\int_\Omega b(x, 0) G_s(u^N) dx &= -\int_{A(s)} b(x, 0) G_s(u^N) dx \leq \|b(x, 0)\|_{L^{q^*}(A(s))} \|G_s(u^N)\|_{L^q(\Omega)} \\ &\leq C_E \|b(x, 0)\|_{L^{q^*}(A(s))} \|G_s(u^N)\|_{H^1(\Omega)}, \end{aligned} \tag{3.76}$$

where $A(s) := \{x \in \Omega : \left|u^N(x)\right| > s\}$. Combining (3.47), (3.74), (3.75), and (3.76), we arrive at the estimate

$$\frac{\epsilon_{\min}}{1 + C_P^2} \|G_s(u^N)\|_{H^1(\Omega)} \leq C_E \|b(x, 0)\|_{L^{q^*}(A(s))}. \tag{3.77}$$

The final step before applying Lemma 2.33 is to estimate the left-hand side of (3.77) from below in terms of $|A(h)|$ for $h > s \geq \|u^L + w\|_{L^\infty(\Omega_{ions})}$ and the right-hand side of (3.77) from above in terms of $|A(s)|$. Again, using the Sobolev embedding $H^1(\Omega) \hookrightarrow L^q(\Omega)$ and Hölder's inequality yields

$$\begin{aligned}\|G_s(u^N)\|_{H^1(\Omega)} &\geq \frac{1}{C_E}\left(\int_\Omega \left|G_s(u^N)\right|^q dx\right)^{\frac{1}{q}} = \frac{1}{C_E}\left(\int_{A(s)} \left|\left|u^N\right| - s\right|^q dx\right)^{\frac{1}{q}} \\ &\geq \frac{1}{C_E}\left(\int_{A(h)} (h-s)^q dx\right)^{\frac{1}{q}} = \frac{1}{C_E}(h-s)\,|A(h)|^{\frac{1}{q}} \end{aligned} \tag{3.78}$$

and

$$\|b(x,0)\|_{L^{q^*}(A(s))} \leq \|b(x,0)\|_{L^2(\Omega)} |A(s)|^{\frac{2-q^*}{2q^*}}. \tag{3.79}$$

Combining (3.77), (3.78), and (3.79), we obtain the following inequality for the nonnegative and nonincreasing function $\varphi(t) := |A(t)|$

$$|A(h)| \leq \left(\frac{C_E^2\left(1+C_P^2\right)}{\epsilon_{\min}}\|b(x,0)\|_{L^2(\Omega)}\right)^q \frac{|A(s)|^{\frac{q-2}{2}}}{(h-s)^q} \quad \text{for all } h > s \geq \|u^L + w\|_{L^\infty(\Omega_{ions})}. \tag{3.80}$$

Since $\frac{q-2}{2} = 2 > 1$, by applying Lemma 2.33 we conclude that there is some $e > 0$ such that

$$\begin{aligned} 0 < e^q &= \left(\frac{C_E^2\left(1+C_P^2\right)}{\epsilon_{\min}}\|b(x,0)\|_{L^2(\Omega)}\right)^q \left|A\left(\|u^L + w\|_{L^\infty(\Omega_{ions})}\right)\right|^{\frac{q-4}{2}} 2^{\frac{q(q-2)}{q-4}} \\ &\leq \left(\frac{C_E^2\left(1+C_P^2\right)}{\epsilon_{\min}}\|b(x,0)\|_{L^2(\Omega)}\right)^q |\Omega|^{\frac{q-4}{2}} 2^{\frac{q(q-2)}{q-4}} =: \overline{e}^q \end{aligned}$$

and $\left|A\left(\|u^L + w\|_{L^\infty(\Omega_{ions})} + \overline{e}\right)\right| = 0$. Hence $\|u^N\|_{L^\infty(\Omega)} \leq \|u^L + w\|_{L^\infty(\Omega_{ions})} + \overline{e}$. □

From Proposition 3.18 it follows that the solution u^N to (3.47) is essentially bounded for all three choices of the test space V. As a result, the nonlinearity evaluated at u^N is also essentially bounded and thus by a standard density argument the weak formulations (3.47) for all three choices of V are equivalent. We summarize the obtained in this section results in the following theorem.

Theorem 3.19
If we assume that

(1) the function $\overline{g}$ specifying the Dirichlet boundary condition in (3.41) *is given as the trace of some function in $H^2(\Omega)$,*

(2) in the case of the 3-term splitting $\Gamma \in C^{0,1}$,

then the unique $u^L \in H^1_{\overline{g}}(\Omega)$ *is also in* $L^\infty(\Omega)$. *Moreover, the variational problem* (3.57) *has a unique minimizer* $u^N_{\min} \in H^1_0(\Omega) \cap L^\infty(\Omega)$ *which coincides with the unique solution of problem* (3.47) *for all three choices of the test space* V. *As a result, problem* (3.41) *has a unique solution* $u = u^N + u^L \in H^1_{\overline{g}}(\Omega) \cap L^\infty(\Omega)$ *for all three choices of the test space* V.

Existence, uniqueness, and boundedness of u without the additional splitting into u^N and u^L

Finally, we note that one can obtain directly the existence of a solution to problem (3.41) by considering the variational problem

$$\text{Find } u_{\min} \in H^1_{\overline{g}}(\Omega) \text{ such that } J(u_{\min}) = \min_{v \in H^1_{\overline{g}}(\Omega)} J(v), \tag{3.81}$$

where the functional $J : H^1_{\overline{g}}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is defined by $J := \frac{1}{2}a(v,v) + \int\limits_\Omega B(x, v+w)dx - \int\limits_\Omega \boldsymbol{f} \cdot \nabla v dx$. Again, here it is understood that $J(v) = +\infty$ whenever $B(x, v+w) \notin L^1(\Omega)$, i.e.,

$$J(v) := \begin{cases} \dfrac{1}{2}a(v,v) + \displaystyle\int_\Omega B(x, v+w)dx - \int_\Omega \boldsymbol{f} \cdot \nabla v dx, \text{ if } B(x, v+w) \in L^1(\Omega), \\ \\ +\infty, \text{ if } B(x, v+w) \notin L^1(\Omega). \end{cases} \tag{3.82}$$

It is easy to see that $dom(J)$ is a convex set in $H^1_{\overline{g}}(\Omega)$ and that J is convex on $dom(J)$ (the argument is similar to the one in Remark 3.23). Since the set $H^1_{\overline{g}}$ is a closed convex (and therefore weakly closed) subset of the reflexive Banach space $H^1(\Omega)$, J is strictly convex, proper, coercive, and sequentially weakly lower semicontinuous functional, from Theorem 2.30 it follows that problem (3.81) has a unique minimizer $u_{\min}$ over $H^1_{\overline{g}}$. That $H^1_{\overline{g}}(\Omega)$ is norm closed in $H^1(\Omega)$ and convex follows easily by the linearity and boundedness of the trace operator γ_2. That J is strictly convex, proper, and s.w.l.s.c. follows by similar arguments to the ones made about J^N. It is left to see that J is coercive over $H^1_{\overline{g}}(\Omega)$. Let $u_{\overline{g}} \in H^1(\Omega)$ be such that $\gamma_2(u_{\overline{g}}) = \overline{g}$ on $\partial\Omega$. For any $v \in H^1_{\overline{g}}(\Omega)$, we have $\gamma_2(v - u_{\overline{g}}) = 0$. Since Ω is a bounded Lipschitz domain, it follows by Theorem 2.16 that $v - u_{\overline{g}} \in H^1_0(\Omega)$. By applying Poincaré's inequality we obtain

$$\begin{aligned} \left| \|v\|_{H^1(\Omega)} - \|u_{\overline{g}}\|_{H^1(\Omega)} \right| \le \|v - u_{\overline{g}}\|_{H^1(\Omega)} &\le \sqrt{1 + C_P^2} \|\nabla(v - u_{\overline{g}})\|_{L^2(\Omega)} \\ &\le \sqrt{1 + C_P^2} \left(\|\nabla v\|_{L^2(\Omega)} + \|\nabla u_{\overline{g}}\|_{L^2(\Omega)} \right). \end{aligned} \tag{3.83}$$

After squaring both sides of (3.83) and using the inequality $2ab \le a^2 + b^2$, $\forall a, b \in \mathbb{R}$ we obtain the estimate

$$\|v\|^2_{H^1(\Omega)} - 2\|v\|_{H^1(\Omega)}\|u_{\overline{g}}\|_{H^1(\Omega)} + \|u_{\overline{g}}\|_{H^1(\Omega)} \le 2\left(1 + C_P^2\right)\left(\|\nabla v\|^2_{L^2(\Omega)} + \|\nabla u_{\overline{g}}\|^2_{L^2(\Omega)} \right). \tag{3.84}$$

Now, the coercivity of J follows by recalling that $B(x,s) \geq 0$, $\forall x \in \Omega$, $\forall s \in \mathbb{R}$ and using (3.84):

$$\begin{aligned} J(v) &= \frac{1}{2}a(v,v) + \int_\Omega B(x, v+w)dx - \int_\Omega \boldsymbol{f}\cdot\nabla v dx \geq \frac{\epsilon_{\min}}{2}\|\nabla v\|^2_{L^2(\Omega)} - \|\boldsymbol{f}\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)} \\ &\geq \frac{\epsilon_{\min}}{4\left(1+C_P^2\right)}\left(\|v\|^2_{H^1(\Omega)} - 2\|v\|_{H^1(\Omega)}\|u_{\overline{g}}\|_{H^1(\Omega)} + \|u_{\overline{g}}\|_{H^1(\Omega)}\right) \\ &\quad - \frac{\epsilon_{\min}}{2}\|\nabla u_{\overline{g}}\|^2_{H^1(\Omega)} - \|\boldsymbol{f}\|_{L^2(\Omega)}\|v\|_{H^1(\Omega)} \to +\infty \text{ whenever } \|v\|_{H^1(\Omega)} \to \infty. \end{aligned} \tag{3.85}$$

We can summarize the above considerations in the following theorem.

Theorem 3.20
Problem (3.81) *has a unique solution* $u_{\min} \in H^1_{\overline{g}}(\Omega)$.

By varying the functional J with directions $v \in H^1_0(\Omega) \cap L^\infty(\Omega)$ one can show in a similar way to the approach for J^N that the unique minimizer $u_{\min}$ of J over $H^1_{\overline{g}}(\Omega)$ is a solution to (3.41) for the test spaces $V = H^1_0(\Omega) \cap L^\infty(\Omega)$ and $V = C^\infty_0(\Omega)$. To show that $u_{\min}$ is also a solution to (3.41) with $V = H^1_0(\Omega)$ one has to show the a priori L^∞ bound on the unique solution of (3.41) with $V = H^1_0(\Omega) \cap L^\infty(\Omega)$ and $V = C^\infty_0(\Omega)$ (see p. 52 for the uniqueness of a solution to (3.41)). This can be done in a similar way to the proof of Proposition 3.18 or by observing that $u = u^N + u^L \in L^\infty(\Omega)$. The a priori L^∞ estimate also follows from the more general Theorem 3.29 which we prove in Section 3.4.

Once we know that the solution u of (3.41) with $V = H^1_0(\Omega) \cap L^\infty(\Omega)$ or $V = C^\infty_0(\Omega)$ is in $L^\infty(\Omega)$, it follows that $b(x, u+w) \in L^\infty(\Omega)$ and therefore by a standard density argument we obtain that u is also the unique solution of (3.41) with $V = H^1_0(\Omega)$. We can summarize the above considerations in the following theorem.

Theorem 3.21
The variational problem (3.81) *has a unique minimizer* $u_{\min} \in H^1_{\overline{g}}(\Omega)$ *which coincides with the unique solution of problem* (3.41) *for the test spaces* $V = C^\infty_0(\Omega)$ *and* $V = H^1_0(\Omega) \cap L^\infty(\Omega)$. *Moreover, if we assume that*

(1) the function $\overline{g}$ *specifying the Dirichlet boundary condition in* (3.41) *is given as the trace of some function in* $H^2(\Omega)$,

(2) in the case of the 3-term splitting $\Gamma \in C^{0,1}$,

then $u_{\min} \in L^\infty(\Omega)$.

Remark 3.22
Denote $F(v) := \int_\Omega B(x, v + u^L + w)dx$. *We have a sequence* $\{v_n\}_{n=1}^\infty$ *such that for each subsequence* $\{v_{n_m}\}_{m=1}^\infty$ *one can find a further subsequence* $\{v_{n_{m_s}}\}_{s=1}^\infty$ *for which*

$F(v) \leq \liminf_{s\to\infty} F(v_{n_{m_s}})$. We claim that $F(v) \leq \liminf_{n\to\infty} F(v_n)$. To see this, suppose to the contrary that $F(v) > \liminf_{n\to\infty} F(v_n)$. Then, there exists a subsequence $\{v_{n_m}\}_{m=1}^{\infty}$ such that $\liminf_{n\to\infty} F(v_n) = \lim_{k\to\infty} F(v_{n_m})$ and therefore $F(v) > \lim_{m\to\infty} F(v_{n_m})$. But for this subsequence, according to the assumptions on $\{v_n\}_{n=1}^{\infty}$, we can find a further subsequence $\{v_{n_{m_s}}\}_{s=1}^{\infty}$ such that

$$F(v) \leq \liminf_{s\to\infty} F(v_{n_{m_s}}) = \lim_{s\to\infty} F(v_{n_{m_s}}) = \lim_{m\to\infty} F(v_{n_m}) < F(v),$$

which is a contradiction with the assumption that $F(v) > \liminf_{n\to\infty} F(v_n)$.

Remark 3.23

It is worth noting that $dom(J^N)$ is a linear subspace of $H_0^1(\Omega)$ for $d \leq 2$ and not a linear subspace of $H_0^1(\Omega)$ if $d \geq 3$. In dimension $d \leq 2$, from [110, 179] we know that $\mathrm{e}^v \in L^2(\Omega)$ for any $v \in H_0^1(\Omega)$ and thus $\mathrm{e}^{\lambda v_1 + \mu v_2} \in L^2(\Omega)$ for any $\lambda, \mu \in \mathbb{R}$ and any $v_1, v_2 \in H_0^1(\Omega)$.

On the other hand, if $d \geq 3$, first observe that $dom(J^N) = \left\{v \in H_0^1(\Omega) : B(x, v + \overline{w}) \in L^1(\Omega)\right\}$, where $\overline{w} := u^L + w \in L^\infty(\Omega_{ions})$. For simplicity we consider the case of the PBE, i.e., $B(x, v + \overline{w}) = \overline{k}^2 \cosh(v + \overline{w})$. Let $\Omega = B(0, 1)$ and let $\Omega_{ions} = B(0, \overline{r})$ for some $\overline{r} < 1$, where $B(0, r)$ denotes the ball in $\mathbb{R}^d$, $d \geq 3$ with radius r and a center at 0. We consider the function $v = \ln \frac{1}{|x|} \in H_0^1(B(0, 1))$. Since $\mathrm{e}^v = \frac{1}{|x|} \in L^1(\Omega_{ions})$ and $\mathrm{e}^{\lambda v} = \frac{1}{|x|^\lambda} \notin L^1(\Omega_{ions})$ for any $\lambda \geq d$ [1], we obtain

$$\int_\Omega \overline{k}^2 \cosh(v + \overline{w}) dx = \int_{\Omega_{ions}} \overline{k}_{ions}^2 \frac{\left(\mathrm{e}^{v+\overline{w}} + \mathrm{e}^{-v-\overline{w}}\right)}{2} dx$$

$$\leq \frac{1}{2}\overline{k}_{ions}^2 \mathrm{e}^{\|\overline{w}\|_{L^\infty(\Omega_{ions})}} \int_{\Omega_{ions}} \left(\mathrm{e}^v + \mathrm{e}^{-v}\right) dx \leq \frac{1}{2}\overline{k}_{ions}^2 \mathrm{e}^{\|\overline{w}\|_{L^\infty(\Omega_{ions})}} \left(\int_{\Omega_{ions}} \mathrm{e}^v dx + |\Omega_{ions}| \right) < \infty$$

and

$$\int_\Omega \overline{k}^2 \cosh(\lambda v + \overline{w}) dx \geq \frac{1}{2} \int_{\Omega_{ions}} \overline{k}_{ions}^2 \mathrm{e}^{\lambda v + \overline{w}} dx \geq \frac{1}{2}\overline{k}_{ions}^2 \mathrm{e}^{-\|\overline{w}\|_{L^\infty(\Omega_{ions})}} \int_{\Omega_{ions}} \mathrm{e}^{\lambda v} dx = \infty$$

for any $\lambda > d$.

This means that $v \in dom(J^N)$, but $\lambda v \notin dom(J^N)$ for any $\lambda \geq d$. Therefore $dom(J^N)$ is not a linear space.

However, $dom(J^N) \subset H_0^1(\Omega)$ is a convex set. To see this, let $v_1, v_2 \in dom(J^N)$, i.e., $B(x, v_1 + \overline{w})$, $B(x, v_2 + \overline{w}) \in L^1(\Omega)$. Since $B(x, \cdot)$ is convex it follows that for almost each $x \in \Omega$ and every $\lambda \in [0, 1]$ we have

$$B(x, \lambda v_1(x) + (1 - \lambda)v_2(x) + \overline{w}(x)) \leq \lambda B(x, v_1(x) + \overline{w}(x)) + (1 - \lambda)B(x, v_2(x) + \overline{w}(x)).$$

[1] For any d, using spherical coordinates, we have $\int_{B(0,1)} \frac{1}{|x|^\lambda} dx \sim \int_0^1 \frac{1}{\rho^\lambda} \rho^{d-1} d\rho = \int_0^1 \frac{1}{\rho^{\lambda - d + 1}} d\rho < \infty$ if and only if $\lambda - d + 1 < 1$, i.e., if and only if $\lambda < d$.

By integrating the above inequality over Ω we obtain

$$\int_\Omega B(x,\lambda v_1+(1-\lambda)v_2+\overline{w})dx \le \lambda\int_\Omega B(x,v_1+\overline{w})dx+(1-\lambda)\int_\Omega B(x,v_2+\overline{w})dx<\infty$$

and thus $\lambda v_1+(1-\lambda)v_2 \in dom(J^N)$ for all $\lambda\in[0,1]$. Hence $dom(J^N)$ is convex.

Remark 3.24

One can also see that the functional $\frac{1}{2}a(v,v)$ is sequentially w.l.s.c. by noting that it is convex and continuous in $H_0^1(\Omega)$. From the continuity in $H_0^1(\Omega)$ it follows that it is also sequentially lower semicontinuous. Now, since $\frac{1}{2}a(v,v)$ is convex and s.l.s.c. it follows that it is also sequentially w.l.s.c. (see, e.g. Corollary 3.9 in [96] or Corollary 2.2 in [65]). To see that $\frac{1}{2}a(v,v)$ is continuous in $H_0^1(\Omega)$ observe that

$$\begin{aligned}&\frac{1}{2}\left(a(v,v)-a(u,u)\right)=\frac{1}{2}\left(a(v,v)+a(v,u)-a(v,u)-a(u,u)\right)\\ =&\frac{1}{2}\left(a(v,v-u)+a(v-u,u)\right)\le \epsilon_{\max}\left(\|v\|_{H^1(\Omega)}\|v-u\|_{H^1(\Omega)}+\|v-u\|_{H^1(\Omega)}\|u\|_{H^1(\Omega)}\right),\end{aligned}$$

where $\epsilon_{\max}:=\|\epsilon\|_{L^\infty(\Omega)}$.

Remark 3.25

In dimension $d=3$, the functional $\int_\Omega B(x,v+\overline{w})dx$ is not Gateaux differentiable at any $u\in H_0^1(\Omega)\cap L^\infty(\Omega)$, where we have denoted $\overline{w}:=u^L+w\in L^\infty(\Omega_{ions})$. In fact $\int_\Omega B(x,v+\overline{w})dx$ is discontinuous at every $u\in H_0^1(\Omega)\cap L^\infty(\Omega)$. This is easy to demonstrate with the paradigm of the following example. Let $\Omega_{ions}=B(0,2)\subset\Omega$ be the ball centered at 0 with radius 2 and for simplicity let $B(x,v+\overline{w})=\overline{k}^2\cosh(v+\overline{w})$. There exists a function $z\in H_0^1(\Omega)$ such that $\int_{\Omega_{ions}} \mathrm{e}^{\lambda z}dx=+\infty$ for any $\lambda>0$. In particular, we can set $z=\phi\,|x|^{-1/3}$, where ϕ is a smooth function equal to 1 in $B(0,1)$ and 0 in $\mathbb{R}^3\setminus B(0,2)$. Then $z\in H_0^1(\Omega)$, but $\mathrm{e}^{\lambda z}\notin L^1(\Omega_{ions})$ for any $\lambda>0$ since $\mathrm{e}^{\lambda z}>|x|^{-3}$ for small enough $|x|$. In this case, for any $u\in H_0^1(\Omega)\cap L^\infty(\Omega)$ and any $\lambda>0$ we have

$$\int_\Omega \overline{k}^2\cosh(u+\lambda z+\overline{w})dx\ge\frac{1}{2}\int_{\Omega_{ions}}\overline{k}_{ions}^2\mathrm{e}^{u+\lambda z+\overline{w}}dx\ge\frac{\overline{k}_{ions}^2\mathrm{e}^{-\|u+\overline{w}\|_{L^\infty(\Omega_{ions})}}}{2}\int_{\Omega_{ions}}\mathrm{e}^{\lambda z}dx=+\infty.$$

Remark 3.26

With respect to the general case of (3.41) which includes both the 2-term and the 3-term splittings, for the nonlinearity $b(x,u+w)$, evaluated at the solution u (if it exists), we have that

$$b(x,u+w)v\in L^1(\Omega)\ \text{ for all } v\in V, \tag{3.86}$$

where V is one of the spaces $H_0^1(\Omega)$, $H_0^1(\Omega)\cap L^\infty(\Omega)$, or $C_0^\infty(\Omega)$. Moreover, from (3.41) and the fact that $a(u,\cdot)$, $\int_\Omega \boldsymbol{f}\cdot\nabla(\cdot)dx$ define bounded linear functionals over $H_0^1(\Omega)$ it follows that

the nonlinearity $b(x, u+w)$ defines a bounded linear functional for all elements in the dense subspace V, i.e.,

$$\left| \int_\Omega b(x, u+w)v \right| \leq C\|v\|_{H^1(\Omega)} \text{ for all } v \in V. \tag{3.87}$$

If we could conclude from (3.86) *and* (3.87) *that $b(x, u+w)$ is in $L^{\frac{6}{5}}(\Omega)$ (the inverse of Hölder's inequality), then a density argument will give us that any solution u of* (3.41) *with $V = H_0^1(\Omega) \cap L^\infty(\Omega)$ and $V = C_0^\infty(\Omega)$ is also a solution with $V = H_0^1(\Omega)$. However, the following example, showed to the author by C. Remling [106], demonstrates that this might not be necessarily true: take the function $z = \frac{1}{1-|x|}$, where Ω is the unit ball $B_1 := B(0,1)$ in $\mathbb{R}^3$. Then $z \notin L^1(\Omega)$, but $\left| \int_{B_1} zvdx \right| \lesssim \|v\|_{H^1(B_1)}$, $\forall v \in H_0^1(B_1)$. Indeed, we have*

$$\begin{aligned} \int_{B_1} |zv|dx &= \int_{B_1} \frac{1}{1-|x|}|v|dx = \int_{B_1} \frac{1}{(1-|x|)^{1/4}} \cdot \frac{|v|}{(1-|x|)^{3/4}}dx \\ &\leq \left(\int_{B_1} \frac{dx}{(1-|x|)^{1/2}} \int_{B_1} \frac{v^2}{(1-|x|)^{3/2}}dx \right)^{1/2} = C_1 \left(\int_{B_1} \frac{v^2}{(1-|x|)^{3/2}}dx \right)^{1/2}, \end{aligned} \tag{3.88}$$

where $C_1 = \left(\int_{B_1} \frac{dx}{(1-|x|)^{1/2}} \right)^{1/2}$. If v is a smooth function in $C_0^\infty(B_1)$, we can obtain the estimate

$$\int_{B_1} \frac{v^2(x)}{(1-|x|)^{3/2}}dx \leq C_2\|v\|^2_{H^1(B_1)} \tag{3.89}$$

for some $C_2 > 0$. Now, from (3.88) *and* (3.89) *it follows that*

$$\int_{B_1} |zv|dx \leq C_1C_2\|v\|_{H^1(B_1)}, \quad \forall v \in C_0^\infty(B_1). \tag{3.90}$$

By the density in $H_0^1(\Omega)$ of compactly supported smooth functions, (3.90) *also holds for all $v \in H_0^1(B_1)$. To see this, let $v \in H_0^1(B_1)$ be an arbitrary function. Then there exists a sequence $\{v_n\}_{n=1}^\infty \subset C_0^\infty(B_1)$ such that $\|v_n - v\|_{H^1(B_1)} \to 0$. Passing to a subsequence we obtain that $v_{n_k}(x) \to v(x)$ a.e $x \in B_1$ and also $\|v_{n_k}\|_{H^1(B_1)} \to \|v\|_{H^1(B_1)}$. For each v_{n_k} we have*

$$\int_{B_1} |zv_{n_k}|dx \leq C_1C_2\|v_{n_k}\|_{H^1(B_1)}.$$

We can apply Fatou's lemma to the measurable and positive functions $|zv_{n_k}|$ to obtain

$$\begin{aligned} \int_{B_1} |z(x)v(x)|dx &= \int_{B_1} \liminf_{k\to\infty} |z(x)v_{n_k}(x)|dx \leq \liminf_{k\to\infty} \int_{B_1} |zv_{n_k}|dx \\ &\leq \liminf_{k\to\infty} C_1C_2\|v_{n_k}\|_{H^1(B_1)} = C_1C_2\|v\|_{H^1(B_1)}. \end{aligned}$$

Showing the estimate in (3.89)***:*** *First observe that for any $x \neq 0$ it holds*

$$v(x) = v(x) - v(x/|x|) = -\int_1^{1/|x|} \nabla v(tx) \cdot x dt.$$

By using the Cauchy-Schwartz inequality and the fact that $0 < |x| \leq 1$, we obtain the estimate

$$v^2(x) \leq \int_1^{1/|x|} |x|^2 \, dt \int_0^{1/|x|} |\nabla v(tx)|^2 \, dt = |x| \, (1 - |x|) \int_0^{1/|x|} |\nabla v(tx)|^2 \, dt \tag{3.91}$$

$$\leq \left(\frac{1}{|x|} - 1 \right) \int_1^{1/|x|} |\nabla v(tx)|^2 \, dt. \tag{3.92}$$

By dividing both sides of (3.91) *by $(1 - |x|)^{3/2}$ and then integrating in the disk $A_{1/2} := B_1 \setminus B_{1/2}$ we obtain*

$$\int_{A_{1/2}} \frac{v^2(x)}{(1-|x|)^{3/2}} dx \leq \int_{A_{1/2}} \left(\frac{1}{|x| \sqrt{1-|x|}} \int_1^{1/|x|} |\nabla v(tx)|^2 \, dt \right) dx \leq$$

$$\int_{A_{1/2}} \left(\frac{2}{\sqrt{1-|x|}} \int_1^{1/|x|} |\nabla v(tx)|^2 \, dt \right) dx.$$

Now we make a spherical change of variables, i.e.,

$$\left| \begin{array}{l} x_1 = r \sin\theta \cos\varphi \\ x_2 = r \sin\theta \sin\varphi \\ x_3 = r \cos\theta \end{array} \right.$$

$$\theta \in [0, \pi], \, \varphi \in [0, 2\pi), \, r \in [1/2, 1]$$

and obtain

$$\int_{A_{1/2}} \left(\frac{2}{\sqrt{1-|x|}} \int_1^{1/|x|} |\nabla v(tx)|^2 \, dt \right) dx$$

$$= \int_{1/2}^1 \int_0^\pi \int_0^{2\pi} \frac{2r^2 \sin\theta}{\sqrt{1-r}} \int_1^{1/r} \Big(|\nabla v(tr\sin\theta\cos\varphi, tr\sin\theta\sin\varphi, tr\cos\theta)|^2 \, dt \Big) d\varphi d\theta dr. \tag{3.93}$$

We make one more change of variables in the integral in t: $tr = l$ and we can continue (3.93)

as follows

$$
\begin{aligned}
&\int\limits_{A_{1/2}} \left(\frac{2}{\sqrt{1-|x|}} \int\limits_{1}^{1/|x|} |\nabla v(tx)|^2 \, dt \right) dx \\
&= \int\limits_{1/2}^{1} \left(\frac{2r^2}{\sqrt{1-r}} \int\limits_{0}^{\pi} \int\limits_{0}^{2\pi} \int\limits_{r}^{1} \frac{\sin\theta}{r} \, |\nabla v(l\sin\theta\cos\varphi, l\sin\theta\sin\varphi, l\cos\theta)|^2 \, dl d\varphi d\theta \right) dr \\
&\leq \int\limits_{1/2}^{1} \left(\frac{2r}{\sqrt{1-r}} \int\limits_{0}^{\pi} \int\limits_{0}^{2\pi} \int\limits_{1/2}^{1} 4l^2 \sin\theta \, |\nabla v(l\sin\theta\cos\varphi, l\sin\theta\sin\varphi, l\cos\theta)|^2 \, dl d\varphi d\theta \right) dr \\
&= \int\limits_{1/2}^{1} \frac{8r}{\sqrt{1-r}} dr \int\limits_{A_{1/2}} |\nabla v(x)|^2 \, dx = C_3 \|\nabla v\|^2_{L^2(A_{1/2})},
\end{aligned}
\tag{3.94}
$$

where $C_3 := \int\limits_{1/2}^{1} \frac{8r}{\sqrt{1-r}} dr$. *Finally, by using* (3.94) *we obtain the following estimate:*

$$
\begin{aligned}
&\int\limits_{B_1} \frac{v^2(x)}{(1-|x|)^{3/2}} dx = \int\limits_{B_{1/2}} \frac{v^2(x)}{(1-|x|)^{3/2}} dx + \int\limits_{A_{1/2}} \frac{v^2(x)}{(1-|x|)^{3/2}} dx \\
&\leq \int\limits_{B_{1/2}} \frac{v^2(x)}{(1-0.5)^{3/2}} dx + C_3 \|\nabla v\|^2_{L^2(A_{1/2})} = \sqrt{8}\|v\|^2_{L^2(B_{1/2})} + C_3 \|\nabla v\|^2_{L^2(A_{1/2})} \\
&\leq \sqrt{8}\|v\|^2_{H^1(B_{1/2})} + C_3 \|v\|^2_{H^1(A_{1/2})} \leq \max\{\sqrt{8}, C_3\} \|v\|^2_{H^1(B_1)}.
\end{aligned}
$$

Therefore, (3.89) *holds with* $C_2 := \max\{\sqrt{8}, C_3\}$.

3.3.4 Regularity of the component u in the 2-term and 3-term splittings

We have proved that if $w \in L^\infty(\Omega_{ions})$, $\boldsymbol{f} \in [L^s(\Omega)]^d$ for some $s > d$ and the function $\overline{g}$, specifying the Dirichlet boundary condition, is given as the trace of some function $u_{\overline{g}} \in H^2(\Omega)$ (for example the function $u^L_{\overline{g}}$ in (3.54) on p. 53), then the regular component u in the 2-term and 3-term splittings defined by (3.41) is essentially bounded.

In the case of the 2-term splitting, the conditions on w and $\boldsymbol{f}$ are obviously satisfied since $w = G \in L^\infty(\Omega_{ions})$ and $\boldsymbol{f} = \chi_{\Omega_s}(\epsilon_m - \epsilon_s)\nabla G \in [L^\infty(\Omega)]^d$.

In the case of the 3-term splitting, we have $w = 0$ and $\boldsymbol{f} = -\chi_{\Omega_m}\epsilon_m \nabla u^H + \chi_{\Omega_s}\epsilon_m \nabla G$ which is in $[L^s(\Omega)]^d$ for some $s > d$ if ∇u^H is in $[L^s(\Omega_m)]^d$. The latter is satisfied, for example, if $\Gamma \in C^{0,1}$ (see the discussion in Section 3.2.3).

From the boundedness of u it follows that $b(x, u_0 + u_{\overline{g}} + w) \in L^\infty(\Omega)$, where $u_0 := u - u_{\overline{g}} \in H_0^1(\Omega) \cap L^\infty(\Omega)$. Then u_0 satisfies the linear problem

$$a(u_0, v) = -\int_\Omega b(x, u_0 + u_{\overline{g}} + w)v dx + \int_\Omega (-\epsilon \nabla u_{\overline{g}} + \boldsymbol{f}) \cdot \nabla v dx \text{ for all } v \in H_0^1(\Omega). \tag{3.95}$$

The right-hand side of (3.95) defines a bounded linear functional over the space $W_0^{1,p}(\Omega)$ for some $p < \frac{d}{d-1}$. Therefore, if we additionally assume that $\Gamma \in C^1$, by Theorem 2.34 it follows that $u_0 \in W_0^{1,q}(\Omega)$ for some $q > d$. As a consequence, u is also in $W^{1,q}(\Omega)$ and thus Hölder continuous. The Hölder continuity of u_0 (and hence of u) also follows from the regularity results of De Giorgi-Nash-Moser applied to (3.95) (see, e.g., Theorem 2.12 in [107], p. 65, Theorem 3.5 in [50]). In the case of the 2-term splitting, no assumptions on Γ are required, and in the case of the 3-term splitting we require $\Gamma \in C^{0,1}$ to ensure that $\nabla u^H \in [L^s(\Omega_m)]^d$ for some $s > d$.

Theorem 3.27
Assume that the function $\overline{g}$ prescribing the Dirichlet boundary condition on $\partial\Omega$ is given as the trace of some function $u_{\overline{g}} \in H^2(\Omega)$. The following statements hold true:

(i) the unique $u \in H^1(\Omega)$ in the 2-term splitting defined by the standard weak formulation (3.50) *is Hölder continuous in $\overline{\Omega}$ and thus belongs to $L^\infty(\Omega)$;*

(ii) if $\Gamma \in C^{0,1}$, then the unique $u \in H^1(\Omega)$ in the 3-term splitting, defined by the standard weak formulation (3.52) *is also Hölder continuous in $\overline{\Omega}$ and thus belongs to $L^\infty(\Omega)$;*

(iii) if we assume additionally that $\Gamma \in C^1$, then for both 2-term and 3-term splittings u is in $W^{1,q}(\Omega)$ for some $q > d$ and hence it is also Hölder continuous;

3.4 A more general semilinear elliptic equation

Results on a priori L^∞ estimates for linear elliptic equations of second order appear for example in [112, 174] and for nonlinear elliptic equations in [19, 20, 26, 120, 178]. Vital techniques in the analysis of these papers are different adaptations of the L^∞ regularity procedure introduced by Stampacchia - for example using different "nonlinear" test functions with respect to $G_k(u)$. In [26], the authors prove an L^∞ estimate on the weak solutions, which are already in $W_0^{1,p}(\Omega) \cap L^\infty(\Omega)$, of the general nonlinear problem

$$A(u) + H(x, u, \nabla u) = f(x) - \operatorname{div} g(x) \quad \text{in } \Omega,\ u = 0 \quad \text{on } \partial\Omega \tag{3.96}$$

where A is a Leray-Lions differential operator and H is a nonlinearity which also depends on the gradient of u and such that the growth condition $|H(x, s, \xi)| \leq C_0 + C_1|\xi|^p$, for a.e $x \in \Omega$, $\forall s \in \mathbb{R}$, $\forall \xi \in \mathbb{R}^d$ is fulfilled. More precisely, the authors show that for a weak solution $u \in W_0^{1,p}(\Omega) \cap L^\infty(\Omega)$ it is true that $\|u\|_{L^\infty(\Omega)} \leq \gamma$, where γ depends only on the data. Then they apply this result to a family of approximate equations for (3.96) in order to prove the

existence of at least one bounded solutiuon of (3.96).

In the result presented below, we assume a linear operator A and a nonlinearity $b(x,s)$ which does not depend on the gradient of the solution. We allow for a nonhomogeneous Dirichlet boundary condition prescribed by a function $g \in H^1(\Omega)$ and we do not assume any growth or sign conditions on $b(x,\cdot)$. It seems that such nonlinearities do not satisfy the growth condition posed on H and thus they are not covered by the result in [26]. With the assumptions we make, we prove that every weak solution $u \in H^1_{\gamma_2(g)}(\Omega)$ must be in $L^\infty(\Omega)$ with $\|u\|_{L^\infty(\Omega)} \leq \gamma$ where γ depends only on the data of the problem. As in [26], our L^∞ result seems to be optimal in the sense that when $b(x,\cdot)$ is a linear term, $u \in L^\infty(\Omega)$ for $s > d$, $r > \frac{d}{2}$ which coincides with the classical (optimal) results of Stampacchia, De Giorgi, and Moser in the linear case (see, e.g. the references in [26]) . In [20, 178], the authors prove L^∞ estimates on the solution of very general nonlinear elliptic equations but with a nonlinear first order term with a growth condition which seems not to cover the case of exponential nonlinearities with respect to u, as it is the case of the general PBE, and homogeneous Dirichlet boundary conditions. In [19, 120], L^∞ estimates are proved for nonlinear elliptic equations with homogeneous Dirichlet boundary conditions and with degenerate coercivity but without a nonlinear first order term.

Definition 3.28 (see, e.g. Definition 3.5 in [55])
Let $\Omega \subset \mathbb{R}^n$ be an open set and let $f : \Omega \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$. Then f is said to be a Carathéodory function if

(i) $s \mapsto f(x,s)$ is continuous for almost every $x \in \Omega$,

(ii) $x \mapsto f(x,s)$ is measurable for every $s \in \mathbb{R}$.

If f is a Carathéodory function and $u : \Omega \to \mathbb{R}$ is measurable, then it follows that the function $g : \Omega \to \mathbb{R} \cup \{+\infty\}$ defined by $g(x) = f(x, u(x))$ is measurable (see, e.g., Proposition 3.7 in [55]).

Theorem 3.29 (A priori L^∞ estimate)
Let $\Omega \subset \mathbb{R}^d$, $d \geq 2$ be a bounded Lipschitz domain and let $b(x,s) : \Omega \times \mathbb{R} \to \mathbb{R}$ be a Carathéodory function such that

$$c_1(x,s) \leq b(x,s) \leq c_2(x,s) \quad \text{for a.e } x \in \Omega, \forall s \in \mathbb{R}, \tag{3.97}$$

where $c_1, c_2 : \Omega \times \mathbb{R} \to \mathbb{R}$ are Carathéodory functions which are nondecreasing in the second argument for a.e $x \in \Omega$. Let $a(u,v) = \int_\Omega \boldsymbol{A}\nabla u \cdot \nabla v dx = \int_\Omega \sum_{i,j=1}^d a_{ij}(x) \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} dx$, where $\boldsymbol{A} = (a_{ij})$, $a_{ij}(x) \in L^\infty(\Omega)$, and $a_{ij}(x)$ satisfy the uniform ellipticity condition

$$\sum_{i,j=1}^d a_{ij}(x)\xi_i\xi_j \geq \mu_1 |\xi|^2, \forall x \in \Omega, \forall \xi \in \mathbb{R}^d$$

for some positive constant μ_1. *Finally, let* $g \in H^1(\Omega)$ *and let*

$$\begin{aligned} &u \in H^1_{\gamma_2(g)}(\Omega) \text{ be such that } b(x,u+w)v \in L^1(\Omega),\ \forall v \in V \text{ and} \\ &a(u,v) + \int_\Omega b(x,u+w)v dx = \int_\Omega \left(f_0 v + \boldsymbol{f}\cdot\nabla v\right) dx,\ \forall v \in V, \end{aligned} \tag{3.98}$$

where $\boldsymbol{f} = (f_1, \ldots, f_d)$ *and the test space* V *can be either* $C_0^\infty(\Omega)$, $H_0^1(\Omega) \cap L^\infty(\Omega)$, *or* $H_0^1(\Omega)$. *Provided that* $g, w \in L^\infty(\Omega)$, ∇g, $\boldsymbol{f} \in [L^s(\Omega)]^d$, f_0, $c_1(x, g+w)$, $c_2(x, g+w) \in L^r(\Omega)$, *where* $s > d$, *and* $r > \frac{d}{2}$, *then* $\|u\|_{L^\infty(\Omega)} \leq \gamma$ *where* γ *depends only on the data, i.e,* μ_1, $|\Omega|$, $\|a_{ij}\|_{L^\infty(\Omega)}$, $\|g\|_{L^\infty(\Omega)}$, $\|w\|_{L^\infty(\Omega)}$, $\|\nabla g\|_{L^s(\Omega)}$, $\|\boldsymbol{f}\|_{L^s(\Omega)}$, $\|f_0\|_{L^r(\Omega)}$, $\|c_1(x, g+w)\|_{L^r(\Omega)}$, $\|c_2(x, g+w)\|_{L^r(\Omega)}$.

Remark 3.30
Note that in Theorem 3.29, we do not assert existence of a solution to problem (3.98).

Remark 3.31
Note that in Theorem 3.29, if $b(x,\cdot)$ *is nondecreasing for almost each* $x \in \Omega$, *then* $c_1(x,\cdot)$ *and* $c_2(x,\cdot)$ *can be taken equal to* $b(x,\cdot)$. *Notice also that there is neither sign condition nor growth condition on the nonlinearity* $b(x,\cdot)$.

Remark 3.32
Note that in Theorem 3.29, it is enough to consider a domain Ω *that is bounded and such that the trace operator is well defined and the Sobolev embeddings that are used in the proof below hold.*

Remark 3.33
Since $c_1(x,\cdot)$ *and* $c_2(x,\cdot)$ *are nondecreasing it follows that*

$$c_1\left(x, -\|g+w\|_{L^\infty(\Omega)}\right) \leq c_1(x, g+w) \leq c_1\left(x, \|g+w\|_{L^\infty(\Omega)}\right)$$

and

$$c_2\left(x, -\|g+w\|_{L^\infty(\Omega)}\right) \leq c_2(x, g+w) \leq c_2\left(x, \|g+w\|_{L^\infty(\Omega)}\right).$$

Then the condition that $c_1(x, g+w)$, $c_2(x, g+w) \in L^r(\Omega)$ *where* $r > \frac{d}{2}$ *can be achieved if* $c_1(x,s)$ *and* $c_2(x,s)$ *define functions in* $L^r(\Omega)$ *for every* $s \in \mathbb{R}$. *For example, this condition will be fulfilled if* $c_1(x,s) = k_1(x)a_1(s)$ *and* $c_2(x,s) = k_2(x)a_2(s)$ *where* $k_1, k_2 \geq 0$, $k_1, k_2 \in L^r(\Omega)$ *and* $a_1, a_2 : \mathbb{R} \to \mathbb{R}$ *are nondecreasing and continuous functions.*

Remark 3.34
From Theorem 3.29 it follows that the nonlinearity evaluated at the solution u *(if it exists), is in* $L^\infty(\Omega)$, *i.e.,* $b(x, u+w) \in L^\infty(\Omega)$. *Therefore, the classical regularity results of De Giorgi-Nash-Moser (see, e.g., Theorem 2.12 in [107], p. 65, Theorem 3.5 in [50]) for linear elliptic equations can be applied to the unique solution* z *(by Lax-Milgram Theorem) of the linear equation*

$$a(z,v) = \int_\Omega \left[\left(-b(x,u+w) + f_0\right)v + \boldsymbol{f}\cdot\nabla v\right] dx,\ \forall v \in H_0^1(\Omega). \tag{3.99}$$

and conclude that $z \equiv u$ is Hölder continuous under the assumption that the trace of the function g on $\partial\Omega$, prescribing the Dirichlet boundary condition, is also Hölder continuous.

Proof of Theorem 3.29. The idea of the proof is based on the proof of Theorem B.2 in [112], where the L^∞ estimate is proved for a linear elliptic problem tested with the space $V = H_0^1(\Omega)$. We recall that the set $H^1_{\gamma_2(g)}(\Omega)$ is defined as all functions in $H^1(\Omega)$ that are equal to $\gamma_2(g)$ on $\partial\Omega$ in the sense of traces. We rewrite (3.98) in the following homogenized form:

$$a(u-g,v) = -\int_\Omega b(x,u+w)v dx - a(g,v) + \int_\Omega (f_0 v + \boldsymbol{f}\cdot\nabla v)\,dx,\ \forall v \in V. \tag{3.100}$$

Similarly to [112], we construct the following test functions

$$G_k(u-g) = (u-g)_k := \begin{cases} u-g-k, & \text{a.e on } \{(u-g)(x) > k\}, \\ 0, & \text{a.e on } \{|(u-g)(x)| \leq k\}, \\ u-g+k, & \text{a.e on } \{(u-g)(x) < -k\}, \end{cases} \tag{3.101}$$

for any $k \geq 0$, where $(u-g)_0 := u-g$. Since $G_k(t) = \text{sign}(t)(|t|-k)^+$ is Lipschitz continuous with $G_k(0) = 0$ and $u-g \in H_0^1(\Omega)$, by Stampacchia's theorem (e.g., see [90, 112]) it follows that $G_k(u-g) \in H_0^1(\Omega)$, $\forall k \geq 0$. Moreover, the weak partial derivatives are given by

$$\frac{\partial (u-g)_k}{\partial x_i} = \begin{cases} \frac{\partial (u-g)}{\partial x_i}, & \text{a.e on } \{(u-g)(x) > k\}, \\ 0, & \text{a.e on } \{|(u-g)(x)| \leq k\}, \\ \frac{\partial (u-g)}{\partial x_i}, & \text{a.e on } \{(u-g)(x) < -k\}. \end{cases} \tag{3.102}$$

We have the Sobolev Embedding $H^1(\Omega) \hookrightarrow L^q(\Omega)$ where $q = \infty$ for $d = 1$, $q < \infty$ for $d = 2$, and $q = \frac{2d}{d-2}$ for $d > 2$. With q^* we will denote the Hölder conjugate to q. Thus $q^* = 1$ for $d = 1$, $q^* = \frac{q}{q-1} > 1$ for $d = 2$, and $q^* = \frac{2d}{d+2}$ for $d > 2$. With C_E we denote the embedding constant in the inequality $\|u\|_{L^q(\Omega)} \leq C_E \|u\|_{H^1(\Omega)}$, which depends only on the domain Ω, d, and q.

Testing with $G_k(u-g)$

By applying again Theorem 2.18, we will show that we can test equation (3.100) with $G_k(u-g)$ for any $k \geq 0$, as well as with $u-g$, which is not obvious because $G_k(u-g)$ and $u-g$ need not be in the test space V. For this observe that

$$\int_\Omega b(x,u+w)v dx = -a(u,v) + \int_\Omega (f_0 v + \boldsymbol{f}\cdot\nabla v)\,dx,\ \forall v \in V \tag{3.103}$$

and that the right-hand side of (3.103) defines a bounded linear functional over $H_0^1(\Omega)$:

$$|a(u,v)| \leq \left(\sum_{i,j=1}^d \|a_{ij}\|_{L^\infty(\Omega)}\right) \|u\|_{H^1(\Omega)}\|v\|_{H^1(\Omega)},\ \forall v \in H_0^1(\Omega) \tag{3.104}$$

and

$$\begin{aligned} &\left|\int_\Omega (f_0 v + \boldsymbol{f}\cdot\nabla v)\,dx\right| \leq \|f_0\|_{L^{q^*}(\Omega)}\|v\|_{L^q(\Omega)} + \|\boldsymbol{f}\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)} \\ &\leq C_E\|f_0\|_{L^{q^*}(\Omega)}\|v\|_{H^1(\Omega)} + \|\boldsymbol{f}\|_{L^2(\Omega)}\|v\|_{H^1(\Omega)},\ \forall v \in H^1(\Omega). \end{aligned} \tag{3.105}$$

From (3.103), (3.104), and (3.105), it is clear that the linear functional T_b defined by the formula $\langle T_b, v\rangle = \int_\Omega b(x,u+w)v dx$, $\forall v \in V$ is bounded in the norm of $H^1(\Omega)$ over the dense subspace V and therefore it can be uniquely extended by continuity to a functional $\overline{T}_b \in H^{-1}(\Omega)$ over the whole space $H_0^1(\Omega)$. By Theorem 2.18, if we show that $b(x,u+w)(u-g)_k \geq f_k(x)$ for some function $f_k \in L^1(\Omega)$, then it will follow that $b(x,u+w)(u-g)_k \in L^1(\Omega)$ and that $\langle \overline{T}_b, (u-g)_k\rangle = \int_\Omega b(x,u+w)(u-g)_k dx$. Since the extension $\overline{T}_b$ is also equal to the right hand side of (3.103), we will also obtain that

$$\begin{aligned} a(u-g,(u-g)_k) = -\int_\Omega b(x,u+w)(u-g)_k dx - a(g,(u-g)_k) \\ + \int_\Omega \left(f_0(u-g)_k + \boldsymbol{f}\cdot\nabla(u-g)_k\right) dx,\ \forall k \geq 0. \end{aligned} \tag{3.106}$$

By using the definition (3.101) of $(u-g)_k$ we can write

$$b(x,u+w)(u-g)_k = \begin{cases} b(x,u+w)(u-g-k), & \text{a.e on } \{(u-g)(x) > k\}, \\ 0, & \text{a.e on } \{|(u-g)(x)| \leq k\}, \\ b(x,u+w)(u-g+k), & \text{a.e on } \{(u-g)(x) < -k\}. \end{cases}$$

Therefore, on the set $\{(u-g)(x) > k\}$ we obtain the estimate

$$b(x,u+w)(u-g-k) \geq c_1(x,u+w)(u-g-k) \geq c_1(x,g+w)(u-g-k), \tag{3.107}$$

and on the set $\{(u-g)(x) < -k\}$ the estimate

$$b(x,u+w)(u-g+k) \geq c_2(x,u+w)(u-g+k) \geq c_2(x,g+w)(u-g+k). \tag{3.108}$$

If we define the function $f_k(x)$ through the equality

$$f_k(x) := \begin{cases} c_1(x,g(x)+w(x))(u(x)-g(x)-k), & \text{a.e on } \{(u-g)(x) > k\}, \\ 0, & \text{a.e on } \{|(u-g)(x)| \leq k\}, \\ c_2(x,g(x)+w(x))(u(x)-g(x)+k), & \text{a.e on } \{(u-g)(x) < -k\}, \end{cases} \tag{3.109}$$

then f_k will be in $L^1(\Omega)$ if $c_1(x,g+w)(u-g-k)$ and $c_2(x,g+w)(u-g+k) \in L^1(\Omega)$, since

$$|f_k(x)| \leq |c_1(x,g(x))(u(x)-g(x)-k)| + |c_2(x,g(x))(u(x)-g(x)+k)| \text{ a.e } x \in \Omega.$$

To ensure that $c_1(x,g+w)(u-g-k) \in L^1(\Omega)$ and $c_2(x,g+w)(u-g+k) \in L^1(\Omega)$, it is enough to require that $c_1(x,g+w)$, $c_2(x,g+w) \in L^{q^*}(\Omega)$. In this case, by Theorem 2.18 it follows that $b(x,u+w)(u-g)_k \in L^1(\Omega)$ for each $k \geq 0$ and that (3.106) holds.

Estimation of the terms in (3.106)

Now the goal is to show that the measure of the set $A(k)$ becomes zero for all $k \geq k_1 > 0$, where for $k \geq 0$ the set $A(k)$ is defined by

$$A(k) := \{x \in \Omega :\ |(u-g)(x)| > k\}.$$

This would mean that $|u-g| \leq k_1$ for almost each $x \in \Omega$. The idea to show this is to obtain an inequality of the form (2.11) in Lemma 2.33 for the nonnegative and nonincreasing function $\varphi(k) := |A(k)|$. To obtain such an inequality we estimate from below the term on the left-hand side of (3.106) and from above the terms on the right-hand side of (3.106).

First, by using (3.107) and (3.108) we observe that

$$\begin{aligned}
&\int_\Omega b(x,u+w)(u-g)_k dx = \int_{A(k)} b(x,u+w)(u-g)_k dx \\
&= \int_{\{u-g>k\}} b(x,u+w)(u-g-k)dx + \int_{\{u-g<-k\}} b(x,u+w)(u-g+k)dx \\
&\geq \int_{\{u-g>k\}} c_1(x,g+w)(u-g-k)dx + \int_{\{u-g<-k\}} c_2(x,g+w)(u-g+k)dx \\
&= \int_{A(k)} c(x,g+w)(u-g)_k dx,
\end{aligned} \tag{3.110}$$

where the function $c : \Omega \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$c(x,g(x)+w(x)) := \begin{cases} c_1(x,g(x)+w(x)), & \text{a.e on } \{(u-g)(x) \geq 0\}, \\ c_2(x,g(x)+w(x)), & \text{a.e on } \{(u-g)(x) < 0\}. \end{cases}$$

Now, we estimate the left-hand side of (3.106) from below. First by using the expression (3.102) for the weak partial derivatives of $G_k(u-g)$, then the coercivity of $a(\cdot,\cdot)$, and finally Poincaré's inequality, we obtain

$$\begin{aligned}
&a(u-g,(u-g)_k) = \int_\Omega \sum_{i,j=1}^d a_{ij} \frac{\partial(u-g)}{\partial x_i} \frac{\partial(u-g)_k}{\partial x_j} dx \\
&= \int_{A(k)} \sum_{i,j=1}^d a_{ij} \frac{\partial(u-g)_k}{\partial x_i} \frac{\partial(u-g)_k}{\partial x_j} dx \\
&= a((u-g)_k,(u-g)_k) \geq \mu_1 \|\nabla(u-g)_k\|^2_{L^2(\Omega)} \geq \frac{\mu_1}{C_P^2+1} \|(u-g)_k\|^2_{H^1(\Omega)}
\end{aligned} \tag{3.111}$$

By combining (3.106) with the estimates (3.110) and (3.111) we obtain the intermediate estimate

$$\begin{aligned}
\frac{\mu_1}{C_P^2+1}\|(u-g)_k\|^2_{H^1(\Omega)} \leq &\left| \int_{A(k)} c(x,g+w)(u-g)_k dx \right| + |a(g,(u-g)_k)| \\
&+ \left| \int_{A(k)} f_0 (u-g)_k dx \right| + \left| \int_{A(k)} \boldsymbol{f} \cdot \nabla(u-g)_k dx \right|.
\end{aligned} \tag{3.112}$$

We continue by estimating from above all terms on the right-hand side of (3.112). By applying Hölder's inequality we obtain

$$\begin{aligned}\left|\int\limits_{A(k)} f_0(u-g)_k dx\right| &\leq \|f_0\|_{L^{q^*}(A(k))}\|(u-g)_k\|_{L^q(\Omega)} \\ &\leq C_E\|f_0\|_{L^{q^*}(A(k))}\|(u-g)_k\|_{H^1(\Omega)}.\end{aligned} \tag{3.113}$$

Thus if $f_0 \in L^r(\Omega)$ with $r > q^*$, again by using Hölder's inequality we obtain

$$\|f_0\|^{q^*}_{L^{q^*}(A(k))} = \int\limits_{A(k)} \underbrace{|f_0|^{q^*}}_{\in L^{\frac{r}{q^*}}(\Omega)} 1dx \leq \left(\int\limits_{A(k)} |f_0|^r\, dx\right)^{\frac{q^*}{r}} \left(\int\limits_{A(k)} 1dx\right)^{\frac{r-q^*}{r}} = \|f_0\|^{q^*}_{L^r(A(k))}|A(k)|^{\frac{r-q^*}{r}}.$$

By combining the last estimate with (3.113), we obtain

$$\left|\int_{A(k)} f_0(u-g)_k dx\right| \leq C_E\|f_0\|_{L^r(\Omega)}|A(k)|^{\frac{r-q^*}{rq^*}}\,\|(u-g)_k\|_{H^1(\Omega)}. \tag{3.114}$$

Similarly, we estimate ($r > q^*$)

$$\begin{aligned}\left|\int\limits_{A(k)} c(x,g+w)(u-g)_k dx\right| &\leq \|c(x,g+w)\|_{L^{q^*}(A(k))}\|(u-g)_k\|_{L^q(\Omega)} \\ &\leq C_E\|c(x,g+w)\|_{L^r(\Omega)}|A(k)|^{\frac{r-q^*}{rq^*}}\,\|(u-g)_k\|_{H^1(\Omega)}.\end{aligned} \tag{3.115}$$

We continue with the estimation of the fourth term in (3.112):

$$\left|\int\limits_{A(k)} \boldsymbol{f}\cdot\nabla(u-g)_k dx\right| \leq \|\boldsymbol{f}\|_{L^2(A(k))}\|(u-g)_k\|_{H^1(\Omega)} \tag{3.116}$$

If $\boldsymbol{f} \in [L^s(\Omega)]^d$ with $s > 2$, by using Hölder's inequality we obtain

$$\|\boldsymbol{f}\|^2_{L^2(A(k))} = \int\limits_{A(k)} \underbrace{|\boldsymbol{f}|^2}_{\in L^{\frac{s}{2}}(\Omega)} 1dx \leq \left(\int\limits_{A(k)} |\boldsymbol{f}|^s\, dx\right)^{\frac{2}{s}} \left(\int\limits_{A(k)} 1dx\right)^{\frac{s-2}{s}} = \|\boldsymbol{f}\|^2_{L^s(A(k))}|A(k)|^{\frac{s-2}{s}},$$

and hence by combining with (3.116), we arrive at the estimate

$$\left|\int\limits_{A(k)} \boldsymbol{f}\cdot\nabla(u-g)_k dx\right| \leq \|\boldsymbol{f}\|_{L^s(\Omega)}|A(k)|^{\frac{s-2}{2s}}\,\|(u-g)_k\|_{H^1(\Omega)}. \tag{3.117}$$

It is left to estimate the second term on the right-hand side of (3.112):

$$|a(g,(u-g)_k)| = \left| \int_\Omega \sum_{i,j=1}^d a_{ij} \frac{\partial g}{\partial x_i} \frac{\partial (u-g)_k}{\partial x_j} dx \right| = \left| \int_{A(k)} \sum_{i,j=1}^d a_{ij} \frac{\partial g}{\partial x_i} \frac{\partial (u-g)}{\partial x_j} dx \right|$$
$$\le \sum_{i,j=1}^d \|a_{ij}\|_{L^\infty(\Omega)} \int_{A(k)} \left| \frac{\partial g}{\partial x_i} \right| \left| \frac{\partial (u-g)}{\partial x_j} \right| dx \le \left(\sum_{i,j=1}^d \|a_{ij}\|_{L^\infty(\Omega)} \right) \int_{A(K)} |\nabla g| \, |\nabla (u-g)| \, dx$$
$$\le \left(\sum_{i,j=1}^d \|a_{ij}\|_{L^\infty(\Omega)} \right) \|\nabla g\|_{L^2(A(k))} \|(u-g)_k\|_{H^1(\Omega)}. \tag{3.118}$$

Similarly to the estimate of $\|\boldsymbol{f}\|_{L^2(A(k))}$, if $\nabla g \in [L^s(\Omega)]^d$ with $s > 2$, by applying Hölder's inequality we can obtain the estimate

$$\|\nabla g\|^2_{L^2(A(k))} = \int_{A(k)} \underbrace{|\nabla g|^2}_{\in L^{\frac{s}{2}}(\Omega)} 1 dx \le \left(\int_{A(k)} |\nabla g|^s \, dx \right)^{\frac{2}{s}} \left(\int_{A(k)} 1 dx \right)^{\frac{s-2}{s}} = \|\nabla g\|^2_{L^s(A(k))} |A(k)|^{\frac{s-2}{s}}$$

and by combining with (3.118), we arrive at the estimate

$$|a(g,(u-g)_k)| \le \left(\sum_{i,j=1}^d \|a_{ij}\|_{L^\infty(\Omega)} \right) \|\nabla g\|_{L^s(\Omega)} |A(k)|^{\frac{s-2}{2s}} \|(u-g)_k\|_{H^1(\Omega)}. \tag{3.119}$$

Combining (3.112) with the estimates (3.114), (3.115), (3.117), and (3.119) for the right-hand side terms in (3.112), and then dividing by $\|(u-g)_k\|_{H^1(\Omega)}$, we obtain

$$\begin{aligned} &\frac{\mu_1}{C_P^2+1} \|(u-g)_k\|_{H^1(\Omega)} \\ &\quad \le C_E \|c(x,g+w)\|_{L^r(\Omega)} |A(k)|^{\frac{r-q^*}{rq^*}} + \left(\sum_{i,j=1}^d \|a_{ij}\|_{L^\infty(\Omega)} \right) \|\nabla g\|_{L^s(\Omega)} |A(k)|^{\frac{s-2}{2s}} \\ &\qquad + C_E \|f_0\|_{L^r(\Omega)} |A(k)|^{\frac{r-q^*}{rq^*}} + \|\boldsymbol{f}\|_{L^s(\Omega)} |A(k)|^{\frac{s-2}{2s}}. \end{aligned} \tag{3.120}$$

Now, it is left to estimate the left-hand side of (3.120) from below in terms of the measure of the set $A(h)$ for $h > k$. We use again the Sobolev embedding theorem and the fact that

$A(k) \supset A(h)$, $\forall h > k \geq 0$:

$$\begin{aligned}
\|(u-g)_k\|_{H^1(\Omega)} &\geq \frac{1}{C_E}\|(u-g)_k\|_{L^q(\Omega)} = \frac{1}{C_E}\left(\int\limits_{\Omega} |(u-g)_k|^q\, dx\right)^{\frac{1}{q}} \\
&= \frac{1}{C_E}\left(\int\limits_{A(k)} \Big|\underbrace{|u-g|-k}_{>0}\Big|^q\, dx\right)^{\frac{1}{q}} \\
&= \frac{1}{C_E}\left(\int\limits_{A(k)\setminus A(h)} (|u-g|-k)^q\, dx + \int\limits_{A(h)} (|u-g|-k)^q\, dx\right)^{\frac{1}{q}} \\
&\geq \frac{1}{C_E}\left(\int\limits_{A(h)} (h-k)^q dx\right)^{\frac{1}{q}} = \frac{1}{C_E}(h-k)\,|A(h)|^{\frac{1}{q}}\,.
\end{aligned} \tag{3.121}$$

From (3.120) and (3.121) it follows that

$$\begin{aligned}
&(h-k)\,|A(h)|^{\frac{1}{q}} \\
&\leq \frac{C_E(C_P^2+1)}{\mu_1}\left[C_E\|c(x,g+w)\|_{L^r(\Omega)}|A(k)|^{\frac{r-q^*}{rq^*}} + \left(\sum_{i,j=1}^{d}\|a_{ij}\|_{L^\infty(\Omega)}\right)\|\nabla g\|_{L^s(\Omega)}|A(k)|^{\frac{s-2}{2s}}\right. \\
&\left.+C_E\|f_0\|_{L^r(\Omega)}|A(k)|^{\frac{r-q^*}{rq^*}} + \|\boldsymbol{f}\|_{L^s(\Omega)}|A(k)|^{\frac{s-2}{2s}}\right] \leq C_M\left(|A(k)|^{\frac{s-2}{2s}} + |A(k)|^{\frac{r-q^*}{rq^*}}\right),
\end{aligned} \tag{3.122}$$

where

$$\begin{aligned}
&C_M := \\
&\frac{C_E(C_P^2+1)}{\mu_1}\max\left\{C_E\left(\|c(x,g+w)\|_{L^r(\Omega)}+\|f_0\|_{L^r(\Omega)}\right), \left(\sum_{i,j=1}^{d}\|a_{ij}\|_{L^\infty(\Omega)}\right)\|\nabla g\|_{L^s(\Omega)}+\|\boldsymbol{f}\|_{L^s(\Omega)}\right\}.
\end{aligned}$$

We have obtained the following inequality for the measure of $A(k)$:

$$(h-k)\,|A(h)|^{\frac{1}{q}} \leq C_M\left(|A(k)|^{\frac{s-2}{2s}} + |A(k)|^{\frac{r-q^*}{rq^*}}\right), \forall h > k \geq 0. \tag{3.123}$$

Since $u-g$ is summable it follows that $|A(k)| = \operatorname{meas}(\{x \in \Omega : |(u-g)(x)| > k\}) \to 0$ monotonically decreasingly as $k \to \infty$. For this reason, there exists a $k_0 > 0$ such that $|A(k)| \leq 1$, $\forall k \geq k_0$ (if $|\Omega| \leq 1$, this is satisfied for all $k \geq 0$). Therefore (3.123) takes the form

$$(h-k)\,|A(h)|^{\frac{1}{q}} \leq 2C_M\,|A(k)|^{\min\{\frac{s-2}{2s}, \frac{r-q^*}{rq^*}\}}, \forall h > k \geq k_0,$$

which is equivalent to the inequality

$$|A(h)| \leq (2C_M)^q \frac{|A(k)|^{\min\{\frac{s-2}{2s}, \frac{r-q^*}{rq^*}\}q}}{(h-k)^q}, \forall h > k \geq k_0. \tag{3.124}$$

However, we want to find a k_0 which depends only on the data of the problem. For this, observe that from (3.112) for $k = 0$, using Hölder's inequality and the embedding $H^1(\Omega) \hookrightarrow L^q(\Omega)$, we have

$$\begin{aligned}
&\frac{\mu_1}{C_P^2+1}\|u-g\|^2_{H^1(\Omega)} \\
\leq& C_E\|c(x,g+w)\|_{L^{q^*}(\Omega)}\|u-g\|_{H^1(\Omega)} + \left(\sum_{i,j=1}^d \|a_{ij}\|_{L^\infty(\Omega)}\right)\|\nabla g\|_{L^2(\Omega)}\|u-g\|_{H^1(\Omega)} \\
&+ C_E\|f_0\|_{L^{q^*}(\Omega)}\|u-g\|_{H^1(\Omega)} + \|\boldsymbol{f}\|_{L^2(\Omega)}\|u-g\|_{H^1(\Omega)}.
\end{aligned} \tag{3.125}$$

By dividing both sides of (3.125) by $\|u-g\|_{H^1(\Omega)}$, for arbitrary $k \geq 0$, we obtain

$$k\,|A(k)|^{\frac{1}{2}} \leq \left(\int_\Omega |u-g|^2\,dx\right)^{\frac{1}{2}} = \|u-g\|_{L^2(\Omega)} \tag{3.126}$$

$$\leq \frac{C_P^2+1}{\mu_1}\left(C_E\|c(x,g+w)\|_{L^{q^*}(\Omega)} + \left(\sum_{i,j=1}^d \|a_{ij}\|_{L^\infty(\Omega)}\right)\|\nabla g\|_{L^2(\Omega)} + C_E\|f_0\|_{L^{q^*}(\Omega)} + \|\boldsymbol{f}\|_{L^2(\Omega)}\right).$$

If we denote by C_D the constant on the right hand side of inequality (3.126), which depends only on the data of the problem (3.98), then a sufficient condition for $|A(k)| \leq 1$ will be

$$\frac{C_D^2}{k^2} \leq 1,$$

which is equivalent to $k \geq C_D =: k_0$. Here we recall that for $d = 2$, q^* can be any number greater than 1 and for $d > 2$ we have $q = \frac{2d}{d-2}$. Since we have required $r > q^*$, the constant C_D is well defined. In order to apply Lemma 2.33 to the nonnegative and nonincreasing function $\varphi(k) = |A(k)|$ we need to ensure that

$$\min\left\{\frac{s-2}{2s}, \frac{r-q^*}{rq^*}\right\} > \frac{1}{q},$$

which is equivalent to

$$\frac{s-2}{2s} > \frac{1}{q} \qquad \text{and} \qquad \frac{r-q^*}{rq^*} > \frac{1}{q}. \tag{3.127}$$

The first inequality in (3.127) is equivalent to $s > \frac{2q}{q-2}$ and the second to $r > \frac{q}{q-2}$. We also recall that in the course of the proof we have required that $s > 2$.

- For $d = 1$, to see that $\|u - g\|_{L^\infty(\Omega)} \leq \gamma$, where γ depends only on the data, it is enough to divide both sides of (3.125) by $\|u - g\|_{H^1(\Omega)}$ and apply the embedding inequality $\|u - g\|_{L^\infty(\Omega)} \leq C_E \|u - g\|_{H^1(\Omega)}$. From (3.125) it is clear that the requirements on s and r would be $s > 2$, $r \geq 1$.
- For $d = 2$, we have $H^1(\Omega) \hookrightarrow L^q(\Omega)$ for any $q < \infty$. In this case the requirements on s and r become $s > 2$, $r > 1$.
- For $d = 3$, we have $H^1(\Omega) \hookrightarrow L^q(\Omega)$ where $q = \frac{2d}{d-2}$ and $q^* = \frac{2d}{d+2}$. In this case the requirements on s and r are $s > d$, $r > \frac{d}{2}$.

We can summarize the conditions on s and r for $d \geq 2$ as $s > d$ and $r > \frac{d}{2}$. Now, if we denote $\beta := \min\{\frac{s-2}{2s}, \frac{r-q^*}{rq^*}\}q$ from Lemma 2.33 it follows that there exists a constant e, defined by $e^q := (2C_M)^q |A(k_0)|^{\beta-1} 2^{\frac{q\beta}{\beta-1}}$ such that $|A(k_0 + e)| = 0$. Since $|A(k_0)| \leq |\Omega|$, we can write $|A(k_1)| = 0$, where $k_1 := k_0 + \left((2C_M)^q |\Omega|^{\beta-1} 2^{\frac{q\beta}{\beta-1}}\right)^{\frac{1}{q}} = C_D + (2C_M) |\Omega|^{\frac{\beta-1}{q}} 2^{\frac{\beta}{\beta-1}}$.

We have proved that $\|u - g\|_{L^\infty(\Omega)} \leq k_1$ and thus $\|u\|_{L^\infty(\Omega)} \leq \gamma := \|g\|_{L^\infty(\Omega)} + k_1$. □

If the function $\overline{g}$ prescribing the Dirichlet boundary condition on $\partial\Omega$ in the general PBE (3.41) is given as the trace of some function $u_{\overline{g}} \in H^2(\Omega)$, then from the Sobolev embeddings for $d \leq 3$ we have $u_{\overline{g}} \in L^\infty(\Omega)$ and $\nabla u_{\overline{g}} \in \left[L^6(\Omega)\right]^d$. Now, from Theorem 3.29 it directly follows that the solution u of (3.41) is in $L^\infty(\Omega)$ (we have $u \in H^1_{\gamma_2(u_{\overline{g}})}(\Omega) = H^1_{\overline{g}}(\Omega)$).

Examples

First observe that when $b(x, \cdot)$ is nondecreasing we can take $c_1(x, s) = c_2(x, s) = b(x, s)$. Additionally, if $b(x, 0) = 0$ and $g + w = 0$, the condition $c_1(x, g + w)$, $c_2(x, g + w) \in L^r(\Omega)$ with $r > \frac{d}{2}$ is trivially satisfied. In particular, $g + w = 0$ when we have a homogeneous Dirichlet boundary condition, i.e., $g = 0$, and $w = 0$.

We give some examples of functions b that satisfy the assumptions in Theorem 3.29.

(A) $b(x, \cdot)$ is not necessarily nondecreasing:

(1) $b(x, s) = k^2(x)\left(\mathrm{e}^s - s^{2p}\right)$, $k^2 \in L^r(\Omega)$, $r > \frac{d}{2}$, $p \geq 0$;

(2) $b(x, s) = k^2(x) s^p \mathrm{e}^{ns}$, $k^2 \in L^r(\Omega)$, $r > \frac{d}{2}$, $p \geq 0$, $n > 0$;

(3) $b(x, s) = k^2(x) s \mathrm{e}^{|s|} |\sin(s)|$ (the example in [187]). If $w + g = 0$, then k^2 can be taken in $L^1(\Omega)$ since there are c_1 and c_2 such that $c_1(x, 0) = c_2(x, 0) = 0$ and $c_1(x, s) \leq b(x, s) \leq c_2(x, s)$ (see Figure 3.2);

(4) $b(x, s) = k^2(x)\left(\sinh(s) + t \sin(s)\right)$ or
$b(x, s) = k^2(x)\left(\sinh(s) + t \cos(s)\right)$, where $k^2 \in L^r(\Omega)$, $r > \frac{d}{2}$, $t \in \mathbb{R}$;

(B) $b(x, \cdot)$ is nondecreasing:

(1) $b(x,s) = k^2(x)s\,|s|^p$, $p \in \mathbb{N}$. If $g + w = 0$, $k^2 \in L^1(\Omega)$, else $k^2 \in L^r(\Omega)$, $r > \frac{d}{2}$;

(2) $b(x,s) = \overline{k}^2(x)\sinh(s)$ where $\overline{k}^2 \in L^\infty(\Omega)$ (the case of the PBE);

(3) $b(x,s) := -\frac{4\pi e_0^2}{k_B T}\sum_{j=1}^{N_{ions}} \overline{M}_j(x)\xi_j \mathrm{e}^{-\xi_j s}$ where $M_j \geq 0$, $M_j \in L^\infty(\Omega)$, $\xi_j \in \mathbb{R}$, $j = 1, \ldots, N_{ions}$ (the case of the general PBE);

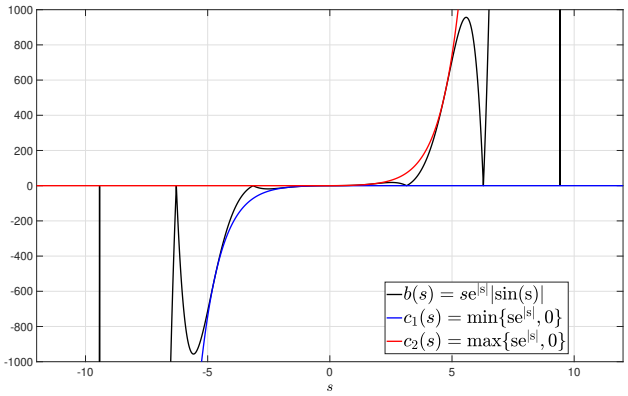


Figure 3.2: $b(x,s) = b(s) = se^{|s|}\,|\sin(s)|$, $c_1(x,s) = c_1(s) = \min\{se^{|s|}, 0\}$, $c_2(x,s) = c_2(s) = \max\{se^{|s|}, 0\}$

Chapter 4

Functional a posteriori error estimates

The purpose of this chapter is to derive a posteriori error estimates for the (nonlinear) PBE which give guaranteed and fully computable bounds on the error measured in global energy norms. More precisely, we derive error estimates for the regular component u in the 2-term and 3-term splittings of the full potential ϕ. In order to treat both splittings, we start in Section 4.1 by considering a more general semilinear elliptic interface problem, where the nonlinearity has the same form as in the PBE. After discussing the existence, uniqueness, and boundedness of the solution to this more general problem by making use of the a priori L^∞ estimate proved in Section 3.4, we present the abstract framework for derivation of functional a posteriori error estimates by following the general approach in [139, 155]. We establish the abstract error identity which defines the error measure natural for the considered class of problems.

The next step is to apply the abstract framework and compute explicit forms of all the respective terms in the error identity. First, we focus on the case of homogeneous Dirichlet boundary condition for which we present in detail all the steps involved in the derivation of the error estimates (Section 4.1.3). We obtain an error identity (4.51) with respect to a certain measure for the error which is the sum of the usual combined energy norm $|||\nabla(v-u)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2$ and a certain nonlinear measure. A main result connected to the explicit computation of this nonlinear measure of the error is given by Proposition 4.5. In the case of a linear elliptic equation of the form $-\operatorname{div}(\boldsymbol{A}\nabla u) + u = f_0$, the nonlinear error measure reduces to $\|v-u\|^2_{L^2(\Omega)} + \|\operatorname{div}(\boldsymbol{y}^* - \boldsymbol{p}^*)\|^2_{L^2(\Omega)}$, where v and $\boldsymbol{y}^*$ are approximations to the exact solution u and the exact flux $\boldsymbol{p}^* = \boldsymbol{A}\nabla u$.

One advantage of the presented error estimate is that it is valid for any conforming approximations of u and $\boldsymbol{A}\nabla u$ and that it does not rely on Galerkin orthogonality or properties specific to the used numerical method. Another advantage is that only the mathematical

structure of the problem is exploited and therefore no mesh dependent constants are present in the estimate. Majorants of the error not only give guaranteed bounds of global (energy) error norms but also generate efficient error indicators. Using only the error majorant, we obtain an analog of Cea's lemma (Proposition 4.12) which forms a basis for the a priori convergence analysis of finite element approximations for this class of semilinear problems. We also derive the explicit form of the terms in the abstract error identity in the case of a more general nonlinearity of the form $k^2(x)b(s)$, where b is a smooth strictly increasing function. We finish Section 4.1.3 with 3 numerical examples that verify the accuracy of error majorants and minorants and confirm efficiency of the error indicator in adaptive mesh procedures.

In Section 4.1.4 we present the case of nonhomogeneous Dirichlet boundary condition and show that all results obtained in the homogeneous case remain valid. Moreover, in Section 4.2 we describe a procedure, based on the patchwise equilibrated flux reconstruction in [30], to obtain a good conforming approximation $\boldsymbol{y}^*$ of the solution $\boldsymbol{p}^*$ of the dual problem. Thus, the evaluation of the error indicator in the adaptive algorithm, based on the derived error estimates, can be realized in a very efficient manner in parallel.

Further, in Section 4.3, we apply the obtained results for the more general semilinear problem to the specific case of the PBE with 2- and 3-term splittings, where in some cases it might be beneficial to make one additional splitting of the regular component u. Section 4.3.1 and Section 4.3.2 focus on the derivation of error majorants and minorants for the individual components of the solution that appear in the different splittings. Since each subproblem that defines a particular component of the solution depends on the the solution of the previous one, this causes a certain perturbation error additionally to the approximation error. To our knowledge, the presented analysis accounts for both sources of error for the first time. This leads to guaranteed and fully computable a posteriori estimates as well as efficient and robust indicators for the overall error.

Finally, in Section 4.3.3, we apply the described in this chapter methodology to study the electrostatic potential in and around the insulin protein with PDB ID 1RWE, the Alexa 488 and 594 dyes, as well as the membrane channel SecYEG. In all these applications we obtain guaranteed bounds on the relative errors in global energy norms.

4.1 General form of the estimates

4.1.1 A more general semilinear interface elliptic problem

First we want to remind the reader that in bold italic font we denote vector or matrix valued functions which are not necessarily constant. The constant vectors and variables in $\mathbb{R}^d$*, as*

well as the scalar functions, we denote in standard math font.

We start by considering a class of nonlinear interface elliptic problems with a nonlinearity of the type $b(x,s) = k^2(x)\sinh(s)$ that also includes the problems defining the regular component in the 2-term and 3-term splitting of the Poisson-Boltzmann equation:

$$-\nabla\cdot(\boldsymbol{A}\nabla u) + k^2\sinh(u+w) = f_0 - \operatorname{div}\boldsymbol{f} \quad \text{in } \Omega, \tag{4.1a}$$

$$u = \overline{g} \quad \text{on } \partial\Omega, \tag{4.1b}$$

which in weak form reads

$$\begin{aligned}&\text{Find } u \in H^1_{\gamma_2(\overline{g})}(\Omega) \text{ such that } k^2\sinh(u+w)v \in L^1(\Omega) \text{ for all } v \in C_0^\infty(\Omega) \text{ and}\\ &\int_\Omega \boldsymbol{A}\nabla u\cdot\nabla v dx + \int_\Omega k^2\sinh(u+w)v dx = \int_\Omega (f_0 v + \boldsymbol{f}\cdot\nabla v)\,dx \text{ for all } v \in C_0^\infty(\Omega).\end{aligned} \tag{4.2}$$

Here $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ is a bounded domain with Lipschitz boundary $\partial\Omega$. We assume that Ω contains an interior subdomain Ω_1 and we denote $\Omega_2 := \Omega \setminus \overline{\Omega}_1$, where in general, Ω_1 may consist of several disconnected parts. Typically, in biophysical applications, Ω_1 is occupied by one or more macromolecules and Ω_2 is occupied by a solution of water and moving ions. This is also the case of the Poisson-Boltzmann equation that we have considered in Chapter 3. There $\Omega_1 = \Omega_m \cup \Omega_{IEL}$ and $\Omega_2 = \Omega_{ions}$ (and if there is no ion exclusion layer Ω_{IEL}, then $\Omega_1 = \Omega_m$ and $\Omega_2 = \Omega_s = \Omega_{ions}$). Concerning the measurable function w and the coefficient $k \in L^\infty(\Omega)$ (not necessarily piecewise constant in Ω), we can identify three main cases:

(a) $k_{\max} \geq k(x) \geq k_{\min} > 0$ in Ω and $w \in L^\infty(\Omega)$;

(b) $k(x) \equiv 0$ in Ω_1, $k_{\max} \geq k(x) \geq k_{\min} > 0$ in Ω_2 and $w \in L^\infty(\Omega_2)$;

(c) $k(x) \equiv 0$ in Ω_2, $k_{\max} \geq k(x) \geq k_{\min} > 0$ in Ω_1 and $w \in L^\infty(\Omega_1)$.

In what follows, the major attention is paid to the case (b), which arrises when solving the PBE and which is most interesting from practical point of view. The cases (a) and (c) can be studied analogously (with some rather obvious modifications). We further assume that the conditions of Theorem 3.29 that ensure the boundedness of the solution u are satisfied. Therefore, the function $\overline{g}$ specifying the Dirichlet boundary condition is such that $\overline{g} \in H^1(\Omega) \cap L^\infty(\Omega)$ and $\nabla\overline{g} \in [L^s(\Omega)]^d$ for some $s > d$, $w \in L^\infty(\Omega_2)$, and $\boldsymbol{f} \in [L^s(\Omega)]^d$. Additionally, we assume that $f_0 \in L^2(\Omega)$ and that $\boldsymbol{A}$ is a symmetric positive definite matrix with bounded entries, i.e., $\boldsymbol{A} \in [L^\infty(\Omega)]^{d\times d}$, such that there exist some $\mu_1, \mu_2 > 0$ for which $\mu_1|\xi|^2 \leq \boldsymbol{A}(x)\xi\cdot\xi \leq \mu_2|\xi|^2$ for a.e. $x \in \Omega$ and all $\xi \in \mathbb{R}^d$. We recall that the space $H^1_{\gamma_2(\overline{g})}(\Omega)$ is defined by

$$H^1_{\gamma_2(\overline{g})}(\Omega) = \left\{v \in H^1(\Omega) : \gamma_2(v) = \gamma_2(\overline{g}) \text{ on } \partial\Omega\right\}$$

with γ_2 being the trace operator from $H^1(\Omega)$ to $H^{\frac{1}{2}}(\partial\Omega)$. Here we note that for the derivation of the a posteriori error estimates in this chapter we do not have to make any assumptions

on the regularity of the boundary of the interior subdomain Ω_1. Moreover, the boundary of Ω_1 is in general not associated with a jump discontinuity of the coefficient matrix $\boldsymbol{A}$ and the set Ω_1 is introduced only to specify the regions where $k \equiv 0$.

First we observe that (4.2) possesses a unique solution u. The existence of a solution can be shown by considering the variational problem

$$\text{Find } u_{\min} \in H^1_{\gamma_2(\overline{g})}(\Omega) \text{ such that } J(u_{\min}) = \min_{v \in H^1_{\gamma_2(\overline{g})}(\Omega)} J(v), \tag{4.3}$$

where the functional $J : H^1_{\gamma_2(\overline{g})}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$J(v) := \begin{cases} \frac{1}{2}a(v,v) + \int\limits_\Omega B(x, v+w)dx - \int\limits_\Omega \left(f_0 v + \boldsymbol{f} \cdot \nabla v\right) dx, \text{ if } B(x,v+w) \in L^1(\Omega), \\ +\infty, \text{ if } B(x,v+w) \notin L^1(\Omega), \end{cases} \tag{4.4}$$

with $a(u,v) = \int\limits_\Omega \boldsymbol{A}\nabla u \cdot \nabla v dx$ and $B(x,s) = k^2(x)\cosh(s)$. It can be shown that there exists a unique minimizer $u_{\min} \in H^1_{\gamma_2(\overline{g})}(\Omega)$ of J and that this minimizer is a solution to (4.2) (see the considerations on p. 61 where $\overline{g}$ is a function defined only on the boundary $\partial\Omega$). The proof of the uniqueness is similar to the proof of the uniqueness of a solution to (3.41) on p. 52. For completeness, we recall it here. Let u_1, u_2 are two solutions of (4.2). Then,

$$a(u_1 - u_2, v) + \int\limits_\Omega \left(b(x, u_1 + w) - b(x, u_2 + w)\right) v dx = 0, \, \forall v \in C_0^\infty(\Omega).$$

Now we want to show that we can test the above equation with $u_1 - u_2 \in H_0^1(\Omega)$. By using the monotonicity of $b(x,\cdot)$ we see that $\left(b(x,u_1+w) - b(x,u_2+w)\right)(u_1-u_2) \geq 0 \in L^1(\Omega)$. Therefore, by applying Theorem 2.18 to the function $u_1 - u_2 \in H_0^1(\Omega)$ and the bounded linear functional T_c defined by the relation

$$\langle T_c, v\rangle = \int\limits_\Omega \underbrace{\left(b(x,u_1+w) - b(x,u_2+w)\right)}_{:=c} v dx, \, \forall v \in C_0^\infty(\Omega),$$

we obtain that $\left(b(x,u_1+w) - b(x,u_2+w)\right)(u_1-u_2) \in L^1(\Omega)$ and that the extension $\overline{T}_c$ of T_c to the whole space $H_0^1(\Omega)$, evaluated at $u_1 - u_2$, is given by

$$\langle \overline{T}_c, u_1 - u_2\rangle = \int\limits_\Omega \left(b(x,u_1+w) - b(x,u_2+w)\right)(u_1-u_2) dx.$$

Therefore it holds

$$a(u_1-u_2, u_1-u_2) + \int\limits_\Omega \left(b(x,u_1+w) - b(x,u_2+w)\right)(u_1-u_2) dx = 0$$

and consequently $u_1 - u_2 = 0$.

From Theorem 3.29 we also know that the solution u of (4.2) is in $L^\infty(\Omega)$, which implies that $b(x, u + w) \in L^\infty(\Omega)$. Therefore, by a standard density argument we see that (4.2) holds for the bigger test space $H_0^1(\Omega)$.

Our goal is to derive a posteriori error estimate for (4.2) of the form

$$|||\nabla(v - u)||| \leq M_\oplus(v, \boldsymbol{y}^*),$$

where $|||\boldsymbol{q}|||^2 = \int\limits_\Omega \boldsymbol{A}\boldsymbol{q} \cdot \boldsymbol{q} dx$, $\forall \boldsymbol{q} \in \left[L^2(\Omega)\right]^d$ is the so-called energy norm, v is an arbitrary conforming approximation of u, i.e., $v \in H^1_{\gamma_2(\overline{g})}(\Omega)$, and $\boldsymbol{y}^*$ is a dual variable which in general lies in the space $\left[L^2(\Omega)\right]^d$ and it is some approximation to the exact flux $\boldsymbol{A}\nabla u$. In particular, if $\boldsymbol{A} = I$, the identity matrix, we have $|||\nabla(v - u)|||^2 = \|\nabla(v - u)\|^2_{L^2(\Omega)}$. Then, we will use the general results obtained for (4.2) and apply them to the PBE with the 2-term and 3-term splittings.

Before we continue with the presentation of the abstract framework for deriving functional a posteriori error estimates, we note that the results we are about to present can be extended for a more general nonlinearity $b(x, s)$ under certain reasonable assumptions on it.

Remark 4.1
Note that the term $\boldsymbol{f} \cdot \nabla v$ in (4.2) can represent a prescribed jump on the quantity $\boldsymbol{A}\nabla u \cdot \boldsymbol{n}_\Gamma$ across some interior interface Γ with a unit outward normal vector $\boldsymbol{n}_\Gamma$. In this case, the exact flux $\boldsymbol{A}\nabla u$ is not in $H(\mathrm{div}; \Omega)$. For example, this is the case with the 2-term and 3-term splittings in the Poisson-Boltzmann equation (see Remark 3.6 and Remark 3.9).

4.1.2 Abstract framework

First, we briefly recall some results from the duality theory ([65,139]) not necessarily in their most general form. Consider a class of variational problems having the following common form:

$$\begin{array}{ll} & \text{Find } u \in V \text{ such that} \\ (P) & J(u) = \inf\limits_{v \in V} J(v), \text{ where } J(v) = G(\Lambda v) + F(v). \end{array} \tag{4.5}$$

Here, V, Y are reflexive Banach spaces with the norms $\|.\|_V$ and $\|.\|_Y$, respectively, $F : V \to \mathbb{R} \cup \{+\infty\}$, $G : Y \to \mathbb{R} \cup \{+\infty\}$ are convex, proper and lower semicontinuous functionals, $\Lambda : V \to Y$ is a bounded linear operator and J is assumed to be coercive. We additionally, assume that $J(0_V) < +\infty$ and G is continuous at $\Lambda 0_V = 0_Y$, where by 0_V and 0_Y we denote the zero elements in V and Y, respectively. In this case, from Theorem 2.30 it follows that Problem (P) has a solution u, which is unique if J is strictly convex. The spaces topologically

dual to V and Y are denoted by V^* and Y^*, respectively. They are endowed with the norms $\|.\|_{V^*}$ and $\|.\|_{Y^*}$. Henceforth, $\langle v^*, v\rangle$ denotes the duality product of $v^* \in V^*$ and $v \in V$. Analogously, (y^*, y) is the duality product of $y^* \in Y^*$ and $y \in Y$. $\Lambda^* : Y^* \to V^*$ is the operator adjoint to Λ and it is defined by the relation

$$\langle \Lambda^* y^*, v\rangle = (y^*, \Lambda v),\ \forall v \in V,\ \forall y^* \in Y^*.$$

Recall that the functional $T^* : V^* \to \overline{\mathbb{R}}$ defined by the relation

$$T^*(v^*) := \sup_{v\in V} \{\langle v^*, v\rangle - T(v)\}$$

is called *dual* (or Fenchel conjugate) to T (see, Definition 2.24) and the functional $T^{**} : V \to \overline{\mathbb{R}}$ defined by the relation

$$T^{**}(v) = \sup_{v^*\in V^*} \{\langle v^*, v\rangle - T^*(v^*)\}$$

is called Fenchel biconjugate of T (see Definition 2.25). Since G is a proper, convex and lower semicontinuous functional, from the Fenchel-Moreau Theorem (Theorem 2.26) it follows that $G = G^{**}$ and we can write

$$J(v) = G(\Lambda v) + F(v) = \sup_{y^*\in Y^*} \{(y^*, \Lambda v) - G^*(y^*) + F(v)\},$$

where $G^* : Y^* \to \overline{\mathbb{R}}$ is the Fenchel conjugate of G and the function $L : V \times Y^* \to \overline{\mathbb{R}}$ defined by

$$L(v, y^*) = (y^*, \Lambda v) - G^*(y^*) + F(v)$$

is called the Lagrangian for J. Now it is clear that problem (P) can be written in the following form

$$J(u) = \inf_{v\in V} J(v) = \inf_{v\in V} \sup_{y^*\in Y^*} L(v, y^*) = \sup_{y^*\in Y^*} L(u, y^*).$$

Analogously to the definition of J, we can also define the functional I^* by the relation

$$\begin{aligned} I^*(y^*) &:= \inf_{v\in V} L(v, y^*) = -G^*(y^*) + \inf_{v\in V} \{(y^*, \Lambda v) + F(v)\} \\ &= -G^*(y^*) - \sup_{v\in V} \{\langle -\Lambda^* y^*, v\rangle - F(v)\} = -G^*(y^*) - F^*(-\Lambda^* y^*). \end{aligned}$$

In accordance with the general duality theory of the calculus of variations, we can define the dual counterpart of the primal Problem (4.5):

$$\begin{aligned} &\text{Find } p^* \in Y^* \text{ such that} \\ (P^*)\quad &I^*(p^*) = \sup_{y^*\in Y^*} I^*(y^*) = \sup_{y^*\in Y^*} \inf_{v\in V} L(v, y^*). \end{aligned} \tag{4.6}$$

Next, we note that from the continuity of G at 0_Y it follows that G^* is coercive (see Remark 4.2). Now, from the coercivity of G^* and the fact that $F(0_V)$ is finite, it follows that

problem (P^*) also has a solution. To see this, note that since G^* and F^* are convex and lower semicontinuous as the pointwise supremum of affine functionals (see, e.g., [65, 139]), it follows that the functional $-I^*(y^*)$ is convex, proper and lower semicontinuous over the reflexive Banach space Y^*. Moreover, since G^* is coercive and $F(0_V)$ is finite we obtain that $-I^*(y^*)$ is also coercive:

$$\begin{aligned} -I^*(y^*) &= G^*(y^*) + F^*(-\Lambda^* y^*) \\ &\geq \quad G^*(y^*) + \langle -\Lambda^* y^*, 0_V \rangle - F(0_V) \to +\infty \ \text{ whenever } \ \|y^*\|_{Y^*} \to \infty. \end{aligned}$$

Thus, $I^* : Y^* \to \mathbb{R} \cup \{-\infty\}$ possesses a maximizer $p^* \in Y^*$, which is unique if at least one of the functionals G^* and F^* is strictly convex. Now, by using the assumption that $J(0_V) < +\infty$ and G is continuous at $\Lambda 0_V = 0_Y$, it follows that strong duality holds for the problems (P) and (P^*) (see, e.g., Chapter III in [65]):

$$J(u) = \inf_{v \in V} J(v) = \inf_{v \in V} \sup_{y^* \in Y^*} L(v, y^*) = \sup_{y^* \in Y^*} \inf_{v \in V} L(v, y^*) = \sup_{y^* \in Y^*} I^*(y^*) = I^*(p^*). \quad (4.7)$$

Furthermore, from (4.7) it follows that the pair (u, p^*) is a saddle point of the Lagrangian L, i.e.,

$$L(u, y^*) \leq L(u, p^*) \leq L(v, p^*), \ \forall v \in V, \ \forall y^* \in Y^*. \quad (4.8)$$

The left-hand side of (4.8) implies that u and p^* satisfy the relations

$$\Lambda u \in \partial G^*(p^*), \quad p^* \in \partial G(\Lambda u), \quad (4.9)$$

and the right-hand side of (4.8) implies the relations

$$\Lambda u \in \partial F^*(-\Lambda^* p^*), \quad -\Lambda^* p^* \in \partial F(u), \quad (4.10)$$

where $\partial G^*(p^*)$ and $\partial G(\Lambda u)$ denote the subdifferentials of G^* at p^* and of G at Λu, respectively, and $\partial F^*(-\Lambda^* p^*)$ and $\partial F(u)$ denote the subdifferentials of F^* at $-\Lambda^* p^*$ and of F at u, respectively.

We have

$$\begin{aligned} J(v) - I^*(y^*) &= G(\Lambda v) + F(v) + G^*(y^*) + F^*(-\Lambda^* y^*) \\ &= D_G(\Lambda v, y^*) + D_F(v, -\Lambda^* y^*) =: M^2_{\oplus}(v, y^*), \end{aligned} \quad (4.11)$$

where

$$D_G(\Lambda v, y^*) := G(\Lambda v) + G^*(y^*) - (y^*, \Lambda v)$$

and

$$D_F(v, -\Lambda^* y^*) := F(v) + F^*(-\Lambda^* y^*) + \langle \Lambda^* y^*, v \rangle$$

are the *compound* functionals for G and F, respectively (see [139]). A compound functional is nonnegative by the definition of Fenchel conjugate. Since $J(v) \geq I^*(y^*)$ for all $v \in V$ and $y^* \in Y^*$ with equality if and only if $v = u$ and $y^* = p^*$, the equality (4.11) shows that D_G and D_F can vanish simultaneously if and only if $v = u$ and $y^* = p^*$. Now, it is clear that solving the primal and dual problems (P) and (P^*) is equivalent to minimizing the duality gap $J(v) - I^*(y^*)$, whose minimum we know is equal to zero. Therefore, if $v \in V$ and $y^* \in Y^*$ are approximations of u and p^*, respectively, the fully computable difference $J(v) - I^*(y^*)$ is a measure for the error between (v, y^*) and (u, p^*). Moreover, by setting $y^* := p^*$ and $v := u$ in (4.11), we obtain analogous identities for the primal and dual parts of the error:

$$J(v) - I^*(p^*) = M^2_{\oplus}(v, p^*) = D_G(\Lambda v, p^*) + D_F(v, -\Lambda^* p^*), \tag{4.12a}$$

$$J(u) - I^*(y^*) = M^2_{\oplus}(u, y^*) = D_G(\Lambda u, y^*) + D_F(u, -\Lambda^* y^*). \tag{4.12b}$$

Using the fact that $J(u) = I^*(p^*)$ and that the above equalities (4.12a), (4.12b) hold, we obtain another important identity (see [139]) which describes the full error $M^2_{\oplus}(v, y^*)$ as the sum of the primal and dual parts of the error:

$$\begin{aligned} &M^2_{\oplus}(v, y^*) = J(v) - I^*(y^*) \\ &= J(v) - I^*(p^*) + J(u) - I^*(y^*) = M^2_{\oplus}(v, p^*) + M^2_{\oplus}(u, y^*). \end{aligned} \tag{4.13}$$

Notice that $M^2_{\oplus}(v, y^*)$ depends on the approximations v and y^* only and, therefore, is fully computable. The right-hand side of (4.13) can be viewed as a certain measure of the distance between (v, y^*) and (u, p^*), which vanishes if and only if $v = u$ and $y^* = p^*$. Hence the relation

$$D_G(\Lambda v, p^*) + D_F(v, -\Lambda^* p^*) + D_G(\Lambda u, y^*) + D_F(u, -\Lambda^* y^*) = M^2_{\oplus}(v, y^*) \tag{4.14}$$

establishes the equality of the computable term $M^2_{\oplus}(v, y^*)$ and an error measure natural for this class of variational problems.

It is worth noting that the identity (4.14) can be represented in terms of norms if G and F are quadratic functionals. For example, if $V = H^1_0(\Omega)$, $V^* = H^{-1}(\Omega)$, $Y = [L^2(\Omega)]^d = Y^*$, $G(\Lambda v) = G(\nabla v) = \int\limits_\Omega \frac{1}{2} \boldsymbol{A} \nabla v \cdot \nabla v dx$ and $F(v) = \int\limits_\Omega \left(\frac{1}{2} v^2 - f_0 v\right) dx$ (where A is a symmetric positive definite matrix with bounded entries and satisfying for some $\mu_1, \mu_2 > 0$ the inequality $\mu_1 |\xi|^2 \leq \boldsymbol{A}(x)\xi \cdot \xi \leq \mu_2 |\xi|^2$ for a.e. $x \in \Omega$ and all $\xi \in \mathbb{R}^d$), then

$$\begin{aligned} D_G(\Lambda v, \boldsymbol{p}^*) &= \frac{1}{2} \int\limits_\Omega \boldsymbol{A} \nabla (v - u) \cdot \nabla (v - u) dx = \frac{1}{2} |\!|\!| \nabla (v - u) |\!|\!|^2, \\ D_G(\Lambda u, \boldsymbol{y}^*) &= \frac{1}{2} \int\limits_\Omega \boldsymbol{A}^{-1} (\boldsymbol{y}^* - \boldsymbol{p}^*) \cdot (\boldsymbol{y}^* - \boldsymbol{p}^*) dx =: \frac{1}{2} |\!|\!| \boldsymbol{y}^* - \boldsymbol{p}^* |\!|\!|^2_*, \\ D_F(v, -\Lambda^* \boldsymbol{p}^*) &= \frac{1}{2} \|v - u\|^2_{L^2(\Omega)}, \quad D_F(u, -\Lambda^* \boldsymbol{y}^*) = \frac{1}{2} \|\mathrm{div}(\boldsymbol{y}^* - \boldsymbol{p}^*)\|^2_{L^2(\Omega)}. \end{aligned} \tag{4.15}$$

In this case, the minimizer of (4.5) solves the linear elliptic problem

$$\begin{aligned} -\operatorname{div}(\boldsymbol{A}\nabla u) + u &= f_0 \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega, \end{aligned} \tag{4.16}$$

and (4.14) is reduced to the error identity

$$\begin{aligned} &|||\nabla(v-u)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2 + \|v-u\|^2_{L^2(\Omega)} + \|\operatorname{div}(\boldsymbol{y}^* - \boldsymbol{p}^*)\|^2_{L^2(\Omega)} \\ &= |||\boldsymbol{A}\nabla v - \boldsymbol{y}^*|||_*^2 + \|v - \operatorname{div}\boldsymbol{y}^* - f_0\|^2_{L^2(\Omega)} = 2M^2_{\oplus}(v, \boldsymbol{y}^*). \end{aligned} \tag{4.17}$$

The sum of the first and the third term in (4.17) represents the primal, the sum of the second and fourth term the dual error.

In what follows, we will present the particular form of the error equality (4.14) for the problem (4.4), or equivalently, (4.2), where the error is measured in a special "nonlinear norm". This measure contains the usual combined energy norm terms, i.e. the sum of the energy norms of the errors for the primal and dual problem, but also two additional nonnegative terms due to the nonlinearity $B(x,\cdot)$ (or equivalently $b(x,\cdot)$) which in some cases may dominate the usual energy norm terms. We start by deriving explicit expressions for G^*, F^* and then we will use these expressions to get an explicit form of the abstract error equality (4.14).

Remark 4.2
Notice that if G is continuous at 0_Y, then G^* is coercive. To prove this, assume to the contrary that there is some $\alpha \in \mathbb{R}$ and a sequence $\{y_n^*\}_{n=1}^{\infty} \subset Y^*$ such that $\|y_n^*\|_{Y^*} \to \infty$ and $G^*(y_n^*) \le \alpha$ for all $n \in \mathbb{N}$. Let $\epsilon > 0$ be fixed. From the continuity of G at 0_Y we can find a $\delta > 0$ such that $|G(y) - G(0)| \le \epsilon$ for all $y \in B_\delta$, where B_δ denotes the ball in Y with center at 0_Y and radius δ. From the definition of Fenchel conjugate, for each $n \in \mathbb{N}$, we obtain

$$\begin{aligned} \alpha \ge G^*(y_n^*) = \sup_{y\in Y}\{(y_n^*, y) - G(y)\} &\ge \sup_{y\in B_\delta}\{(y_n^*, y) - G(y)\} \\ &\ge \sup_{y\in B_\delta}(y_n^*, y) - |G(0)| - \epsilon. \end{aligned} \tag{4.18}$$

We will show that $\sup_{y\in B_\delta}(y_n^*, y) \to +\infty$ when $n \to \infty$. For this, we will use a corollary of the Hahn-Banach theorem (see, e.g., Corollary 3.1.3 in [64]) and the reflexivity of Y. In particular, the corollary says that for any $y_0 \in Y$ with $\|y_0\|_Y = 1$ there exists a functional $y_0^* \in Y^*$ with $\|y_0^*\|_{Y^*} = 1$ such that $(y_0^*, y_0) = 1$. This means that for an arbitrary $Y \ni y_0 \neq 0_Y$ there is $y_0^* \in Y^*$ with $\|y_0^*\|_{Y^*} = 1$ such that $(y_0^*, y_0) = \|y_0\|_Y$ (since $\frac{y_0}{\|y_0\|_Y}$ has a unit norm). Thus, if we define $y_1^* := \|y_0\|_Y y_0^*$, we will have $\|y_1^*\|_{Y^*} = \|y_0\|_Y$ and $(y_1^*, y_0) = \|y_0\|_Y^2$.

We apply the above property for each y_n^* in (4.18). Therefore, we find $y_n^{**} \in Y^{**}$ with $\|y_n^{**}\|_{Y^{**}} = \|y_n^*\|_{Y^*}$ such that $(y_n^{**}, y_n^*) = \|y_n^*\|_{Y^*}^2$ for every $n \in \mathbb{N}$. Now, from the reflexivity of

Y, it follows that there is a unique $y_n \in Y$ such that $(y_n^{**}, y_n^*) = (y_n^*, y_n)$ and $\|y_n^{**}\|_{Y^{**}} = \|y_n\|_Y$. Therefore, from (4.18) *we obtain*

$$\begin{aligned} \alpha \geq G^*(y_n^*) \geq \sup_{y\in B_\delta} (y_n^*, y) - |G(0)| - \epsilon \geq & \frac{\delta}{2}\left(y_n^*, \frac{y_n}{\|y_n\|_Y}\right) \\ = & \frac{\delta}{2\|y_n\|_Y}(y_n^{**}, y_n^*) = \frac{\delta}{2\|y_n^*\|_{Y^*}}\|y_n^*\|_{Y^*}^2 \to +\infty, \end{aligned} \tag{4.19}$$

which is a contradiction with the assumption that G^* is not coercive.

In fact, the opposite is also true, i.e., from the coercivity of G^* and the fact that G is a proper l.s.c. convex functional follows the continuity of G at 0_Y (see Theorem 7A in [160]).

4.1.3 Homogeneous Dirichlet boundary condition

We start by considering (4.2) with a homogeneous Dirichlet boundary condition, i.e., $\overline{g} = 0$ in Ω. We set $V := H_0^1(\Omega)$, $Y := [L^2(\Omega)]^d$ $(d = 2, 3)$, and Λ the gradient operator $\nabla : H_0^1(\Omega) \to [L^2(\Omega)]^d$. We further denote $g : \Omega \times \mathbb{R}^3 \to \mathbb{R}$, $\quad g(x, \xi) := \frac{1}{2}\boldsymbol{A}(x)\xi \cdot \xi$. With this notation, we have

$$\begin{aligned} G(\Lambda v) &:= \int_\Omega g(x, \nabla v(x))dx = \int_\Omega \frac{1}{2}\boldsymbol{A}(x)\nabla v \cdot \nabla v dx, \\ F(v) &:= \int_\Omega B(x, v + w)dx - \int_\Omega (f_0 v + \boldsymbol{f} \cdot \nabla v)\, dx, \end{aligned} \tag{4.20}$$

where we recall that $B(x, s) = k^2(x)\cosh(s)$, and the functional J, defined by (4.4), can be written in the form $J(v) = G(\Lambda v) + F(v)$. For any $v \in V$ the functional $G(\Lambda v)$ is finite, while $F : V \to \mathbb{R} \cup \{+\infty\}$ may take the value $+\infty$ for some $v \in V$ if $d \geq 3$ (e.g $v = \log \frac{1}{|x|^\alpha}$, $\alpha \geq d$ on the unit ball in $\mathbb{R}^d$). However, if $d \leq 2$, then $\exp(v) \in L^1(\Omega)$, $\forall v \in H_0^1(\Omega)$ and $F : V \to \mathbb{R}$ (see [110, 179]). Also, $F(0_V)$ is obviously finite since $w \in L^\infty(\Omega_2)$ and it is satisfied that $J(0_V) < +\infty$ and G is continuous at $\Lambda 0_V = 0_Y$. Moreover, G and F are proper, convex and sequentially lower semicontinuous functionals. That G is proper and convex is obvious. That G is s.l.s.c., follows form its continuity in Y. The fact that F is proper follows from the fact that $B(x, s) \geq 0$ for a.e. $x \in \Omega$ and all $s \in \mathbb{R}$. The convexity of F follows from the convexity of $B(x, \cdot)$, and the (weak) sequential lower semicontinuity of F follows from the fact that it is the sum of the s.l.s.c. functionals $\int_\Omega B(x, v + w)dx$ and $-\int_\Omega (f_0 v + \boldsymbol{f} \cdot \nabla v)\, dx$. The sequential lower semicontinuity of the functional $\int_\Omega B(x, v + w)dx$ follows by applying Fatou's lemma and we note that this argument is similar to the one made about the functional $\int_\Omega B(x, v + u^L + w)dx$ on p. 55.

We continue by setting $V^* = H^{-1}(\Omega)$ and $Y^* = Y = [L^2(\Omega)]^d$. In this case, Λ^* coincides with $-\operatorname{div}$ considered as an operator from $[L^2(\Omega)]^d$ to $H^{-1}(\Omega)$. It is clear that G is coercive

in Y since $G(\boldsymbol{y}) = \int\limits_{\Omega} \frac{1}{2}\boldsymbol{A}\boldsymbol{y}\cdot\boldsymbol{y}dx \geq \frac{1}{2}\mu_1\|\boldsymbol{y}\|^2_{L^2(\Omega)} \to +\infty$ whenever $\|\boldsymbol{y}\|_{L^2(\Omega)} \to \infty$. We will see that $G^*(\boldsymbol{y}^*) = \int\limits_{\Omega} \frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{y}^*\cdot\boldsymbol{y}^*dx \geq \frac{1}{2\mu_2}\|\boldsymbol{y}^*\|^2_{L^2(\Omega)}$ and hence it is also coercive in Y^* (this also follows from the continuity of G at 0_Y). Therefore, based on the discussion in Section 4.1.2 and the properties of G and F that we verified above, all assumptions that guarantee existence of a solution to the primal and dual problem, as well as validity of the strong duality relation (4.7) are fulfilled. We conclude that the primal problem (P) and its dual (P^*) in our particular case have solutions $u \in H^1_0(\Omega)$ and $\boldsymbol{p}^* \in \left[L^2(\Omega)\right]^d$, which are unique since G and G^* are strictly convex, and moreover, the strong duality relation (4.7) holds. In addition, the optimality conditions (4.9) and (4.10) hold. Since G and G^* are Gateaux differentiable, they have unique subgradients at Λu and $\boldsymbol{p}^*$ respectively, and therefore, we obtain the relation

$$\boldsymbol{p}^* = \boldsymbol{A}\nabla u \tag{4.21}$$

between the solution of the primal and dual problem.

Fenchel Conjugates of the functionals G and F

Fenchel conjugate of G

It is easy to find that $G^*(\boldsymbol{y}^*) = \int\limits_{\Omega} \frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{y}^*\cdot\boldsymbol{y}^*dx$. Indeed, we have

$$G^*(\boldsymbol{y}^*) = \sup_{\boldsymbol{y}\in Y}\left\{\int\limits_{\Omega}\left(\boldsymbol{y}^*\cdot\boldsymbol{y} - \frac{1}{2}\boldsymbol{A}\boldsymbol{y}\cdot\boldsymbol{y}\right)dx\right\} \leq \int\limits_{\Omega}\sup_{\xi\in\mathbb{R}^d}\left\{\boldsymbol{y}^*(x)\cdot\xi - \frac{1}{2}\boldsymbol{A}(x)\xi\cdot\xi\right\}dx. \tag{4.22}$$

The supremum in $\xi \in \mathbb{R}^d$ is actually achieved for a.e. $x \in \Omega$ at some $\boldsymbol{\xi}_0(x)$ since the function

$$-\left(\boldsymbol{y}^*(x)\cdot\xi - \frac{1}{2}\boldsymbol{A}(x)\xi\cdot\xi\right)$$

is strictly convex, continuous and coercive over $\mathbb{R}^d$ for a.e. $x \in \Omega$. The necessary conditions for a maximum are

$$\frac{\partial}{\partial\xi_i}\left(\boldsymbol{y}^*(x)\cdot\xi - \frac{1}{2}\boldsymbol{A}\xi\cdot\xi\right) = 0,\ i = 1, 2, \ldots, d,$$

from which we obtain $\boldsymbol{\xi}_0(x) = \boldsymbol{A}^{-1}(x)\boldsymbol{y}^*(x)$ for a.e. $x \in \Omega$. Thus, substituting the expression for $\boldsymbol{\xi}_0$ in (4.22) we obtain

$$G^*(\boldsymbol{y}^*) = \sup_{\boldsymbol{y}\in Y}\left\{\int\limits_{\Omega}\left(\boldsymbol{y}^*\cdot\boldsymbol{y} - \frac{1}{2}\boldsymbol{A}\boldsymbol{y}\cdot\boldsymbol{y}\right)dx\right\} \leq \int\limits_{\Omega}\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{y}^*\cdot\boldsymbol{y}^*dx. \tag{4.23}$$

Since $\boldsymbol{\xi}(x) = \boldsymbol{A}^{-1}(x)\boldsymbol{y}^*(x) \in \left[L^2(\Omega)\right]^d$, we see that the supremum in the definition of Fenchel conjugate is actually achieved and thus

$$G^*(\boldsymbol{y}^*) = \int\limits_{\Omega}\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{y}^*\cdot\boldsymbol{y}^*dx. \tag{4.24}$$

Fenchel conjugate of F

First, we observe that the exact flux $\boldsymbol{p}^* = \boldsymbol{A}\nabla u \in \left[L^2(\Omega)\right]^d$ can be represented in the form

$$\boldsymbol{p}^* = \boldsymbol{f} + \boldsymbol{p}_0^*, \text{ where } \boldsymbol{p}_0^* \in H(\mathrm{div};\Omega) \text{ and } \mathrm{div}\,\boldsymbol{p}_0^* = b(x, u+w) - f_0. \tag{4.25}$$

To see this, notice that $\boldsymbol{p}^*$ satisfies the equation

$$\int_\Omega \boldsymbol{p}^* \cdot \nabla v dx + \int_\Omega b(x, u+w) v dx = \int_\Omega \left(f_0 v + \boldsymbol{f} \cdot \nabla v\right) dx. \tag{4.26}$$

By substituting $\boldsymbol{p}^* = \boldsymbol{f} + \boldsymbol{p}_0^*$ in (4.26), where a priori $\boldsymbol{p}_0^*$ is only in $\left[L^2(\Omega)\right]^d$, we obtain

$$\int_\Omega \boldsymbol{p}_0^* \cdot \nabla v dx = \int_\Omega \left(-b(x, u+w) + f_0\right) v dx. \tag{4.27}$$

Now, since $u \in L^\infty(\Omega)$, it follows that $-b(x, u+w) + f_0 \in L^2(\Omega)$, and therefore $\boldsymbol{p}_0^* \in H(\mathrm{div};\Omega)$ with a weak divergence given by

$$\mathrm{div}\,\boldsymbol{p}_0^* = b(x, u+w) - f_0. \tag{4.28}$$

In particular, since $b(x,\cdot) = 0$ for a.e. $x \in \Omega_1$, it follows that $\mathrm{div}\,\boldsymbol{p}_0^* + f_0 = 0$ in Ω_1.

Since the exact flux $\boldsymbol{p}^*$ has the form (4.25), it is enough to find an explicit form of $F^*(-\Lambda^*\boldsymbol{y}^*)$ only for such functions $\boldsymbol{y}^* \in \left[L^2(\Omega)\right]^d$ that have the form

$$\boldsymbol{y}^* = \boldsymbol{f} + \boldsymbol{y}_0^* \text{ for some } \boldsymbol{y}_0^* \in H(\mathrm{div};\Omega). \tag{4.29}$$

For $\boldsymbol{q}^* \in H(\mathrm{div};\Omega)$ and an arbitrary measurable function $z : \Omega_2 \to \mathbb{R}$, we introduce the functional

$$I_{\boldsymbol{q}^*}(z) := \int_{\Omega_2} \left[(\mathrm{div}\,\boldsymbol{q}^* + f_0) z - B(x, z+w)\right] dx. \tag{4.30}$$

Recalling that in the particular case (b) for the functions k and w, the nonlinearity B is

supported on Ω_2, for any $\boldsymbol{y}^*$ of the form (4.29) we have

$$\begin{aligned}
F^*(-\Lambda^*\boldsymbol{y}^*) &= \sup_{z\in H_0^1(\Omega)} \{\langle -\Lambda^*\boldsymbol{y}^*, z\rangle - F(z)\} = \sup_{z\in H_0^1(\Omega)} \{(-\boldsymbol{y}^*, \Lambda z) - F(z)\} \\
&= \sup_{z\in H_0^1(\Omega)} \int_\Omega (-\boldsymbol{y}^*\cdot\nabla z - B(x, z+w) + f_0 z + \boldsymbol{f}\cdot\nabla z)\, dx \\
&= \sup_{z\in H_0^1(\Omega)} \int_\Omega (-\boldsymbol{y}_0^*\cdot\nabla z - B(x, z+w) + f_0 z)\, dx \\
&= \sup_{z\in H_0^1(\Omega)} \int_\Omega (\operatorname{div}\boldsymbol{y}_0^* z - B(x, z+w) + f_0 z)\, dx \quad (\text{finite if } \operatorname{div}\boldsymbol{y}_0^* + f_0 = 0 \text{ in } \Omega_1) \\
&= \sup_{z\in H_0^1(\Omega)} I_{\boldsymbol{y}_0^*}(z) \le \int_{\Omega_2} \sup_{\xi\in\mathbb{R}} \Big\{ (\operatorname{div}\boldsymbol{y}_0^*(x) + f_0(x))\,\xi - B(x, \xi + w(x)) \Big\} dx \\
&= \int_{\Omega_2} \Big((\operatorname{div}\boldsymbol{y}_0^*(x) + f_0(x))\,\xi_0(x) - B(x, \xi_0(x) + w(x)) \Big) dx = I_{\boldsymbol{y}_0^*}(\xi_0).
\end{aligned} \tag{4.31}$$

Here $\xi_0 : \Omega_2 \to \mathbb{R}$ is computed from the condition

$$\frac{d}{d\xi}\left[(\operatorname{div}\boldsymbol{y}_0^*(x) + f_0(x))\,\xi - B(x, \xi + w(x))\right] = 0, \text{ for a.e. } x \in \Omega_2, \tag{4.32}$$

which is equivalent to

$$\operatorname{div}\boldsymbol{y}_0^*(x) + f_0(x) - k^2(x)\sinh(\xi + w(x)) = 0 \text{ for a.e. } x \in \Omega_2.$$

We notice that (4.32) is a necessary condition for a maximum which is also sufficient since $B(x,\cdot)$ is convex. The solution of the last equation exists, is unique, and is given by

$$\begin{aligned}
\xi_0(x) &= \operatorname{arsinh}\left(\frac{\operatorname{div}\boldsymbol{y}_0^*(x) + f_0(x)}{k^2(x)}\right) - w(x) = \ln\left(\rho(\boldsymbol{y}_0^*) + \sqrt{\rho^2(\boldsymbol{y}_0^*) + 1}\right) - w(x) \\
&= \ln\left(\Theta\left(\rho(\boldsymbol{y}_0^*)\right)\right) - w, \text{ for a.e. } x \in \Omega_2,
\end{aligned} \tag{4.33}$$

where

$$\rho(\boldsymbol{y}_0^*) := \frac{\operatorname{div}\boldsymbol{y}_0^*(x) + f_0(x)}{k^2(x)} \quad \text{and} \quad \Theta(s) := s + \sqrt{s^2+1} \text{ for } s \in \mathbb{R},$$

and we have used the formula $\operatorname{arsinh}(s) = \ln\left(s + \sqrt{s^2+1}\right)$, $\forall s \in \mathbb{R}$.

In Proposition 4.5, we will prove that we have not overestimated the supremum over $z \in H_0^1(\Omega)$ in (4.31) and that we actually have equalities everywhere. By using the expression for $\xi_0(x)$, for any $\boldsymbol{y}^* \in [L^2(\Omega)]^d$ of the form (4.29) with $\operatorname{div}\boldsymbol{y}_0^* + f_0 = 0$ in Ω_1 we obtain an explicit

formula for $F^*(-\Lambda^*\boldsymbol{y}^*)$:

$$\begin{aligned} F^*(-\Lambda^*\boldsymbol{y}^*) &= \int_{\Omega_2} \left[k^2 \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) \left(\operatorname{arsinh} \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) - w \right) \right. \\ &\quad \left. - k^2 \cosh \left(\operatorname{arsinh} \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) \right) \right] dx \qquad (4.34) \\ &= \int_{\Omega_2} \left[k^2 \rho(\boldsymbol{y}_0^*) \left(\ln \left(\Theta \left(\rho(\boldsymbol{y}_0^*) \right) \right) - w \right) - k^2 \sqrt{\rho^2(\boldsymbol{y}_0^*) + 1} \right] dx, \end{aligned}$$

where we have used the formula $\cosh(\operatorname{arsinh}(s)) = \sqrt{s^2+1}$, $\forall s \in \mathbb{R}$.

Remark 4.3
Since $\left| \ln \left(t + \sqrt{t^2+1} \right) \right| \leq |t|$, $\forall t \in \mathbb{R}$, *the function* $\ln(\Theta(f(x))) - w(x)$ *belongs to* $L^2(\Omega_2)$ *for any* $f \in L^2(\Omega_2)$. *Since* $\rho(\boldsymbol{y}_0^*) = \frac{\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x)}{k^2(x)}$ *and* $f_0 \in L^2(\Omega_2)$, $k^2 \geq k_{\min}^2 > 0$ *in* Ω_2, *we conclude that* $\xi_0(x) \in L^2(\Omega_2)$ *if* $\boldsymbol{y}_0^* \in H(\operatorname{div}; \Omega)$. *Therefore the integrals in* (4.34) *are well defined.*

Remark 4.4
Note that the set Ω_1 *where the coefficient* k *is zero (and where the problem is linear) dictates where the dual variable* $\boldsymbol{y}_0^*$ *has to be exactly equilibrated, i.e.,* $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ *(see the derivation of* $F^*(-\Lambda^*\boldsymbol{y}^*)$ *in* (4.31)*).*

Now our goal is to prove that the inequality $\sup_{z \in H_0^1(\Omega)} I_{\boldsymbol{y}_0^*}(z) \leq I_{\boldsymbol{y}_0^*}(\xi_0)$ holds as equality. In other words, we want to prove that the error estimate remains sharp and that the computed majorant $M_{\oplus}^2(v, \boldsymbol{y}^*)$ will be indeed zero if approximations $(v, \boldsymbol{y}^*)$ coincide with the exact solution $(u, \boldsymbol{p}^*)$.

Proposition 4.5
For any $\boldsymbol{y}_0^* \in H(\operatorname{div}; \Omega)$ *with* $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ *in* Ω_1 *it holds*

$$\sup_{z \in H_0^1(\Omega)} I_{\boldsymbol{y}_0^*}(z) = I_{\boldsymbol{y}_0^*}(\xi_0) < \infty.$$

Proof. The idea is to approximate $\rho(\boldsymbol{y}_0^*) = \frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \in L^2(\Omega_2)$ and $w_{\restriction \Omega_2} \in L^\infty(\Omega_2)$ by $C_0^\infty(\Omega_2)$ functions (in the a.e. sense) and use the Lebesgue dominated convergence theorem. From the density of $C_0^\infty(\Omega_2)$ in $L^2(\Omega_2)$ we can find a sequence $\{\psi_n\}_{n=1}^\infty \subset C_0^\infty(\Omega_2)$ such that $\psi_n(x) \to \rho(\boldsymbol{y}_0^*(x))$, *a.e.* in Ω_2 and $|\psi_n(x)| \leq h(x) \in L^2(\Omega_2)$ (see Theorem 4.9 in [96]). Again, from the density of $C_0^\infty(\Omega_2)$ in $L^2(\Omega_2)$ we can find a sequence $\{w_n\}_{n=1}^\infty \subset C_0^\infty(\Omega_2)$ such that $w_n(x) \to w(x)$, *a.e.* in Ω_2 and $|w_n(x)| \leq m + 2$, where $m := \|w\|_{L^\infty(\Omega_2)}$ (see Remark 4.6). Then

$$z_n(x) := \ln(\Theta(\psi_n(x))) - w_n(x) \to \xi_0(x), \text{ a.e. in } \Omega_2$$

and $z_n \in C_0^\infty(\Omega_2) \subset H_0^1(\Omega_2) \subset H_0^1(\Omega)$ (by extending the functions by zero in Ω_1). Since $B(x,\cdot)$ is continuous, we have the pointwise a.e. in Ω_2 convergence

$$(\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x))\, z_n(x) - B\left(x, z_n + w(x)\right) \to (\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x))\, \xi_0(x) - B(x, \xi_0(x) + w(x)).$$

Now we search for a function in $L^1(\Omega_2)$ that majorates the function
$|(\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x))\, z_n(x) - B\left(x, z_n + w(x)\right)|$:

$$\begin{aligned} &\left|(\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x))\, z_n(x) - k^2(x) \cosh\left(z_n(x) + w(x)\right)\right| \\ &\leq |\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x)|\, |z_n(x)| + k^2(x) \mathrm{e}^{\|w\|_{L^\infty(\Omega_2)}} \mathrm{e}^{|z_n(x)|} \end{aligned} \tag{4.35}$$

Our next goal is to bound $|z_n(x)|$ in (4.35). For the first summand, by using Remark 4.3, we obtain

$$|z_n(x)| = |\ln\left(\Theta\left(\psi_n(x)\right)\right) - w_n(x)| \leq |\psi_n(x)| + m + 2 \leq h(x) + m + 2 \in L^2(\Omega_2).$$

However, this bound cannot be used in the second term because e^h might not belong even to $L^1(\Omega_2)$. In order to find an L^1-majorant for the second summand in (4.35), we distinguish the following two cases:
Case 1: $\psi_n(x) > 0$. Then $|\ln\left(\Theta\left(\psi_n(x)\right)\right)| \leq |\ln\left(\Theta\left(h(x)\right)\right)|$.

Case 2: $\psi_n(x) \leq 0$. We have $\Theta\left(\psi_n(x)\right) \leq 1$. Therefore, $0 \geq \ \psi_n(x) \geq -h(x)$. Since $\Theta(s)$ is a monotonically increasing function,

$$\Theta(0) = 1 \geq \Theta(\psi_n(x)) \geq \Theta(-h(x)) > 0.$$

From here we obtain

$$\ln(1) = 0 \geq \ln\left(\Theta\left(\psi_n(x)\right)\right) \geq \ln\left(\Theta\left(-h(x)\right)\right)$$

and using the relation $\Theta(-h) = \frac{1}{\Theta(h)}$ we conclude that

$$|\ln\left(\Theta\left(\psi_n(x)\right)\right)| \leq |\ln\left(\Theta\left(-h(x)\right)\right)| = |\ln\left(\Theta\left(h(x)\right)\right)|.$$

Finally, for almost all $x \in \Omega_2$ we have

$$\begin{aligned} &|z_n(x)| = |\ln\left(\Theta\left(\psi_n(x)\right)\right) - w_n(x)| \leq |\ln\left(\Theta\left(h(x)\right)\right)| + m + 2 \\ &= \ln\left(\Theta\left(h(x)\right)\right) + m + 2, \text{ because } h(x) \geq 0, \text{ for a.e. } x \in \Omega_2. \end{aligned}$$

Therefore,

$$\begin{aligned} &\left|(\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x))\, z_n(x) - k^2(x) \cosh\left(z_n(x) + w(x)\right)\right| \\ &\leq |\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x)|\, \left(h(x) + \|w\|_{L^\infty(\Omega_2)} + 2\right) \\ &+ k^2(x) \mathrm{e}^{2\|w\|_{L^\infty(\Omega_2)} + 2} \Theta\left(h(x)\right) := H(x) \in L^2(\Omega_2), \end{aligned}$$

where in the last line we have used the fact that $\Theta\left(h(x)\right) \in L^2(\Omega_2)$. All the conditions of the Lebesgue's dominated convergence theorem are satisfied and we see that $I_{\boldsymbol{y}_0^*}(z_n) \to I_{\boldsymbol{y}_0^*}(\xi_0)$ and, consequently, $\sup\limits_{z \in H_0^1(\Omega)} I_{\boldsymbol{y}_0^*}(z) = I_{\boldsymbol{y}_0^*}(\xi_0)$. □

Remark 4.6
Since $|\Omega_2| < \infty$ it follows that $L^\infty(\Omega_2) \subset L^p(\Omega_2)$ for all $1 \leq p < \infty$. Therefore, if we fix one such p, by the density of $C_0^\infty(\Omega_2)$ in $L^p(\Omega_2)$ we can find a sequence $\{\psi_n\}_{n=1}^\infty \subset C_0^\infty(\Omega_2)$ such that $\psi_n \to w$ in $L^p(\Omega)$. From Theorem 4.9 in [96] it follows that there is a subsequence (not relabeled) such that $\psi_n(x) \to w(x)$ for a.e. $x \in \Omega$. Now, if $m = \|w\|_{L^\infty(\Omega_2)}$, let $\varphi : \mathbb{R} \to \mathbb{R}$ be a smooth function such that

$$\varphi(t) = \begin{cases} t, & |t| \leq m+1, \\ m+2, & t > m+2, \\ -m-2, & t < -m-2. \end{cases}$$

Then for the sequence $\{w_n\}_{n=1}^\infty$ defined by $w_n := \varphi \circ \psi_n$, $\forall n \in \mathbb{N}$, we have $w_n \in C_0^\infty(\Omega_2)$, $\forall n \in \mathbb{N}$, $\|w_n\|_{L^\infty(\Omega_2)} \leq m+2$, $\forall n \in \mathbb{N}$, and $w_n(x) = \varphi(\psi_n(x)) \to \varphi(w(x)) = w(x)$ for a.e. $x \in \Omega_2$.

Remark 4.7
Let us denote $f_x(s) = f(x,s) := B(x, s + w(x)) - f_0(x)s$, $\forall s \in \mathbb{R}$ and for a.e. $x \in \Omega$. If we define the functional $\tilde{F} : L^2(\Omega) \to \mathbb{R} \cup \{+\infty\}$ by the relation

$$\tilde{F}(z) := \int_\Omega f_x(z(x))dx = \int_\Omega f(x, z(x))dx = \int_\Omega (B(x, z+w) - f_0 z)\, dx,$$

then the Fenchel conjugate functional of $\tilde{F}$ (with respect to the pairing of $L^2(\Omega)$ and $L^2(\Omega)$ given by the inner product in $L^2(\Omega)$), evaluated at $\operatorname{div} \boldsymbol{y}_0^* \in L^2(\Omega)$, is given by the relation

$$\tilde{F}^*(\operatorname{div} \boldsymbol{y}_0^*) = \sup_{z \in L^2(\Omega)} \left\{ \int_\Omega \operatorname{div} \boldsymbol{y}_0^* z dx - \int_\Omega f(x, z(x))dx \right\} = \sup_{z \in L^2(\Omega)} I_{\boldsymbol{y}_0^*}(z). \tag{4.36}$$

By a well known theorem due to Rockafellar (Theorem 2 in [161]) it follows that the Fenchel conjugate functional of $\tilde{F}$ evaluated at $\operatorname{div} \boldsymbol{y}_0^* \in L^2(\Omega)$ (with respect to the pairing of $L^2(\Omega)$ and $L^2(\Omega)$ given by the inner product in $L^2(\Omega)$) is given by the relation

$$\tilde{F}^*(\operatorname{div} \boldsymbol{y}_0^*) = \int_\Omega f_x^*(\operatorname{div} \boldsymbol{y}_0^*(x))dx = \int_\Omega f^*(x, \operatorname{div} \boldsymbol{y}_0^*(x))dx,$$

where $f_x^*(\cdot)$ is the conjugate to $f_x(\cdot)$ for a.e. $x \in \Omega$ (with respect to the pairing between $\mathbb{R}$ and $\mathbb{R}$ given by $(a^*, a) = a^* a$ for all $a^*, a \in \mathbb{R}$).

The purpose of this remark is to pay attention to the fact that the result of Rockafellar is applicable when one finds Fenchel conjugates of integral functionals defined over spaces that are "decomposable" (such as L^p-spaces). However, $H_0^1(\Omega)$ is not of this type, and therefore the computations in (4.31) *and Prposition 4.5 are required.*

Remark 4.8
Notice that since

$$\xi_0(x) = \ln\left(\Theta\left(\rho(\boldsymbol{y}_0^*)\right)\right) - w \in L^2(\Omega_2), \text{ for a.e. } x \in \Omega_2,$$

then ξ_0 is in $L^2(\Omega)$ when extended by zero in Ω_1. Now, taking into account the fact that

$$\sup_{z\in L^2(\Omega)} I_{\boldsymbol{y}_0^*}(z) \leq \int_{\Omega_2} \sup_{\xi\in\mathbb{R}} \Big\{ \left(\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x)\right)\xi - B\left(x, \xi + w(x)\right) \Big\} dx = I_{\boldsymbol{y}_0^*}(\xi_0)$$

we see that the supremum over $L^2(\Omega)$ above is actually achieved and $\sup_{z\in L^2(\Omega)} I_{\boldsymbol{y}_0^*}(z) = I_{\boldsymbol{y}_0^*}(\xi_0)$. Therefore, by Proposition 4.5 it follows that

$$\sup_{z\in H_0^1(\Omega)} I_{\boldsymbol{y}_0^*}(z) = I_{\boldsymbol{y}_0^*}(\xi_0) = \sup_{z\in L^2(\Omega)} I_{\boldsymbol{y}_0^*}(z). \tag{4.37}$$

One may ask if we could exploit the density of $H_0^1(\Omega)$ in $L^2(\Omega)$ to show the left equality in (4.37) *and not having to go through Proposition 4.5. If $I_{\boldsymbol{y}_0^*}$ was continuous, then the left equality in* (4.37) *follows from the density of $H_0^1(\Omega)$ in $L^2(\Omega)$. However, the functional $I_{\boldsymbol{y}_0^*}$ is only upper semicontinuous and not continuous over $L^2(\Omega)$. Hence, a density argument is not applicable.*

Error measures

In this section, we apply the abstract framework from Section 4.1.2 and derive explicit form of relation (4.14) adapted to our problem. For any $\boldsymbol{y}^* \in \left[L^2(\Omega)\right]^d$ of the form (4.29) with $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1, the quantity $M_\oplus^2(v, \boldsymbol{y}^*)$ is fully computable and is given by the relation (4.11). To give an explicit expression for $M_\oplus^2(v, \boldsymbol{y}^*)$ we first have to compute $D_G(\Lambda v, \boldsymbol{y}^*)$ and $D_F(v, -\Lambda^*\boldsymbol{y}^*)$. For the first compound functional we obtain

$$\begin{aligned} D_G(\Lambda v, \boldsymbol{y}^*) =& G(\Lambda v) + G^*(\boldsymbol{y}^*) - (\boldsymbol{y}^*, \Lambda v) \\ =& \int_\Omega \frac{1}{2}\boldsymbol{A}\nabla v \cdot \nabla v dx + \int_\Omega \frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{y}^* \cdot \boldsymbol{y}^* dx - \int_\Omega \boldsymbol{y}^* \cdot \nabla v dx \\ =& \int_\Omega \frac{1}{2}\boldsymbol{A}^{-1}(\boldsymbol{A}\nabla v - \boldsymbol{y}^*) \cdot (\boldsymbol{A}\nabla v - \boldsymbol{y}^*) dx = \frac{1}{2}|||\boldsymbol{A}\nabla v - \boldsymbol{y}^*|||_*^2. \end{aligned} \tag{4.38}$$

For the second compound functional we have

$$\begin{aligned} D_F(v, -\Lambda^*\boldsymbol{y}^*) =& F(v) + F^*(-\Lambda^*\boldsymbol{y}^*) + \langle \Lambda^*\boldsymbol{y}^*, v\rangle \\ =& \int_\Omega \left(k^2 \cosh(v+w) - f_0 v - \boldsymbol{f}\cdot\nabla v\right) dx \\ &+ \int_{\Omega_2} \left[k^2 \left(\frac{\operatorname{div}\boldsymbol{y}_0^* + f_0}{k^2}\right)\left(\operatorname{arsinh}\left(\frac{\operatorname{div}\boldsymbol{y}_0^* + f_0}{k^2}\right) - w\right)\right. \\ &\left. - k^2 \cosh\left(\operatorname{arsinh}\left(\frac{\operatorname{div}\boldsymbol{y}_0^* + f_0}{k^2}\right)\right)\right] dx + \int_\Omega \boldsymbol{y}^*\cdot\nabla v dx, \end{aligned}$$

from which by using the relation $\boldsymbol{y}^* = \boldsymbol{f} + \boldsymbol{y}_0^*$ and the fact that $\boldsymbol{y}_0^*$ is in $H(\mathrm{div};\Omega)$ with $\mathrm{div}\,\boldsymbol{y}_0^* + f_0 = 0$ in Ω_1 we obtain

$$\begin{aligned} D_F(v,-\Lambda^*\boldsymbol{y}^*) = \int\limits_{\Omega_2} k^2 \Bigg[&\cosh(v+w) + \left(\frac{\mathrm{div}\,\boldsymbol{y}_0^*+f_0}{k^2}\right)\mathrm{arsinh}\left(\frac{\mathrm{div}\,\boldsymbol{y}_0^*+f_0}{k^2}\right) \\ &- \cosh\left(\mathrm{arsinh}\left(\frac{\mathrm{div}\,\boldsymbol{y}_0^*+f_0}{k^2}\right)\right) - \left(\frac{\mathrm{div}\,\boldsymbol{y}_0^*+f_0}{k^2}\right)(v+w)\Bigg]\,dx. \end{aligned} \tag{4.39}$$

The fully computable majorant $M_\oplus^2(v,\boldsymbol{y}^*)$ is given by

$$\begin{aligned} M_\oplus^2(v,\boldsymbol{y}^*) &= D_G(\Lambda v,\boldsymbol{y}^*) + D_F(v,-\Lambda^*\boldsymbol{y}^*) \\ &= \int\limits_\Omega \eta^2(x)dx = \frac{1}{2}|||\boldsymbol{A}\nabla v - \boldsymbol{y}^*|||_*^2 + D_F(v,-\Lambda^*\boldsymbol{y}^*), \end{aligned} \tag{4.40}$$

where $D_G(\Lambda v,\boldsymbol{y}^*)$ and $D_F(v,-\Lambda^*\boldsymbol{y}^*)$ are given by (4.38) and (4.39), respectively, and

$$\eta^2(x) = \begin{cases} \frac{1}{2}\boldsymbol{A}^{-1}(\boldsymbol{A}\nabla v - \boldsymbol{f} - \boldsymbol{y}_0^*)\cdot(\boldsymbol{A}\nabla v - \boldsymbol{f} - \boldsymbol{y}_0^*), \ \text{ for } x\in\Omega_1, \\ \\ \frac{1}{2}\boldsymbol{A}^{-1}(\boldsymbol{A}\nabla v - \boldsymbol{f} - \boldsymbol{y}_0^*)\cdot(\boldsymbol{A}\nabla v - \boldsymbol{f} - \boldsymbol{y}_0^*) \\ \quad +k^2\Big[\cosh(v+w) + \left(\frac{\mathrm{div}\,\boldsymbol{y}_0^*+f_0}{k^2}\right)\mathrm{arsinh}\left(\frac{\mathrm{div}\,\boldsymbol{y}_0^*+f_0}{k^2}\right) \\ \quad -\cosh\left(\mathrm{arsinh}\left(\frac{\mathrm{div}\,\boldsymbol{y}_0^*+f_0}{k^2}\right)\right) - \left(\frac{\mathrm{div}\,\boldsymbol{y}_0^*+f_0}{k^2}\right)(v+w)\Big], \ \text{ for } x\in\Omega_2. \end{cases} \tag{4.41}$$

It is clear that $\eta^2(x)\geq 0$ since it is the sum of the compound functionals (which are nonnegative by the definiton of a Fenchel conjugate) generated by $g_x(\xi) := g(x,\xi) = \frac{1}{2}\boldsymbol{A}(x)\xi\cdot\xi$ and $f_x(s) := B(x,s+w(x)) - f_0(x)s$ and evaluated at $(\nabla v(x),\boldsymbol{y}^*(x))$ and $(v(x),\mathrm{div}\,\boldsymbol{y}_0^*(x))$, respectively. It therefore qualifies as an error indicator, provided that $\boldsymbol{y}_0^*$ is chosen appropriately, which we demonstrate with numerical experiments in the next sections.

To give an explicit form of the principal error identity (4.14), we should also compute the quantities $D_G(\Lambda v,\boldsymbol{p}^*)$, $D_G(\Lambda u,\boldsymbol{y}^*)$, $D_F(v,-\Lambda^*\boldsymbol{p}^*)$, and $D_F(u,-\Lambda^*\boldsymbol{y}^*)$. By using (4.38) and the relation $\boldsymbol{p}^* = \boldsymbol{A}\nabla u$, we obtain

$$D_G(\Lambda v,\boldsymbol{p}^*) = \frac{1}{2}\int\limits_\Omega \boldsymbol{A}\nabla(v-u)\cdot\nabla(v-u)dx = \frac{1}{2}|||\nabla(v-u)|||^2, \tag{4.42}$$

$$D_G(\Lambda u,\boldsymbol{y}^*) = \frac{1}{2}\int\limits_\Omega \boldsymbol{A}^{-1}(\boldsymbol{y}^*-\boldsymbol{p}^*)\cdot(\boldsymbol{y}^*-\boldsymbol{p}^*)dx = \frac{1}{2}|||\boldsymbol{y}^*-\boldsymbol{p}^*|||_*^2. \tag{4.43}$$

Now, we find explicit expressions for the nonlinear measures $D_F(v,-\Lambda^*\boldsymbol{p}^*)$ and $D_F(u,-\Lambda^*\boldsymbol{y}^*)$ similar to the ones for the case of quadratic F in (4.15) for the linear elliptic equation $-\mathrm{div}(\boldsymbol{A}\nabla u) + u = f_0$. We will need the following assertion:

Proposition 4.9
For all $s,t\in\mathbb{R}$ it holds

$$\frac{(t-s)^2}{2} \leq U(s,t) \leq \frac{(\sinh(t)-\sinh(s))^2}{2}, \tag{4.44}$$

where $U(s,t) = \cosh(t) - \cosh(s) + s\sinh(s) - t\sinh(s)$.

Proof. For the first inequality, denote

$$U_1(s,t) := U(s,t) - \frac{(t-s)^2}{2}.$$

We prove that for any fixed $s \in \mathbb{R}$, $U_1(s,t) \geq 0$ for all $t \in \mathbb{R}$. If $s = 0$, we have $\cosh(t) - 1 \geq \frac{t^2}{2}$ for all $t \in \mathbb{R}$. If $s \neq 0$, the necessary condition for a minimum in t is $\frac{\partial U_1}{\partial t}(s,t) = 0$ which is equivalent to $\sinh(t) - \sinh(s) - t + s = 0$. The only solution of this equation is $t = s$ because the function $\sinh(t) - t$ is strictly monotonically increasing. It is left to observe that at $t = s$ we have $\frac{\partial^2 U_1}{\partial t^2} = \cosh(s) - 1 > 0$ and that $U_1(s, t = s) = 0$.

For the second inequality, denote

$$U_2(s,t) := \frac{(\sinh(t) - \sinh(s))^2}{2} - U(s,t).$$

If $t = 0$, the inequality $U_2(s,0) \geq 0$ reduces to the inequality $q(s) := \frac{\sinh^2(s)}{2} - 1 + \cosh(s) - s\sinh(s) \geq 0$ which is true since the minimum of the function $q(s)$ is 0. If $t \neq 0$, the necessary condition for a minimum in s is $\frac{\partial U_2}{\partial s} = 0$ which is equivalent to $\cosh(s)(\sinh(s) - \sinh(t) - s + t) = 0$. The only solution of this equation is $s = t$. Now, it is left to observe that at $s = t$ we have $\frac{\partial^2 U_2}{\partial s^2} = \cosh(t)(\cosh(t) - 1) > 0$ and that $U_2(s = t, t) = 0$. □

Since for the exact solution u of (4.2) we have

$$\frac{\operatorname{div} \boldsymbol{p}_0^* + f_0}{k^2} = \sinh(u + w) \quad \text{and} \quad u = \operatorname{arsinh}\left(\frac{\operatorname{div} \boldsymbol{p}_0^* + f_0}{k^2}\right) - w \text{ a.e. in } \Omega_2,$$

by using (4.39) we obtain

$$D_F(v, -\Lambda^* \boldsymbol{p}^*) = \int_{\Omega_2} k^2 \left(\cosh(v + w) - \cosh(u + w) + u\sinh(u + w) - v\sinh(u + w)\right) dx. \tag{4.45}$$

Similarly,

$$D_F(u, -\Lambda^* \boldsymbol{y}^*) = \int_{\Omega_2} k^2 \left(\cosh(T) - \cosh(S) + S\sinh(S) - T\sinh(S)\right) dx, \tag{4.46}$$

where

$$T := \operatorname{arsinh}\left(\frac{\operatorname{div} \boldsymbol{p}_0^* + f_0}{k^2}\right) \quad \text{and} \quad S := \operatorname{arsinh}\left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2}\right).$$

The nonlinear quantities $D_F(v, -\Lambda^* \boldsymbol{p}^*)$ and $D_F(u, -\Lambda^* \boldsymbol{y}^*)$ measure the error over Ω_2 in v and in $\operatorname{div} \boldsymbol{y}_0^*$, respectively. Using inequality (4.44), we can represent these two measures in a

form, which resembles the corresponding estimates in the case (4.15) of a quadratic functional F, namely,

$$\int\limits_{\Omega_2} \frac{k^2}{2}(v-u)^2 dx \le D_F(v, -\Lambda^* \boldsymbol{p}^*) \le \int\limits_{\Omega_2} \frac{k^2}{2}(\sinh(v+w) - \sinh(u+w))^2 dx \tag{4.47}$$

and

$$\int\limits_{\Omega_2} \frac{k^2}{2}(S-T)^2 dx \le D_F(u, -\Lambda^* \boldsymbol{y}^*) \le \int\limits_{\Omega_2} \frac{1}{2k^2}(\operatorname{div} \boldsymbol{y}_0^* - \operatorname{div} \boldsymbol{p}_0^*)^2 dx. \tag{4.48}$$

Note that if the coefficient k satisfies case (a), i.e., $k_{\max} \ge k \ge k_{\min} > 0$ in Ω, the equivalences $\int\limits_{\Omega} \frac{k^2}{2}(v-u)^2 dx \eqsim \|v-u\|^2_{L^2(\Omega)}$ and $\int\limits_{\Omega} \frac{1}{2k^2}(\operatorname{div} \boldsymbol{y}_0^* - \operatorname{div} \boldsymbol{p}_0^*)^2 dx \eqsim \|\operatorname{div} \boldsymbol{y}_0^* - \operatorname{div} \boldsymbol{p}_0^*\|^2_{L^2(\Omega)}$ hold (in this case the requirement $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1 is not needed: look at the derivation of $F^*(-\Lambda^* \boldsymbol{y}^*)$ in (4.31)). Moreover, replacing the nonlinear term $k^2 \sinh(u+w)$ with u, the inequalities (4.47) and (4.48) reduce to the equalities for $D_F(v, -\Lambda^* \boldsymbol{p}_0^*)$ and $D_F(u, -\Lambda^* \boldsymbol{y}^*)$ in (4.15) because in this case the inverse function of $f(s) = s$ is again $f(s)$. The functions on the left-hand side, in the middle, and on the right-hand side in the inequality (4.44) are depicted on Figure 4.1. Further, if v is in a δ_1-neighborhood of u in $L^\infty(\Omega)$ norm, since sinh is a locally Lipschitz function, we can find a constant $C_1\left(\delta_1, \|u\|_{L^\infty(\Omega)}\right) > 1$ such that

$$\int\limits_{\Omega_2} \frac{k^2}{2}(\sinh(v+w) - \sinh(u+w))^2 dx \le C_1\left(\delta_1, \|u\|_{L^\infty(\Omega)}\right) \int\limits_{\Omega_2} \frac{k^2}{2}(v-u)^2 dx. \tag{4.49}$$

Analogously, if $f_0 \in L^\infty(\Omega_2)$ and $\|\operatorname{div}(\boldsymbol{y}_0^* - \boldsymbol{p}_0^*)\|_{L^\infty(\Omega_2)} \le \delta_2$ (recall that when $f_0 \in L^\infty(\Omega_2)$, $\operatorname{div} \boldsymbol{p}_0^*$ is in $L^\infty(\Omega_2)$ and if $f_0 \in L^\infty(\Omega)$, then $\operatorname{div} \boldsymbol{p}_0^*$ is in $L^\infty(\Omega)$), then we can find a constant $C_2\left(\delta_2, \|\operatorname{div} \boldsymbol{p}_0^*\|_{L^\infty(\Omega_2)}\right) < 1$ such that

$$C_2\left(\delta_2, \|\operatorname{div} \boldsymbol{p}_0^*\|_{L^\infty(\Omega_2)}\right) \int\limits_{\Omega_2} \frac{1}{2k^2}(\operatorname{div} \boldsymbol{y}_0^* - \operatorname{div} \boldsymbol{p}_0^*)^2 dx \le \int\limits_{\Omega_2} \frac{k^2}{2}(S-T)^2 dx. \tag{4.50}$$

Here, the constant C_2 is again a Lipschitz constant for the locally Lipschitz function sinh. Notice that if $k^2_{\max} \ge k^2 \ge k^2_{\min} > 0$ in Ω, then everywhere in (4.47), (4.48), (4.49), and (4.50), the integrals are taken over the entire domain Ω. Now, the abstract error identity (4.14) takes the form

$$\begin{aligned}
&\frac{1}{2}|||\nabla(v-u)|||^2 + \frac{1}{2}|||\boldsymbol{y}^* - \boldsymbol{p}^*|||^2_* \\
&+ \int\limits_{\Omega_2} \frac{k^2}{2}(v-u)^2 dx + C_2\left(\delta_2, \|\operatorname{div} \boldsymbol{p}_0^*\|_{L^\infty(\Omega_2)}\right) \int\limits_{\Omega_2} \frac{1}{2k^2}(\operatorname{div} \boldsymbol{y}_0^* - \operatorname{div} \boldsymbol{p}_0^*)^2 dx \\
&\le \frac{1}{2}|||\nabla(v-u)|||^2 + \frac{1}{2}|||\boldsymbol{y}^* - \boldsymbol{p}^*|||^2_* + D_F(v, -\Lambda^* \boldsymbol{p}^*) + D_F(u, -\Lambda^* \boldsymbol{y}^*) = M^2_{\oplus}(v, \boldsymbol{y}^*) \\
&\le \frac{1}{2}|||\nabla(v-u)|||^2 + \frac{1}{2}|||\boldsymbol{y}^* - \boldsymbol{p}^*|||^2_* \\
&+ C_1\left(\delta_1, \|u\|_{L^\infty(\Omega)}\right) \int\limits_{\Omega_2} \frac{k^2}{2}(v-u)^2 dx + \int\limits_{\Omega_2} \frac{1}{2k^2}(\operatorname{div} \boldsymbol{y}_0^* - \operatorname{div} \boldsymbol{p}_0^*)^2 dx.
\end{aligned} \tag{4.51}$$

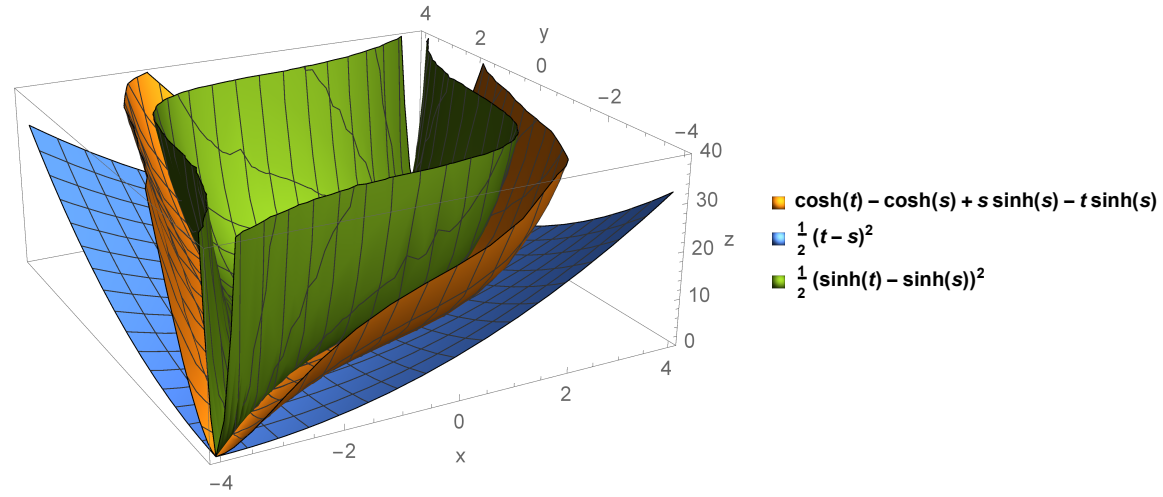


Figure 4.1: Functions in the inequality (4.44).

Relation (4.51) shows that the computable majorant $M^2_{\oplus}(v, \boldsymbol{y}^*)$ is bounded from below and above by a multiple of one and the same error norm. Since $D_F(v, -\Lambda^* \boldsymbol{p}^*) \geq 0$ and $D_F(u, -\Lambda^* \boldsymbol{y}^*) \geq 0$ we also obtain a guaranteed bound on the error in the combined energy norm:

$$|||\nabla(u-v)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2 \leq 2M^2_{\oplus}(v, \boldsymbol{y}^*). \tag{4.52}$$

Moreover, from the pointwise equality

$$\begin{aligned} &\boldsymbol{A}^{-1}(\boldsymbol{A}\nabla v - \boldsymbol{y}^*) \cdot (\boldsymbol{A}\nabla v - \boldsymbol{y}^*) \\ =&\boldsymbol{A}^{-1}(\boldsymbol{A}\nabla(v-u) - (\boldsymbol{y}^* - \boldsymbol{p}^*)) \cdot (\boldsymbol{A}\nabla(v-u) - (\boldsymbol{y}^* - \boldsymbol{p}^*)) \\ =&\boldsymbol{A}\nabla(v-u) \cdot \nabla(v-u) + \boldsymbol{A}^{-1}(\boldsymbol{y}^* - \boldsymbol{p}^*) \cdot (\boldsymbol{y}^* - \boldsymbol{p}^*) - 2(\boldsymbol{y}^* - \boldsymbol{p}^*) \cdot \nabla(v-u), \end{aligned} \tag{4.53}$$

after applying Young's inequality and integrating over Ω, we obtain a lower bound for the error in combined energy norm:

$$\frac{1}{2}|||\boldsymbol{A}\nabla v - \boldsymbol{y}^*|||_*^2 \leq |||\nabla(v-u)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2. \tag{4.54}$$

Remark 4.10

Integrating (4.53) *over* Ω *we obtain the algebraic identity*

$$|||\boldsymbol{A}\nabla v - \boldsymbol{y}^*|||_*^2 = |||\nabla(v-u)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2 - 2\int_\Omega (\boldsymbol{y}^* - \boldsymbol{p}^*) \cdot \nabla(v-u)dx, \tag{4.55}$$

from which the Prager-Synge identity is derived. Indeed, since $\boldsymbol{y}^* = \boldsymbol{f} + \boldsymbol{y}_0^*$ *and* $\boldsymbol{p}^* = \boldsymbol{f} + \boldsymbol{p}_0^*$ *with* $\boldsymbol{y}_0^*$, $\boldsymbol{p}_0^* \in H(\mathrm{div}; \Omega)$, *we obtain*

$$|||\boldsymbol{A}\nabla v - \boldsymbol{y}^*|||_*^2 = |||\nabla(v-u)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2 + 2\int_\Omega \mathrm{div}(\boldsymbol{y}_0^* - \boldsymbol{p}_0^*)(v-u)dx. \tag{4.56}$$

Now, if $\operatorname{div} \boldsymbol{y}_0^* = \operatorname{div} \boldsymbol{p}_0^* = -f_0 + k^2 \sinh(u+w)$ *in* Ω*, i.e., if* $\boldsymbol{y}_0^*$ *is exactly equilibrated, the above equality implies the Prager-Synge equality.*

Comparing (4.55) *with* (4.51)*, by using the fact that*

$$M_{\oplus}(v, \boldsymbol{y}^*)^2 = \frac{1}{2} |\!|\!| \boldsymbol{A}\nabla v - \boldsymbol{y}^* |\!|\!|_*^2 + D_F(v, -\Lambda^* \boldsymbol{y}^*),$$

we arrive at the relation

$$D_F(v, -\Lambda^* \boldsymbol{y}^*) = D_F(v, -\Lambda^* \boldsymbol{p}^*) + D_F(u, -\Lambda^* \boldsymbol{y}^*) + \int_\Omega (\boldsymbol{y}^* - \boldsymbol{p}^*) \cdot \nabla(v-u) dx, \qquad (4.57)$$

which is an analogue of the Prager-Synge identity in the case when $\boldsymbol{y}_0^*$ *is exactly equilibrated. From here, it is seen that if the integral on the right-hand side of* (4.57) *is small compared to the other terms, then the error in* v *and* $\operatorname{div} \boldsymbol{y}_0^*$ *measured with* $D_F(v, -\Lambda^* \boldsymbol{p}^*) + D_F(u, -\Lambda^* \boldsymbol{y}^*)$ *is controlled mainly by the computable term* $D_F(v, -\Lambda^* \boldsymbol{y}^*)$ *in the majorant* $M_{\oplus}^2(v, \boldsymbol{y}^*)$*. Moreover,* (4.55) *enables us to give a practical estimation of the error in combined energy norm, which is very close to the real error in all of the experiments that we have conducted.*

Remark 4.11 (Convergence order under uniform refinement with Lagrange P_1 elements) *Note that for a smooth coefficient matrix* $\boldsymbol{A}$ *with* $\boldsymbol{f} = 0$*,* $f_0 \in L^2(\Omega)$*,* $b(x,s) = s$*,* Ω *a bounded open convex polyhedral domain, and a homogeneous Dirichlet boundary condition, we have that the solution* u *of the problem*

$$\begin{aligned} -\operatorname{div}(\boldsymbol{A}\nabla u) + u &= f_0 \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega \end{aligned}$$

is in $H^2(\Omega)$*. In this case,* $\boldsymbol{y}^* = \boldsymbol{y}_0^* \in H(\operatorname{div}; \Omega)$ *(note that we do not require the condition* $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ *in* Ω_1 *since in this case* $k^2 = 1 > 0$ *everywhere in* Ω*) and we have* $\|u_h - u\|_{L^2(\Omega)} = O(h^2)$*, where* $\{u_h\}$ *are the Galerkin approximations of* u *in the finite element spaces* $\{V_h\}_{h\to 0}$ *of continuous piecewise linear functions defined over a regular family of triangulations* $\{\mathscr{T}_h\}_{h\to 0}$*. Moreover, let* L_h *be the finite element space of piecewise constant functions over* $\mathscr{T}_h$ *and let* Π_{L_h} *be the* $L^2(\Omega)$ *projection onto the space* $L_h \subset L^2(\Omega)$*. Then, if* $\boldsymbol{y}^*$ *is partially equilibrated, i.e., such that* $\operatorname{div} \boldsymbol{y}^* = \Pi_{L_h}(-f_0 + u_h)$ *(in particular,* $\Pi_{L_h}(-f_0) = -f_0$ *when* f_0 *is piecewise constant over* $\mathscr{T}_h$*), then we find*

$$\begin{aligned} \|\operatorname{div}(\boldsymbol{y}^* - \boldsymbol{p}^*)\|_{L^2(\Omega)} &= \|\Pi_{L_h}(-f_0 + u_h) - (-f_0 + u)\|_{L^2(\Omega)} \\ &\le \|f_0 - \Pi_{L_h}(f_0)\|_{L^2(\Omega)} + \|\Pi_{L_h}(u_h) - u\|_{L^2(\Omega)}. \end{aligned} \qquad (4.58)$$

To estimate the term $\|f_0 - \Pi_{L_h}(f_0)\|_{L^2(\Omega)}$*, let us additionally assume that* $f_0 \in H^1(K)$ *for all* $K \in \mathscr{T}_h$*. In this case, by using Poincaré's inequality (for functions with zero mean) on each*

element $K \in \mathscr{T}_h$ we obtain

$$\begin{aligned}\|f_0 - \Pi_{L_h}(f_0)\|^2_{L^2(\Omega)} &= \int_\Omega \left(f_0 - \Pi_{L_h}(f_0)\right)^2 dx = \sum_{K\in\mathscr{T}_h} \int_K \left(f_0 - \frac{1}{|K|}\int_K f_0 dx\right)^2 dx \\ &\leq \sum_{K\in\mathscr{T}_h} C_P^2(K)\|\nabla f_0\|^2_{L^2(K)} \leq \frac{h^2}{\pi^2}\sum_{K\in\mathscr{T}_h}\|\nabla f_0\|^2_{L^2(K)} = O(h^2),\end{aligned} \tag{4.59}$$

where we have used the fact that on the convex domains K the Poincaré constant $C_P(K)$ is bounded by $\frac{h_K}{\pi} \leq \frac{h}{\pi}$ (see [149]) with h being the maximum diameter among all diameters h_K of elements K in the triangulation $\mathscr{T}_h$. For the term $\|\Pi_{L_h}(u_h) - u\|_{L^2(\Omega)}$, since $u_h \in H^1(\Omega)$, similarly to (4.59)*, we have*

$$\begin{aligned}\|\Pi_{L_h}(u_h) - u\|_{L^2(\Omega)} &\leq \|\Pi_{L_h}(u_h) - u_h\|_{L^2(\Omega)} + \|u_h - u\|_{L^2(\Omega)} \\ &\leq \frac{h}{\pi}\|\nabla u_h\|_{L^2(\Omega)} + O(h^2) = O(h),\end{aligned} \tag{4.60}$$

where we have used the fact that $\|\nabla u_h\|_{L^2(\Omega)}$ is bounded due to the fact that $\|\nabla(u_h - u)\|_{L^2(\Omega)} \to 0$ as $h \to 0$. By taking into account (4.58)*,* (4.59)*, and* (4.60) *we obtain*

$$\int_\Omega (\boldsymbol{y}^* - \boldsymbol{p}^*)\cdot\nabla(u_h - u)dx = -\int_\Omega \operatorname{div}(\boldsymbol{y}^* - \boldsymbol{p}^*)(u_h - u)dx = O(h^3),$$

$$|||\nabla(u_h - u)|||^2 = O(h^2), \qquad D_F(u_h, -\Lambda^*\boldsymbol{p}^*) = \frac{1}{2}\|u_h - u\|^2_{L^2(\Omega)} = O(h^4),$$
$$D_F(u, -\Lambda^*\boldsymbol{y}^*) = \frac{1}{2}\|\operatorname{div}(\boldsymbol{y}^* - \boldsymbol{p}^*)\|^2_{L^2(\Omega)} = O(h^2).$$

Depending on the approach used to find an approximation $\boldsymbol{y}^*$ of the exact flux $\boldsymbol{p}^* = \boldsymbol{A}\nabla u$, we may also have $|||\boldsymbol{y}^* - \boldsymbol{p}^*|||^2_* = O(h^2)$.

Guaranteed lower and upper bounds for the primal part of the error

From Section 4.1.2 we know that in abstract form the primal part of the error is given by

$$J(v) - J(u) = D_G(\Lambda v, p^*) + D_F(v, -\Lambda^* p^*). \tag{4.12a}$$

By using (4.13) and the fact that u is a minimizer of J, for any approximation $v \in V$ of u and any $w \in V$, $y^* \in Y^*$ we obtain

$$M^2_\ominus(v, w) := J(v) - J(w) \leq J(v) - J(u) \leq M^2_\oplus(v, y^*), \tag{4.61}$$

where the left-hand side of (4.61) makes sense only if $J(v) - J(w) \geq 0$. In particular, for the case of the linear problem (4.16), the primal part of the error is given by the expression:

$$2\left(J(v) - J(u)\right) = |||\nabla(v - u)|||^2 + \|v - u\|^2_{L^2(\Omega)}, \tag{4.62}$$

where we have used the expressions for $D_G(\Lambda v, p^*)$ and $D_F(v, -\Lambda^* p^*)$ in (4.15). Thus, for any approximation $v \in H_0^1(\Omega)$ of u and any $w \in V = H_0^1(\Omega)$, $\boldsymbol{y}^* \in H(\mathrm{div};\Omega) \subset \left[L^2(\Omega)\right]^d = Y^*$ (4.61) takes the explicit form

$$\begin{aligned} 2\left(J(v) - J(w)\right) &\leq |||\nabla(v-u)|||^2 + \|v-u\|^2_{L^2(\Omega)} \\ &\leq |||\boldsymbol{A}\nabla v - \boldsymbol{y}^*|||^2_* + \|v - \mathrm{div}\,\boldsymbol{y}^* - f_0\|^2_{L^2(\Omega)}. \end{aligned} \tag{4.63}$$

Obviously, when $\boldsymbol{A} = I$, the identity matrix, then (4.62) coincides with the squared $H^1(\Omega)$ norm of the error. As we have seen by (4.47) and (4.49) in the nonlinear case, the term $D_F(v, -\Lambda^*\boldsymbol{p}^*)$ is equivalent to the squared $L^2(\Omega_2)$-norm of the error and hence the "nonlinear" measure of the error (4.12a) is an analogue of the squared H^1 norm of the error in the linear case. In a similar way one can also derive lower and upper bounds for the dual part of the error given by (4.12b).

Near best approximation result

Here we present a near best approximation result which is a byproduct of the functional a posteriori error estimates that we have derived above. Contrary to the result in [48, Theorem 6.2], we do not make any restrictive assumptions on the meshes to ensure that the finite element approximations u_h are uniformly bounded in $L^\infty(\Omega)$ norm. In our analysis, $V_h \subset L^\infty(\Omega)$ is a closed subspace of $H_0^1(\Omega)$ (not necessarily finite dimensional) and u_h is the (unique) minimizer of J (J is defined by (4.4)) over V_h, which is the unique solution of the Galerkin problem:

$$\begin{aligned} &\text{Find } u_h \in V_h \text{ such that} \\ &a(u_h, v) + \int_\Omega b(x, u_h + w)v\,dx = \int_\Omega \left(f_0 v + \boldsymbol{f}\cdot\nabla v\right)dx \;\text{ for all } v \in V_h. \end{aligned} \tag{4.64}$$

Then, using (4.12a) and the expression (4.42) for $D_G(\Lambda v, \boldsymbol{p}^*)$, for any $v \in V_h$ we can write

$$\begin{aligned} |||\nabla(u_h - u)|||^2 + 2D_F(u_h, -\Lambda^*\boldsymbol{p}^*) &= 2\left(J(u_h) - J(u)\right) \\ \leq 2\left(J(v) - J(u)\right) &= |||\nabla(v-u)|||^2 + 2D_F(v, -\Lambda^*\boldsymbol{p}^*). \end{aligned}$$

Since $2D_F(u_h, -\Lambda^*\boldsymbol{p}^*) \geq 0$, by using (4.47), we obtain the following generalization of Cea's Lemma to the case of our nonlinear problem.

Proposition 4.12
Let $V_h \subset L^\infty(\Omega)$ be a closed subspace of $H_0^1(\Omega)$ and $u_h \in V_h$ be the Galerkin approximation of u defined by (4.64)*. Then*

$$|||\nabla(u_h - u)|||^2 \leq \inf_{v\in V_h}\left\{|||\nabla(v-u)|||^2 + \int_{\Omega_2} k^2(\sinh(v+w) - \sinh(u+w))^2 dx\right\}. \tag{4.65}$$

For example, if Ω is a polyhedral domain and we use the finite element method with P_1 Lagrange elements, let V_h^1 be the corresponding space where h refers to the maximum element size of a triangulation of Ω into tetrahedrons. By $I_h(\psi)$ we denote the Lagrange finite element interpolant of $\psi \in C^0(\Omega)$. Using (4.65) we can show unqualified convergence of the finite element approximations u_h to u as $h \to 0$. Let $\varepsilon > 0$ and $\bar{u} \in C_0^\infty(\Omega)$ be such that $\|\nabla(\bar{u}-u)\|_{L^2(\Omega)} \leq \varepsilon$ and $\|\bar{u}\|_{L^\infty(\Omega)} \leq \|u\|_{L^\infty(\Omega)}+2$ (see Remark 4.13). Also, let L be the Lipschitz constant in the inequality $|\sinh(s)-\sinh(t)| \leq L\,|s-t|$ for all $s,t \in \left[-\|u\|_{L^\infty(\Omega)}-\|w\|_{L^\infty(\Omega_2)}-2, \|u\|_{L^\infty(\Omega)}+\|w\|_{L^\infty(\Omega_2)}+2\right]$. Then, by applying first Proposition 4.12 with $v = I_h(\bar{u})$ and then the triangle inequality together with Young's inequality, we obtain

$$\begin{aligned} &|||\nabla(u_h-u)|||^2 \leq 2\left(|||\nabla(I_h(\bar{u})-\bar{u})|||^2 + |||\nabla(\bar{u}-u)|||^2\right) \qquad (4.66)\\ &+2\left(\int_{\Omega_2} k^2(\sinh(I_h(\bar{u})+w)-\sinh(\bar{u}+w))^2dx + \int_{\Omega_2} k^2(\sinh(\bar{u}+w)-\sinh(u+w))^2dx\right). \end{aligned}$$

For the first term in (4.66), by assuming mesh regularity, we have

$$|||\nabla(I_h(\bar{u})-\bar{u})|||^2 + |||\nabla(\bar{u}-u)|||^2 \leq \mu_2\left(C|\bar{u}|_2^2h^2+\varepsilon^2\right),$$

where $|\bar{u}|_2$ denotes the H^2 seminorm of $\bar{u}$ and $C>0$ is a constant depending on the mesh regularity. Using the fact that $\|I_h(\bar{u})\|_{L^\infty(\Omega)} \leq \|\bar{u}\|_{L^\infty(\Omega)} \leq \|u\|_{L^\infty(\Omega)}+2$, for the second term in (4.66), we obtain the upper bound

$$\begin{aligned} &2k_{\max}^2L^2\left(\|I_h(\bar{u})-\bar{u}\|_{L^2(\Omega)}^2+\|\bar{u}-u\|_{L^2(\Omega)}^2\right)\\ &\leq 2k_{\max}^2L^2C_P^2\left(\|\nabla(I_h(\bar{u})-\bar{u})\|_{L^2(\Omega)}^2+\|\nabla(\bar{u}-u)\|_{L^2(\Omega)}^2\right) \leq 2k_{\max}^2L^2C_P^2\left(C|\bar{u}|_2^2h^2+\varepsilon^2\right), \end{aligned}$$

where C_P is the constant in the Poincaré's inequality $\|v\|_{L^2(\Omega)} \leq C_P\|\nabla v\|_{L^2(\Omega)}$, $\forall v \in H_0^1(\Omega)$. This inequality shows that the right-hand side of (4.66) can be made as small as desired provided that we choose ε and h small enough and therefore $|||\nabla(u_h-u)||| \to 0$ when $h \to 0$. Moreover, (4.65) can be also used to obtain qualified convergence of u_h in the energy norm under additional assumptions on the regularity of $\boldsymbol{A}$, the meshes, and the regularity of u.

Remark 4.13
Note that if $u \in H_0^1(\Omega)\cap L^\infty(\Omega)$ one can find a sequence $\{u_n\}_{n=1}^\infty \subset C_0^\infty(\Omega)$ such that $u_n \to u$ in $H^1(\Omega)$ and $\|u_n\|_{L^\infty(\Omega)} \leq m+2$, where $m := \|u\|_{L^\infty(\Omega)}$. Indeed, since $C_0^\infty(\Omega)$ is dense in $H_0^1(\Omega)$, we can find a sequence $\{\psi_n\}_{n=1}^\infty \subset C_0^\infty(\Omega)$ such that $\psi_n \to u$ in $H^1(\Omega)$. Up to a subsequence (not relabeled) we have also that $\psi_n(x) \to u(x)$, a.e $x \in \Omega$ and $\nabla\psi_n(x) \to \nabla u(x)$, a.e $x \in \Omega$. Now, let $\varphi : \mathbb{R} \to \mathbb{R}$ be the smooth function defined in Remark 4.6. One can choose φ such that $0 \leq \varphi'(t) \leq 1$, $\forall t \in \mathbb{R}$. The functions $u_n := \varphi \circ \psi_n$ are in $C_0^\infty(\Omega)$ with $\|u_n\|_{L^\infty(\Omega)} \leq m+2$. Now, by using the fact that $|\varphi(t_1)-\varphi(t_2)| \leq |t_1-t_2|$, $\forall t_1,t_2 \in \mathbb{R}$ and the Lebesgue dominated convergence theorem one can show that $\|u_n-u\|_{H^1(\Omega)} \to 0$.

Computing $F^*(-\Lambda^* \boldsymbol{y}^*)$ for a more general nonlinearity of the form $k^2(x)b(s)$

If we consider the more general problem

$$\begin{aligned} &\text{Find } u \in H_0^1(\Omega) \text{ such that for all } v \in H_0^1(\Omega) \\ &\int_\Omega \boldsymbol{A}\nabla u \cdot \nabla v dx + \int_\Omega k^2(x) b(u+w) v dx = \int_\Omega \left(f_0 v + \boldsymbol{f} \cdot \nabla v\right) dx. \end{aligned} \tag{4.67}$$

where $\boldsymbol{A} \in [L^\infty(\Omega)]^{d\times d}$ is as in (4.2), k and w satisfy, for example, case (b), i.e., w is measurable in Ω with $w \in L^\infty(\Omega_2)$ and $k \in L^\infty(\Omega)$ with $k = 0$ in Ω_1 and $k^2_{\max} \geq k^2 \geq k^2_{\min} > 0$ in Ω_2 (in our particular case of the PBE we have $k = 0$ in $\Omega_m \cup \Omega_{IEL}$ and $k = \overline{k}_{ions} = const$ in Ω_{ions}), $b(s)$ is a strictly increasing continuous function on $\mathbb{R}$, $w \in L^\infty(\Omega_2)$, $f_0 \in L^2(\Omega)$, $\boldsymbol{f} \in [L^s(\Omega)]^d$, $s > d$. From Theorem 3.29 we have $u \in L^\infty(\Omega)$. Let $F : H_0^1(\Omega) \to \mathbb{R} \cup \{+\infty\}$ be defined as usual by

$$F(v) = \int_\Omega \left(k^2 B(v+w) - f_0 v - \boldsymbol{f} \cdot \nabla v\right) dx, \tag{4.68}$$

where $B(s) = \int_0^s b(q)dq$. Under appropriate additional conditions on $b(\cdot)$ that also ensure the unique solvability of the primal and dual variational problems (P) and (P^*), as well as the validity of the strong duality relation (4.7), one can compute $F^*(-\Lambda^* \boldsymbol{y}^*)$ explicitly for certain $\boldsymbol{y}^* \in \left[L^2(\Omega)\right]^d$. Following the ideas in (4.31) and Proposition 4.5 we can find an explicit expression of $F^*(-\Lambda^* \boldsymbol{y}^*)$ for any $\boldsymbol{y}^* \in [L^2(\Omega)]^d$ of the form $\boldsymbol{y}^* = \boldsymbol{f} + \boldsymbol{y}_0^*$ with $\boldsymbol{y}_0^* \in H(\mathrm{div}; \Omega)$ and $\mathrm{div}\, \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1:

$$\begin{aligned} F^*(-\Lambda^* \boldsymbol{y}^*) = \int_{\Omega_2} \Bigg[& k^2 \left(\frac{\mathrm{div}\, \boldsymbol{y}_0^* + f_0}{k^2}\right) \left(b^{-1}\left(\frac{\mathrm{div}\, \boldsymbol{y}_0^* + f_0}{k^2}\right) - w\right) \\ & - k^2 B\left(b^{-1}\left(\frac{\mathrm{div}\, \boldsymbol{y}_0^* + f_0}{k^2}\right)\right)\Bigg] dx, \end{aligned} \tag{4.69}$$

where $b^{-1} : \mathbb{R} \to \mathbb{R}$ is the inverse function of $b(\cdot)$. Then we can compute

$$\begin{aligned} D_F(v, -\Lambda^* \boldsymbol{y}^*) = \int_{\Omega_2} k^2 \Bigg[& B(v+w) + \left(\frac{\mathrm{div}\, \boldsymbol{y}_0^* + f_0}{k^2}\right) b^{-1}\left(\frac{\mathrm{div}\, \boldsymbol{y}_0^* + f_0}{k^2}\right) \\ & - B\left(b^{-1}\left(\frac{\mathrm{div}\, \boldsymbol{y}_0^* + f_0}{k^2}\right)\right) - \left(\frac{\mathrm{div}\, \boldsymbol{y}_0^* + f_0}{k^2}\right)(v+w)\Bigg] dx. \end{aligned} \tag{4.70}$$

By using the relation $\frac{\mathrm{div}\, \boldsymbol{p}_0^* + f_0}{k^2} = b(u+w)$, where $\boldsymbol{p}^* = \boldsymbol{f} + \boldsymbol{p}_0^*$, we find

$$\begin{aligned} D_F(v, -\Lambda^* \boldsymbol{p}^*) &= F(v) + F^*(-\Lambda^* \boldsymbol{p}^*) + \langle \Lambda^* \boldsymbol{p}^*, v\rangle \\ &= \int_{\Omega_2} k^2 \left[B(v+w) - B(u+w) + ub(u+w) - vb(u+w)\right] dx. \end{aligned} \tag{4.71}$$

Next, by using the monotonicity of $b(\cdot)$ we observe that for any $s, t \in \mathbb{R}$ the following inequality holds:

$$\begin{aligned} B(t) - B(s) + sb(s) - tb(s) &= \int_s^t b(q)dq - b(s)(t-s) \\ &\leq b(t)(t-s) - b(s)(t-s) = (b(t) - b(s))\,(t-s). \end{aligned} \tag{4.72}$$

Therefore,

$$D_F(v, -\Lambda^* \boldsymbol{p}^*) \leq \int_{\Omega_2} k^2 \left(b(v+w) - b(u+w)\right)(v-u)dx \tag{4.73}$$

and we obtain an analogue of the near best approximation result in Proposition 4.12

$$|||\nabla(u_h - u)|||^2 \leq \inf_{v \in V_h} \left\{ |||\nabla(v-u)|||^2 + \int_{\Omega_2} k^2 \left(b(v+w) - b(u+w)\right)(v-u)dx \right\}, \tag{4.74}$$

where u_h is the Galerkin approximation of u in the closed subspace $V_h \subset H_0^1(\Omega) \cap L^\infty(\Omega)$. Now, if $b(\cdot)$ is Lipschitz continuous with a Lipschitz constant L, then the second term in the infimum in (4.74) is bounded by

$$L \int_{\Omega_2} k^2 (v-u)^2 dx \leq L C_P^2 \|\nabla(v-u)\|^2_{L^2(\Omega)},$$

where C_P is Poincaré's constant in the inequality $\|v\|_{L^2(\Omega)} \leq C_P \|\nabla v\|_{L^2(\Omega)}$, $\forall v \in H_0^1(\Omega)$. If $b(\cdot)$ is only locally Lipsctiz, then qualified and unqualified convergence can be shown by using the local Lipschitz constant of $b(\cdot)$ for the interval $\left[-\|u\|_{L^\infty(\Omega)} - \|w\|_{L^\infty(\Omega_2)} - 2, \|u\|_{L^\infty(\Omega)} + \|w\|_{L^\infty(\Omega_2)} + 2\right]$ similarly to the considerations after Proposition 4.12.

Remark 4.14
When $b(s) = s$, we have $B(s) = \frac{s^2}{2}$ and $b^{-1}(s) = s$. One can easily obtain the following expression for D_F:

$$D_F(v, -\Lambda^* \boldsymbol{y}^*) = \int_{\Omega_2} \frac{1}{2k^2} \left(k^2(v+w) - \operatorname{div} \boldsymbol{y}_0^* - f_0\right)^2 dx, \tag{4.75}$$

where $\boldsymbol{y}^* = \boldsymbol{f} + \boldsymbol{y}_0^*$ with $\boldsymbol{y}_0^* \in H(\operatorname{div}; \Omega)$ and $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1. For any $v \in H_0^1(\Omega)$, the main error identity (4.14) *takes the form*

$$\begin{aligned} &|||\nabla(v-u)|||^2 + \int_{\Omega_2} k^2 (v-u)^2 dx + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2 + \int_{\Omega_2} \frac{1}{k^2} \left(\operatorname{div} \boldsymbol{y}_0^* - \operatorname{div} \boldsymbol{p}_0^*\right)^2 dx \\ =& M_\oplus^2(v, \boldsymbol{y}^*) = |||\boldsymbol{A}\nabla v - \boldsymbol{y}^*|||_*^2 + \int_{\Omega_2} \frac{1}{k^2} \left(k^2(v+w) - \operatorname{div} \boldsymbol{y}_0^* - f_0\right)^2 dx, \end{aligned} \tag{4.76}$$

where $\boldsymbol{p}^* = \boldsymbol{A}\nabla u$ and $\boldsymbol{p}^* = \boldsymbol{f} + \boldsymbol{p}_0^*$ with $\boldsymbol{p}_0^* \in H(\operatorname{div}; \Omega)$ and $\operatorname{div} \boldsymbol{p}_0^* = k^2(u+w) - f_0$ in Ω. In the case of the LPBE with 2-term or 3-term splitting we have $w = G$ or $w = 0$ and $\boldsymbol{f} = \chi_{\Omega_s}(\epsilon_m - \epsilon_s)\nabla G$ or $\boldsymbol{f} = \chi_{\Omega_m}\epsilon_m \nabla u^H + \chi_{\Omega_s}\epsilon_m \nabla G$, respectively, $k = \overline{k}$, $\Omega_1 = \Omega_m \cup \Omega_{IEL}$, $\Omega_2 = \Omega_{ions}$, $f_0 = 0$, $\boldsymbol{A} = \epsilon I$.

Effect of data oscillation

In the case of the PBE, we have $f_0 = 0$ and therefore the equilibration condition $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in the domain $\Omega_1 = \Omega_m \cup \Omega_{IEL}$ can be satisfied exactly. However, for general functions f_0 this condition can only be satisfied approximately. Let V_h^1 and L_h be the finite element spaces of continuous piecewise linear functions and of piecewise constant functions, respectively, defined on a triangulation $\mathscr{T}_h$ of Ω. We assume that we apply a partial equilibration of $\boldsymbol{y}_0^*$ that satisfies $\operatorname{div} \boldsymbol{y}_0^* + \Pi_{L_h}(f_0) = 0$ in Ω_1, where Π_{L_h} is the $L^2(\Omega)$-projection operator that maps onto the space L_h. For example, this is the case when $\boldsymbol{y}_0^*$ is found in the lowest order Raviart-Thomas space RT_0 by the patchwise flux reconstruction in [30, 31] (see Section 4.2). Let us denote by $\overline{f}_0 \in L_h$ the projection $\Pi_{L_h}(f_0)$ and by $\overline{u}$ the solution of (4.2) with $\overline{f}_0$ instead of f_0. Then for all $v \in H_0^1(\Omega)$ we have

$$\int_\Omega \boldsymbol{A}\nabla u \cdot \nabla v dx + \int_\Omega k^2 \sinh(u+w) v dx = \int_\Omega (f_0 v + \boldsymbol{f} \cdot \nabla v)\, dx, \tag{4.77a}$$

$$\int_\Omega \boldsymbol{A}\nabla \overline{u} \cdot \nabla v dx + \int_\Omega k^2 \sinh(\overline{u}+w) v dx = \int_\Omega \left(\overline{f}_0 v + \boldsymbol{f} \cdot \nabla v\right) dx, \tag{4.77b}$$

where we recall that $u, \overline{u} \in L^\infty(\Omega)$ and hence the test space can be taken $H_0^1(\Omega)$. By subtracting the second equation in (4.77) from the first one and by taking $v := u - \overline{u}$ we obtain

$$|||\nabla(u-\overline{u})|||^2 + \int_\Omega k^2 \left(\sinh(u+w) - \sinh(\overline{u}+w)\right)(u-\overline{u}) dx = \int_\Omega \left(f_0 - \overline{f}_0\right)(u-\overline{u}) dx. \tag{4.78}$$

From (4.78), by using the monotonicity of sinh and the fact that $\int_K (f_0 - \overline{f}_0) dx = 0$, $\forall K \in \mathscr{T}_h$, for $c_K := \frac{1}{|K|}\int_K (u - \overline{u}) dx$, $\forall K \in \mathscr{T}_h$ we obtain in a standard way:

$$\begin{aligned} |||\nabla(u-\overline{u})|||^2 &\le \sum_{K\in\mathscr{T}_h} (f_0 - \overline{f}_0)(u - \overline{u} - c_K) dx \\ &\le \frac{1}{\mu_1} \sum_{K\in\mathscr{T}_h} C_P(K) \|f_0 - \overline{f}_0\|_{L^2(K)} |||\nabla(u-\overline{u})|||_K \\ &\le \frac{1}{\mu_1} \left(\sum_{K\in\mathscr{T}_h} \frac{h_K^2}{\pi^2} \|f_0 - \overline{f}_0\|^2_{L^2(K)} \right)^{\frac{1}{2}} |||\nabla(u-\overline{u})|||, \end{aligned} \tag{4.79}$$

where $|||\boldsymbol{q}|||_K^2 := \int_K \boldsymbol{A}\boldsymbol{q} \cdot \boldsymbol{q} dx$ and we have also applied Poincaré's inequality on each element K together with the fact that the constants $C_P(K)$ are bounded by $\frac{h_K}{\pi}$ (see [149]). Therefore, we see that

$$|||\nabla(u-\overline{u})||| \le \frac{1}{\mu_1} \left(\sum_{K\in\mathscr{T}_h} \frac{h_K^2}{\pi^2} \|f_0 - \overline{f}_0\|^2_{L^2(K)} \right)^{\frac{1}{2}}, \tag{4.80}$$

where the righ-hand side of (4.80) is fully computable.

Now, our goal is to estimate the distance between an approximation $\overline{u}_h$ and u, where $\overline{u}_h \in V_h^1$ is an approximation of $\overline{u}$. By using the error identity (4.51) applied to the problem (4.77b) defining $\overline{u}$, we find

$$\begin{aligned} |||\nabla(\overline{u}_h - u)||| &\leq |||\nabla(\overline{u}_h - \overline{u})||| + |||\nabla(\overline{u} - u)||| \\ &\leq \sqrt{2}\overline{M}_{\oplus}(\overline{u}_h, \boldsymbol{y}^*) + \frac{1}{\mu_1}\left(\sum_{K\in\mathscr{T}_h} \frac{h_K^2}{\pi^2}\|f_0 - \overline{f}_0\|^2_{L^2(K)}\right)^{\frac{1}{2}}, \end{aligned} \tag{4.81}$$

where $\overline{M}^2_{\oplus}(\overline{u}_h, \boldsymbol{y}^*)$ is the error majorant for problem (4.77b) $\boldsymbol{y}^* = \boldsymbol{f} + \boldsymbol{y}_0^*$ with $\boldsymbol{y}_0^* \in H(\mathrm{div};\Omega)$ and $\mathrm{div}\,\boldsymbol{y}_0^* + \overline{f}_0 = 0$ in Ω_1. In other words, if f_0 is not a piecewise constant function, we find an approximation $\overline{u}_h \in H_0^1(\Omega)$ of $\overline{u}$. Then $\overline{u}_h$ is also an approximation to u, for which the error estimate (4.81) holds. Moreover, if $f_0 \in H^1(\Omega)$, then by using again Poincaré's inequality on each element K, we see that the second term in the upper bound for $|||\nabla(\overline{u}_h - u)|||$ in (4.81) is $O(h^2)$. Here we note that the approximation $\overline{u}_h$ could be any element in $H_0^1(\Omega)$ and not necessarily in V_h^1. This is due to the fact that the majorant $\overline{M}^2_{\oplus}(\overline{u}_h, \boldsymbol{y}^*)$ is well defined for any conforming approximation of $\overline{u}$.

Numerical experiments for a homogeneous interface condition

In this section we present numerical examples which illustrate the error identity (4.51) and performance of functional a posteriori error estimates for the case $\boldsymbol{f} = 0$ and Ω_1 with a Lipschitz boundary Γ whose unit outward normal vector is $\boldsymbol{n}_\Gamma$. Numerical examples for the case $\boldsymbol{f} \neq 0$ will be given directly for the PBE on real biomolecular structures in Section 4.3. All numerical experiments are carried out in FreeFem++ developed and maintained by Frederich Hecht [98] and all pictures are generated in VisIt [52]. We solve adaptively the homogeneous nonlinear interface Problem (4.2) with $\boldsymbol{A} = \epsilon I$, $\epsilon : \Omega \to \mathbb{R}$, a smooth function in each subdomain Ω_1 and Ω_2, and $w := w_{h_{ref}} = \zeta - z_{h_{ref}}$ where $z_{h_{ref}}$ is a good Galerkin finite element approximation of the solution z of

$$-\nabla\cdot(\epsilon\nabla z) = -k^2\sinh(\zeta) + f_0 \quad \text{in } \Omega_1 \cup \Omega_2, \tag{4.82a}$$

$$[z]_\Gamma = 0, \tag{4.82b}$$

$$[\epsilon\nabla z \cdot \boldsymbol{n}_\Gamma]_\Gamma = 0, \tag{4.82c}$$

$$z = 0, \quad \text{on } \partial\Omega, \tag{4.82d}$$

for given functions ζ and f_0. We compare the accuracy of the adaptively computed solution u_h of (4.2) for $w = w_{h_{ref}}$ to the reference solution $z_{h_{ref}}$. The adaptive mesh refinement (AMR) is based on the error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ for subdomains O_i where the function η is defined in (4.41) and η^2 is the integrand of the majorant $M^2_{\oplus}(v, \boldsymbol{y}^*)$. The factor $\sqrt{2}$ accounts

for the factor 2 in (4.52). More precisely, we find approximations u_h to the exact solution $u \in H_0^1(\Omega)$ of the problem

$$\int_\Omega \epsilon \nabla u \cdot \nabla v dx + \int_\Omega b(x, u + w_{h_{ref}}) v dx = \int_\Omega f_0 v dx = 0, \, \forall v \in H_0^1(\Omega). \tag{4.83}$$

In all examples, we use piecewise constant parameters ϵ and k, and for $\boldsymbol{y}^* \in H(\mathrm{div}; \Omega)$ ($\boldsymbol{y}^* = \boldsymbol{y}_0^*$ when $\boldsymbol{f} = 0$), we use a patchwise equilibrated reconstruction of the numerical flux $\epsilon \nabla u_h$ based on [30, 31] (see Section 4.2). More precisely, we find $\boldsymbol{y}^*$ in the lowest order Raviart-Thomas space RT_0 over the same mesh, such that its divergence is equal to the L^2 orthogonal projection of $k^2 \sinh(u_h + w) + f_0$ onto the space L_h of piecewise constant functions.

Recall that

$$M_\oplus^2(v, \boldsymbol{y}^*) = M_\oplus^2(v, \boldsymbol{p}^*) + M_\oplus^2(u, \boldsymbol{y}^*),$$

where

$$M_\oplus^2(v, \boldsymbol{y}^*) = \frac{1}{2} |\!|\!| \epsilon \nabla v - \boldsymbol{y}^* |\!|\!|_*^2 + D_F(v, -\Lambda^* \boldsymbol{y}^*)$$

is fully computable and

$$M_\oplus^2(v, \boldsymbol{p}^*) = J(v) - J(u) = \frac{1}{2} |\!|\!| \nabla (v - u) |\!|\!|^2 + D_F(v, -\Lambda^* \boldsymbol{p}^*)$$

is the primal error, whereas

$$M_\oplus^2(u, \boldsymbol{y}^*) = I^*(\boldsymbol{p}^*) - I^*(\boldsymbol{y}^*) = \frac{1}{2} |\!|\!| \boldsymbol{y}^* - \boldsymbol{p}^* |\!|\!|_*^2 + D_F(u, -\Lambda^* \boldsymbol{y}^*)$$

is the dual error. Further, we use v for the approximate solution u_h and u for the reference solution $z_{h_{ref}}$ and define the efficiency index of the lower bound for the error in combined energy norm (4.54) by

$$I_{\mathrm{Eff}}^{\mathrm{CEN,Low}} := \frac{\frac{\sqrt{2}}{2} |\!|\!| \epsilon \nabla v - \boldsymbol{y}^* |\!|\!|_*}{\sqrt{|\!|\!| \nabla (v - u) |\!|\!|^2 + |\!|\!| \boldsymbol{y}^* - \boldsymbol{p}^* |\!|\!|_*^2}}.$$

Similarly,

$$I_{\mathrm{Eff}}^{\mathrm{CEN,Up}} := \frac{\sqrt{2 M_\oplus^2(v, \boldsymbol{y}^*)}}{\sqrt{|\!|\!| \nabla (v - u) |\!|\!|^2 + |\!|\!| \boldsymbol{y}^* - \boldsymbol{p}^* |\!|\!|_*^2}}$$

defines the efficiency index of the upper bound (4.52) for the error in combined energy norm. Finally,

$$I_{\mathrm{Eff}}^{\mathrm{E}} := \frac{\sqrt{2 M_\oplus^2(v, \boldsymbol{y}^*)}}{|\!|\!| \nabla (v - u) |\!|\!|} \quad \text{and} \quad P_{\mathrm{rel}}^{\mathrm{CEN}} := \frac{|\!|\!| \epsilon \nabla v - \boldsymbol{y}^* |\!|\!|_*}{\sqrt{|\!|\!| \nabla v |\!|\!|^2 + |\!|\!| \boldsymbol{y}^* |\!|\!|_*^2}}$$

define the efficiency index of the upper bound for the error in energy norm and the practical estimate of the relative error in combined energy norm, respectively.

Example 1 (2D problem)
In the first example, the domain Ω is a square with a side 20 with Ω_1 being a regular 15-sided polygon with a radius of its circumscribed circle equal to 2. The coefficients ϵ and k are defined by the relations

$$\epsilon(x) = \begin{cases} \epsilon_1 = 1, & x \in \Omega_1, \\ \epsilon_2 = 100, & x \in \Omega_2, \end{cases} \qquad k(x) = \begin{cases} k_1 = 0.15, & x \in \Omega_1, \\ k_2 = 0.4, & x \in \Omega_2, \end{cases}$$

and

$$\zeta = L\left(\exp\left(-b_1\left(\frac{(x_1 - c_1))^2}{\sigma_1^2} - 1\right)\right) - \exp\left(-b_2\left(\frac{(x_2 - c_2)^2}{\sigma_2^2} - 1\right)\right)\right),$$

$f_0 = 0$, where $b_1 = 2 = b_2 = 2$, $c_1 = -1$, $c_2 = 6$, $\sigma_1 = \sigma_2 = 1.5$, $L = 0.8$. The reference solution $z_{h_{ref}}$ is computed on a multiply refined mesh with 50 086 142 triangles. Note that $k^2 = 0.0225$ in Ω_1 and $k^2 = 0.16$ in Ω_2. The mesh adaptation is done with the built in function "adaptmesh" of FreeFem++. The localized error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$, computed on each vertex patch O_i of the mesh, is compared to its average value over all patches and the local mesh size is divided by two if this average is smaller then the local value.

Table 4.12 illustrates the main error identity (4.13) and the convergence of its constituent parts. Further, it is seen that the dual error $2M_{\oplus}^2(u, \boldsymbol{y}^*)$ dominates the primal error in this example. This is due to the fact that the term $2D_F(u, -\Lambda^*\boldsymbol{y}^*)$, measuring the error in $\operatorname{div} \boldsymbol{y}^*$ (cf. (4.48) and (4.50)), is much larger than $|||\nabla(v-u)|||^2 + D_F(v, -\Lambda^*\boldsymbol{p}^*)$, where $D_F(v, -\Lambda^*\boldsymbol{p}^*)$ behaves like $\|v-u\|^2_{L^2(\Omega_2)}$ (cf. (4.47) and (4.49)). As we mentioned earlier, for $\boldsymbol{y}^*$ we use a partially equilibrated reconstruction of the numerical flux $\epsilon\nabla v$ which is the reason why the integral term in (4.55) is negligible compared to the combined energy norm of the error. This fact is confirmed by the values of the efficiency index of the lower bound (4.54).

In Table 4.14 we can see that $I_{\text{Eff}}^{\text{CEN,Low}}$ is approximately equal to $0.7071 \approx \frac{\sqrt{2}}{2}$. The value of the efficiency index with respect to the combined energy norm and the value of the ratio $D_F(v, -\Lambda^*\boldsymbol{y}^*)/M_{\oplus}^2(v, \boldsymbol{y}^*)$ are also coupled in the sense that if we have only one of these two quantities, we can estimate the other one by using the main error equality (4.51). This estimation is accurate because the integral term in (4.57) is very close to zero and therefore $D_F(v, -\Lambda^*\boldsymbol{y}^*) \approx D_F(v - \Lambda^*\boldsymbol{p}^*) + D_F(u - \Lambda^*\boldsymbol{y}^*)$. One more consequence of using a partially equilibrated flux is that we obtain a very accurate practical estimate of the absolute and relative error in combined energy norm as illustrated in the last two columns of Table 4.14. Figure 4.4 depicts a mesh that is a part of a sequence of meshes obtained by mesh adaptation using the localized functional error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$. Figure 4.5 depicts a mesh with approximately the same number of elements but obtained by mesh adaptation using the error

Table 4.1: Example 1 (2D)
AMR with the indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with patchwise flux equilibration for $\boldsymbol{y}^*$.
Recall that $2M^2_\oplus(v,\boldsymbol{p}^*) + 2M^2_\oplus(u,\boldsymbol{y}^*) = 2M^2_\oplus(v,\boldsymbol{y}^*)$.

Example 1 (2D): $k_1 = 0.15$, $k_2 = 0.4$, $\epsilon_1 = 1$, $\epsilon_2 = 100$						
#elts	$\frac{\|v-u\|_{L^2(\Omega)}}{\|u\|_{L^2(\Omega)}}$[%]	$\frac{\|\|\nabla(v-u)\|\|}{\|\|\nabla u\|\|}$[%]	$\frac{\|\|\boldsymbol{y}^*-\boldsymbol{p}^*\|\|_*}{\|\|\boldsymbol{p}^*\|\|_*}$[%]	$2M^2_\oplus(v,\boldsymbol{y}^*)$	$2M^2_\oplus(v,\boldsymbol{p}^*)$	$2M^2_\oplus(u,\boldsymbol{y}^*)$
196	15.0077	51.5582	86.1021	1778.14	66.5980	1711.54
347	5.69339	30.8534	41.7241	703.594	20.7780	682.816
630	4.20384	21.7715	31.4858	217.719	10.2201	207.498
1 315	2.39552	15.8532	23.1244	76.8018	5.37574	71.4261
2 865	1.87075	11.7353	17.1655	33.9310	2.94414	30.9869
5 938	0.64611	7.93001	11.4692	16.0812	1.33874	14.7425
12 006	0.36985	5.64786	8.23544	7.75232	0.67872	7.07360
24 571	0.16023	3.94241	5.76054	3.85268	0.33039	3.52229
48 483	0.08909	2.80265	4.09366	1.90043	0.16682	1.73361
97 423	0.03961	1.97875	2.88455	0.96275	0.08304	0.87970
192 905	0.02230	1.39832	2.03200	0.47524	0.04136	0.43388
386 185	0.01015	0.99471	1.44616	0.24134	0.02082	0.22052

Table 4.2: Example 1 (2D)
AMR with the indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with patchwise flux equilibration for $\boldsymbol{y}^*$.
Recall that $\|\|\nabla(v-u)\|\|^2 + \|\|\boldsymbol{y}^* - \boldsymbol{p}^*\|\|^2_* + 2D_F(v,-\Lambda^*\boldsymbol{p}^*) + 2D_F(u,-\Lambda^*\boldsymbol{y}^*) = 2M^2_\oplus(v,\boldsymbol{y}^*)$.

Example 1 (2D): $k_1 = 0.15$, $k_2 = 0.4$, $\epsilon_1 = 1$, $\epsilon_2 = 100$				
#elts	$\|\|\nabla(v-u)\|\|^2$	$\|\|\boldsymbol{y}^* - \boldsymbol{p}^*\|\|^2_*$	$2D_F(v,-\Lambda^*\boldsymbol{p}^*)$	$2D_F(u,-\Lambda^*\boldsymbol{y}^*)$
196	56.5057	157.588	10.0923	1553.95
347	20.2350	37.0058	0.54296	645.811
630	10.0756	21.0729	0.14450	186.425
1 315	5.34235	11.3668	0.03338	60.0593
2 865	2.92742	6.26338	0.01671	24.7235
5 938	1.33673	2.79619	0.00200	11.9462
12 006	0.67805	1.44169	0.00067	5.63191
24 571	0.33038	0.70538	0.00001	2.81691
48 483	0.16696	0.35622	0.00000	1.37739
97 423	0.08323	0.17687	0.00000	0.70283
192 905	0.04156	0.08777	0.00000	0.34611
386 185	0.02103	0.04445	0.00000	0.17606

Table 4.3: Example 1 (2D)
AMR with the indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with patchwise flux equilibration for $\boldsymbol{y}^*$.

Example 1 (2D): $k_1 = 0.15$, $k_2 = 0.4$, $\epsilon_1 = 1$, $\epsilon_2 = 100$						
#elts	$\frac{D_F(v,-\Lambda^*\boldsymbol{y}^*)}{M^2_\oplus(v,\boldsymbol{y}^*)}$ [%]	$I^{\text{CEN,Low}}_{\text{Eff}}$	$I^{\text{CEN,Up}}_{\text{Eff}}$	$I^{\text{E,Up}}_{\text{Eff}}$	$P^{\text{CEN}}_{\text{rel}}$ [%]	True rel. error in CEN [%]
196	89.0701	0.67371	2.88191	5.60966	74.6973	70.9641
347	92.4942	0.67919	3.50597	5.89671	36.2638	36.6935
630	85.9525	0.70066	2.64380	4.64848	27.1574	27.0680
1 315	78.2616	0.70681	2.14392	3.79158	19.9383	19.8250
2 865	72.8992	0.70729	1.92142	3.40452	14.7523	14.7032
5 938	74.3009	0.70708	1.97256	3.46846	9.87419	9.85973
12 006	72.6473	0.70722	1.91238	3.38130	7.06762	7.06119
24 571	73.1176	0.70708	1.92864	3.41485	4.93753	4.93591
48 483	72.4826	0.70694	1.90588	3.37371	3.50789	3.50805
97 423	73.0084	0.70678	1.92392	3.40108	2.47256	2.47347
192 905	72.8486	0.70629	1.91692	3.38145	1.74226	1.74418
386 185	72.9912	0.70546	1.91972	3.38748	1.23829	1.24114

indicator

$$|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)} = \left(\int_{O_i} \frac{1}{\epsilon} \left|\epsilon\nabla v - \boldsymbol{y}^*\right|^2 dx \right)^{\frac{1}{2}}.$$

The mesh in Figure 4.4 is refined mainly where the error in $\operatorname{div} \boldsymbol{y}^*$ is the dominant part of the error $M^2_\oplus(v, -\Lambda^*\boldsymbol{p}^*) + M^2_\oplus(u, -\Lambda^*\boldsymbol{y}^*)$. On the other hand, the mesh in Figure 4.5 is refined most around the extrema of the solution. Figure 4.7 depicts the minimal set of elements K of a mesh $\mathscr{T}_h$ that contains at least 30% of the total indicated error $\sum_{K\in\mathscr{T}_h} |||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(K)}$ (greedy algorithm with a bulk factor of 0.3), where $\mathscr{T}_h$ is part of the same sequence as the mesh illustrated in Figure 4.5.

Figure 4.9 depicts the elements marked by the greedy algorithm using a bulk factor of 0.5 and employing the true error $\sqrt{2M^2_\oplus(v,\boldsymbol{p}^*) + 2M^2_\oplus(u,\boldsymbol{y}^*)}$ as indicator. Figure 4.8 depicts elements which are marked additionally or failed to be marked by the same greedy algorithm when employing the functional error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ for the same bulk factor. The ratio of the number of these differently marked elements, that is, elements which are marked by one of the two methods but not by the other one, and the total number of elements is 0.022 and the ratio of the number of differently marked elements to the number of marked elements using the true error is 0.048 (see Table 4.4). Comparing the indicated error and the true error elementwise, one finds that the error indicator generated by the majorant $M^2_\oplus(v,\boldsymbol{y}^*)$ reproduces the local distribution of the error with a very high accuracy. This is also confirmed by Figure 4.3 where it can be seen that all error measures are almost

identical in both cases of adaptive mesh refinement. Mesh adaptation based on the functional error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ instead of the error indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)}$ (see Figure 4.2) yields approximately twice smaller efficiency indexes in energy and combined energy norms and approximately twice smaller values for the full error $M^2_\oplus(v, \boldsymbol{p}^*) + M^2_\oplus(u, \boldsymbol{y}^*)$ on meshes with a comparable number of elements. The reason for the higher efficiency indexes is that no adaptive control is applied on the nonlinear part of the error measure in (4.51), and consequently, the ratio $D_F(v, -\Lambda^*\boldsymbol{y}^*)/M^2_\oplus(v, \boldsymbol{y}^*)$ is increasing, reaching values close to 100% on fine meshes. However, the error in $|||\nabla(v-u)|||$ and $|||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*$ might be a little higher in the case of the functional error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$. For example, on the mesh from Figure 4.7 with 24 122 elements, $M^2_\oplus(v, \boldsymbol{p}^*) + M^2_\oplus(u, \boldsymbol{y}^*) = 3.8314$, $|||\nabla(v-u)||| = 0.4674$, $|||\boldsymbol{y}^* - \boldsymbol{p}^*|||_* = 0.6540$, whereas on a mesh with 24 571 elements from the sequence adapted with the indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$, we obtained a value of 1.9263 for $M^2_\oplus(v, \boldsymbol{y}^*)$, and 0.574791 and 0.8399 for $|||\nabla(v-u)|||$ and $|||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*$, respectively. This shows that by reducing the error in $\operatorname{div}\boldsymbol{y}^*$ the functional error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ provides a better approximation for the primal and dual problem together.

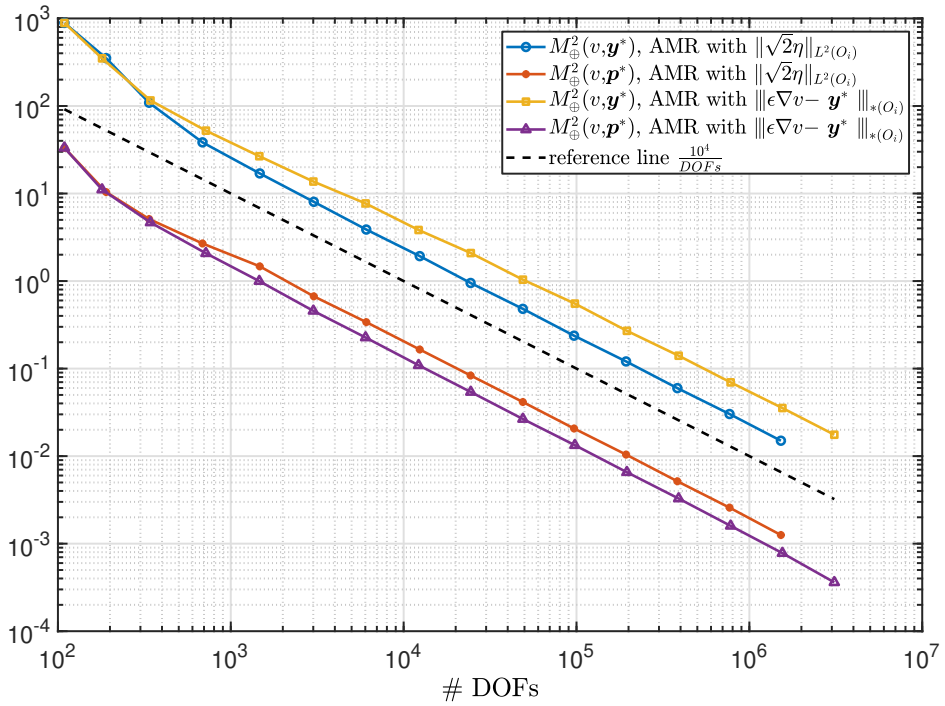


Figure 4.2: Comparison of errors for AMR based on the functional error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ versus AMR based on the indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)}$ (cf. Remark 4.11).

Now, we want to demonstrate that flux equilibration is indeed an important subtask to make the proposed error bounds reliable and efficient. For this purpose, we use a simple global gradient averaging procedure, i.e. project the numerical flux $\epsilon\nabla v \in L^2(\Omega)$ onto the subspace $[V_h]^2$, where V_h is the finite element space of continuous piecewise linear functions. Then, the

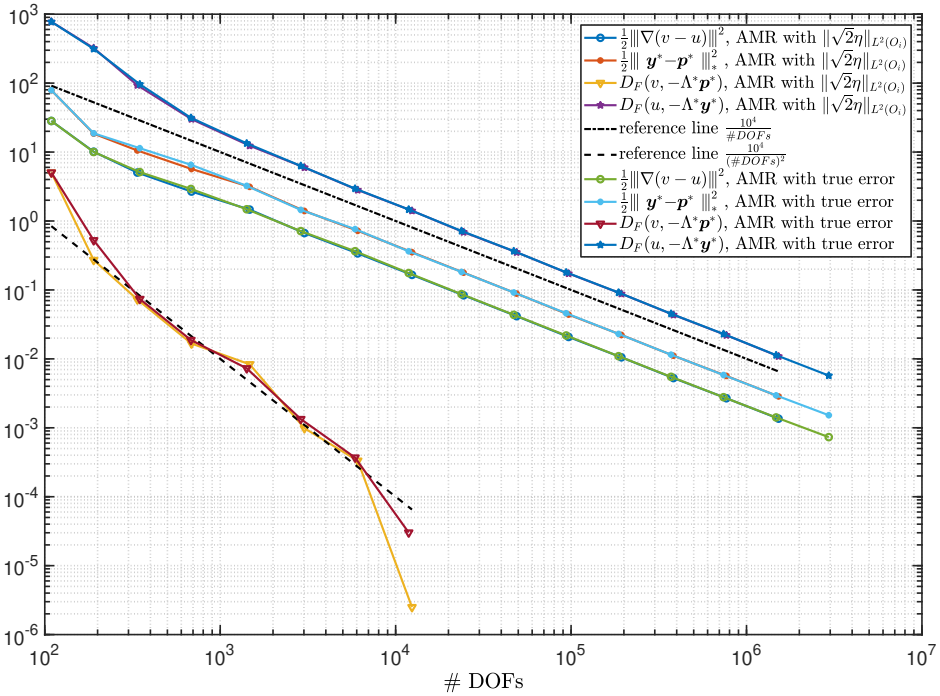


Figure 4.3: Comparison of errors for AMR based on the functional error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ versus AMR based on the indicator generated by the true error $\sqrt{2M_{\oplus}^2(v,\boldsymbol{p}^*)+2M_{\oplus}^2(u,\boldsymbol{y}^*)}$ (cf. Remark 4.11).

Table 4.4: Example 1 (2D)

Example 1 (2D): $k_1 = 0.15$, $k_2 = 0.4$, $\epsilon_1 = 1$, $\epsilon_2 = 100$			
#elts	#marked elts with true error	#differently marked elts	differently marked elts in % of all mesh elts
196	62	6	3.06122
347	150	10	2.88184
630	288	14	2.22222
1 315	632	39	2.96578
2 865	1439	113	3.94415
5 938	2949	216	3.63759
12 006	5981	534	4.44778
24 571	12099	961	3.91111
48 483	24194	2233	4.60574
97 423	47784	4012	4.11812

problem in Example 1 is solved adaptively once by applying the functional error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ and next by applying the error indicator $|||\epsilon\nabla v-\boldsymbol{y}^*|||_{*(O_i)}$. Figure 4.10 shows the adapted mesh with 563 965 elements, which is a part of a sequence of meshes obtained by

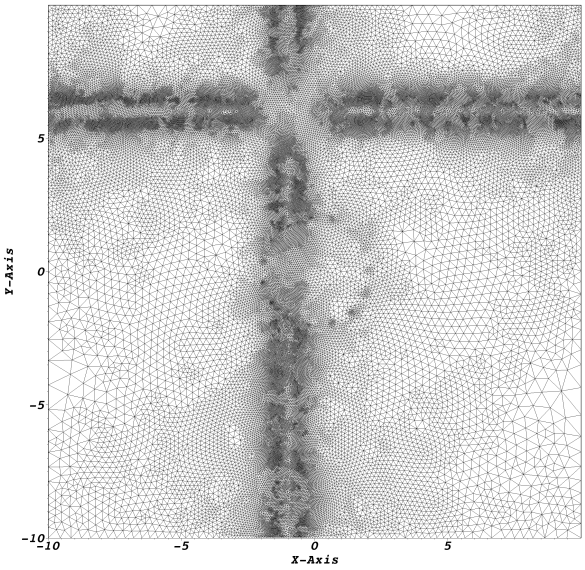


Figure 4.4: Mesh on the 9-th level of AMR (97 423 elements) based on the error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with flux equilibration for $\boldsymbol{y}^*$.

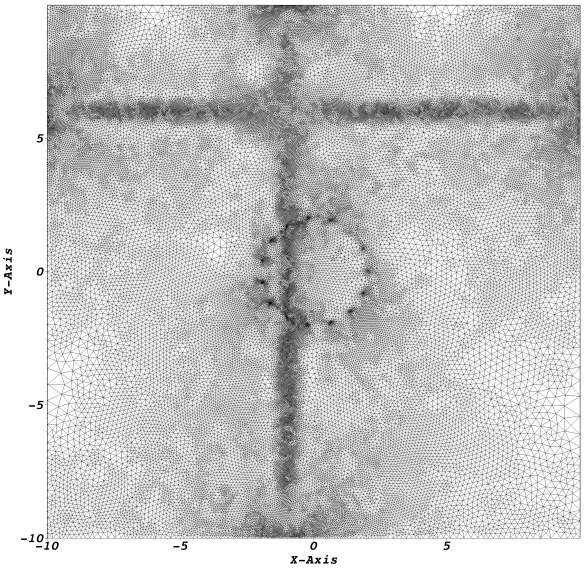


Figure 4.5: Mesh on the 9-th level of AMR (97 353 elements) based on the error indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)}$ with flux equilibration for $\boldsymbol{y}^*$.

applying the functional error indicator with gradient averaging for $\boldsymbol{y}^*$. Figure 4.11 shows a mesh with 444 092 elements, which is part of a sequence of meshes adapted using the second indicator with gradient averaging for $\boldsymbol{y}^*$. By comparing with the results based on flux equilibration for $\boldsymbol{y}^*$ it can be seen that the mesh in Ω_2 close to the interface Γ is refined too much for both error indicators. Apart from that, the meshes on Figures 4.11 and 4.5 look quite similar, unlike the meshes on Figures 4.10 and 4.4. For meshes with a comparable number of elements, by applying the indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)}$ using gradient averaging instead of flux equilibration we obtained $\sim 30\%$ larger values for the error $|||\nabla(v-u)|||$ and 60% larger values for the error $|||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*$. The difference in the errors when applying the functional indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with gradient averaging for $\boldsymbol{y}^*$ instead of flux equilibration resulted in an even more drastic increase of the error, namely between 40% and 180% for $|||\nabla(v-u)|||$ and between 64% and 66% for $|||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*$, where the meshes had between 21 528 and 563 965 elements. In both cases we obtained an increasing sequence of efficiency indexes with respect to energy and combined energy norms reaching values of 133 and 107 with the functional error indicator on a mesh with 2 089 022 elements, and 570 and 269 with the error indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)}$ on a mesh with 2 954 218 elements (see Table 4.5). This is due to the fact that the nonlinear term $D_F(u, -\Lambda^*\boldsymbol{y}^*)$, which measures the error in $\operatorname{div}\boldsymbol{y}^*$ (see (4.48) and (4.50)), dominates the other terms in the nonlinear measure $M^2_\oplus(v, \boldsymbol{p}^*) + M^2_\oplus(u, \boldsymbol{y}^*)$ for the error, reaching more than 99.99% of it in both cases. In both experiments with gradient averaging for $\boldsymbol{y}^*$, increasing values of $D_F(u, -\Lambda^*\boldsymbol{y}^*)$ are in correspondence with increasing error $\|\operatorname{div}\boldsymbol{y}^* - \operatorname{div}\boldsymbol{p}^*\|_{L^2(\Omega)}$ and increasing efficiency indexes.

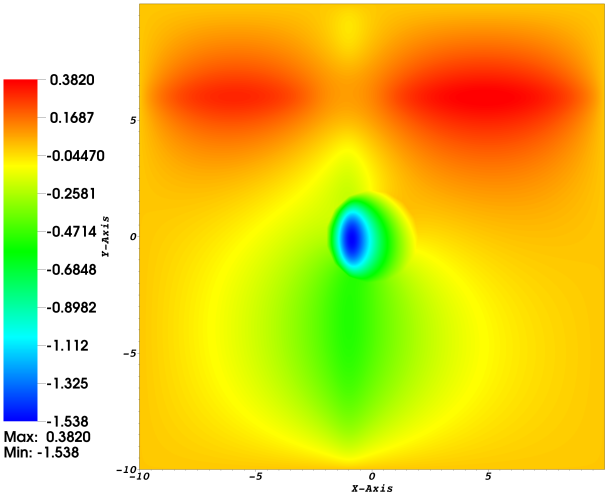


Figure 4.6: Reference solution for Example 1 (2D).

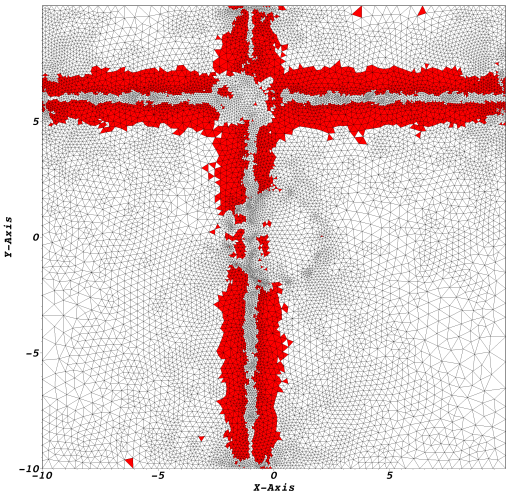


Figure 4.7: Mesh on the 7th level of AMR (24 122 elements) based on the error indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)}$ with flux equilibration for $\boldsymbol{y}^*$. The elements are marked by applying the error indicator $\|\sqrt{2}\eta\|_{L^2(K)}$ and using the greedy algorithm with bulk factor 0.3.

Table 4.5: Example 1 (2D)
AMR with the indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*O_i}$ with simple gradient averaging for $\boldsymbol{y}^*$.
Recall that $|||\nabla(v-u)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2 + 2D_F(v, -\Lambda^*\boldsymbol{p}^*) + 2D_F(u, -\Lambda^*\boldsymbol{y}^*) = 2M_\oplus^2(v, \boldsymbol{y}^*)$.

Example 1 (2D): $k_1 = 0.15$, $k_2 = 0.4$, $\epsilon_1 = 1$, $\epsilon_2 = 100$						
#elts	$\|\|\|\nabla(v-u)\|\|\|^2$	$\|\|\|\boldsymbol{y}^* - \boldsymbol{p}^*\|\|\|_*^2$	$2D_F(v, -\Lambda^*\boldsymbol{p}^*)$	$2D_F(u, -\Lambda^*\boldsymbol{y}^*)$	$2M_\oplus^2(v, \boldsymbol{y}^*)$	$I_{\text{Eff}}^{\text{E,Up}}$
196	56.5057	119.481	10.0923	3271.95	3458.03	7.8229
400	27.6244	64.5618	0.94063	2210.53	2303.66	9.1319
739	13.3498	32.5272	0.26263	1533.74	1579.88	10.8786
1 399	6.90615	15.8220	0.05236	1186.27	1209.05	13.2314
2 723	3.27272	9.22987	0.01196	916.267	928.782	16.8462
4 903	1.94379	4.89744	0.00388	804.389	811.234	20.4291
9 463	0.98523	2.63409	0.00087	815.815	819.435	28.8395
17 907	0.52683	1.47142	0.00012	775.161	777.159	38.4076
34 040	0.27330	0.79663	0.00002	767.024	768.094	53.0131

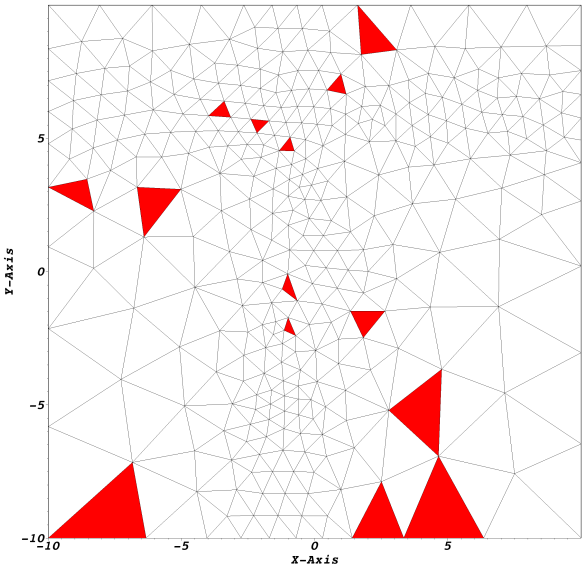


Figure 4.8: Mesh on the 2nd level of AMR (630 elements) based on the error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with flux equilibration for $\boldsymbol{y}^*$. Here we mark red those elements, which differ in the markings based on the indicator $\|\sqrt{2}\eta\|_{L^2(K)}$ and on the true error $\sqrt{2M^2_{\oplus}(v,\boldsymbol{p}^*)+2M^2_{\oplus}(u,\boldsymbol{y}^*)}$. Marking is done by greedy algorithm with bulk factor 0.5.

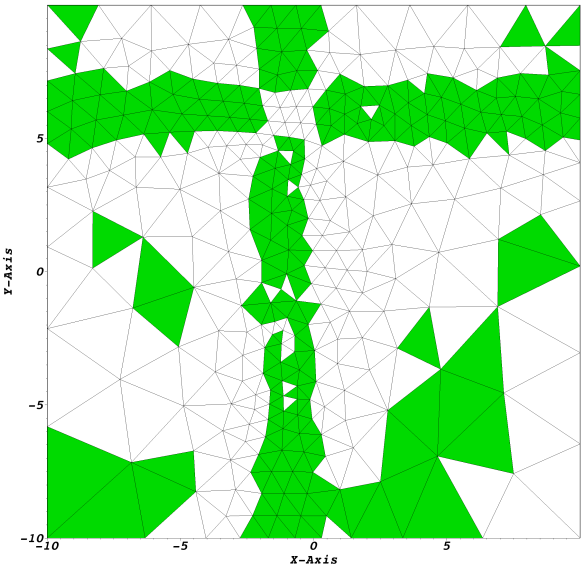


Figure 4.9: Mesh on the 2nd level of AMR (630 elements) based on the error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with flux equilibration for $\boldsymbol{y}^*$. The elements are marked by applying the true error $\sqrt{2M^2_{\oplus}(v,\boldsymbol{p}^*)+2M^2_{\oplus}(u,\boldsymbol{y}^*)}$ as an indicator using greedy algorithm with bulk factor 0.5.

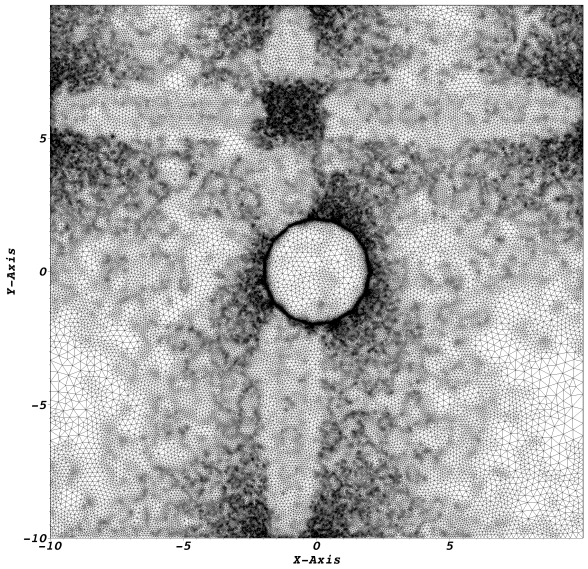


Figure 4.10: Mesh with 563 965 elements, adapted using the error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with gradient averaging for $\boldsymbol{y}^*$.

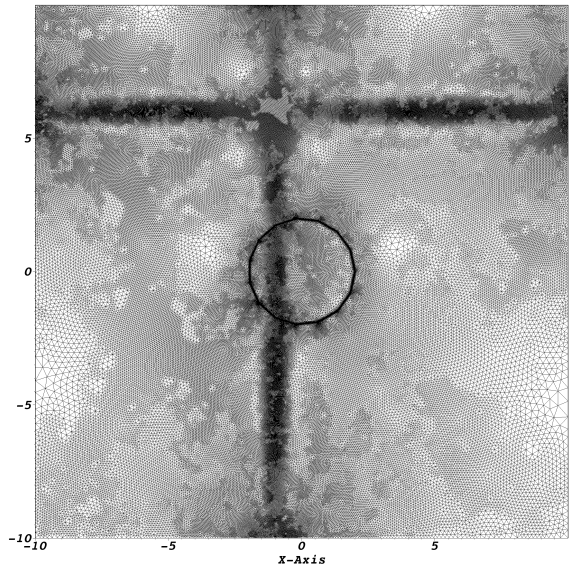


Figure 4.11: Mesh with 444 092 elements, adapted using the error indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)}$ with gradient averaging for $\boldsymbol{y}^*$.

Example 2 (2D problem)
Figures 4.13 and 4.15 show how meshes depend on the indicator in another example, where $\epsilon_1 = 1\,\epsilon_2 = 100$, $k_1 = 0.2$, $k_2 = 0.3$. The function $\zeta = \exp\left(-b_1\left(\frac{|x-c_1|^2}{\sigma_1^2} - 1\right)\right) - \exp\left(-b_2\left(\frac{|x-c_2|^2}{\sigma_2^2} - 1\right)\right)$ and $f_0 = \exp\left(-b_3\left(\frac{|x|^2}{\sigma_3^2} - 1\right)\right)\sin\left(\frac{x_1 x_2}{4}\right)$, where $b_1 = 2.2$, $b_2 = 2.5$, $b_3 = 6$, $c_1 = (-1, 0)$, $c_2 = (5, 5)$, $\sigma_1 = \sigma_2 = 2$, $\sigma_3 = 10$. The indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)}$ correctly approximates elementwise errors in the combined energy norm but does not capture the rest of the error, which results from the nonlinearity $k^2\sinh(u + w)$ and the right-hand side f_0 in (4.83). On the other hand, the term $D_F(v, -\Lambda^*\boldsymbol{y}^*)$ controls the error $D_F(v, -\Lambda^*\boldsymbol{p}^*) + D_F(u, -\Lambda^*\boldsymbol{y}^*)$ and this is the reason why the mesh on Figure 4.13 resembles the wavy features of the function $f := -k^2\sinh(u + w) + f_0$. The isolines of the reference solution and of the function f are depicted on Figures 4.14 and 4.12.

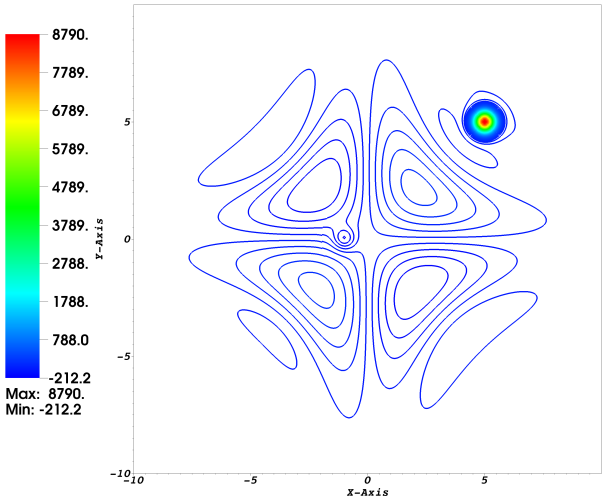


Figure 4.12: Function $f = -k^2 \sinh(u+w)+ f_0$.

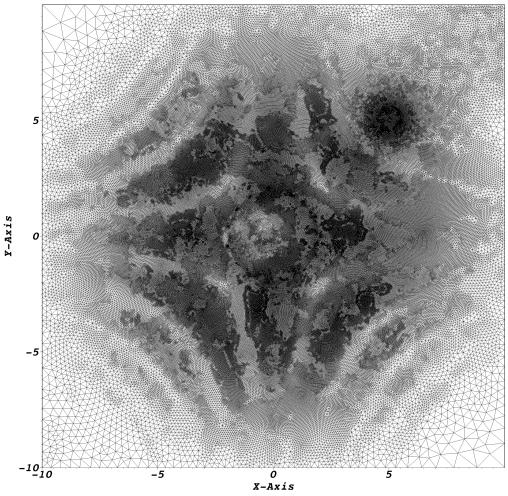


Figure 4.13: Mesh with 395 935 elements, obtained by AMR using the error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with flux equilibration for $\boldsymbol{y}^*$.

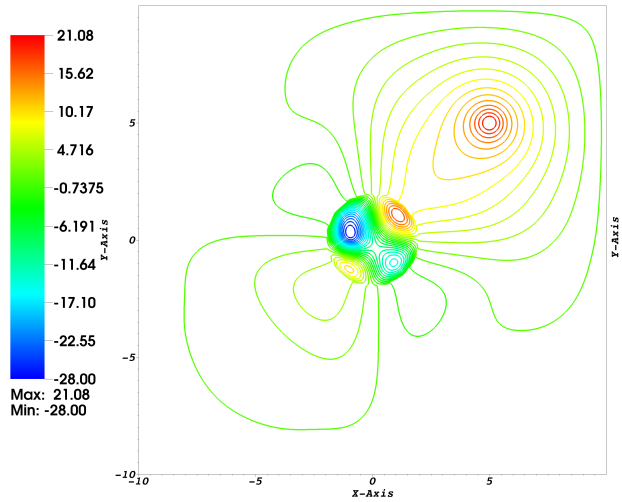


Figure 4.14: Reference solution.

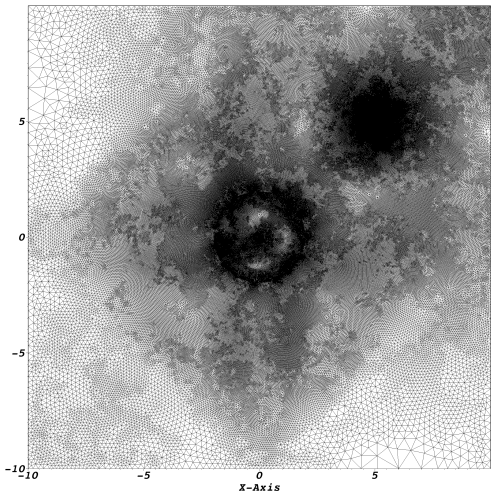


Figure 4.15: Mesh with 555 489 elements, obtained by AMR using the error indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)}$ with flux equilibration for $\boldsymbol{y}^*$.

Example 3 (3D problem)

Here we consider an example close to a real physical problem. The computational domain Ω is a cube of side length 20 Angstroms with a triangulated water molecule Ω_1 in it. The diameter of the water molecule, which is positioned in the center of the cube, is about 2.75 Angstroms. Its shape is not changed during the mesh adaptation process. The surface mesh

of the water molecule is taken from [1]. Figure 4.16 illustrates the initial tetrahedral mesh, which consists of 60 222 elements. It is generated using TetGen [170] and adaptively refined with the help of mmg3d [62]. Using the localized error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ computed on each vertex patch O_i of the mesh, a new local mesh size at each vertex is defined by the formula

$$h_i^{\text{new}} = h_i^{\text{old}} \left(\max\left\{ \min\left\{ \frac{\text{AM}\left\{ \|\sqrt{2}\eta\|_{L^2(O_j)} \right\}}{\|\sqrt{2}\eta\|_{L^2(O_i)}}, 1 \right\}, 0.35 \right\} \right) \tag{4.84}$$

and supplied to mmg3d, where $\text{AM}\left\{ \|\sqrt{2}\eta\|_{L^2(O_j)} \right\}$ is the arithmetic mean of $\left\{ \|\sqrt{2}\eta\|_{L^2(O_j)} \right\}$ over all vertex patches O_j. The coefficients ϵ and k for this example are typical for electrostatic computations in biophysics using the PBE and are given by

$$\epsilon(x) = \begin{cases} \epsilon_1 = 2, & x \in \Omega_1, \\ \epsilon_2 = 80, & x \in \Omega_2, \end{cases} \qquad k(x) = \begin{cases} k_1 = 0, & x \in \Omega_1, \\ k_2 = 0.84, & x \in \Omega_2. \end{cases}$$

We consider the homogeneous problem, i.e., $f_0 = 0$, and

$$\begin{aligned} \zeta = &\exp\left(-b_1 \left(\frac{|x - c_1|^2}{\sigma_1^2} - 1 \right) \right) - \exp\left(-b_2 \left(\frac{|x - c_2|^2}{\sigma_2^2} - 1 \right) \right) \\ &+ \exp\left(-b_3 \left(\frac{|x - c_3|^2}{\sigma_3^2} - 1 \right) \right) + \exp\left(-b_4 \left(\frac{|x - c_4|^2}{\sigma_4^2} - 1 \right) \right), \end{aligned}$$

where $b_1 = b_2 = b_3 = b_4 = 2.3$, $c_1 = (1, 1, 0)$, $c_2 = (4, 4, 0)$, $c_2 = (0, 6, 0)$, $c_2 = (-5, 0, 0)$, $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 2$. The reference solution $z_{h_{ref}}$ is computed on a very fine mesh, obtained after a sequence of adaptive mesh refinements, that contains 79 917 007 tetrahedrons.

Since $f_0 = 0$ in Ω_1 is a constant function, the patchwise reconstruction from [30] produces a flux $\boldsymbol{y}^*$ with zero divergence in Ω_1 and, therefore, the reliability of our majorant is guaranteed. In this example, we achieved a very tight guaranteed bound on the error in combined energy norm, as well as in energy norm. The efficiency index $I_{\text{Eff}}^{\text{CEN,Up}}$ settles at around 1.05 and the efficiency index $I_{\text{Eff}}^{\text{E,Up}}$ decreases to 1.30 (see Table 4.26). This is in a good agreement with the fact that in this example the ratio $D_F(v, -\Lambda^*\boldsymbol{y}^*)/M_\oplus^2(v, \boldsymbol{y}^*)$ is well controlled and decreases to around 10% (see column 2 in Table 4.27). We also note that in this example we obtained very similar results with the error indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_{*(O_i)}$. For the efficiency index $I_{\text{Eff}}^{\text{CEN,Low}}$ of the lower bound on the combined energy norm of the error we obtain values converging to approximately 0.7071 which is the approximate value of $\frac{\sqrt{2}}{2}$ (see column 3 in Table 4.26). This means that the combined energy norm of the error $\sqrt{|||\nabla(v-u)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2}$ is practically equal to $|||\epsilon\nabla v - \boldsymbol{y}^*|||_*$. Another consequence of this fact is the good accuracy of the practical estimation $P_{\text{rel}}^{\text{CEN}}$ of the relative error in combined energy norm (see columns 6 and 7 in Table 4.26). The tight bounds on the error also enable us to compute tight and

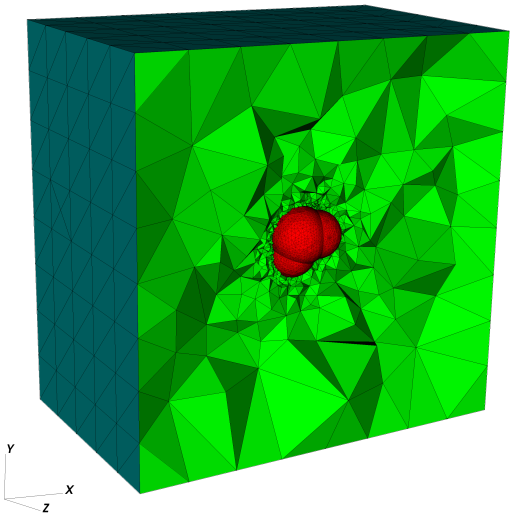


Figure 4.16: Initial mesh in Example 3 consisting of 60 222 tetrahedrons.

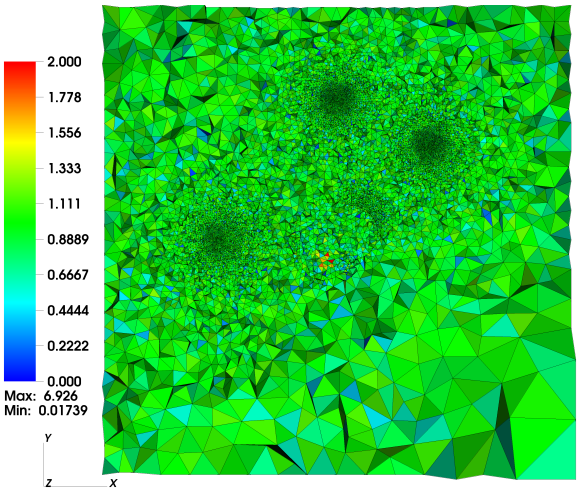


Figure 4.17: Ratio of the error indicator $|||\epsilon\nabla v - \boldsymbol{y}^*|||_*$ and combined energy norm of the error, elementwise. Mesh on the 4th level of AMR ($1.1736e+06$ elements) in Example 3 using the error indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with flux equilibration for $\boldsymbol{y}^*$.

guaranteed upper bounds on the relative error in energy norm.

$$\frac{|||\nabla(v-u)|||}{|||\nabla u|||} \leq \frac{\sqrt{2M^2_\oplus(v,\boldsymbol{y}^*)}}{|||\nabla v||| - \sqrt{2M^2_\oplus(v,\boldsymbol{y}^*)}} =: \mathrm{REN}^{\mathrm{Up}} \tag{4.85}$$

where (4.85) is valid if $|||\nabla v||| - \sqrt{2M^2_\oplus(v,\boldsymbol{y}^*)} > 0$.

Table 4.6: Example 3 (3D)
AMR with the indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with patchwise flux equilibration for $\boldsymbol{y}^*$.
Recall that $2M^2_\oplus(v,\boldsymbol{p}^*) + 2M^2_\oplus(u,\boldsymbol{y}^*) = 2M^2_\oplus(v,\boldsymbol{y}^*)$.

Example 3 (3D): $k_1 = 0$, $k_2 = 0.84$, $\epsilon_1 = 2$, $\epsilon_2 = 80$						
#elts	$\frac{\|v-u\|_{L^2(\Omega)}}{\|u\|_{L^2(\Omega)}}$ [%]	$\frac{\|\|\nabla(v-u)\|\|}{\|\|\nabla u\|\|}$ [%]	$\frac{\|\|\boldsymbol{y}^*-\boldsymbol{p}^*\|\|_*}{\|\|\boldsymbol{p}^*\|\|_*}$ [%]	$2M^2_\oplus(v,\boldsymbol{y}^*)$	$2M^2_\oplus(v,\boldsymbol{p}^*)$	$2M^2_\oplus(u,\boldsymbol{y}^*)$
60 222	76.8320	108.015	167.589	425569	117373	308196
103 236	11.9257	46.3306	55.1210	47104.5	17845.0	29259.5
222 118	1.09233	17.7353	14.9578	4484.44	2224,69	2259.75
552 936	0.49820	8.67222	7.09062	965.067	513.706	451.361
1 173 598	0.25609	6.58075	5.33661	539.734	295.254	244.481
2 056 678	0.17094	5.37625	4.18207	350.648	197.016	153.631
2 973 146	0.12317	4.73466	3.53852	265.167	152.783	112.385
3 906 919	0.10071	4.32886	3.12966	216.336	127.703	88.6336

Table 4.7: Example 3 (3D)
AMR with the indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with patchwise flux equilibration for $\boldsymbol{y}^*$.
Recall that $|||\nabla(v-u)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2 + 2D_F(v, -\Lambda^*\boldsymbol{p}^*) + 2D_F(u, -\Lambda^*\boldsymbol{y}^*) = 2M_\oplus^2(v, \boldsymbol{y}^*)$.

Example 3 (3D): $k_1 = 0$, $k_2 = 0.84$, $\epsilon_1 = 2$, $\epsilon_2 = 80$					
#elts	$\|\|\|\nabla(v-u)\|\|\|^2$	$\|\|\|\boldsymbol{y}^* - \boldsymbol{p}^*\|\|\|_*^2$	$2D_F(v, -\Lambda^*\boldsymbol{p}^*)$	$2D_F(u, -\Lambda^*\boldsymbol{y}^*)$	$\mathrm{REN}^{\mathrm{Up}}$[%]
60 222	79487.0	191346	37886.0	116850	-
103 236	14623.9	20699.7	3221.12	8559.78	310.049
222 118	2142.92	1524.28	81.7757	735.474	33.9219
552 936	512.376	342.528	1.32980	108.833	13.4714
1 173 598	295.039	194.026	0.21458	50.455	9.75647
2 056 678	196.919	119.155	0.09743	34.4762	7.72193
2 973 146	152.724	85.3044	0.05857	27.0805	6.64970
3 906 919	127.666	66.7303	0.03663	21.9033	5.96873

As a remark, we note that the efficiency indexes with respect to the energy and combined energy norms of the error can be improved if we use a flux reconstruction in a bigger space, say RT_1, which has better approximation properties. In this way the error in $\operatorname{div}\boldsymbol{y}^*$ will decrease and as a result, the term $D_F(v, -\Lambda^*\boldsymbol{y}^*)$ and consequently the dual part of the error $2M_\oplus^2(u, \boldsymbol{y}^*) = |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2 + D_F(u, -\Lambda^*\boldsymbol{y}^*)$ will constitute a smaller part of the whole majorant and the error, respectively. Even better, we can minimize the majorant with respect to $\boldsymbol{y}^*$ in a subspace of $H(\operatorname{div};\Omega)$ like RT_0, possibly on another finer mesh. Note that in the limit case we have

$$\inf_{\boldsymbol{y}^*\in H(\operatorname{div};\Omega)} M_\oplus^2(v, \boldsymbol{y}^*) = M_\oplus^2(v, \boldsymbol{p}^*) = \frac{1}{2}|||\nabla(v-u)|||^2 + D_F(v, -\Lambda^*\boldsymbol{p}^*)$$

and the dual error completely vanishes. In this case,

$$I_{\mathrm{Eff}}^{\mathrm{CEN,Up}} = I_{\mathrm{Eff}}^{\mathrm{E}} = \frac{\sqrt{2M_\oplus^2(v, \boldsymbol{p}^*)}}{|||\nabla(v-u)|||} = \frac{\sqrt{|||\nabla(v-u)|||^2 + 2D_F(v, -\Lambda^*\boldsymbol{p}^*)}}{|||\nabla(v-u)|||},$$

where the last ratio tends to 1 because by (4.49) the term $D_F(v, -\Lambda^*\boldsymbol{p}^*) \sim \|v-u\|_{L^2(\Omega)}^2$ and has a higher order of convergence than $|||\nabla(v-u)|||^2$. In practice, we can minimize the majorant with respect to $\boldsymbol{y}^*$ only once on a sufficiently big subspace of $H(\operatorname{div};\Omega)$ to find some good approximation $\overline{\boldsymbol{y}}^*$ of $\boldsymbol{p}^*$ and then reuse this $\overline{\boldsymbol{y}}^*$ and obtain guaranteed and tight bounds on the error in energy and combined energy norm.

A key factor that determines the efficiency index is the ratio $\frac{D_F(v, -\Lambda^*\boldsymbol{y}^*)}{M_\oplus^2(v, \boldsymbol{y}^*)}$. Assuming that

$$D_F(v, -\Lambda^*\boldsymbol{y}^*) \approx D_F(v, -\Lambda^*\boldsymbol{p}^*) + D_F(u, -\Lambda^*\boldsymbol{y}^*),$$

Table 4.8: Example 3 (3D)
AMR with the indicator $\|\sqrt{2}\eta\|_{L^2(O_i)}$ with patchwise flux equilibration for $\boldsymbol{y}^*$.

Example 3 (3D): $k_1 = 0$, $k_2 = 0.84$, $\epsilon_1 = 2$, $\epsilon_2 = 80$						
#elts	$\frac{D_F(v,-\Lambda^*\boldsymbol{y}^*)}{M^2_\oplus(v,\boldsymbol{y}^*)}$[%]	$I^{\mathrm{CEN,Low}}_{\mathrm{Eff}}$	$I^{\mathrm{CEN,Up}}_{\mathrm{Eff}}$	$I^{\mathrm{E,Up}}_{\mathrm{Eff}}$	$P^{\mathrm{CEN}}_{\mathrm{rel}}$ [%]	True rel. error in CEN [%]
60 222	40.0541	0.68627	1.25353	2.31386	92.8434	140.985
103 236	20.4500	0.72828	1.15478	1.79473	47.6870	50.9159
222 118	16.1172	0.71615	1.10583	1.44661	16.4040	16.4054
552 936	11.2249	0.70786	1.06248	1.37241	7.90966	7.92099
1 173 598	9.33477	0.70731	1.05053	1.35254	5.98505	5.99106
2 056 678	9.82289	0.70725	1.05327	1.33442	4.81343	4.81632
2 973 146	10.2057	0.70722	1.05547	1.31767	4.17784	4.17960
3 906 919	10.1194	0.70719	1.05492	1.30175	3.77592	3.77716

which means that the last term in (4.57) is close to zero, we obtain from (4.51) the estimate

$$I^{\mathrm{CEN,Up}}_{\mathrm{Eff}} \approx \frac{1}{\sqrt{1 - \frac{D_F(v,-\Lambda^*\boldsymbol{y}^*)}{M^2_\oplus(v,\boldsymbol{y}^*)}}}.$$

From what we have observed, the efficiency index $I^{\mathrm{E,Up}}_{\mathrm{Eff}}$ with respect to the energy norm usually is no more than twice bigger than $I^{\mathrm{CEN,Up}}_{\mathrm{Eff}}$ (assuming we have a good approximation $\boldsymbol{y}^*$ to $\boldsymbol{p}^*$). Therefore, if during the computations we detect that this ratio is increasing we can apply the so-called estimation with one step delay, i.e compute the value of the majorant $M^2_\oplus(v,\boldsymbol{y}^*)$ for the current mesh level with the reconstructed $\boldsymbol{y}^*$ from the next level.

Recomputing the majorant with the last reconstructed y^*

To illustrate these ideas, for the first example we recomputed the value of the majorant $M^2_\oplus(v,\boldsymbol{y}^*)$ on all mesh levels (sequence of meshes is the same one from Tables 4.12, 4.13, 4.14) using the flux $\overline{\boldsymbol{y}}^*$ that we obtained through the patchwise reconstruction with equilibration at the last level 11, where the mesh consists of 386 185 elements. This $\overline{\boldsymbol{y}}^*$ gives a very good approximation to the exact flux $\boldsymbol{p}^*$ and thus the error in $\operatorname{div}\overline{\boldsymbol{y}}^*$ at all adaptation levels before level 11 is much smaller relative to the error measured in energy or combined energy norm. As a consequence, the ratio $D_F(v,-\Lambda^*\overline{\boldsymbol{y}}^*)/M^2_\oplus(v,\overline{\boldsymbol{y}}^*)$ is small and increases from around 4% to its final value of 73% at level 11. The respective efficiency indexes with respect to the energy and combined energy norms are given in Table 4.18. This time, the majorant $M^2_\oplus(v,\overline{\boldsymbol{y}}^*)$ gives a much tighter bound on the error in energy and combined energy norm and the efficiency indexes increase from around 1 to their final values at level 11 of 3.3889 and 1.9206, respectively.

Table 4.9: Example 1 (2D)

Example 1 (2D): $k_1 = 0.15$, $k_2 = 0.4$, $\epsilon_1 = 1$, $\epsilon_2 = 100$						
#elts	$\frac{\|v-u\|_{L^2(\Omega)}}{\|u\|_{L^2(\Omega)}}$[%]	$\frac{\|\|\nabla(v-u)\|\|}{\|\|\nabla u\|\|}$[%]	$\frac{\|\|\overline{\boldsymbol{y}}^*-\boldsymbol{p}^*\|\|_*}{\|\|\boldsymbol{p}^*\|\|_*}$[%]	$2M^2_\oplus(v,\overline{\boldsymbol{y}}^*)$	$2M^2_\oplus(v,\boldsymbol{p}^*)$	$2M^2_\oplus(u,\overline{\boldsymbol{y}}^*)$
196	15.0077	51.5582	1.44602	66.8185	66.5980	0.22050
347	5.69339	30.8534	1.44602	20.9985	20.7780	0.22050
630	4.20384	21.7715	1.44602	10.4406	10.2201	0.22050
1 315	2.39552	15.8532	1.44602	5.59620	5.37574	0.22050
2 865	1.87075	11.7353	1.44602	3.16460	2.94414	0.22050
5 938	0.64611	7.93001	1.44602	1.55920	1.33874	0.22050
12 006	0.36985	5.64786	1.44602	0.89920	0.67872	0.22050
24 571	0.16023	3.94241	1.44602	0.55090	0.33039	0.22050
48 483	0.08909	2.80265	1.44602	0.38750	0.16682	0.22050
97 423	0.03961	1.97875	1.44602	0.30370	0.08304	0.22050
192 905	0.02230	1.39832	1.44602	0.26210	0.04136	0.22050
386 185	0.01015	0.99471	1.44602	0.24150	0.02082	0.22050

Table 4.10: Example 1 (2D)

Example 1 (2D): $k_1 = 0.15$, $k_2 = 0.4$, $\epsilon_1 = 1$, $\epsilon_2 = 100$				
#elts	$\|\|\|\nabla(v-u)\|\|\|^2$	$\|\|\|\overline{\boldsymbol{y}}^*-\boldsymbol{p}^*\|\|\|_*^2$	$2D_F(v,-\Lambda^*\boldsymbol{p}^*)$	$2D_F(u,-\Lambda^*\overline{\boldsymbol{y}}^*)$
196	56.5057	0.04444	10.0923	0.17606
347	20.2350	0.04444	0.54296	0.17606
630	10.0756	0.04444	0.14450	0.17606
1 315	5.34235	0.04444	0.03338	0.17606
2 865	2.92742	0.04444	0.01671	0.17606
5 938	1.33673	0.04444	0.00200	0.17606
12 006	0.67805	0.04444	0.00067	0.17606
24 571	0.33038	0.04444	0.00001	0.17606
48 483	0.16696	0.04444	0.00000	0.17606
97 423	0.08323	0.04444	0.00000	0.17606
192 905	0.04156	0.04444	0.00000	0.17606
386 185	0.02103	0.04444	0.00000	0.17606

Table 4.11: Example 1 (2D)

Example 1 (2D): $k_1 = 0.15$, $k_2 = 0.4$, $\epsilon_1 = 1$, $\epsilon_2 = 100$						
# elements	$\frac{D_F(v,-\Lambda^*\overline{\boldsymbol{y}}^*)}{M^2_{\oplus}(v,\overline{\boldsymbol{y}}^*)}$ [%]	$I_{\text{Eff}}^{\text{CEN,Low}}$	$I_{\text{Eff}}^{\text{CEN,Up}}$	$I_{\text{Eff}}^{\text{E,Up}}$	$P_{\text{rel}}^{\text{CEN}}$ [%]	True rel. error in CEN [%]
196	15.8135	0.70520	1.08700	1.08740	38.5074	36.4137
347	3.61970	0.70640	1.01760	1.01870	22.1410	21.8386
630	3.24520	0.70650	1.01570	1.01800	15.5098	15.4285
1 315	2.99700	0.70980	1.01930	1.02350	11.3338	11.2565
2 865	5.11630	0.71080	1.03190	1.03970	8.41663	8.36086
5 938	9.91240	0.71310	1.06250	1.08000	5.75210	5.69982
12 006	19.9535	0.70580	1.11560	1.15160	4.11607	4.12246
24 571	35.1659	0.69030	1.21230	1.29130	2.89890	2.96931
48 483	45.0879	0.70940	1.35380	1.52340	2.23724	2.23000
97 423	59.5529	0.69360	1.54240	1.91030	1.69993	1.73298
192 905	68.6293	0.69130	1.74560	2.51110	1.39059	1.42237
386 185	73.0132	0.70550	1.92060	3.38890	1.23821	1.24105

4.1.4 Nonhomogeneous Dirichlet boundary condition

Now, we consider problem (4.2) in the case when $\gamma_2(\overline{g})$ is not identically zero on $\partial\Omega$. The unique solution $u \in H^1_{\gamma_2(\overline{g})}(\Omega) \cap L^\infty(\Omega)$ of (4.2) is the unique minimizer over $H^1_{\gamma_2(\overline{g})}(\Omega)$ of the functional J defined by (4.4). Then $u_0 := u - \overline{g} \in H^1_0(\Omega) \cap L^\infty(\Omega)$ is the unique solution of the problem

$$\text{Find } u_0 \in H^1_0(\Omega) \text{ such that } k^2 \sinh(\overline{g} + u_0 + w)v \in L^1(\Omega) \text{ for all } v \in H^1_0(\Omega) \text{ and} \tag{4.86}$$
$$\int_\Omega \boldsymbol{A}\nabla(\overline{g} + u_0) \cdot \nabla v dx + \int_\Omega k^2 \sinh(\overline{g} + u_0 + w)v dx = \int_\Omega (f_0 v + \boldsymbol{f} \cdot \nabla v)\, dx \text{ for all } v \in H^1_0(\Omega)$$

as well as of the variational problem

$$(P) \quad \text{Find } u_0 \in H^1_0(\Omega) \text{ such that } \overline{J}(u_0) = \min_{v \in H^1_0(\Omega)} \overline{J}(v), \tag{4.87}$$

where $\overline{J} : H^1_0(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is defined by $\overline{J}(v) := J(\overline{g} + v)$, i.e.,

$$\overline{J}(v) = \frac{1}{2} a(\overline{g} + v, \overline{g} + v) + \int_\Omega B(x, \overline{g} + v + w) dx - \int_\Omega (f_0(\overline{g} + v) + \boldsymbol{f} \cdot \nabla(\overline{g} + v))\, dx. \tag{4.88}$$

We recall that in (4.88) it is understood that $\overline{J}(v) = +\infty$ whenever $B(x, \overline{g} + v + w) \notin L^1(\Omega)$. Now problem (4.87) falls in the abstract framework considered in Section 4.1.2 and we can proceed as we did in the case of homogeneous Dirichlet boundary condition.

Remark 4.15
Another equivalent approach to the derivation of functional a posteriori error estimates for problem (4.86) *is to rewrite it in the following homogenized form*

$$\begin{aligned} &\textit{Find } u_0 \in H^1_0(\Omega) \textit{ such that } k^2 \sinh(\overline{g} + u_0 + w)v \in L^1(\Omega) \textit{ for all } v \in H^1_0(\Omega) \textit{ and} \\ &\int_\Omega \boldsymbol{A}\nabla u_0 \cdot \nabla v dx + \int_\Omega k^2 \sinh(\overline{g} + u_0 + w)v dx = \int_\Omega \Big(f_0 v + (\boldsymbol{f} - \boldsymbol{A}\nabla\overline{g}) \cdot \nabla v\Big) dx \\ &\textit{for all } v \in H^1_0(\Omega). \end{aligned} \tag{4.89}$$

Now, by recalling that $\overline{g} \in H^1_0(\Omega) \cap L^\infty(\Omega)$ and $\nabla\overline{g} \in [L^s(\Omega)]^d$ for some $s > d$, we can make the substitution $w \to \overline{g} + w \in L^\infty(\Omega_2)$ and $\boldsymbol{f} \to \boldsymbol{f} - \boldsymbol{A}\nabla\overline{g} \in [L^s(\Omega)]^d$ and apply all the results obtained in Section 4.1.3 for the case of homogeneous Dirichlet boundary condition.

We set $V := H^1_0(\Omega)$, $V^* := H^{-1}(\Omega)$, $Y := \big[L^2(\Omega)\big]^d$, $Y^* := \big[L^2(\Omega)\big]^d$ $(d = 2, 3)$, Λ the gradient operator $\nabla : H^1_0(\Omega) \to [L^2(\Omega)]^d$, and $\Lambda^* : Y^* \to V^*$ the operator adjoint to Λ. By noting that $\Lambda\overline{g} \in Y$, for any $v \in H^1_0(\Omega)$ and any $\boldsymbol{y} \in \big[L^2(\Omega)\big]^d$, we have (see also Chapter 7.3

in [139])

$$\begin{aligned}
\overline{G}(\boldsymbol{y}) &= G(\Lambda\overline{g}+\boldsymbol{y}) = \int\limits_{\Omega} \frac{1}{2}\boldsymbol{A}\left(\nabla\overline{g}+\boldsymbol{y}\right)\cdot\left(\nabla\overline{g}+\boldsymbol{y}\right)dx,\\
\overline{G}(\Lambda v) &:= G\left(\Lambda(\overline{g}+v)\right) = \int\limits_{\Omega} \frac{1}{2}\boldsymbol{A}\nabla(\overline{g}+v)\cdot\nabla(\overline{g}+v)dx,\\
\overline{F}(v) &:= F(\overline{g}+v) = \int\limits_{\Omega} B(x,\overline{g}+v+w)dx - \int\limits_{\Omega}\left(f_0(\overline{g}+v)+\boldsymbol{f}\cdot\nabla(\overline{g}+v)\right)dx,
\end{aligned} \tag{4.90}$$

where G and F are defined by (4.20). Then $\overline{J}(v)$ can be written in the form

$$\overline{J}(v) = \overline{G}(\Lambda v) + \overline{F}(v). \tag{4.91}$$

The dual counterpart to the primal variational problem (P) is given by

$$(P^*) \quad \text{Find } \boldsymbol{p}^* \in \left[L^2(\Omega)\right]^d \text{ such that } \overline{I}^*(\boldsymbol{p}^*) = \sup_{\boldsymbol{y}^*\in[L^2(\Omega)]^d} \overline{I}^*(\boldsymbol{y}^*), \tag{4.92}$$

where $\overline{I}^* : \left[L^2(\Omega)\right]^d \to \mathbb{R}\cup\{-\infty\}$ is defined by the relation

$$\overline{I}^*(\boldsymbol{y}^*) = -\overline{G}^*(\boldsymbol{y}^*) - \overline{F}^*(-\Lambda^*\boldsymbol{y}^*). \tag{4.93}$$

It is easy to see that all conditions on $\overline{G}$, $\overline{F}$, and $\overline{J}$ posed in Section 4.1.2 that ensure the existence and uniqueness of the primal and dual problems (P) and (P^*), as well as the validity of the strong duality relations (4.7), are satisfied. Hence we have

$$\overline{J}(u_0) = \inf_{v\in H_0^1(\Omega)} \overline{J}(v) = \sup_{\boldsymbol{y}^*\in[L^2(\Omega)]^d} \overline{I}^*(\boldsymbol{y}^*) = \overline{I}^*(\boldsymbol{p}^*).$$

Moreover, the optimality conditions (4.9) are satisfied, i.e.,

$$\Lambda u_0 \in \partial\overline{G}^*(\boldsymbol{p}^*), \quad \boldsymbol{p}^* \in \partial\overline{G}(\Lambda u_0). \tag{4.94}$$

The functional $\overline{G}$ is Gateaux-differentiable at any $\boldsymbol{y}\in\left[L^2(\Omega)\right]^d$ with a Gateaux differential given by $\boldsymbol{A}\left(\nabla\overline{g}+\boldsymbol{y}\right)$. Indeed, if $t>0$ and $\boldsymbol{q}\in\left[L^2(\Omega)\right]^d$, by using the symmetry of $\boldsymbol{A}$ we obtain

$$\begin{aligned}
\lim_{t\to0^+}\frac{\overline{G}(\boldsymbol{y}+t\boldsymbol{q})-\overline{G}(\boldsymbol{y})}{t} &= \lim_{t\to0^+}\frac{\frac{1}{2}\left(\boldsymbol{A}(\nabla\overline{g}+\boldsymbol{y}+t\boldsymbol{q}),\nabla\overline{g}+\boldsymbol{y}+t\boldsymbol{q}\right)-\frac{1}{2}\left(\boldsymbol{A}(\nabla\overline{g}+\boldsymbol{y}),\nabla\overline{g}+\boldsymbol{y}\right)}{t}\\
&= \lim_{t\to0^+}\frac{t\left(\boldsymbol{A}(\nabla\overline{g}+\boldsymbol{y}),\boldsymbol{q}\right)+\frac{t^2}{2}\left(\boldsymbol{A}\boldsymbol{q},\boldsymbol{q}\right)}{t}\\
&= \left(\boldsymbol{A}(\nabla\overline{g}+\boldsymbol{y}),\boldsymbol{q}\right).
\end{aligned}$$

This means that $\overline{G}$ has a unique subdifferential at $\boldsymbol{y}$ which coincides with its Gateaux differential $\boldsymbol{A}(\nabla\overline{g}+\boldsymbol{y})$. Therefore, from the second optimality condition in (4.94), it follows that

$$\boldsymbol{p}^* = \boldsymbol{A}\nabla(\overline{g}+u_0) = \boldsymbol{A}\nabla u. \tag{4.95}$$

Fenchel Conjugates of the functionals $\overline{G}$ and $\overline{F}$

Fenchel conjugate of $\overline{G}$

For any $\boldsymbol{y}^* \in Y^*$ we obtain

$$\begin{aligned}\overline{G}^*(\boldsymbol{y}^*) &= \sup_{\boldsymbol{y}\in Y}\left\{(\boldsymbol{y}^*,\boldsymbol{y}) - \overline{G}(\boldsymbol{y})\right\}\\ &= \sup_{\boldsymbol{y}\in Y}\left\{(\boldsymbol{y}^*,\Lambda\overline{g}+\boldsymbol{y}) - G(\Lambda\overline{g}+\boldsymbol{y}) - (\boldsymbol{y}^*,\Lambda\overline{g})\right\}\\ &= \sup_{\boldsymbol{\sigma}\in Y}\left\{(\boldsymbol{y}^*,\boldsymbol{\sigma}) - G(\boldsymbol{\sigma})\right\} - (\boldsymbol{y}^*,\Lambda\overline{g})\\ &= G^*(\boldsymbol{y}^*) - (\boldsymbol{y}^*,\Lambda\overline{g}),\end{aligned} \tag{4.96}$$

where we have used the fact that $\Lambda\overline{g} \in Y$.

Fenchel conjugate of $\overline{F}$

From (4.86) and (4.95) it follows that the exact solution $\boldsymbol{p}^* \in \left[L^2(\Omega)\right]^d$ of the dual problem (P^*) satisfies

$$\int_\Omega \boldsymbol{p}^*\cdot\nabla v dx + \int_\Omega k^2\sinh(\overline{g}+u_0+w)v dx = \int_\Omega (f_0 v + \boldsymbol{f}\cdot\nabla v)\,dx \ \text{ for all } v\in H_0^1(\Omega). \tag{4.97}$$

Therefore, $\boldsymbol{p}^*$ has the form

$$\boldsymbol{p}^* = \boldsymbol{f} + \boldsymbol{p}_0^*, \ \text{ where } \ \boldsymbol{p}_0^* \in H(\mathrm{div};\Omega) \ \text{ with } \ \mathrm{div}\,\boldsymbol{p}_0^* = b(x,\overline{g}+u_0+w) - f_0 \tag{4.98}$$

and it makes sense to compute $\overline{F}^*(-\Lambda^*\boldsymbol{y}^*)$ only for such $\boldsymbol{y}^* \in \left[L^2(\Omega)\right]^d$ that have the form

$$\boldsymbol{y}^* = \boldsymbol{f} + \boldsymbol{y}_0^* \ \text{ with } \ \boldsymbol{y}_0^* \in H(\mathrm{div};\Omega). \tag{4.99}$$

Notice that we do not impose the additional condition that $\boldsymbol{p}_0^*$ satisfies. However, in the process of computing $\overline{F}^*(-\Lambda^*\boldsymbol{y}^*)$ we will see that in order for $\overline{F}^*(-\Lambda^*\boldsymbol{y}^*)$ to be finite, $\boldsymbol{y}_0^*$ has to satisfy the condition $\mathrm{div}\,\boldsymbol{y}_0^* + f_0 = 0$ in the region where $k = 0$, i.e., in Ω_1 (the same reasoning was used for the computation of $\overline{F}^*(-\Lambda^*\boldsymbol{y}^*)$ in the case of homogeneous Dirichlet condition).

Recalling that in the particular case (b) for the functions k and w, the nonlinearity B is supported on Ω_2, for any $\boldsymbol{y}^*$ of the form (4.99), similarly to the case of homogeneous Dirichlet boundary condition, we have

$$\begin{aligned}\overline{F}^*(-\Lambda^*\boldsymbol{y}^*) &= \sup_{z\in H_0^1(\Omega)}\{\langle -\Lambda^*\boldsymbol{y}^*, z\rangle - F(\overline{g}+z)\} = \sup_{z\in H_0^1(\Omega)}\{(-\boldsymbol{y}^*,\Lambda z) - F(\overline{g}+z)\}\\ &= \sup_{z\in H_0^1(\Omega)}\int_\Omega \Big(-\boldsymbol{y}^*\cdot\nabla z - B(x,\overline{g}+z+w) + f_0(\overline{g}+z) + \boldsymbol{f}\cdot\nabla(\overline{g}+z)\Big)dx\\ &= \sup_{z\in H_0^1(\Omega)}\int_\Omega \Big(-\boldsymbol{y}_0^*\cdot\nabla z - B(x,\overline{g}+z+w) + f_0(\overline{g}+z) + \boldsymbol{f}\cdot\nabla\overline{g}\Big)dx\\ &= \sup_{z\in H_0^1(\Omega)}\int_\Omega \Big(\mathrm{div}\,\boldsymbol{y}_0^* z - B(x,\overline{g}+z+w) + f_0 z\Big)dx + \int_\Omega (f_0\overline{g} + \boldsymbol{f}\cdot\nabla\overline{g})\,dx.\end{aligned}$$

The last supremum is finite if $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1. For $\boldsymbol{q}^* \in H(\operatorname{div}; \Omega)$ and an arbitrary measurable function $z : \Omega_2 \to \mathbb{R}$, we introduce the functional

$$\overline{I}_{\boldsymbol{q}^*}(z) := \int_{\Omega_2} \left[(\operatorname{div} \boldsymbol{q}^* + f_0) z - B(x, \overline{g} + z + w) \right] dx. \tag{4.100}$$

Then, for $\boldsymbol{y}^*$ of the form (4.99) with $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1 we obtain

$$\begin{aligned}
&\overline{F}^*(-\Lambda^* \boldsymbol{y}^*) = \sup_{z \in H_0^1(\Omega)} \overline{I}_{\boldsymbol{y}_0^*}(z) + \int_{\Omega} (f_0 \overline{g} + \boldsymbol{f} \cdot \nabla \overline{g})\, dx \\
&\leq \int_{\Omega_2} \sup_{\xi \in \mathbb{R}} \Big\{ (\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x))\, \xi - B(x, \overline{g} + \xi + w(x)) \Big\} dx + \int_{\Omega} (f_0 \overline{g} + \boldsymbol{f} \cdot \nabla \overline{g})\, dx \\
&= \int_{\Omega_2} \Big((\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x))\, \xi_0(x) - B(x, \overline{g} + \xi_0(x) + w(x)) \Big) dx + \int_{\Omega} (f_0 \overline{g} + \boldsymbol{f} \cdot \nabla \overline{g})\, dx \\
&= \overline{I}_{\boldsymbol{y}_0^*}(\xi_0) + \int_{\Omega} (f_0 \overline{g} + \boldsymbol{f} \cdot \nabla \overline{g})\, dx.
\end{aligned} \tag{4.101}$$

Here $\xi_0 : \Omega_2 \to \mathbb{R}$ is computed from the condition

$$\frac{d}{d\xi} \left[(\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x))\, \xi - B(x, \overline{g} + \xi + w(x)) \right] = 0, \text{ for a.e. } x \in \Omega_2, \tag{4.102}$$

which is equivalent to

$$\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x) - k^2(x) \sinh(\overline{g}(x) + \xi + w(x)) = 0 \text{ for a.e. } x \in \Omega_2.$$

The solution of the last equation is given by

$$\xi_0(x) = \operatorname{arsinh}\left(\frac{\operatorname{div} \boldsymbol{y}_0^*(x) + f_0(x)}{k^2(x)} \right) - (\overline{g}(x) + w(x)) \text{ for a.e. } x \in \Omega_2, \tag{4.103}$$

If we substitute w with $\overline{g} + w$ in Proposition 4.5, we see that $\sup_{z \in H_0^1(\Omega)} \overline{I}_{\boldsymbol{y}_0^*}(z) = \overline{I}_{\boldsymbol{y}_0^*}(\xi_0)$ and hence we have equalities everywhere in (4.101). By using the expression for $\xi_0(x)$, for any $\boldsymbol{y}^* \in [L^2(\Omega)]^d$ of the form (4.29) with $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1 we obtain an explicit formula for $\overline{F}^*(-\Lambda^* \boldsymbol{y}^*)$:

$$\begin{aligned}
\overline{F}^*(-\Lambda^* \boldsymbol{y}^*) = \int_{\Omega_2} \Bigg[& k^2 \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) \left(\operatorname{arsinh}\left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) - (\overline{g} + w) \right) \\
& - k^2 \cosh\left(\operatorname{arsinh}\left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) \right) \Bigg] dx + \int_{\Omega} (f_0 \overline{g} + \boldsymbol{f} \cdot \nabla \overline{g})\, dx.
\end{aligned} \tag{4.104}$$

Error measures

For any $\boldsymbol{y}^* \in \left[L^2(\Omega)\right]^d$ of the form (4.99) with $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1, the quantity $\overline{M}^2_\oplus(v, \boldsymbol{y}^*)$ is fully computable and is given by the relation (4.11). To give an explicit expression for $\overline{M}^2_\oplus(v, \boldsymbol{y}^*)$ we compute $D_{\overline{G}}(\Lambda v, \boldsymbol{y}^*)$ and $D_{\overline{F}}(v, -\Lambda^* \boldsymbol{y}^*)$, where $v \in H^1_0(\Omega)$ is an approximation of u_0. For the first compound functional by using (4.90) and (4.96) we obtain

$$\begin{aligned} D_{\overline{G}}(\Lambda v, \boldsymbol{y}^*) =& \overline{G}(\Lambda v) + \overline{G}^*(\boldsymbol{y}^*) - (\boldsymbol{y}^*, \Lambda v) \\ =& G\left(\Lambda(\overline{g} + v)\right) + G^*(\boldsymbol{y}^*) - (\boldsymbol{y}^*, \Lambda \overline{g}) - (\boldsymbol{y}^*, \Lambda v) \\ =& D_G(\Lambda(\overline{g} + v), \boldsymbol{y}^*) = \frac{1}{2} |\!|\!| \boldsymbol{A} \nabla(\overline{g} + v) - \boldsymbol{y}^* |\!|\!|^2_*. \end{aligned} \tag{4.105}$$

For the second compound functional by using (4.90) and (4.104) we have

$$\begin{aligned} & D_{\overline{F}}(v, -\Lambda^* \boldsymbol{y}^*) = \overline{F}(v) + \overline{F}^*(-\Lambda^* \boldsymbol{y}^*) + \langle \Lambda^* \boldsymbol{y}^*, v \rangle \\ =& \int_\Omega \left(k^2 \cosh(\overline{g} + v + w) - f_0(\overline{g} + v) - \boldsymbol{f} \cdot \nabla(\overline{g} + v) \right) dx \\ +& \int_{\Omega_2} \left[k^2 \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) \left(\operatorname{arsinh} \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) - (\overline{g} + w) \right) \right. \\ & \left. - k^2 \cosh \left(\operatorname{arsinh} \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) \right) \right] dx + \int_\Omega (f_0 \overline{g} + \boldsymbol{f} \cdot \nabla \overline{g})\, dx + \int_\Omega \boldsymbol{y}^* \cdot \nabla v dx, \end{aligned}$$

from which by using the relation $\boldsymbol{y}^* = \boldsymbol{f} + \boldsymbol{y}_0^*$ and the fact that $\boldsymbol{y}_0^*$ is in $H(\operatorname{div}; \Omega)$ with $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1 we obtain

$$\begin{aligned} D_{\overline{F}}(v, -\Lambda^* \boldsymbol{y}^*) =& \int_{\Omega_2} k^2 \left[\cosh(\overline{g} + v + w) + \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) \operatorname{arsinh} \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) \right. \\ & \left. - \cosh \left(\operatorname{arsinh} \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) \right) - \left(\frac{\operatorname{div} \boldsymbol{y}_0^* + f_0}{k^2} \right) (\overline{g} + v + w) \right] dx \\ =& D_F(\overline{g} + v, -\Lambda^* \boldsymbol{y}^*). \end{aligned} \tag{4.106}$$

The fully computable majorant $\overline{M}^2_\oplus(v, \boldsymbol{y}^*)$ is given by

$$\begin{aligned} \overline{M}^2_\oplus(v, \boldsymbol{y}^*) &= D_{\overline{G}}(\Lambda v, \boldsymbol{y}^*) + D_{\overline{F}}(v, -\Lambda^* \boldsymbol{y}^*) \\ &= D_G(\Lambda(\overline{g} + v), \boldsymbol{y}^*) + D_F(\overline{g} + v, -\Lambda^* \boldsymbol{y}^*) \\ &= \int_\Omega \overline{\eta}^2(x) dx \\ &= \frac{1}{2} |\!|\!| \boldsymbol{A} \nabla(\overline{g} + v) - \boldsymbol{y}^* |\!|\!|^2_* + D_F(\overline{g} + v, -\Lambda^* \boldsymbol{y}^*) = M^2_\oplus(\overline{g} + v, \boldsymbol{y}^*), \end{aligned} \tag{4.107}$$

where $D_{\overline{G}}(\Lambda v, \boldsymbol{y}^*)$ and $D_{\overline{F}}(v, -\Lambda^* \boldsymbol{y}^*)$ are given by (4.105) and (4.106), respectively, and D_G, D_F, $M^2_\oplus$ are the compound functionals and majorant from the case with homogeneous Dirichlet boundary condition. If we denote $\overline{v} = \overline{g} + v$, we obtain

$$\overline{M}^2_\oplus(v, \boldsymbol{y}^*) = M^2_\oplus(\overline{v}, \boldsymbol{y}^*)$$

and

$$\overline{\eta}^2(x) = \begin{cases} \frac{1}{2}\boldsymbol{A}^{-1}\left(\boldsymbol{A}\nabla\overline{v} - \boldsymbol{f} - \boldsymbol{y}_0^*\right)\cdot\left(\boldsymbol{A}\nabla\overline{v} - \boldsymbol{f} - \boldsymbol{y}_0^*\right), \ \text{for } x\in\Omega_1, \\ \\ \frac{1}{2}\boldsymbol{A}^{-1}\left(\boldsymbol{A}\nabla\overline{v} - \boldsymbol{f} - \boldsymbol{y}_0^*\right)\cdot\left(\boldsymbol{A}\nabla\overline{v} - \boldsymbol{f} - \boldsymbol{y}_0^*\right) \\ \quad +k^2\left[\cosh(\overline{v}+w) + \left(\frac{\operatorname{div}\boldsymbol{y}_0^*+f_0}{k^2}\right)\operatorname{arsinh}\left(\frac{\operatorname{div}\boldsymbol{y}_0^*+f_0}{k^2}\right)\right. \\ \quad \left. -\cosh\left(\operatorname{arsinh}\left(\frac{\operatorname{div}\boldsymbol{y}_0^*+f_0}{k^2}\right)\right) - \left(\frac{\operatorname{div}\boldsymbol{y}_0^*+f_0}{k^2}\right)(\overline{v}+w)\right], \ \text{for } x\in\Omega_2. \end{cases} \tag{4.108}$$

Since v is an arbitrary approximation of u_0 in $H_0^1(\Omega)$, we can consider $\overline{v}$ as an arbitrary approximation of u in $H^1_{\gamma_2(\overline{g})}(\Omega)$. We recall that $\overline{\eta}^2(x)\geq 0$ since it is the sum of the compound functionals (which are nonnegative by the definiton of a Fenchel conjugate) generated by $g_x(\xi) := g(x,\xi) = \frac{1}{2}\boldsymbol{A}(x)\xi\cdot\xi$ and $f_x(s) := B(x, s+w(x)) - f_0(x)s$ and evaluated at $(\nabla\overline{v}(x), \boldsymbol{y}^*(x))$ and $(\overline{v}(x), \operatorname{div}\boldsymbol{y}_0^*(x))$, respectively.

To give an explicit form of the principal error identity (4.14), we should also compute the quantities $D_{\overline{G}}(\Lambda v, \boldsymbol{p}^*)$, $D_{\overline{G}}(\Lambda u_0, \boldsymbol{y}^*)$, $D_{\overline{F}}(v, -\Lambda^*\boldsymbol{p}^*)$, and $D_{\overline{F}}(u_0, -\Lambda^*\boldsymbol{y}^*)$. By using (4.105) and the relation $\boldsymbol{p}^* = \boldsymbol{A}\nabla(\overline{g}+u_0) = \boldsymbol{A}\nabla u$, we obtain

$$\begin{aligned} D_{\overline{G}}(\Lambda v, \boldsymbol{p}^*) &= \frac{1}{2}\int_\Omega \boldsymbol{A}\nabla(v-u_0)\cdot\nabla(v-u_0)dx \\ &= \frac{1}{2}|||\nabla(v-u_0)|||^2 = \frac{1}{2}|||\nabla\left(\overline{v}-u\right)|||^2, \end{aligned} \tag{4.109}$$

$$D_{\overline{G}}(\Lambda u_0, \boldsymbol{y}^*) = \frac{1}{2}\int_\Omega \boldsymbol{A}^{-1}(\boldsymbol{y}^*-\boldsymbol{p}^*)\cdot(\boldsymbol{y}^*-\boldsymbol{p}^*)dx = \frac{1}{2}|||\boldsymbol{y}^*-\boldsymbol{p}^*|||_*^2. \tag{4.110}$$

Notice that $D_{\overline{G}}(\Lambda v, \boldsymbol{p}^*)$ measures the error in energy norm in both approximations v and $\overline{v}$ of u_0 and u, respectively.

Next, we obtain explicit expressions for the nonlinear measures $D_{\overline{F}}(v, -\Lambda^*\boldsymbol{p}^*)$ and $D_{\overline{F}}(u_0, -\Lambda^*\boldsymbol{y}^*)$ in a similar fashion to the case of homogeneous Dirichlet boundary condition. Since for the exact solution u_0 of the primal problem (P) and the part $\boldsymbol{p}_0^*$ of the exact solution of the dual problem (P^*) we have the relations

$$\frac{\operatorname{div}\boldsymbol{p}_0^*+f_0}{k^2} = \sinh(\overline{g}+u_0+w) \quad \text{and} \quad u_0 = \operatorname{arsinh}\left(\frac{\operatorname{div}\boldsymbol{p}_0^*+f_0}{k^2}\right) - (\overline{g}+w) \ \text{ a.e. in } \Omega_2,$$

by using (4.106) we obtain

$$\begin{aligned}
&D_{\overline{F}}(v,-\Lambda^*\boldsymbol{p}^*)\\
=&\int_{\Omega_2} k^2\left(\cosh(\overline{g}+v+w)-\cosh(\overline{g}+u_0+w)+u_0\sinh(\overline{g}+u_0+w)-v\sinh(\overline{g}+u_0+w)\right)dx\\
=&\int_{\Omega_2} k^2\left(\cosh(\overline{v}+w)-\cosh(u+w)+u\sinh(u+w)-\overline{v}\sinh(u+w)\right)dx\\
=&D_F(\overline{v},-\Lambda^*\boldsymbol{p}^*).
\end{aligned}\tag{4.111}$$

Similarly,

$$\begin{aligned}
D_{\overline{F}}(u_0,-\Lambda^*\boldsymbol{y}^*)&=\int_{\Omega_2} k^2\left(\cosh(T)-\cosh(S)+S\sinh(S)-T\sinh(S)\right)dx\\
&=D_F(u,-\Lambda^*\boldsymbol{y}^*),
\end{aligned}\tag{4.112}$$

where

$$T:=\operatorname{arsinh}\left(\frac{\operatorname{div}\boldsymbol{p}_0^*+f_0}{k^2}\right)\quad\text{and}\quad S:=\operatorname{arsinh}\left(\frac{\operatorname{div}\boldsymbol{y}_0^*+f_0}{k^2}\right).$$

As before, the nonlinear quantities $D_{\overline{F}}(v,-\Lambda^*\boldsymbol{p}^*)$ and $D_{\overline{F}}(u_0,-\Lambda^*\boldsymbol{y}^*)$ measure the error over Ω_2 in $v,\overline{v}$ and in $\operatorname{div}\boldsymbol{y}_0^*$, respectively. As before, by using inequality (4.44), we can obtain

$$\int_{\Omega_2}\frac{k^2}{2}(v-u_0)^2dx\le D_{\overline{F}}(v,-\Lambda^*\boldsymbol{p}^*)\le\int_{\Omega_2}\frac{k^2}{2}(\sinh(\overline{g}+v+w)-\sinh(\overline{g}+u_0+w))^2dx\tag{4.113}$$

and

$$\int_{\Omega_2}\frac{k^2}{2}(S-T)^2dx\le D_{\overline{F}}(u_0,-\Lambda^*\boldsymbol{y}^*)\le\int_{\Omega_2}\frac{1}{2k^2}(\operatorname{div}\boldsymbol{y}_0^*-\operatorname{div}\boldsymbol{p}_0^*)^2dx.\tag{4.114}$$

By taking into account (4.109), (4.110) and the relations $D_{\overline{F}}(v,-\Lambda^*\boldsymbol{p}^*)=D_F(\overline{v},-\Lambda^*\boldsymbol{p}^*)$, $D_{\overline{F}}(u_0,-\Lambda^*\boldsymbol{y}^*)=D_F(u,-\Lambda^*\boldsymbol{y}^*)$, $M^2_{\oplus}(v,\boldsymbol{y}^*)=M^2_{\oplus}(\overline{v},\boldsymbol{y}^*)$, the abstract error identity (4.14) can be written in an analogous form to (4.51) (we only give it in terms of $\overline{v}=\overline{g}+v$ and $u=\overline{g}+u_0$)

$$\begin{aligned}
&\frac{1}{2}|||\nabla(\overline{v}-u)|||^2+\frac{1}{2}|||\boldsymbol{y}^*-\boldsymbol{p}^*|||_*^2\\
&+\int_{\Omega_2}\frac{k^2}{2}(\overline{v}-u)^2dx+C_2\left(\delta_2,\|\operatorname{div}\boldsymbol{p}_0^*\|_{L^\infty(\Omega_2)}\right)\int_{\Omega_2}\frac{1}{2k^2}(\operatorname{div}\boldsymbol{y}_0^*-\operatorname{div}\boldsymbol{p}_0^*)^2dx\\
&\le\frac{1}{2}|||\nabla(\overline{v}-u)|||^2+\frac{1}{2}|||\boldsymbol{y}^*-\boldsymbol{p}^*|||_*^2+D_F(\overline{v},-\Lambda^*\boldsymbol{p}^*)+D_F(u,-\Lambda^*\boldsymbol{y}^*)=M^2_{\oplus}(\overline{v},\boldsymbol{y}^*)\\
&\le\frac{1}{2}|||\nabla(\overline{v}-u)|||^2+\frac{1}{2}|||\boldsymbol{y}^*-\boldsymbol{p}^*|||_*^2\\
&+C_1\left(\delta_1,\|u\|_{L^\infty(\Omega)}\right)\int_{\Omega_2}\frac{k^2}{2}(\overline{v}-u)^2dx+\int_{\Omega_2}\frac{1}{2k^2}(\operatorname{div}\boldsymbol{y}_0^*-\operatorname{div}\boldsymbol{p}_0^*)^2dx,
\end{aligned}\tag{4.115}$$

where $\delta_1 > 0$, $\delta_2 > 0$ are arbitrarily fixed numbers and C_1, C_2 are local Lipschitz constants for the function sinh that depend on δ_1, $\|u\|_{L^\infty(\Omega)}$ and δ_2, $\|\operatorname{div} \boldsymbol{p}_0^*\|_{L^\infty(\Omega_2)}$, respectively. We note that the inequalities in (4.115) are valid for $\overline{v}$ and $\boldsymbol{y}_0^*$ such that $\|\overline{v} - u\|_{L^\infty(\Omega)} \leq \delta_1$ (or equivalently, $\|v - u_0\|_{L^\infty(\Omega)} \leq \delta_1$) and for $f_0 \in L^\infty(\Omega_2)$, $\|\operatorname{div}(\boldsymbol{y}_0^* - \boldsymbol{p}_0^*)\|_{L^\infty(\Omega_2)} \leq \delta_2$. We also pay attention to the fact that the underlined equality in (4.115) is valid without the conditions mentioned above (remember that $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1 is always assumed).

We also have the following lower bound for the combined energy norm of the error:

$$\frac{1}{2}|||\boldsymbol{A}\nabla\overline{v} - \boldsymbol{y}^*|||_*^2 \leq |||\nabla(\overline{v} - u)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2. \tag{4.116}$$

Guaranteed lower and upper bounds for the primal part of the error

Recall from Section 4.1.2 that the primal part of the error is given by

$$\overline{J}(v) - \overline{J}(u_0) = D_{\overline{G}}(\Lambda v, \boldsymbol{p}^*) + D_{\overline{F}}(v, -\Lambda^*\boldsymbol{p}^*), \tag{4.117}$$

or equivalently, in terms of $\overline{v} = \overline{g} + v$,

$$J(\overline{v}) - J(u) = D_G(\Lambda\overline{v}, \boldsymbol{p}^*) + D_F(\overline{v}, -\Lambda^*\boldsymbol{p}^*). \tag{4.118}$$

By using (4.13) and the fact that u_0 is a minimizer of $\overline{J}$ (or equivalently, u is a minimizer of J), for any approximation $v \in H_0^1(\Omega)$ of u_0 and any $w \in H_0^1(\Omega)$, $\boldsymbol{y}^* \in Y^*$ of the form (4.99) with $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1 we obtain

$$\overline{M}_{\ominus}^2(v, w) := \overline{J}(v) - \overline{J}(w) \leq \overline{J}(v) - \overline{J}(u_0) \leq \overline{M}_{\oplus}^2(v, \boldsymbol{y}^*). \tag{4.119}$$

Equivalently, by using the relation $\overline{J}(v) = J(\overline{g} + v)$ for any $v \in H_0^1(\Omega)$, we can write (4.119) in terms of $\overline{v}$, $\overline{w} := \overline{g} + w$, $u = \overline{g} + u_0$ as follows

$$M_{\ominus}^2(\overline{v}, \overline{w}) := J(\overline{v}) - J(\overline{w}) \leq J(\overline{v}) - J(u) \leq M_{\oplus}^2(\overline{v}, \boldsymbol{y}^*), \tag{4.120}$$

where we remind that the left-hand sides of (4.119) and (4.120) make sense only if $\overline{J}(v) - \overline{J}(w) \geq 0$ and $J(\overline{v}) - J(\overline{w}) \geq 0$.

Near best approximation result

Let $V_h \subset L^\infty(\Omega)$ be a closed subspace of $H_0^1(\Omega)$ (not necessarily finite dimensional) and $u_{0,h}$ be the minimizer of $\overline{J}$ over V_h (this minimizer is unique), which is the unique solution of the Galerkin problem:

$$\begin{aligned}&\text{Find } u_{0,h} \in V_h \text{ such that}\\ &a(u_{0,h}, v) + \int\limits_\Omega b(x, \overline{g} + u_{0,h} + w)v dx = \int\limits_\Omega \Big(f_0 v + (\boldsymbol{f} - \boldsymbol{A}\nabla\overline{g}) \cdot \nabla v\Big) dx \ \text{ for all } v \in V_h.\end{aligned} \tag{4.121}$$

Then, using (4.12b) and the expression (4.109) for $D_{\overline{G}}(\Lambda v, \boldsymbol{p}^*)$, for any $v \in V_h$ we can write

$$\begin{aligned} &|||\nabla(u_{0,h} - u_0)|||^2 + 2D_{\overline{F}}(u_{0,h}, -\Lambda^*\boldsymbol{p}^*) = 2\left(\overline{J}(u_{0,h}) - \overline{J}(u_0)\right) \\ &\leq 2\left(\overline{J}(v) - \overline{J}(u_0)\right) = |||\nabla(v - u_0)|||^2 + 2D_{\overline{F}}(v, -\Lambda^*\boldsymbol{p}^*). \end{aligned}$$

Since $2D_{\overline{F}}(u_{0,h}, -\Lambda^*\boldsymbol{p}^*) \geq 0$, by using (4.113), we obtain the following generalization of Cea's Lemma.

Proposition 4.16
Let $V_h \subset L^\infty(\Omega)$ be a closed subspace of $H_0^1(\Omega)$ and $u_{0,h} \in V_h$ be the Galerkin approximation of u_0 defined by (4.121)*. Then*

$$|||\nabla(u_{0,h} - u_0)|||^2 \leq \inf_{v \in V_h} \left\{ |||\nabla(v - u_0)|||^2 + \int_{\Omega_2} k^2(\sinh(\overline{g} + v + w) - \sinh(\overline{g} + u_0 + w))^2 dx \right\}, \tag{4.122}$$

or equivalently,

$$|||\nabla(\overline{u}_h - u)|||^2 \leq \inf_{v \in \overline{g} + V_h} \left\{ |||\nabla(\overline{v} - u)|||^2 + \int_{\Omega_2} k^2(\sinh(\overline{v} + w) - \sinh(u + w))^2 dx \right\}, \tag{4.123}$$

where as usual $\overline{v} = \overline{g} + v$, $\overline{u}_h = \overline{g} + u_{0,h}$, and $u = \overline{g} + u_0$.

Effect of data oscillation

The effect of data oscillation, i.e., the effect of partial equilibration of $\boldsymbol{y}_0^*$ is quite analogous to the case of homogeneous Dirichlet boundary condition and we do not repeat it here. We only note that one has first to obtain an analogue of (4.81) for u_0 and $u_{0,h}^{\mathrm{osc}}$, an approximation of the solution u_0^{osc} of the homogenized problem (4.89) with $f_0^{\mathrm{osc}} = \Pi_{L_h}(f_0)$ instead of f_0. Then the obtained estimate is rewritten in terms of $u = \overline{g} + u_0$, $u_h^{\mathrm{osc}} := \overline{g} + u_{0,h}^{\mathrm{osc}}$ to obtain

$$|||\nabla(u_h^{\mathrm{osc}} - u)||| \leq \sqrt{2} M_{\oplus,\mathrm{osc}}(u_h^{\mathrm{osc}}, \boldsymbol{y}^*) + \frac{1}{\mu_1} \left(\sum_{K \in \mathscr{T}_h} \frac{h_K^2}{\pi^2} \|f_0 - f_0^{\mathrm{osc}}\|_{L^2(K)}^2 \right)^{\frac{1}{2}}, \tag{4.124}$$

where $M_{\oplus,\mathrm{osc}}^2(u_h^{\mathrm{osc}}, \boldsymbol{y}^*)$ is the majorant for the nonhomogeneous problem with f_0^{osc} instead of f_0 and has the form (4.107).

4.2 Finding a good approximation of the dual variable

Minimization of the majorant

The problem of finding a good approximation $\boldsymbol{y}^*$ of $\boldsymbol{p}^*$ is the same for both cases of homogeneous and nonhomogeneous Dirichlet boundary conditions. Thus, we only consider the homogeneous case and use the respective notation. Our goal is to find for a given $v \in H_0^1(\Omega)$

a good approximation $\boldsymbol{y}^* \in \left[L^2(\Omega)\right]^d$ of the exact flux $\boldsymbol{p}^*$ of the form $\boldsymbol{y}^* = \boldsymbol{f} + \boldsymbol{y}_0^*$ with $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1, i.e., over the set where $k = 0$. For this purpose we can minimize the majorant $M_{\oplus}^2(v, \boldsymbol{f} + \boldsymbol{y}_0^*)$ or, equivalently, minimize the functional

$$-I^*(\boldsymbol{f} + \boldsymbol{y}_0^*) = G^*(\boldsymbol{f} + \boldsymbol{y}_0^*) + F^*\left(-\Lambda^*(\boldsymbol{f} + \boldsymbol{y}_0^*)\right)$$

in $\boldsymbol{y}_0^*$ over a subspace of $H(\operatorname{div};\Omega)$ by additionally enforcing the condition that $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1. For example one can minimize $M_{\oplus}^2(v, \boldsymbol{f} + \boldsymbol{y}_0^*)$ over a finite dimensional subspace of $H(\operatorname{div};\Omega)$ such as the lowest order Raviart-Thomas space RT_0. Let $\mathscr{T}_h$ be a partition of $\overline{\Omega}$ into triangles in 2D and tetrahedrons in 3D and let f_0 be a piecewise constant of degree s over each element $K \in \mathscr{T}_h$. Then one can minimize the majorant $M_{\oplus}^2(v, \boldsymbol{f} + \boldsymbol{y}_0^*)$ over the subspace RT_0 by additionally enforcing the condition $\operatorname{div} \boldsymbol{y}_0^* + f_0 = 0$ in Ω_1 (if f_0 is not piecewise constant, then one also has the data oscillation effect described in Section 4.1.3 on p. 108). Obviously, this approach gives the best possible $\boldsymbol{y}_0^*$ in the respective subspace of $H(\operatorname{div};\Omega)$, but it essentially requires the solution of the dual problem which can be quite costly.

Patchwise equilibrated flux reconstruction

Another, computationally favorable, approach, which can be easily realized in parallel, is to apply a patchwise flux reconstruction that will also yield a partial equilibration of $\boldsymbol{y}_0^*$ (see [29–31]). Since the computations on each patch are independent from the computations on the rest of the patches, this reconstruction is easy to implement in parallel. We assume that Ω is a polyhedral domain. Let $\mathscr{T}_h := \{K_i\}_{i=1}^{N_E}$ be a partition of $\overline{\Omega}$ into triangles in 2D and tetrahedrons in 3D, which is a part of a family of shape-regular triangulations $\{\mathscr{T}_h\}_{h\to 0}$. We will present this equilibrated patchwise flux reconstruction in the case where u_h is the Galerkin approaximation of u in the finite element space $V_h^1 \subset H_0^1(\Omega)$ of continuous piecewise linear functions over the triangulation $\mathscr{T}_h$, i.e., u_h satisfies the problem

$$\begin{aligned} &\text{Find } u_h \in V_h^1 \text{ such that} \\ &a(u_h, v) + \int_\Omega b(x, u_h + w) v dx = \int_\Omega \left(f_0 v + \boldsymbol{f} \cdot \nabla v\right) dx \ \text{ for all } v \in V_h^1. \end{aligned} \tag{4.64}$$

Additionally, let $L_h \subset L^2(\Omega)$ denote the space of piecewise constant functions with respect to the triangulation $\mathscr{T}_h$. Then this approach generates a $\boldsymbol{y}_0^*$ such that $\operatorname{div} \boldsymbol{y}_0^* + \Pi_{L_h}(f_0) = 0$ in Ω_1, where Π_{L_h} denotes the $L^2(\Omega)$-orthogonal projection operator onto the subspace L_h.
First, we briefly present the method in the case of a linear problem with a homogeneous interface condition, i.e., $b = 0$, $\boldsymbol{f} = 0$ and $\epsilon \nabla u \in H(\operatorname{div};\Omega)$ by following the original work in [29–31]. We consider the problem

$$\begin{aligned} -\nabla \cdot (\epsilon \nabla u) &= f_0 \ \text{ in } \Omega, \\ u &= 0 \ \text{ on } \partial\Omega, \end{aligned} \tag{4.125}$$

where ϵ is a piecewise constant function with respect to the triangulation $\mathscr{T}_h$ such that $\epsilon_{\max} \geq \epsilon(x) \geq \epsilon_{\min} > 0$ for a.e. $x \in \Omega$. Here we note that the case of nonhomogeneous Dirichlet boundary condition is treated in the exact same way. The goal is to find a dual variable $\boldsymbol{\sigma} \in RT_0 \subset H(\mathrm{div};\Omega)$ (here $\boldsymbol{\sigma}$ plays the role of $\boldsymbol{y}^*$) with $\mathrm{div}\,\boldsymbol{\sigma} + \Pi_{L_f}(f_0) = 0$ in Ω. We search it in the form $\boldsymbol{\sigma} = \epsilon\nabla u_h + \boldsymbol{\sigma}^\Delta$, where $u_h \in H_0^1(\Omega)$ is the Galerkin finite element approximation of (4.125) in the space V_h^1 and

$$\begin{aligned} RT_{0,-1} &:= \left\{ \boldsymbol{\tau} \in \left[L^2(\Omega)\right]^d : \boldsymbol{\tau}_{\restriction_K} = a_K + b_K x,\, a_K \in \mathbb{R}^d,\, b_K \in \mathbb{R},\, \forall K \in \mathscr{T}_h \right\}, \\ RT_0 &:= RT_{0,-1} \cap H(\mathrm{div};\Omega). \end{aligned} \tag{4.126}$$

Obviously, $\epsilon\nabla u_h$ is a piecewise constant function with respect to $\mathscr{T}_h$ and in general it is not in $H(\mathrm{div};\Omega)$. This means that $\boldsymbol{\sigma}^\Delta$ is also not in $H(\mathrm{div};\Omega)$ and that $\boldsymbol{\sigma}^\Delta$ is only in the broken Raviart-Thomas space $RT_{0,-1}$. The conditions $\boldsymbol{\sigma}^\Delta + \epsilon\nabla u_h \in H(\mathrm{div};\Omega)$ and $\mathrm{div}(\boldsymbol{\sigma}^\Delta + \epsilon\nabla u_h)$ can be rewritten in the form

$$\begin{aligned} \mathrm{div}\,\boldsymbol{\sigma}^\Delta &= -f_0 \ \text{ in all } K \in \mathscr{T}_h, \\ \left[\boldsymbol{\sigma}^\Delta \cdot \boldsymbol{n}\right] &= -\left[\epsilon\nabla u_h \cdot \boldsymbol{n}\right] \ \text{ on all internal faces, } F, \text{ i.e., } F \not\subset \partial\Omega. \end{aligned} \tag{4.127}$$

For a given vertex V of the mesh let $\psi_V \in V_h^1$ be the hat basis function with $\psi_V(V) = 1$ and $\psi_V(x) = 0$ for $x \in \Omega \setminus \omega_V$, and let $\omega_V := \bigcup \{K : V \in \partial K\}$ be its associated patch. Then the correction $\boldsymbol{\sigma}^\Delta \in RT_{0,-1}$ is defined as

$$\boldsymbol{\sigma}^\Delta = \sum_{i=1}^{N_V} \boldsymbol{\sigma}_{\omega_V}, \tag{4.128}$$

where N_V is the number of vertices in $\mathscr{T}_h$ and $\{\boldsymbol{\sigma}_{\omega_V}\} \subset RT_{0,-1}$ with $\mathrm{supp}\,\boldsymbol{\sigma}_{\omega_V} \subset \omega_V$ for all vertices V, are the solutions of the following local problems:

$$\begin{aligned} \mathrm{div}\,\boldsymbol{\sigma}_{\omega_V} &= -\frac{1}{|K|}\int\limits_K f_0\psi_V dx \ \text{ in all } K \subset \omega_V, \\ [\boldsymbol{\sigma}_{\omega_V} \cdot \boldsymbol{n}] &= -\frac{1}{d}\left[\epsilon\nabla u_h \cdot \boldsymbol{n}\right] \ \text{ on all } F \subset \omega_V \setminus \partial\omega_V, \\ \boldsymbol{\sigma}_{\omega_V} \cdot \boldsymbol{n} &= 0 \ \text{ on } \partial\omega_V. \end{aligned} \tag{4.129}$$

We note that the last equation in (4.129) is enforced only for vertices $V \notin \partial\Omega$ and that all local problems have a solution (see [30, 31]). In 2D there is also a constructive solution for the local problems (see, e.g. [28, 30]).

Now, we consider the general problem (4.64) where ϵ is not necessarily a piecewise constant function. Our considerations are motivated by the observation that the exact $\boldsymbol{p}_0^* := \boldsymbol{p}^* - \boldsymbol{f}$ satisfies the integral identity

$$\int\limits_\Omega \boldsymbol{p}_0^* \cdot \nabla v dx = \int\limits_\Omega \left(-b(x, u+w) + f_0\right) v dx \ \text{ for all } v \in H_0^1(\Omega), \tag{4.130}$$

where $\boldsymbol{p}^* = \epsilon \nabla u$. If we define the function $\boldsymbol{q} := \epsilon \nabla u_h - \boldsymbol{f}$, then by using (4.64) we see that

$$\int_\Omega \boldsymbol{q} \cdot \nabla v dx = \int_\Omega \left(-b(x, u_h + w) + f_0\right) v dx \text{ for all } v \in V_h^1.$$

Since $\nabla v \in L_h$ for all $v \in V_h^1$ it follows that $\Pi_{L_h}(\boldsymbol{q}) \in L_h$ (by abusing the notation we use L_h and Π_{L_h} for both the scalar and vectorial cases) satisfies the problem

$$\int_\Omega \Pi_{L_h}(\boldsymbol{q}) \cdot \nabla v dx = \int_\Omega \left(-b(x, u_h + w) + f_0\right) v dx \text{ for all } v \in V_h^1. \tag{4.131}$$

We define $\boldsymbol{y}_0^* \in RT_0$ by applying the described above patchwise equilibrated flux reconstruction to the numerical flux $\Pi_{L_h}(\boldsymbol{q}) \in L_h$, where $-b(x, u_h + w) + f_0$ plays the role of f_0 in (4.125). Notice that since $k = 0$ in Ω_1, the obtained $\boldsymbol{y}_0^*$ satisfies the relation $\operatorname{div} \boldsymbol{y}_0^* + \Pi_{L_h}(f_0) = 0$ in Ω_1. Therefore, if f_0 is a piecewise constant function in Ω_1, then $\Pi_{L_h}(f_0) = f_0$ and $\boldsymbol{y}_0^*$ is exactly equilibrated in Ω_1. This is the case of the PBE where $f_0 = 0$ in Ω for both 2-term and 3-term splittings.

4.3 Poisson-Boltzmann equation

First we recall that a function ϕ is called a weak solution of the general Poisson-Boltzmann problem (3.35) (see Definition 3.12) if $\phi \in \bigcap_{p < \frac{d}{d-1}} W_g^{1,p}(\Omega)$ is such that $b(x, \phi)v \in L^1(\Omega)$ for all $v \in \bigcup_{q > d} W_0^{1,q}(\Omega)$ and

$$\int_\Omega \epsilon \nabla \phi \cdot \nabla v dx + \int_\Omega b(x, \phi) v dx = \langle \mathcal{F}, v \rangle, \ \forall v \in \bigcup_{q>d} W_0^{1,q}(\Omega). \tag{3.40}$$

In this section we will consider only the special case of a symmetric 1-1 ionic solution where the nonlinearity is given by $b(x, s) = \overline{k}^2 \sinh(s)$. We note that the general case is treated in a similar way, the only difference being the expressions for F, F^*, D_F in the main error identity (4.51). In the general case, the nonlinearity has the form

$$b(x, s) := \chi_{\Omega_{ions}}(x) \frac{4\pi e_0^2}{k_B T} \sum_{j=1}^{N_{ions}} -M_j \xi_j \mathrm{e}^{-\xi_j s}, \ \forall x \in \Omega, \ \forall s \in \mathbb{R}$$

and hence it can be represented in the form $b(x, s) = k^2(x) b(s)$ with $k^2(x) = \chi_{\Omega_{ions}}(x)$ and $b(s) = \frac{4\pi e_0^2}{k_B T} \sum_{j=1}^{N_{ions}} -M_j \xi_j \mathrm{e}^{-\xi_j s}$. Then F, F^*, and D_F can be computed by using the formulas (4.68), (4.69), and (4.70) in Section 4.1.3.

4.3.1 2-term splitting

With the 2-term splitting the dimensionless potential ϕ is decomposed into G and u, i.e, $\phi = G + u$, where G is given by (3.11). As we have shown in Section 3.3.1 and Section 3.3.3, one can define a particular solution of (3.40) by considering the problem

$$\begin{aligned} &\text{Find } u \in H^1_{g-G}(\Omega) \text{ such that } b(x, u+G)v \in L^1(\Omega) \text{ for all } v \in H^1_0(\Omega) \text{ and} \\ &a(u,v) + \int_\Omega b(x, u+G)v dx = \langle \mathcal{G}_2, v\rangle \text{ for all } v \in H^1_0(\Omega), \end{aligned} \tag{3.50}$$

where

$$\langle \mathcal{G}_2, v\rangle = \int_\Omega \boldsymbol{f}_{\mathcal{G}_2} \cdot \nabla v = \int_{\Omega_s} (\epsilon_m - \epsilon_s)\nabla G \cdot \nabla v dx$$

and $\boldsymbol{f}_{\mathcal{G}_2} := \chi_{\Omega_s}(\epsilon_m - \epsilon_s)\nabla G$. From Theorem 3.13 we know that this problem possesses a unique solution in $H^1(\Omega) \cap L^\infty(\Omega)$.

Additional splitting of u into u^L and u^N

First, we analyze the case where the regular component u is split into u^L and u^N, i.e., $u = u^L + u^N$, where u^L satisfies the nonhomogeneous linear interface problem (3.46) and u^N satisfies the homogeneous nonlinear interface problem (3.47) which take the particular form

$$\begin{aligned} &\text{Find } u^L \in H^1_{g-G}(\Omega) \text{ such that} \\ &a(u^L, v) = \int_{\Omega_s} (\epsilon_m - \epsilon_s)\nabla G \cdot \nabla v dx \text{ for all } v \in H^1_0(\Omega) \end{aligned} \tag{4.132}$$

and

$$\begin{aligned} &\text{Find } u^N \in H^1_0(\Omega) \text{ such that } b(x, u^N + u^L + G)v \in L^1(\Omega) \text{ for all } v \in H^1_0(\Omega) \text{ and} \\ &a(u^N, v) + \int_\Omega b(x, u^N + u^L + G)v dx = 0 \text{ for all } v \in H^1_0(\Omega). \end{aligned} \tag{4.133}$$

Notice that the solution of (4.133) depends on the solution of (4.132) and recall that u^L, $u^N \in L^\infty(\Omega)$ (see Theorem 3.19 and Theorem 3.27). Let $u^L_{h_L}$ denote an approximation of u^L that is computed by some conforming FEM based on the weak formulation (4.132) on some mesh $\mathscr{T}_{h_L}$ for which we use a subscript h_L to distinguish the finite element functions associated with it. Such an approximation could be computed by solving the homogenized version of (4.132) with some function $u_{g-G} \in H^1(\Omega) \cap L^\infty(\Omega)$ and $\nabla(u_{g-G}) \in L^s(\Omega)$, $s > d$ such that $\gamma_2(u_{g-G}) = g - G$ on $\partial\Omega$ (see, e.g., the considerations in Section 3.2.3). The exact solution of (4.132) with $u^L_{h_L}$ instead of u^L is denoted by $\tilde{u}^N$ and satisfies the problem

$$\begin{aligned} &\text{Find } \tilde{u}^N \in H^1_0(\Omega) \text{ such that } b(x, \tilde{u}^N + u^L_{h_L} + G)v \in L^1(\Omega) \text{ for all } v \in H^1_0(\Omega) \text{ and} \\ &a(\tilde{u}^N, v) + \int_\Omega b(x, \tilde{u}^N + u^L_{h_L} + G)v dx = 0 \text{ for all } v \in H^1_0(\Omega). \end{aligned} \tag{4.134}$$

Since $u^L_{h_L} \in L^\infty(\Omega)$, from the considerations in Section 3.3.3 it follows that there is a unique solution $\tilde{u}^N \in H^1_0(\Omega) \cap L^\infty(\Omega)$ of (4.134). Further, let $\tilde{u}^N_{h_N}$ denote an approximation of $\tilde{u}^N$ computed by some conforming FEM based on the weak formulation (4.134) on some mesh $\mathscr{T}_{h_N}$ which might be different from $\mathscr{T}_{h_L}$. Then the function $u_h := u^L_{h_L} + \tilde{u}^N_{h_N} \in H^1_{g-G}(\Omega)$ is a conforming approximation of the solution u of (3.50), where the subscript h signifies that u_h is a discrete approximation. Our goal is to estimate the error $|||\nabla(u_h - u)|||$. We have

$$\begin{aligned} |||\nabla(u_h - u)||| &= |||\nabla(u^L_{h_L} + \tilde{u}^N_{h_N} - u^L - u^N)||| \\ &\leq |||\nabla(u^L_{h_L} - u_L)||| + |||\nabla(\tilde{u}^N_{h_N} - u^N)|||. \end{aligned} \tag{4.135}$$

For the second term on the RHS we have that

$$|||\nabla(\tilde{u}^N_{h_N} - u^N)||| \leq |||\nabla(\tilde{u}^N_{h_N} - \tilde{u}^N)||| + |||\nabla(\tilde{u}^N - u^N)|||. \tag{4.136}$$

Now, the first term on the RHS in (4.136) we estimate by the functional a posteriori error estimate (4.51) where $\boldsymbol{A} = \epsilon I$, $k = \overline{k}$, $w = u^L_{h_L} + G \in L^\infty(\Omega_{ions})$, $f_0 = 0$, and $\boldsymbol{f} = 0$. For any $\tilde{\boldsymbol{y}}^*_N \in H(\mathrm{div}; \Omega)$ with $\mathrm{div}\, \tilde{\boldsymbol{y}}^*_N = 0$ in $\Omega_m \cup \Omega_{IEL}$ we have

$$|||\nabla(\tilde{u}^N_{h_N} - \tilde{u}^N)||| \leq \sqrt{2} M_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N), \tag{4.137}$$

where the fully computable majorant $M^2_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N)$ is given by

$$2M^2_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N) = |||\epsilon \nabla \tilde{u}^N_{h_N} - \tilde{\boldsymbol{y}}^*_N|||^2_* + 2D_{F,N}(\tilde{u}^N_{h_N}, -\Lambda^* \tilde{\boldsymbol{y}}^*_N), \tag{4.138}$$

and

$$\begin{aligned} D_{F,N}(\tilde{u}^N_{h_N}, -\Lambda^* \tilde{\boldsymbol{y}}^*_N) = \int\limits_{\Omega_{ions}} \overline{k}^2 \Bigg[&\cosh(\tilde{u}^N_{h_N} + u^L_{h_L} + G) + \left(\frac{\mathrm{div}\, \tilde{\boldsymbol{y}}^*_N}{\overline{k}^2}\right) \mathrm{arsinh}\left(\frac{\mathrm{div}\, \tilde{\boldsymbol{y}}^*_N}{\overline{k}^2}\right) \\ &- \cosh\left(\mathrm{arsinh}\left(\frac{\mathrm{div}\, \tilde{\boldsymbol{y}}^*_N}{\overline{k}^2}\right)\right) - \left(\frac{\mathrm{div}\, \tilde{\boldsymbol{y}}^*_N}{\overline{k}^2}\right)(\tilde{u}^N_{h_N} + u^L_{h_L} + G) \Bigg] dx. \end{aligned} \tag{4.139}$$

The second term in (4.136) we estimate as follows:

$$\left| \begin{array}{l} a(u^N, v) + (b(x, u^N + u^L + G), v) = 0, \; \forall v \in H^1_0(\Omega) \\ a(\tilde{u}^N, v) + (b(x, \tilde{u}^N + u^L_{h_L} + G), v) = 0, \; \forall v \in H^1_0(\Omega) \end{array} \right. \tag{4.140}$$

Subtracting the second from the first equation above, we get

$$a\left(u^N - \tilde{u}^N, v\right) = \left(b(x, \tilde{u}^N + u^L_{h_L} + G) - b(x, u^N + u^L + G), v\right), \; \forall v \in H^1_0(\Omega). \tag{4.141}$$

Now by taking $v := u^N + u^L - \tilde{u}^N - u^L_{h_L} \in H^1_0(\Omega)$ we obtain

$$\begin{aligned} &a\left(u^N - \tilde{u}^N, u^N + u^L - \tilde{u}^N - u^L_{h_L}\right) \\ &= \left(b(x, \tilde{u}^N + u^L_{h_L} + G) - b(x, u^N + u^L + G), u^N + u^L - \tilde{u}^N - u^L_{h_L}\right) \leq 0, \end{aligned} \tag{4.142}$$

where we have used the monotonicity of the nonlinearity: $(b(x,w)-b(x,z),w-z) \geq 0$, $\forall w,z \in H^1(\Omega)\cap L^\infty(\Omega)$. Using the boundedness of the bilinear form $a(\cdot,\cdot)$ we get

$$\begin{aligned}\left|\!\left|\!\left|\nabla(u^N-\tilde{u}^N)\right|\!\right|\!\right|^2 &= a\left(u^N-\tilde{u}^N, u^N-\tilde{u}^N\right)\\ &= a\left(u^N-\tilde{u}^N, u^N+u^L-\tilde{u}^N-u^L_{h_L}\right)+a\left(u^N-\tilde{u}^N, u^L_{h_L}-u^L\right)\\ &\leq 0+\left|\!\left|\!\left|\nabla(u^N-\tilde{u}^N)\right|\!\right|\!\right|\left|\!\left|\!\left|\nabla(u^L_{h_L}-u^L)\right|\!\right|\!\right|.\end{aligned}$$

Thus,

$$\left|\!\left|\!\left|\nabla(u^N-\tilde{u}^N)\right|\!\right|\!\right| \leq \left|\!\left|\!\left|\nabla(u^L_{h_L}-u^L)\right|\!\right|\!\right|. \tag{4.143}$$

Combining the estimates (4.135), (4.136), (4.137), and (4.143), for any $\tilde{\boldsymbol{y}}^*_N \in H(\mathrm{div};\Omega)$ with $\mathrm{div}\,\tilde{\boldsymbol{y}}^*_N = 0$ in $\Omega_m \cup \Omega_{IEL}$ we obtain the estimate

$$\left|\!\left|\!\left|\nabla(u_h-u)\right|\!\right|\!\right| \leq 2\left|\!\left|\!\left|\nabla(u^L_{h_L}-u_L)\right|\!\right|\!\right| + \sqrt{2}M_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N). \tag{4.144}$$

It is left to estimate the term $\left|\!\left|\!\left|\nabla(u^L_{h_L}-u_L)\right|\!\right|\!\right|$. Let us denote $\boldsymbol{p}^*_L := \epsilon\nabla u^L$. Then from (4.132) it is clear that $\boldsymbol{p}^*_L \in \left[L^2(\Omega)\right]^d$ has the form $\boldsymbol{p}^*_L = \boldsymbol{f}_{\mathcal{G}_2} + \boldsymbol{p}^*_{L,0}$ for some $\boldsymbol{p}^*_{L,0} \in H(\mathrm{div};\Omega)$ with $\mathrm{div}\,\boldsymbol{p}^*_{L,0} = 0$ in Ω. Now, for any $\psi \in H^1_0(\Omega)$, any approximation $v \in H^1_{g-G}(\Omega)$ of u^L, and any $\boldsymbol{y}^*_L \in \left[L^2(\Omega)\right]^d$ of the form $\boldsymbol{y}^*_L = \boldsymbol{f}_{\mathcal{G}_2} + \boldsymbol{y}^*_{L,0}$ with $\boldsymbol{y}^*_{L,0} \in H(\mathrm{div};\Omega)$ we obtain

$$\begin{aligned}(\epsilon\nabla(u^L-v),\nabla\psi) &= \int_\Omega \boldsymbol{f}_{\mathcal{G}_2}\cdot\nabla\psi dx - (\epsilon\nabla v,\nabla\psi) + (\boldsymbol{y}^*_L,\nabla\psi) - (\boldsymbol{f}_{\mathcal{G}_2}+\boldsymbol{y}^*_{L,0},\nabla\psi)\\ &= \left(\mathrm{div}\,\boldsymbol{y}^*_{L,0},\psi\right) - \left(\epsilon\nabla v - \boldsymbol{y}^*_L,\nabla\psi\right).\end{aligned}$$

Now, taking $\psi = u^L - v \in H^1_0(\Omega)$ we have that

$$\left|\!\left|\!\left|\nabla(u^L-v)\right|\!\right|\!\right|^2 \leq \frac{C_{P\Omega}}{\sqrt{\epsilon_{\min}}}\|\mathrm{div}\,\boldsymbol{y}^*_{L,0}\|_{L^2(\Omega)}\left|\!\left|\!\left|\nabla(u^L-v)\right|\!\right|\!\right| + \left|\!\left|\!\left|\epsilon\nabla v-\boldsymbol{y}^*_L\right|\!\right|\!\right|_*\left|\!\left|\!\left|\nabla(u^L-v)\right|\!\right|\!\right|$$

and thus by dividing by $\left|\!\left|\!\left|\nabla(u^L-v)\right|\!\right|\!\right|$ we obtain

$$\left|\!\left|\!\left|\nabla(u^L-v)\right|\!\right|\!\right| \leq \frac{C_{P\Omega}}{\sqrt{\epsilon_{\min}}}\|\mathrm{div}\,\boldsymbol{y}^*_{L,0}\|_{L^2(\Omega)} + \left|\!\left|\!\left|\epsilon\nabla v-\boldsymbol{y}^*_L\right|\!\right|\!\right|_* =: M_{\oplus,L}\left(v,\boldsymbol{y}^*_L\right), \tag{4.145}$$

where $C_{P,\Omega}$ is Poincaré's constant in the inequality $\|v\|_{L^2(\Omega)} \leq C_{P,\Omega}\|\nabla v\|_{L^2(\Omega)}$, $\forall v \in H^1_0(\Omega)$.

Remark 4.17
Notice that we have only used the implication that $u^L_{h_L}$ and $\tilde{u}^N_{h_N}$ are in $L^\infty(\Omega)$ and have not used anywhere the fact that $u^L_{h_L}$ and $\tilde{u}^N_{h_N}$ satisfy a Galerkin formulation. Thus, all considerations above are valid for any conforming approximations of u^L and u^N that are also in $L^\infty(\Omega)$.

Finally, for the 2-term splitting, by using (4.144) and (4.145) we arrive at the following proposition.

Proposition 4.18
Let $u^L_{h_L} \in H^1_{g-G}(\Omega)$ be an approximation of u^L and let $\tilde{u}^N_{h_N} \in H^1_0(\Omega)$ be an approximation of the solution $\tilde{u}^N$ of (4.134). Then $u_h := u^L_{h_L} + \tilde{u}^N_{h_N} \in H^1_{g-G}(\Omega)$ is a conforming approximation of the exact solution u of (3.50). Moreover, for any $\boldsymbol{y}^*_L$ of the form $\boldsymbol{y}^*_L = \boldsymbol{f}_{\mathcal{G}_2} + \boldsymbol{y}^*_{L,0}$ with $\boldsymbol{y}^*_{L,0} \in H(\mathrm{div};\Omega)$ and for any $\tilde{\boldsymbol{y}}^*_N \in H(\mathrm{div};\Omega)$ with $\mathrm{div}\,\tilde{\boldsymbol{y}}^*_N = 0$ in $\Omega_m \cup \Omega_{IEL}$ the following guaranteed estimate holds

$$|||\nabla(u_h - u)||| \leq 2M_{\oplus,L}\left(u^L_{h_L}, \boldsymbol{y}^*_L\right) + \sqrt{2}M_{\oplus,N}\left(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N\right), \tag{4.146}$$

where $M_{\oplus,L}\left(u^L_{h_L}, \boldsymbol{y}^*_L\right)$ and $M_{\oplus,N}\left(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N\right)$ are defined by (4.145) and (4.138), respectively.

Note that we can also estimate the H^1-seminorm and the full H^1-norm of the difference $u_h - u$ by introducing Poincaré's constant and the minimum and maximum values of the dielectric coefficient ϵ.

Remark 4.19 (Choosing $\boldsymbol{y}^*_{L,0}$)
In practice, to find a sharp bound on the error $\left|\left|\left|\nabla(u^L_{h_L} - u^L)\right|\right|\right|$ we can do a minimization of the majorant $M_{\oplus,L}(u^L_{h_L}, \boldsymbol{y}^*_L)$ in $\boldsymbol{y}^*_{L,0}$ over a finite dimensional subspace of $H(\mathrm{div};\Omega)$. However, it is more convenient to minimize the squared majorant $M^2_{\oplus,L}(u^L_{h_L}, \boldsymbol{y}^*_L;\alpha)$ simultaneously over $\alpha \in \mathbb{R}_{>0} := \{x \in \mathbb{R} : x > 0\}$ and $\boldsymbol{y}^*_{L,0}$ in a finite dimensional subspace of $H(\mathrm{div};\Omega)$, cf. [156]:

$$\begin{aligned}
|||\nabla(u^L_{h_L} - u^L)|||^2 &\leq M^2_{\oplus,L}(u^L_{h_L}, \boldsymbol{y}^*_L) \\
&\leq (1+\alpha)\frac{C^2_{P\Omega}}{\epsilon_{\min}}\|\mathrm{div}\,\boldsymbol{y}^*_{L,0}\|^2_{L^2(\Omega)} + \left(1 + \frac{1}{\alpha}\right)\left|\left|\left|\epsilon\nabla u^L_{h_L} - \boldsymbol{f}_{\mathcal{G}_2} - \boldsymbol{y}^*_{L,0}\right|\right|\right|^2_* \\
&:= M^2_{\oplus,L}(u^L_{h_L}, \boldsymbol{y}^*_L;\alpha).
\end{aligned} \tag{4.147}$$

Another computationally favorable approach that we use in applications and that gives practically the same estimate and error indicator as the minimization of the majorant is to apply the patchwise flux reconstruction described in Section 4.2 on p. 137. In this case it is necessary that $u^L_{h_L}$ is obtained by a Galerkin FEM. More precisely, $u^L_{h_L} = u^L_{g-G} + u^L_{0,h_L}$ where $u^L_{g-G} \in H^1(\Omega) \cap L^\infty(\Omega)$ with $\nabla u^L_{g-G} \in [L^s(\Omega)]^d$ for some $s > d$ is such that $\gamma_2(u^L_{g-G}) = g - G$ on $\partial\Omega$ and $u^L_{0,h_L} \in V^1_{h_L}$ satisfies the Galerkin problem

$$\int_\Omega \epsilon\nabla\left(u^L_{g-G} + u^L_{0,h_L}\right)\cdot\nabla v dx = \int_\Omega \boldsymbol{f}_{\mathcal{G}_2}\cdot\nabla v dx,\ \forall v \in V^1_{h_L}, \tag{4.148}$$

where $V^1_{h_L} \subset H^1_0(\Omega)$ is the space of continuous piecewise linear functions over a mesh $\mathscr{T}_{h_L}$. Then $\boldsymbol{q} := \epsilon\nabla u^L_{h_L} - \boldsymbol{f}_{\mathcal{G}_2}$ satisfies

$$\int_\Omega \boldsymbol{q}\cdot\nabla v dx = 0,\ \forall v \in V^1_{h_L}$$

and $\Pi_{L_{h_L}}(\boldsymbol{q})$ *satisfies*

$$\int_\Omega \Pi_{L_{h_L}}(\boldsymbol{q}) \cdot \nabla v dx = 0, \, \forall v \in V^1_{h_L}.$$

Finally, $\boldsymbol{y}^*_{L,0} \in RT_0$ *is defined as the patchwise flux reconstruction of* $\Pi_{L_{h_L}}(\boldsymbol{q})$.

Remark 4.20 (Choosing $\tilde{\boldsymbol{y}}^*_N$)
To choose the best possible $\tilde{\boldsymbol{y}}^*_N$ *in some finite dimensional subspace of* $H(\mathrm{div};\Omega)$ *one can minimize the majorant* $M^2_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N)$ *in* $\tilde{\boldsymbol{y}}^*_N$. *In applications, we again use the patchwise flux reconstruction from Section 4.2. More precisely, let* $\tilde{u}^N_{h_N}$ *be a Galerkin approximation of* $\tilde{u}^N$ *in the space* $V^1_{h_N}$ *of continuous piecewise linear functions over a mesh* $\mathscr{T}_{h_N}$, *i.e.,*

$$\int_\Omega \epsilon \nabla \tilde{u}^N_{h_N} \cdot \nabla v dx + \int_\Omega \overline{k}^2 \sinh(\tilde{u}^N_{h_N} + u^L_{h_L} + G) v dx, \, \forall v \in V^1_{h_N}.$$

Then, if ϵ *is piecewise constant with respect to* $\mathscr{T}_{h_N}$, *we apply the flux reconstruction to the piecewise constant numerical flux* $\epsilon \nabla \tilde{u}^N_{h_N}$ *with the right-hand side*

$$f_0 := -\overline{k}^2 \sinh(\tilde{u}^N_{h_N} + u^L_{h_L} + G).$$

If ϵ *is not piecewise constant, we apply the flux reconstruction to the piecewise constant flux* $\Pi_{L_{h_N}}(\epsilon \nabla \tilde{u}^N_{h_N})$.

Remark 4.21
Notice that the near best approximation result in Propostion 4.12 holds for $\tilde{u}^N_{h_N}$ *if* $\tilde{u}^N_{h_N}$ *is chosen as the Galerkin approximation of* $\tilde{u}^N$ *in a closed subspace* V_{h_N} *of* $H^1_0(\Omega) \cap L^\infty(\Omega)$.

No additional splitting of u into u^L and u^N

One can solve (3.50) directly without the additional splitting into u^L and u^N. In this case, we can apply directly the a posteriori error estimates from Section 4.1.4 where $\overline{g} = u_{g-G} \in H^1(\Omega) \cap L^\infty(\Omega)$ with $\nabla(u_{g-G}) \in [L^s(\Omega)]^d$, $u = u_{g-G} + u_0$, $\Omega_1 = \Omega_m \cup \Omega_{IEL}$, $\Omega_2 = \Omega_{ions}$, $\boldsymbol{A} = \epsilon I$, $k = \overline{k}$, $w = G$, $f_0 = 0$, $\boldsymbol{f} = \boldsymbol{f}_{\mathcal{G}_2}$. Let $u_h \in H^1_{g-G}(\Omega)$ be a conforming approximation of u and let, as usual, $\boldsymbol{p}^* = \epsilon \nabla u$. Then for any $\boldsymbol{y}^*$ of the form $\boldsymbol{y}^* = \boldsymbol{f}_{\mathcal{G}_2} + \boldsymbol{y}^*_0$ with $\boldsymbol{y}^*_0 \in H(\mathrm{div};\Omega)$ and $\mathrm{div}\, \boldsymbol{y}^*_0 = 0$ in $\Omega_m \cup \Omega_{IEL}$ the error identity (4.115) takes the form

$$\frac{1}{2}|||\nabla(u_h - u)|||^2 + \frac{1}{2}|||\boldsymbol{y}^* - \boldsymbol{p}^*|||^2_* + D_F(u_h, -\Lambda^* \boldsymbol{p}^*) + D_F(u, -\Lambda^* \boldsymbol{y}^*) = M^2_\oplus(u_h, \boldsymbol{y}^*), \quad (4.149)$$

where

$$2M^2_\oplus(u_h, \boldsymbol{y}^*) = |||\epsilon \nabla u_h - \boldsymbol{y}^*|||^2_* + 2D_F(u_h, -\Lambda^* \boldsymbol{y}^*) \quad (4.150)$$

and $D_F(u_h, -\Lambda^* \boldsymbol{p}^*)$, $D_F(u, -\Lambda^* \boldsymbol{y}^*)$, and $D_F(u_h, -\Lambda^* \boldsymbol{y}^*)$ are given by (4.111), (4.112), and (4.106), respectively. Also, we have the near best approximation result in Proposition 4.16 provided that $u_h = u_{g-G} + u_{0,h}$ where $u_{0,h}$ is the solution of the Galerkin problem

$$\begin{aligned} &\text{Find } u_{0,h} \in V_h \text{ such that} \\ &a(u_{g-G} + u_{0,h}, v) + \int_\Omega \overline{k}^2 \sinh(u_{0,h} + u_{g-G} + G) v dx = \int_\Omega \boldsymbol{f}_{\mathcal{G}_2} \cdot \nabla v dx, \, \forall v \in V_h \end{aligned} \tag{4.151}$$

for some closed subspace $V_h \subset L^\infty(\Omega)$ of $H_0^1(\Omega)$.

Remark 4.22 (Choosing $\boldsymbol{y}_0^*$)
To define a good $\boldsymbol{y}_0^*$ we proceed similarly to Remark 4.20 and we only comment on the patchwise flux reconstruction approach. Let $u_{0,h}$ be the Galerkin solution of (4.151) *in the space $V_h^1 \subset H_0^1(\Omega)$ of continuous piecewise linear functions over a mesh $\mathscr{T}_h$ and let $u_h := u_{g-G} + u_{0,h}$. Then $\boldsymbol{q} := \epsilon \nabla u_h - \boldsymbol{f}_{\mathcal{G}_2}$ satisfies*

$$\int_\Omega \boldsymbol{q} \cdot \nabla v dx = \int_\Omega -\overline{k}^2 \sinh(u_h + G) v dx, \, \forall v \in V_h^1.$$

Thus $\Pi_{L_h}(\boldsymbol{q})$, where Π_{L_h} is the L^2-projection operator onto the space of piecewise constant functions over $\mathscr{T}_h$, also satisfies the above equation and we define $\boldsymbol{y}_0^*$ as the patchwise flux reconstruction from Section 4.2 of $\Pi_{L_h}(\boldsymbol{q})$. Therefore, $\operatorname{div} \boldsymbol{y}_0^* + \Pi_{L_h}(-\overline{k}^2 \sinh(u_h + G)) = 0$. *Since $\overline{k} = 0$ in $\Omega_m \cup \Omega_{IEL}$ we see that* $\operatorname{div} \boldsymbol{y}_0^*$ *is exactly equilibrated in $\Omega_m \cup \Omega_{IEL}$ and the reliability of the majorant $M_\oplus^2(u_h, \boldsymbol{y}^*)$ is guaranteed.*

4.3.2 3-term splitting

With the 3-term splitting the dimensionless potential ϕ is decomposed into G, u^H, and u, i.e, $\phi = G + u^H + u$, where G is given by (3.11) and $u^H = -G$ in Ω_s and satisfies the following problem in Ω_m:

$$u^H \in H^1_{-G}(\Omega_m) \quad \text{and} \quad \int_{\Omega_m} \nabla u^H \cdot \nabla v dx = 0 \text{ for all } v \in H_0^1(\Omega_m). \tag{3.29}$$

As we have shown in Section 3.3.2 and Section 3.3.3, one can define a particular solution of (3.40) by considering the problem

$$\begin{aligned} &\text{Find } u \in H_g^1(\Omega) \text{ such that } b(x,u)v \in L^1(\Omega) \text{ for all } v \in H_0^1(\Omega) \text{ and} \\ &\int_\Omega \epsilon \nabla u \cdot \nabla v dx + \int_\Omega b(x,u) v dx = \langle \mathcal{G}_3, v \rangle \text{ for all } v \in H_0^1(\Omega), \end{aligned} \tag{3.52}$$

where

$$\langle \mathcal{G}_3, v \rangle = \int_\Omega \boldsymbol{f}_{\mathcal{G}_3} \cdot \nabla v = -\int_{\Omega_m} \epsilon_m \nabla u^H \cdot \nabla v dx + \int_{\Omega_s} \epsilon_m \nabla G \cdot \nabla v dx$$

with

$$\boldsymbol{f}_{\mathcal{G}_3} := \chi_{\Omega_m}\epsilon_m\nabla u^H + \chi_{\Omega_s}\epsilon_m\nabla G.$$

From Theorem 3.15 we know that the problem (3.52) possesses a unique solution in $H^1(\Omega)\cap L^\infty(\Omega)$. Notice that the solution of (3.52) depends on the solution of (3.29). For this reason we will need first to derive a posteriori error estimates for the approximate solution of (3.29).

Harmonic component u^H

Let $\tilde{u}^H \in H^1_{-G}(\Omega_m)$ denote a conforming approximation of u^H on Ω_m and let $T(\nabla\tilde{u}^H) \in H(\mathrm{div};\Omega_m)$ with T being a regularization operator that maps the numerical flux $\nabla\tilde{u}^H$ into $H(\mathrm{div};\Omega)$. If $\nabla\tilde{u}^H$ is in $H(\mathrm{div};\Omega_m)$, then T can be chosen to be the identity mapping. To obtain fully reliable error bounds for approximate solutions of problem (3.52) with the 3-term splitting, we need first to derive a posteriori estimates for the quantities

$$\|\nabla(\tilde{u}^H - u^H)\|_{L^2(\Omega_m)} \quad \text{and} \quad \|T\left(\nabla\tilde{u}^H\right) - \nabla u^H\|_{L^2(\Omega_m)}.$$

For the first quantity, similarly to (4.145), we have (see also [156])

$$\|\nabla\left(\tilde{u}^H - u^H\right)\|_{L^2(\Omega_m)} \leq C_{P\Omega_m}\|\mathrm{div}\left(T\left(\nabla\tilde{u}^H\right)\right)\|_{L^2(\Omega_m)} + \|\nabla\tilde{u}^H - T(\nabla\tilde{u}^H)\|_{L^2(\Omega_m)}, \quad (4.152)$$

where C_{P,Ω_m} is Poincaré's constant in the inequality $\|v\|_{L^2(\Omega_m)} \leq C_{P,\Omega_m}\|\nabla v\|_{L^2(\Omega_m)}$, $\forall v \in H^1_0(\Omega_m)$. For the second quantity, we proceed as follows:

$$\begin{aligned}\|\nabla\tilde{u}^H - T\left(\nabla\tilde{u}^H\right)\|^2_{L^2(\Omega_m)} &= \|\nabla\left(\tilde{u}^H - u^H\right)\|^2_{L^2(\Omega_m)} \\ &\quad + \|\nabla u^H - T\left(\nabla\tilde{u}^H\right)\|^2_{L^2(\Omega_m)} \\ &\quad + 2\int_{\Omega_m}\nabla\left(\tilde{u}^H - u^H\right)\cdot\left(\nabla u^H - T\left(\nabla\tilde{u}^H\right)\right)dx.\end{aligned} \quad (4.153)$$

Thus, using the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned}&\|\nabla\left(\tilde{u}^H - u^H\right)\|^2_{L^2(\Omega_m)} + \|\nabla u^H - T\left(\nabla\tilde{u}^H\right)\|^2_{L^2(\Omega_m)} \\ &\leq \|\nabla\tilde{u}^H - T\left(\nabla\tilde{u}^H\right)\|^2_{L^2(\Omega_m)} + 2\|\mathrm{div}\left(T\left(\nabla\tilde{u}^H\right)\right)\|_{L^2(\Omega_m)}C_{P\Omega_m}\|\nabla\left(\tilde{u}^H - u^H\right)\|_{L^2(\Omega_m)} \\ &\leq \|\nabla\tilde{u}^H - T\left(\nabla\tilde{u}^H\right)\|^2_{L^2(\Omega_m)} + \|\nabla\left(\tilde{u}^H - u^H\right)\|^2_{L^2(\Omega_m)} + C^2_{P\Omega_m}\|\mathrm{div}\left(T\left(\nabla\tilde{u}^H\right)\right)\|^2_{L^2(\Omega_m)}.\end{aligned}$$

Finally,

$$\begin{aligned}&\|\nabla u^H - T\left(\nabla\tilde{u}^H\right)\|_{L^2(\Omega_m)} \\ &\leq \left(\|\nabla\tilde{u}^H - T\left(\nabla\tilde{u}^H\right)\|^2_{L^2(\Omega_m)} + C^2_{P\Omega_m}\|\mathrm{div}\left(T\left(\nabla\tilde{u}^H\right)\right)\|^2_{L^2(\Omega_m)}\right)^{\frac{1}{2}} \\ &=: M_{\oplus,H}\left(\tilde{u}^H, T\left(\nabla\tilde{u}^H\right)\right)\end{aligned} \quad (4.154)$$

If $T\left(\nabla\tilde{u}^H\right)$ is additionally equilibrated (for example using the patchwise equilibration technique from Section 4.2) then both estimates (4.152) and (4.154) follow from (4.153) (Prager-Synge estimate).

Additional splitting of u into u^L and u^N

We start with the case where the regular component u is further split into u^L and u^N, i.e., $u = u^L + u^N$, where u^L satisfies the nonhomogeneous linear interface problem (3.46) and u^N satisfies the homogeneous nonlinear interface problem (3.47) which take the particular form

$$\begin{aligned} &\text{Find } u^L \in H^1_g(\Omega) \text{ such that} \\ &a(u^L, v) = -\int_{\Omega_m} \epsilon_m \nabla u^H \cdot \nabla v dx + \int_{\Omega_s} \epsilon_m \nabla G \cdot \nabla v dx \ \text{ for all } v \in H^1_0(\Omega) \end{aligned} \tag{4.155}$$

and

$$\begin{aligned} &\text{Find } u^N \in H^1_0(\Omega) \text{ such that } b(x, u^N + u^L)v \in L^1(\Omega) \ \text{ for all } v \in H^1_0(\Omega) \text{ and} \\ &a(u^N, v) + \int_{\Omega} b(x, u^N + u^L) v dx = 0 \ \text{ for all } v \in H^1_0(\Omega). \end{aligned} \tag{4.156}$$

Notice that the solution of (4.156) depends on the solution of (4.155) which depends on the solution of (3.29). We also recall that u^L and u^N are in $L^\infty(\Omega)$ (see Theorem 3.19 and Theorem 3.27).

Let $\tilde{u}^L$ denote the exact solution of problem (4.155) with $T\left(\nabla \tilde{u}^H\right)$ instead of ∇u^H on the right-hand side, i.e.,

$$\begin{aligned} &\text{Find } \tilde{u}^L \in H^1_g(\Omega) \text{ such that} \\ &a(\tilde{u}^L, v) = -\int_{\Omega_m} \epsilon_m T\left(\nabla \tilde{u}^H\right) \cdot \nabla v dx + \int_{\Omega_s} \epsilon_m \nabla G \cdot \nabla v dx =: \langle \tilde{\mathcal{G}}_3, v \rangle \ \text{ for all } v \in H^1_0(\Omega), \end{aligned} \tag{4.157}$$

where $\langle \tilde{\mathcal{G}}_3, v \rangle$ can also be written as $\langle \tilde{\mathcal{G}}_3, v \rangle = \int_{\Omega} \boldsymbol{f}_{\tilde{\mathcal{G}}_3} \cdot \nabla v dx$ with

$$\boldsymbol{f}_{\tilde{\mathcal{G}}_3} := \chi_{\Omega_m} \epsilon_m T\left(\nabla \tilde{u}^H\right) + \chi_{\Omega_s} \epsilon_m \nabla G.$$

Here we note that $\tilde{u}^L$ will be in $L^\infty(\Omega)$ if $T\left(\nabla \tilde{u}^H\right) \in [L^s(\Omega)]^d$ for some $s > d$. This will be ensured if, for example, $\tilde{u}^H$ is a continuous piecewise linear finite element approximation and the operator T corresponds to the patchwise flux reconstruction procedure from Section 4.2. Additionally, let $\tilde{u}^L_{h_L}$ denote some conforming approximation of $\tilde{u}^L$ computed on some partition $\mathscr{T}_{h_L}$ of $\overline{\Omega}$ (actually $\tilde{u}^L_{h_L}$ can be any conforming approximation as long as it is in $L^\infty(\Omega)$). Next, let $\tilde{u}^N$ denote the exact solution of problem (4.156) but with $\tilde{u}^L_{h_L}$ instead of u^L in it, i.e.,

$$\begin{aligned} &\text{Find } \tilde{u}^N \in H^1_0(\Omega) \text{ such that } b(x, \tilde{u}^N + \tilde{u}^L_{h_L})v \in L^1(\Omega) \ \text{ for all } v \in H^1_0(\Omega) \text{ and} \\ &a(\tilde{u}^N, v) + \int_{\Omega} b(x, \tilde{u}^N + \tilde{u}^L_{h_L}) v dx = 0 \ \text{ for all } v \in H^1_0(\Omega). \end{aligned} \tag{4.158}$$

Finally, by $\tilde{u}^N_{h_N}$ we denote a conforming approximation of $\tilde{u}^N$ computed on some mesh $\mathscr{T}_{h_N}$ (again $\tilde{u}^N_{h_N}$ can be any conforming approximation as long as it is in $L^\infty(\Omega)$). With the above notation, $u_h := \tilde{u}^L_{h_L} + \tilde{u}^N_{h_N} \in H^1_g(\Omega)$ is a conforming approximation of the solution u of (3.52). Our goal is to estimate the error $|||\nabla(u_h - u)|||$. In other words, we estimate the effect of using an approximation of u^H in (4.155) to compute an approximation of u^L which is then used in equation (4.156) to compute an approximation of u^N on the quality of the total approximate regular part of the potential u_h. The arguments are similar to the ones we made in the case of the 2-term splitting.

We have

$$\begin{aligned} |||\nabla(u_h - u)||| &= |||\nabla(\tilde{u}^N_{h_N} + \tilde{u}^L_{h_L} - u^N - u^L)||| \\ &\leq |||\nabla(\tilde{u}^N_{h_N} - u^N)||| + |||\nabla(\tilde{u}^L_{h_L} - u^L)||| \end{aligned} \tag{4.159}$$

For the first term on the RHS in (4.159) we have

$$|||\nabla(\tilde{u}^N_{h_N} - u^N)||| \leq |||\nabla(\tilde{u}^N_{h_N} - \tilde{u}^N)||| + |||\nabla(\tilde{u}^N - u^N)|||. \tag{4.160}$$

The first term on the RHS in (4.160) we estimate by the functional a posteriori error estimate (4.51), where $\boldsymbol{A} = \epsilon I$, $k = \overline{k}$, $w = \tilde{u}^L_{h_L} \in L^\infty(\Omega_{ions})$, $f_0 = 0$, and $\boldsymbol{f} = 0$. For any $\tilde{\boldsymbol{y}}^*_N \in H(\mathrm{div};\Omega)$ with $\mathrm{div}\,\tilde{\boldsymbol{y}}^*_N = 0$ in $\Omega_m \cup \Omega_{IEL}$ we have

$$|||\nabla(\tilde{u}^N_{h_N} - \tilde{u}^N)||| \leq \sqrt{2} M_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N), \tag{4.161}$$

where the fully computable majorant $M^2_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N)$ is given by

$$2M^2_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N) = \left|||\epsilon\nabla\tilde{u}^N_{h_N} - \tilde{\boldsymbol{y}}^*_N|||^2_* + 2D_{F,N}(\tilde{u}^N_{h_N}, -\Lambda^*\tilde{\boldsymbol{y}}^*_N), \right. \tag{4.162}$$

and

$$\begin{aligned} D_{F,N}(\tilde{u}^N_{h_N}, -\Lambda^*\tilde{\boldsymbol{y}}^*_N) = \int\limits_{\Omega_{ions}} \overline{k}^2 \Bigg[&\cosh(\tilde{u}^N_{h_N} + \tilde{u}^L_{h_L}) + \left(\frac{\mathrm{div}\,\tilde{\boldsymbol{y}}^*_N}{\overline{k}^2}\right) \mathrm{arsinh}\left(\frac{\mathrm{div}\,\tilde{\boldsymbol{y}}^*_N}{\overline{k}^2}\right) \\ &- \cosh\left(\mathrm{arsinh}\left(\frac{\mathrm{div}\,\tilde{\boldsymbol{y}}^*_N}{\overline{k}^2}\right)\right) - \left(\frac{\mathrm{div}\,\tilde{\boldsymbol{y}}^*_N}{\overline{k}^2}\right)(\tilde{u}^N_{h_N} + \tilde{u}^L_{h_L}) \Bigg] dx. \end{aligned} \tag{4.163}$$

The second term in (4.160) we estimate as follows:

$$\left| \begin{array}{l} a\left(u^N, v\right) + (b(x, u^N + u^L), v) = 0,\ \forall v \in H^1_0(\Omega) \\ a\left(\tilde{u}^N, v\right) + (b(x, \tilde{u}^N + \tilde{u}^L_{h_L}), v) = 0,\ \forall v \in H^1_0(\Omega) \end{array} \right. \tag{4.164}$$

By subtracting the second from the first equation above, we obtain

$$a\left(u^N - \tilde{u}^N, v\right) = \left(b\left(x, \tilde{u}^N + \tilde{u}^L_{h_L}\right) - b\left(x, u^N + u^L\right), v\right),\ \forall v \in H^1_0(\Omega). \tag{4.165}$$

Now take $v := u^N + u^L - \tilde{u}^N - \tilde{u}^L_{h_L} \in H^1_0(\Omega)$ and we obtain

$$\begin{aligned} &a\left(u^N - \tilde{u}^N, u^N + u^L - \tilde{u}^N - \tilde{u}^L_{h_L}\right) \\ &= \left(b\left(x, \tilde{u}^N + \tilde{u}^L_{h_L}\right) - b\left(x, u^N + u^L\right), u^N + u^L - \tilde{u}^N - \tilde{u}^L_{h_L}\right) \leq 0, \end{aligned} \tag{4.166}$$

where we have used the monotonicity of the nonlinearity: $(b(x,w)-b(x,z),w-z)\geq 0$, $\forall w,z\in H^1(\Omega)\cap L^\infty(\Omega)$. Using the boundedness of the bilinear form $a(\cdot,\cdot)$ we obtain

$$\begin{aligned}\left|\!\left|\!\left|\nabla\left(u^N-\tilde{u}^N\right)\right|\!\right|\!\right|^2 &= a\left(u^N-\tilde{u}^N,u^N-\tilde{u}^N\right)\\ &= a\left(u^N-\tilde{u}^N,u^N+u^L-\tilde{u}^N-\tilde{u}^L_{h_L}\right)+a\left(u^N-\tilde{u}^N,\tilde{u}^L_{h_L}-u^L\right)\\ &\leq 0+\left|\!\left|\!\left|\nabla\left(u^N-\tilde{u}^N\right)\right|\!\right|\!\right|\left|\!\left|\!\left|\nabla\left(\tilde{u}^L_{h_L}-u^L\right)\right|\!\right|\!\right|.\end{aligned}$$

Thus,

$$\left|\!\left|\!\left|\nabla(u^N-\tilde{u}^N)\right|\!\right|\!\right|\leq\left|\!\left|\!\left|\nabla(\tilde{u}^L_{h_L}-u^L)\right|\!\right|\!\right|. \tag{4.167}$$

Next, we estimate the term $\left|\!\left|\!\left|\nabla(\tilde{u}^L_{h_L}-u^L)\right|\!\right|\!\right|$ that also appears in (4.159). By the triangle inequality, we have

$$\left|\!\left|\!\left|\nabla(\tilde{u}^L_{h_L}-u^L)\right|\!\right|\!\right|\leq\left|\!\left|\!\left|\nabla\left(\tilde{u}^L_{h_L}-\tilde{u}^L\right)\right|\!\right|\!\right|+\left|\!\left|\!\left|\nabla\left(\tilde{u}^L-u^L\right)\right|\!\right|\!\right|. \tag{4.168}$$

The second term in (4.168) is bounded as follows: subtract equation (4.155) from (4.157), take for a test function $v=\tilde{u}^L-u^L$ and use Cauchy-Schwartz inequality to obtain

$$\left|\!\left|\!\left|\nabla\left(\tilde{u}^L-u^L\right)\right|\!\right|\!\right|^2\leq\sqrt{\epsilon_m}\|T\left(\nabla\tilde{u}^H\right)-\nabla u^H\|_{L^2(\Omega_m)}\left|\!\left|\!\left|\nabla(\tilde{u}^L-u^L)\right|\!\right|\!\right|$$

Thus, we get

$$\left|\!\left|\!\left|\nabla\left(\tilde{u}^L-u^L\right)\right|\!\right|\!\right|\leq\sqrt{\epsilon_m}\|T\left(\nabla\tilde{u}^H\right)-\nabla u^H\|_{L^2(\Omega_m)}\leq\sqrt{\epsilon_m}M_{\oplus,H}\left(\tilde{u}^H,T\left(\nabla\tilde{u}^H\right)\right). \tag{4.169}$$

It is left to estimate the first term in (4.168). Let us denote $\tilde{\boldsymbol{p}}^*_L:=\epsilon\nabla\tilde{u}^L$. Then from (4.157) it follows that $\tilde{\boldsymbol{p}}^*_L\in\left[L^2(\Omega)\right]^d$ has the form $\tilde{\boldsymbol{p}}^*_L=\boldsymbol{f}_{\tilde{\mathcal{G}}_3}+\tilde{\boldsymbol{p}}^*_{L,0}$ for some $\tilde{\boldsymbol{p}}^*_{L,0}\in H(\mathrm{div};\Omega)$ with $\mathrm{div}\,\tilde{\boldsymbol{p}}^*_{L,0}=0$ in Ω. Similarly to the estimate (4.145) in the case of the 2-term splitting, for any $v\in H^1_g(\Omega)$ and any $\tilde{\boldsymbol{y}}^*_L\in\left[L^2(\Omega)\right]^d$ of the form $\tilde{\boldsymbol{y}}^*_L=\boldsymbol{f}_{\tilde{\mathcal{G}}_3}+\tilde{\boldsymbol{y}}^*_{L,0}$ with $\tilde{\boldsymbol{y}}^*_{L,0}\in H(\mathrm{div};\Omega)$ we obtain

$$\left|\!\left|\!\left|\nabla(\tilde{u}^L-v)\right|\!\right|\!\right|\leq\frac{C_{P\Omega}}{\sqrt{\epsilon_{\min}}}\|\mathrm{div}\,\tilde{\boldsymbol{y}}^*_{L,0}\|_{L^2(\Omega)}+\left|\!\left|\!\left|\epsilon\nabla v-\tilde{\boldsymbol{y}}^*_L\right|\!\right|\!\right|_*=:M_{\oplus,L}\left(v,\tilde{\boldsymbol{y}}^*_L\right). \tag{4.170}$$

Finally, if we want to compute an approximation u_h of u with a prescribed error tolerance δ, by combining (4.159), (4.160), (4.161), (4.167), (4.168), (4.169), and (4.170) we obtain

$$\begin{aligned}\left|\!\left|\!\left|\nabla\left(u_h-u\right)\right|\!\right|\!\right| &\leq \left|\!\left|\!\left|\nabla\left(\tilde{u}^N_{h_N}-\tilde{u}^N\right)\right|\!\right|\!\right|+\left|\!\left|\!\left|\nabla\left(\tilde{u}^L_{h_L}-u^L\right)\right|\!\right|\!\right|+\left|\!\left|\!\left|\nabla\left(\tilde{u}^L_{h_L}-u^L\right)\right|\!\right|\!\right|\\ &\leq \sqrt{2}M_{\oplus,N}\left(\tilde{u}^N_{h_N},\tilde{\boldsymbol{y}}^*_N\right)+2\left(\left|\!\left|\!\left|\nabla\left(\tilde{u}^L_{h_L}-\tilde{u}^L\right)\right|\!\right|\!\right|+\left|\!\left|\!\left|\nabla\left(\tilde{u}^L-u^L\right)\right|\!\right|\!\right|\right)\\ &\leq \sqrt{2}M_{\oplus,N}\left(\tilde{u}^N_{h_N},\tilde{\boldsymbol{y}}^*_N\right)+2\left(M_{\oplus,L}\left(\tilde{u}^L_{h_L},\tilde{\boldsymbol{y}}^*_L\right)+\sqrt{\epsilon_m}\|T\left(\nabla\tilde{u}^H\right)-\nabla u^H\|_{L^2(\Omega_m)}\right)\\ &\leq 2\sqrt{\epsilon_m}M_{\oplus,H}\left(\tilde{u}^H,T\left(\nabla\tilde{u}^H\right)\right)+2M_{\oplus,L}\left(\tilde{u}^L_{h_L},\tilde{\boldsymbol{y}}^*_L\right)+\sqrt{2}M_{\oplus,N}\left(\tilde{u}^N_{h_N},\tilde{\boldsymbol{y}}^*_N\right)\leq\delta.\end{aligned} \tag{4.171}$$

Proposition 4.23
Let $\tilde{u}^H$ be a conforming approximation of the solution u^H of (3.29) and let $T\left(\nabla\tilde{u}^H\right) \in H(\mathrm{div};\Omega_m)$ be a reconstruction of the numerical flux $\nabla\tilde{u}^H$. Next, let $\tilde{u}^L_{h_L} \in H^1_g(\Omega)$ be an approximation of the solution $\tilde{u}^L$ of (4.157) and let $\tilde{u}^N_{h_N} \in H^1_0(\Omega)$ be an approximation of the solution $\tilde{u}^N$ of (4.158). Then $u_h := \tilde{u}^L_{h_L} + \tilde{u}^N_{h_N} \in H^1_g(\Omega)$ is a conforming approximation of the exact solution u of (3.52). Moreover, for any $\tilde{\boldsymbol{y}}^*_L$ of the form $\tilde{\boldsymbol{y}}^*_L = \boldsymbol{f}_{\tilde{\mathcal{G}}_3} + \tilde{\boldsymbol{y}}^*_{L,0}$ with $\tilde{\boldsymbol{y}}^*_{L,0} \in H(\mathrm{div};\Omega)$ and for any $\tilde{\boldsymbol{y}}^*_N \in H(\mathrm{div};\Omega)$ with $\mathrm{div}\,\tilde{\boldsymbol{y}}^*_N = 0$ in $\Omega_m \cup \Omega_{IEL}$ the following guaranteed estimate holds

$$\left|\!\left|\!\left|\nabla\left(u_h - u\right)\right|\!\right|\!\right| \leq 2\sqrt{\epsilon_m} M_{\oplus,H}\left(\tilde{u}^H, T\left(\nabla\tilde{u}^H\right)\right) + 2M_{\oplus,L}\left(\tilde{u}^L_{h_L}, \tilde{\boldsymbol{y}}^*_L\right) + \sqrt{2} M_{\oplus,N}\left(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N\right), \tag{4.172}$$

where $M_{\oplus,L}\left(\tilde{u}^L_{h_L}, \tilde{\boldsymbol{y}}^*_L\right)$ and $M_{\oplus,N}\left(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N\right)$ are defined by (4.170) and (4.162), respectively.

Remark 4.24
Similarly to Remark 3.9, the functional $\tilde{\mathcal{G}}_3$ generates a jump condition on the normal component of the flux $\epsilon\nabla\tilde{u}^L$. When $\tilde{u}^L$ is smooth in each subdomain and $T(\nabla\tilde{u}^H)$ is regular enough, we have $\left[\epsilon\nabla\tilde{u}^L\cdot\boldsymbol{n}_\Gamma\right]_\Gamma = -\epsilon_m\left(T(\nabla\tilde{u}^H) + \nabla G\right)\cdot\boldsymbol{n}_\Gamma$, where $\boldsymbol{n}_\Gamma$ is the unit outward normal vector with respect to Ω_m. Indeed, by applying the divergence theorem (Theorem 2.20) and using the fact that ϵ_m is constant and G is harmonic in a neighborhood of Ω_s, one can obtain

$$\begin{aligned}
\langle\tilde{\mathcal{G}}_3, v\rangle &= -\int_{\Omega_m} \epsilon_m T\left(\nabla\tilde{u}^H\right)\cdot\nabla v dx + \int_{\Omega_s}\epsilon_m\nabla G\cdot\nabla v dx \\
&= -\left\langle\gamma_{\boldsymbol{n}_\Gamma,\Omega_m}\left(\epsilon_m T\left(\nabla\tilde{u}^H\right)\right), \gamma_{2,\Gamma}(v)\right\rangle_{H^{-1/2}(\Gamma)\times H^{1/2}(\Gamma)} \\
&\quad + \left\langle\gamma_{\boldsymbol{n}_\Gamma,\Omega_s}\left(\epsilon_m\nabla G\right), \gamma_{2,\Gamma}(v)\right\rangle_{H^{-1/2}(\Gamma)\times H^{1/2}(\Gamma)} \\
&\quad + \int_{\Omega_m}\mathrm{div}\left(\epsilon_m T\left(\nabla\tilde{u}^H\right)\right) v dx,\ \forall v\in H^1_0(\Omega),
\end{aligned} \tag{4.173}$$

where $\gamma_{\boldsymbol{n}_\Gamma,\Omega_m}$ and $\gamma_{\boldsymbol{n}_\Gamma,\Omega_s}$ are the normal trace operators in $H(\mathrm{div};\Omega_m)$ and $H(\mathrm{div};\Omega_s)$, respectively, and $\gamma_{2,\Gamma}(v)$ is the trace of v on Γ. Thus, if $T\left(\nabla\tilde{u}^H\right)$ is more regular, then (4.173) can be represented in terms of integrals over Γ:

$$\langle\tilde{\mathcal{G}}_3, v\rangle = \int_\Gamma -\epsilon_m\left(T\left(\nabla\tilde{u}^H\right) + \nabla G\right)\cdot\boldsymbol{n}_\Gamma v ds + \int_{\Omega_m}\mathrm{div}\left(\epsilon_m T\left(\nabla\tilde{u}^H\right)\right) v dx. \tag{4.174}$$

The jump condition is now perturbed since we only have an approximation $T\left(\nabla\tilde{u}^H\right)$ to ∇u^H.

Remark 4.25
Notice that the near best approximation result in Propostion 4.12 holds for $\tilde{u}^N_{h_N}$ if $\tilde{u}^N_{h_N}$ is chosen as the Galerkin approximation of $\tilde{u}^N$ in a closed subspace V_{h_N} of $H^1_0(\Omega)\cap L^\infty(\Omega)$.

Remark 4.26
To define good $\tilde{\boldsymbol{y}}^*_{L,0}$ and $\tilde{\boldsymbol{y}}^*_N$ one proceeds similarly to Remark 4.19 and Remark 4.20.

No additional splitting of u into u^L and u^N

One can also solve (3.52) directly without the additional splitting into u^L and u^N. More precisely, for a given approximation $\tilde{u}^H$ of u^H let $\tilde{u}$ be the exact solution of the following problem:

$$\begin{aligned}&\text{Find } \tilde{u} \in H^1_g(\Omega) \text{ such that } b(x,\tilde{u})v \in L^1(\Omega) \text{ for all } v \in H^1_0(\Omega) \text{ and}\\ &\int\limits_\Omega \epsilon\nabla\tilde{u}\cdot\nabla v dx + \int\limits_\Omega b(x,\tilde{u})v dx = \langle\tilde{\mathcal{G}}_3, v\rangle \text{ for all } v \in H^1_0(\Omega).\end{aligned} \tag{4.175}$$

In this case, if $\tilde{u}_h \in H^1_g(\Omega)$ denotes a conforming approximation of $\tilde{u}$, we can apply directly the a posteriori error estimates from Section 4.1.4 where $\overline{g} = \tilde{u}_g \in H^1(\Omega)\cap L^\infty(\Omega)$ with $\nabla\tilde{u}_g \in [L^s(\Omega)]^d$, $\tilde{u} = \tilde{u}_g + \tilde{u}_0$, $\Omega_1 = \Omega_m \cup \Omega_{IEL}$, $\Omega_2 = \Omega_{ions}$, $\boldsymbol{A} = \epsilon I$, $k = \overline{k}$, $w = 0$, $f_0 = 0$, $\boldsymbol{f} = \boldsymbol{f}_{\tilde{\mathcal{G}}_3}$. Let, as usual, $\tilde{\boldsymbol{p}}^* = \epsilon\nabla\tilde{u}$. Then for any $\tilde{y}^*$ of the form $\tilde{\boldsymbol{y}}^* = \boldsymbol{f}_{\tilde{\mathcal{G}}_3} + \tilde{\boldsymbol{y}}^*_0$ with $\tilde{\boldsymbol{y}}^*_0 \in H(\mathrm{div};\Omega)$ and $\mathrm{div}\,\tilde{\boldsymbol{y}}^*_0 = 0$ in $\Omega_m \cup \Omega_{IEL}$ the error identity (4.115) takes the form

$$\frac{1}{2}|||\nabla(\tilde{u}_h - \tilde{u})|||^2 + \frac{1}{2}|||\tilde{\boldsymbol{y}}^* - \tilde{\boldsymbol{p}}^*|||^2_* + D_F(\tilde{u}_h, -\Lambda^*\tilde{\boldsymbol{p}}^*) + D_F(\tilde{u}, -\Lambda^*\tilde{\boldsymbol{y}}^*) = M^2_\oplus(\tilde{u}_h, \tilde{\boldsymbol{y}}^*), \tag{4.176}$$

where

$$2M^2_\oplus(\tilde{u}_h, \tilde{\boldsymbol{y}}^*) = |||\epsilon\nabla\tilde{u}_h - \tilde{\boldsymbol{y}}^*|||^2_* + 2D_F(\tilde{u}_h, -\Lambda^*\tilde{\boldsymbol{y}}^*) \tag{4.177}$$

and $D_F(\tilde{u}_h, -\Lambda^*\tilde{\boldsymbol{p}}^*)$, $D_F(\tilde{u}, -\Lambda^*\tilde{\boldsymbol{y}}^*)$, and $D_F(\tilde{u}_h, -\Lambda^*\tilde{\boldsymbol{y}}^*)$ are given by (4.111), (4.112), and (4.106), respectively, (or by (4.45), (4.46), and (4.39) in the case of homogeneous Dirichlet boundary condition $g = 0$). Also, we have the near best approximation result in Proposition 4.16 provided that $\tilde{u}_h = \tilde{u}_g + \tilde{u}_{0,h}$ where $\tilde{u}_{0,h}$ is the solution of the Galerkin problem

$$\begin{aligned}&\text{Find } \tilde{u}_{0,h} \in V_h \text{ such that}\\ &a(\tilde{u}_g + \tilde{u}_{0,h}, v) + \int\limits_\Omega \overline{k}^2\sinh(\tilde{u}_{0,h} + \tilde{u}_g)v dx = \int\limits_\Omega \boldsymbol{f}_{\tilde{\mathcal{G}}_3}\cdot\nabla v dx,\ \forall v \in V_h\end{aligned} \tag{4.178}$$

for some closed subspace $V_h \subset L^\infty(\Omega)$ of $H^1_0(\Omega)$.

Now the overall estimate for the error $|||\nabla(\tilde{u}_h - u)|||$ can be easily obtained. We have

$$|||\nabla(\tilde{u}_h - u)||| \le |||\nabla(\tilde{u}_h - \tilde{u})||| + |||\nabla(\tilde{u} - u)|||. \tag{4.179}$$

By using (4.176), the first term on the right-hand side of (4.179) is estimated by $\sqrt{2}M_\oplus(\tilde{u}_h, \tilde{\boldsymbol{y}}^*)$. For the second term, after subtracting equation (4.175) from (3.52), we obtain

$$a(u - \tilde{u}, v) = (b(x,\tilde{u}) - b(x,u)), v) + \int\limits_{\Omega_m} \epsilon_m\left(T\left(\nabla\tilde{u}^H\right) - \nabla u^H\right)\cdot\nabla v dx \text{ for all } v \in H^1_0(\Omega).$$

Set here $v := u - \tilde{u} \in H^1_0(\Omega)$. Using the monotonicity of $b(x,\cdot)$, we see that

$$|||\nabla(u - \tilde{u})||| \le \sqrt{\epsilon_m}\|T\left(\nabla\tilde{u}^H\right) - \nabla u^H\|_{L^2(\Omega_m)}. \tag{4.180}$$

By combining (4.179) and (4.180) we can formulate the following proposition for the overall error estimate of the component u.

Proposition 4.27
Let $\tilde{u}^H$ be a conforming approximation of the solution u^H of (3.29) and let $T\left(\nabla\tilde{u}^H\right) \in H(\mathrm{div};\Omega_m)$ be a reconstruction of the numerical flux $\nabla\tilde{u}^H$. Next, let $\tilde{u}_h \in H^1_g(\Omega)$ be an approximation of the solution $\tilde{u}$ of (4.175). Then, for any $\tilde{\boldsymbol{y}}^*$ of the form $\tilde{\boldsymbol{y}}^* = \boldsymbol{f}_{\tilde{\mathcal{G}}_3} + \tilde{\boldsymbol{y}}_0^*$ with $\tilde{\boldsymbol{y}}_0^* \in H(\mathrm{div};\Omega)$ and $\mathrm{div}\,\tilde{\boldsymbol{y}}_0^* = 0$ in $\Omega_m \cup \Omega_{IEL}$ the following guaranteed error estimate holds:

$$|||\nabla(\tilde{u}_h - u)||| \leq \sqrt{\epsilon_m} M_{\oplus,H}\left(\tilde{u}^H, T\left(\nabla\tilde{u}^H\right)\right) + \sqrt{2} M_{\oplus}\left(\tilde{u}_h, \tilde{\boldsymbol{y}}^*\right), \tag{4.181}$$

where $M_{\oplus,H}\left(\tilde{u}^H, T\left(\nabla\tilde{u}^H\right)\right)$ and $M_{\oplus}\left(\tilde{u}_h, \tilde{\boldsymbol{y}}^*\right)$ are defined by (4.154) and (4.177), respectively.

Remark 4.28 (Choosing $\tilde{\boldsymbol{y}}_0^*$)
We proceed similarly to Remark 4.22. Let $\tilde{u}_{0,h}$ be the Galerkin solution of (4.178) in the space $V_h^1 \subset H_0^1(\Omega)$ of continuous piecewise linear functions over a mesh $\mathscr{T}_h$ and let $\tilde{u}_h := \tilde{u}_g + u_{0,h}$. Then $\boldsymbol{q} := \epsilon\nabla\tilde{u}_h - \boldsymbol{f}_{\tilde{\mathcal{G}}_3}$ satisfies

$$\int_\Omega \boldsymbol{q}\cdot\nabla v dx = \int_\Omega -\overline{k}^2 \sinh(\tilde{u}_h) v dx,\ \forall v \in V_h^1.$$

Thus $\Pi_{L_h}(\boldsymbol{q})$, where Π_{L_h} is the L^2-projection operator onto the space of piecewise constant functions over $\mathscr{T}_h$, also satisfies the above equation and we define $\tilde{\boldsymbol{y}}_0^*$ as the patchwise flux reconstruction from Section 4.2 of $\Pi_{L_h}(\boldsymbol{q})$. Therefore, $\mathrm{div}\,\tilde{\boldsymbol{y}}_0^* + \Pi_{L_h}(-\overline{k}^2\sinh(\tilde{u}_h)) = 0$. Since $\overline{k} = 0$ in $\Omega_m \cup \Omega_{IEL}$ we see that $\mathrm{div}\,\tilde{\boldsymbol{y}}_0^*$ is exactly equilibrated in $\Omega_m \cup \Omega_{IEL}$ and the reliability of the majorant $M^2_{\oplus}(\tilde{u}_h, \tilde{\boldsymbol{y}}^*)$ is guaranteed.

4.3.3 Applications

In this section we present six numerical examples based on the two term and three term splittings. They show that the nonlinear mathematical model in question can be studied by fully reliable computer simulation methods that provide results with guaranteed and explicitly known accuracy. In the first four examples, we consider the system of two chromophores Alexa 488 and Alexa 594, which are used for protein labeling in biophysical experiments. The fifth experiment is conducted on an insulin protein (PDB ID: 1RWE). The sixth one is performed on the protein membrane channel SecYEG. In the first five examples, we assume a solvent consisting of NaCl with $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, corresponding to ionic strength of $I_s \approx 1.178\,M$ at $T = 298\,K$. The ground state charges are obtained by CHARMM 32. In the first four examples, we assume dielectric constants $\epsilon_m = 2$, $\epsilon_s = 80$ and in the fifth example $\epsilon_m = 20$, $\epsilon_s = 80$. In the sixth example, we again assume a solvent consisting of NaCl, but this time with an ionic strength $I_s = 0.1\,M$ and an absolute temperature $T = 300\,K$. These conditions correspond to $\overline{k}^2_{ions} = 0.843\,\mathring{A}^{-2}$. The dielectric coefficient inside the protein and membrane regions is $\epsilon_m = 4$, while $\epsilon_s = \epsilon_s(x)$ is variable in the part of the solution region which is inside the channel.

The numerical experiments are carried out in FreeFem++ developed and maintained by Frederich Hecht [98]. All figures below are generated with the help of VisIt [52]. The computational domain Ω for the first four examples is a cube with a side length of $A = 6 \times a_{\max} + 24\,\mathring{A} = 295.85\,\mathring{A}$, where $a_{\max}$ is the maximum side length of the smallest bounding box for the molecules with edges parallel to the coordinate axes. The computational domain Ω for the fifth example is a cube with a side length of $20 \times a_{\max} = 814.02\,\mathring{A}$ and the computational domain in the sixth example is a cube with a side length of $1000\,\mathring{A}$. The molecules are positioned in the center of Ω. For the Poincaré's constant $C_{P\Omega}$ on a cube we have that $C_{P\Omega} \leq \frac{A\sqrt{3}}{3\pi}$ (see [135]). The Dirichlet boundary condition for all experiments is given by $g = 0$ on $\partial\Omega$. The discretizations used in the numerical tests to find conforming approximations of u^L, u^N, and u are based on standard linear P_1 finite elements although the derived estimates apply to any conforming approximations that are in $L^\infty(\Omega)$, which could for example also be obtained from higher order finite element methods (hpFEM) or isogeometric analysis (IGA).

The surface mesh for the first four examples is constructed with TMSmesh 2.1 [49, 129] which produces a Gaussian molecular surface. The surface mesh of the two chromophores is additionally optimized with the help of mmgs [58]. The surface meshes of the insulin protein in the fifth example and of the membrane channel SecYEG in the sixth example are generated with NanoShaper [61] for a grid parameter *GridScale*=2 and represent the solvent-accessible surface (Connolly surface) with a radius of the probe sphere equal to $1.4\,\mathring{A}$. In the sixth example, an ion exclusion layer with a thickness of $2\,\mathring{A}$ is added around the channel SecYEG and the membrane in which it is embedded.

The initial tetrahedral meshes are generated using TetGen [170] and then they are adaptively refined with the help of mmg3d [62]. The shape of the molecules is not changed during adaptation. This is justified, since the molecule structure is only known with a certain precision from X-ray crystallography. It is also possible to use isoparametric elements to represent the molecular surface exactly. Then, in the mesh refining procedure new points will be inserted on the surface by splitting the curved elements on the interface Γ.

Remark 4.29
In order to compute a true conforming approximation of u^L in the 2-term splitting, one has to solve the homogenized problem (4.148) *with some function $u^L_{g-G} \in H^1(\Omega) \cap L^\infty(\Omega)$ with $\nabla u^L_{g-G} \in [L^s(\Omega)]^d$ for some $s > d$ and $\gamma_2(u^L_{g-G}) = g - G$ on $\partial\Omega$. In practice, even if $g = 0$, one cannot just take the function $-G$ defined over Ω a.e. since it is not even in $H^1(\Omega)$. However, a simple truncation $T_s(-G) := \max\{-s, \min\{s, -G\}\}$ for $s \geq \|G\|_{L^\infty(\partial\Omega)}$ already ensures that $T_s(-G) \in H^1(\Omega) \cap L^\infty(\Omega)$ with $\nabla T_s(-G) \in [L^\infty(\Omega)]^d$ and $\gamma_2\left(T_s(-G)\right) = -G$ on $\partial\Omega$. Thus, one can solve the homogenized problem* (4.148) *with $u^L_{-G} := T_s(-G)$, i.e.,*

$$\int_\Omega \epsilon \nabla \left(u^L_{-G} + u^L_{0,h_L}\right) \cdot \nabla v dx = \int_\Omega \boldsymbol{f}_{\mathcal{G}_2} \cdot \nabla v dx,\ \forall v \in V^1_{0,h_L},$$

where $V^1_{0,h_L} \subset H^1_0(\Omega)$ is a finite element space of continuous piecewise linear functions on a triangulation $\mathscr{T}_{h_L}$. However, in the examples that we present below, we find $u^L_{h_L}$ by simply solving the problem

$$\begin{aligned} &\textit{Find } \tilde{u}^L_{h_L} \textit{ in } V^1_{-G,h_L} \textit{ such that} \\ &\int_\Omega \nabla \tilde{u}^L_{h_L} \cdot \nabla v dx = \int_\Omega \boldsymbol{f}_{\mathcal{G}_2} \cdot \nabla v dx, \, \forall v \in V^1_{0,h_L}, \end{aligned}$$

where $V^1_{-G,h_L} \subset H^1(\Omega_m)$ is the finite element space of continuous piecewise linear functions which are equal to $\Pi_h(-G)$ on $\partial\Omega$ with Π_h being the standard nodal interpolation operator. Therefore, in this case $\gamma_2\left(u^L_{h_L}\right) \approx -G$ on $\partial\Omega$. We proceed in a similar way when dealing with the harmonic component u^H.

Example 1: System 1 (Alexa 488 and Alexa 594), 2-term splitting with additional decomposition into u^L and u^N

The first system consists of the two chromophores (dyes) Alexa 594 and Alexa 488, frequently used for protein labeling in biophysical experiments, with a total of 171 atoms in aqueous solution of NaCl. The parameters of the force fields of Alexa chromophores were created by an analogy approach from that of similar chemical groups in the CHARMM force field (version v35b3). The coordinates of the molecules are taken from a time frame of molecular dynamic simulations. In the all-atom MD simulations the dyes were attached to a polyproline 11 and dissolved in water box with NaCl [172]. The parameters of this example are $\epsilon_m = 2$, $\epsilon_s = 80$, $I_s = 1.178\, M$, which corresponds to $\overline{k}^2_{ions} = 10\, \mathring{A}^{-2}$ at $T = 298\, K$. In this example, $\Omega_{IEL} = \emptyset$ and thus $\Omega_s \equiv \Omega_{ions}$, $\Omega_m \cup \Omega_{IEL} = \Omega_m$.

Finding u^L

First, we solve adaptively (4.132) to find an approximation $u^L_{h_L}$ of u^L. As an error indicator we use the second term in the error estimate (4.145) computed over each element K

$$\eta^L_K = \left|\left|\left| \epsilon \nabla u^L_{h_L} - \boldsymbol{y}^*_L \right|\right|\right|_{*(K)} = \left(\int_K \frac{1}{\epsilon} \left| \epsilon \nabla u^L_{h_L} - \boldsymbol{y}^*_L \right|^2 dx \right)^{\frac{1}{2}}, \tag{4.182}$$

where $\boldsymbol{y}^*_L = \chi_{\Omega_s}(\epsilon_m - \epsilon_s)\nabla G + \boldsymbol{y}^*_{L,0}$. In this example, to find $\boldsymbol{y}^*_{L,0} \in H(\mathrm{div}; \Omega)$, we perform a minimization of the squared majorant $M^2_{\oplus,L}(u^L_{h_L}, \boldsymbol{y}^*_L; \alpha)$ defined in (4.147) over $\alpha \in \mathbb{R}_{>0}$ and $\boldsymbol{y}^*_{L,0} \in RT_0$ defined over the same mesh. This procedure gives a very sharp bound from above for the error. Moreover, we have a simple and efficient lower bound for the energy norm of the error $\nabla(u^L_{h_L} - u^L)$. Indeed, let us denote by J^L the quadratic functional whose unique minimizer over $H^1_{g-G}(\Omega)$ is the solution u^L of (4.132) and which is defined by $J^L(v) = \int_\Omega \frac{\epsilon}{2} \left| \nabla u^L \right| dx - \langle \mathcal{G}_2, v \rangle$. Then, assuming that $u^L_{h_L} \in H^1_{g-G}(\Omega)$ (for example when $u^L_{h_L} = u^L_{g-G} + u^L_{0,h_L}$, where u^L_{0,h_L} is the finite element solution of the homogenized equation

(4.148)), from the equality

$$\left|\left|\left|\nabla\left(u_{h_L}^L - u^L\right)\right|\right|\right|^2 = 2\left(J^L(u_{h_L}^L) - J^L(u^L)\right) \tag{4.183}$$

it follows that for all $w \in H^1_{g-G}(\Omega)$

$$\left|\left|\left|\nabla\left(u_{h_L}^L - u^L\right)\right|\right|\right|^2 \geq 2\left(J^L\left(u_{h_L}^L\right) - J^L(w)\right) =: M^2_{\ominus,L}\left(u_{h_L}^L, w\right). \tag{4.184}$$

For w we always take the last available approximation $u_{h_L}^L$ from the adaptive procedure and compute the lower bound for the error on all previous levels. For convenience, we will denote the approximation $u_{h_L}^L$ on mesh level i by u_i^L. Instead $\boldsymbol{y}_L^*$ and $\boldsymbol{y}_{L,0}^*$ we write $\boldsymbol{y}_{L,i}^*$ and $\boldsymbol{y}_{L,0,i}^*$ where $i = 0, 1, ..\bar{p}$, $i = 0$ corresponds to the initial mesh, and $i = \bar{p}$ corresponds to the last mesh. The results after solving adaptively for u^L are shown in the tables below where $\|v\|_0$ denotes the $L^2(\Omega)$ norm of the function v and $\bar{p} = 7$.

Table 4.12: Example 1. System 1

System 1: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\|u_i^L\|_0$	$\left\|\left\|\left\|\nabla u_i^L\right\|\right\|\right\|$	$M_{\ominus,L}(u_i^L, u_{\bar{p}}^L)$	$M_{\oplus,L}(u_i^L, \boldsymbol{y}_{L,i}^*)$	$J^L(u_i^L)$
0	667 008	63584.27	15679.01	3318.24	3358.51	-122925209.65
1	1 695 251	64014.26	15977.11	1267.71	1369.07	-127627031.79
2	3 803 582	64064.47	16006.50	819.032	968.374	-128095169.28
3	7 238 416	64094.10	16018.35	543.720	749.503	-128282760.11
4	10 268 886	64109.01	16022.61	401.463	653.047	-128349989.97
5	13 164 899	64115.19	16024.94	295.404	593.411	-128386944.42
6	16 124 993	64119.55	16026.50	194.810	549.973	-128411600.81
7	19 531 518	64122.38	16027.69	-	514.037	-128430576.31

We can also find guaranteed lower and upper bounds on the relative errors in energy and combined energy norm, as well as practical estimations for these quantities. The combined energy norm of the pair $(v, \boldsymbol{q}) \in H_0^1(\Omega) \times \left[L^2(\Omega)\right]^d$ is defined by

$$|||(v, \boldsymbol{q})|||_{\mathrm{CEN}} := \sqrt{|||\nabla v|||^2 + |||\boldsymbol{q}|||_*^2}.$$

By $\mathrm{REN}^{\mathrm{L,Up}}_{i,j,k,s}$ we denote the guaranteed upper bound for the relative error in energy norm, by $\mathrm{RCEN}^{\mathrm{L,Up}}_{i,j,s}$ the guaranteed upper bound on the relative error in combined energy norm, by $\mathrm{REN}^{\mathrm{L,Low}}_{i,j,k,s}$ the guaranteed lower bound on the relative energy norm error, and by $\mathrm{RCEN}^{\mathrm{L,Low}}_{i,j,s}$ the guaranteed lower bound for the relative error in combined energy norm where the indices i, j, k, s correspond to the refinement levels from which approximations for u^L and $\boldsymbol{p}_L^*$ are taken. For any $i, j, k, s \in \{0, 1, 2, ..., \bar{p}\}$ we have

$$\left|\left|\left|\nabla u_j^L\right|\right|\right| - M_{\oplus,L}\left(u_j^L, \boldsymbol{y}_{L,s}^*\right) \leq \left|\left|\left|\nabla u_j^L - \nabla\left(u_j^L - u^L\right)\right|\right|\right| \leq \left|\left|\left|\nabla u_j^L\right|\right|\right| + M_{\oplus,L}\left(u_j^L, \boldsymbol{y}_{L,s}^*\right).$$

Therefore,

$$\frac{|||\nabla(u_i^L - u^L)|||}{|||\nabla u^L|||} \le \frac{M_{\oplus,L}(u_i^L, \boldsymbol{y}_{L,k}^*)}{\left|\left|\left|\nabla u_j^L\right|\right|\right| - M_{\oplus,L}(u_j^L, \boldsymbol{y}_{L,s}^*)} =: \mathrm{REN}_{i,j,k,s}^{\mathrm{L,Up}}, \tag{4.185a}$$

$$\mathrm{REN}_{i,j,k,s}^{\mathrm{L,Low}} := \frac{M_{\ominus,L}(u_i^L, u_k^L)}{\left|\left|\left|\nabla u_j^L\right|\right|\right| + M_{\oplus,L}(u_j^L, \boldsymbol{y}_{L,s}^*)} \le \frac{|||\nabla(u_i^L - u^L)|||}{|||\nabla u^L|||}, \tag{4.185b}$$

where (4.185a) is valid if $\left|\left|\left|\nabla u_j^L\right|\right|\right| - M_{\oplus,L}(u_j^L, \boldsymbol{y}_{L,s}^*) > 0$. For any level i, the above bounds are expected to be the sharpest when we take $j, k, s = \bar{p}$. In practice, on each level i, the best one can do is to take for $\mathrm{REN}_{i,j,k,s}^{\mathrm{L,Up}}$ $j = i, s = i, k = i$. Optionally, once the computations are done, i.e., we have reached level $\bar{p}$, one can return and recompute slightly sharper upper bounds for each $i = 0, 1, ..., \bar{p}$ with $j = \bar{p}, s = \bar{p}$ and $k = i$. This results in only 2 arithmetic operations per level, provided that $\left|\left|\left|\nabla u_{\bar{p}}^L\right|\right|\right|$ and $M_{\oplus,L}(u_i^L, \boldsymbol{y}_{L,i}^*)$, $i = 0, 1, ..., \bar{p}$ are stored. On the other hand, $\mathrm{REN}_{i,j,k,s}^{\mathrm{L,Low}}$ is equal to zero if $k = i$ and it is expected that $\mathrm{REN}_{i,j,k,s}^{\mathrm{Low}}$ will be negative if $k < i$. Therefore, on each level i, to compute the best possible lower bounds for all previous levels $0, 1, ..., i-1$, we take $k = i, j = i, s = i$. Further, we have the equality

$$|||\epsilon\nabla v - \boldsymbol{y}^*|||_*^2 = |||\nabla(v-u)|||^2 + |||\boldsymbol{y}^* - \boldsymbol{p}^*|||_*^2 - 2\int_\Omega (\boldsymbol{y}^* - \boldsymbol{p}^*)\cdot\nabla(v-u)dx, \tag{4.186}$$

which holds for any $v, u \in H^1(\Omega)$ and any $\boldsymbol{y}^*, \boldsymbol{p}^* \in \left[L^2(\Omega)\right]^d$.Taking into account that $\boldsymbol{y}_L^* = \boldsymbol{f}_{\mathcal{G}_2} + \boldsymbol{y}_{L,0}^*$ and $\boldsymbol{p}_L^* = \boldsymbol{f}_{\mathcal{G}_2} + \boldsymbol{p}_{L,0}^*$ with $\operatorname{div}\boldsymbol{p}_{L,0}^* = 0$ in Ω and $\boldsymbol{y}_{L,0}^* \in H(\operatorname{div};\Omega)$, we also have

$$\begin{aligned}\left|\int_\Omega (\boldsymbol{y}_L^* - \boldsymbol{p}_L^*)\cdot\nabla(u_{h_L}^L - u^L)dx\right| &= \left|\int_\Omega \operatorname{div}\boldsymbol{y}_{L,0}^*(u_{h_L}^L - u^L)dx\right| \\ &\le \frac{C_{P\Omega}\|\operatorname{div}\boldsymbol{y}_{L,0}^*\|_{L^2(\Omega)}}{\sqrt{\epsilon_{\min}}} M_{\oplus,L}(u_{h_L}^L, \boldsymbol{y}_L^*),\end{aligned} \tag{4.187}$$

where we have used Poincaré's inequality $\|v\|_{L^2(\Omega)} \le C_{P\Omega}\|\nabla v\|)L^2(\Omega)$, $\forall v \in H_0^1(\Omega)$ together with the estimate (4.145). Therefore, we obtain the estimates

$$\begin{aligned}\left(M_{\ominus,L}^{\mathrm{CEN}}(u_{h_L}^L, \boldsymbol{y}_L^*)\right)^2 &:= |||\epsilon\nabla u_{h_L}^L - \boldsymbol{y}_L^*|||_*^2 - 2\frac{C_{P\Omega}\|\operatorname{div}\boldsymbol{y}_{L,0}^*\|_{L^2(\Omega)}}{\sqrt{\epsilon_{\min}}} M_{\oplus,L}(u_{h_L}^L, \boldsymbol{y}_L^*) \\ &\le |||(u_{h_L}^L - u^L, \boldsymbol{y}_L^* - \boldsymbol{p}_L^*)|||_{\mathrm{CEN}}^2 \\ &\le |||\epsilon\nabla u_{h_L}^L - \boldsymbol{y}_L^*|||_*^2 + 2\frac{C_{P\Omega}\|\operatorname{div}\boldsymbol{y}_{L,0}^*\|_{L^2(\Omega)}}{\sqrt{\epsilon_{\min}}} M_{\oplus,L}(u_{h_L}^L, \boldsymbol{y}_L^*) \\ &=: \left(M_{\oplus,L}^{\mathrm{CEN}}(u_{h_L}^L, \boldsymbol{y}_L^*)\right)^2.\end{aligned} \tag{4.188}$$

Since in this example $\boldsymbol{y}_{L,0}^*$ is found by minimization of $M_{\oplus,L}^2(u_{h_L}^L, \boldsymbol{y}_L^*; \alpha)$, in our experiments $\|\operatorname{div}\boldsymbol{y}_{L,0}^*\|_{L^2(\Omega)}$ is usually of the order 10^{-5} to 10^{-4} and the above estimate is very sharp.

On the other hand, if we apply the patchwise flux reconstruction from Remark 4.19, then $\boldsymbol{y}^*_{L,0}$ is exactly equilibrated in Ω and $\|\operatorname{div} \boldsymbol{y}^*_{L,0}\|_{L^2(\Omega)}$ is exactly equal to zero. For any level $i = 0, 1, ..., \bar{p}$, we can bound the relative error in the combined energy norm as follows

$$\frac{\left|\!\left|\!\left| \left(u_i^L - u^L, \boldsymbol{y}^*_{L,i} - \boldsymbol{p}^*_L\right) \right|\!\right|\!\right|_{\text{CEN}}}{\left|\!\left|\!\left| (u^L, \boldsymbol{p}^*_L) \right|\!\right|\!\right|_{\text{CEN}}} \leq \frac{M^{\text{CEN}}_{\oplus,L}(u_i^L, \boldsymbol{y}^*_{L,i})}{\sqrt{2}\left(\left|\!\left|\!\left| \nabla u_j^L \right|\!\right|\!\right| - M_{\oplus,L}(u_j^L, \boldsymbol{y}^*_{L,s})\right)} := \text{RCEN}^{\text{L,Up}}_{i,j,s}, \tag{4.189}$$

$$\text{RCEN}^{\text{L,Low}}_{i,j,s} := \frac{M^{\text{CEN}}_{\ominus,L}(u_i^L, \boldsymbol{y}^*_{L,i})}{\sqrt{2}\left(\left|\!\left|\!\left| \nabla u_j^L \right|\!\right|\!\right| + M_{\oplus,L}(u_j^L, \boldsymbol{y}^*_{L,s})\right)} \leq \frac{\left|\!\left|\!\left| \left(u_i^L - u^L, \boldsymbol{y}^*_{L,i} - \boldsymbol{p}^*_L\right) \right|\!\right|\!\right|_{\text{CEN}}}{\left|\!\left|\!\left| (u^L, \boldsymbol{p}^*_L) \right|\!\right|\!\right|_{\text{CEN}}}.$$

For every level i, the sharpest estimates $\text{RCEN}^{\text{L,Up}}_{i,j,s}$ and $\text{RCEN}^{\text{L,Low}}_{i,j,s}$ are expected to be obtained when $j = \bar{p}$, $s = \bar{p}$. In the table below, we also present the practical estimation $\text{P}^{\text{L,CEN}}_{\text{rel},i}$ for the relative error in combined energy norm given by

$$\text{P}^{\text{L,CEN}}_{\text{rel},i} := \frac{\left|\!\left|\!\left| \epsilon \nabla u_i^L - \boldsymbol{y}^*_{L,i} \right|\!\right|\!\right|_*}{\sqrt{2}\left|\!\left|\!\left| \nabla u_i^L \right|\!\right|\!\right|} \text{ for all } i = 0, 1, ..., \bar{p}. \tag{4.190}$$

Table 4.13: Example 1. System 1.

System 1: $\bar{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\text{REN}^{\text{L,Low}}_{i,\bar{p},\bar{p},\bar{p}}$[%]	$\text{REN}^{\text{L,Up}}_{i,\bar{p},i,\bar{p}}$[%]	$\text{RCEN}^{\text{L,Low}}_{i,\bar{p},\bar{p}}$[%]	$\text{P}^{\text{L,CEN}}_{\text{rel},i}$[%]	$\text{RCEN}^{\text{L,Up}}_{i,\bar{p},\bar{p}}$[%]
0	667 008	20.0598	21.6487	14.3546	15.1455	15.3079
1	1 695 251	7.66370	8.82493	5.85216	6.05907	6.24017
2	3 803 582	4.95130	6.24207	4.13936	4.27784	4.41381
3	7 238 416	3.28696	4.83125	3.20376	3.30850	3.41621
4	10 268 886	2.42697	4.20949	2.79147	2.88196	2.97656
5	13 164 899	1.78581	3.82509	2.53655	2.61840	2.70474
6	16 124 993	1.17768	3.54509	2.35088	2.42650	2.50675
7	19 531 518	0.00000	3.31345	2.19726	2.26777	2.34296

Finding $\tilde{u}^N$

Once we have obtained an approximation $u^L_{h_L}$ for u^L, we solve adaptively (4.134) with $u^L_{h_L}$ in it to find approximations $\tilde{u}^N_{h_N}$ of $\tilde{u}^N$. For $u^L_{h_L}$ we take the approximation u^L_2 from level 2. In this case, we have $M_{\oplus,L}(u^L_2, \boldsymbol{y}^*_{L,2}) = 968.374$, see Table 4.12, and $\text{REN}^{\text{L,Up}}_{2,2,2,2} = 6.43946\%$. For $\tilde{\boldsymbol{y}}^*_N \in H(\text{div}; \Omega)$ with $\text{div}(\tilde{\boldsymbol{y}}^*_N) = 0$ in $\Omega_m \cup \Omega_{IEL}$, we use the patchwise equilibrated reconstruction of the numerical flux $\epsilon \nabla \tilde{u}^N_{h_N}$ described in Remark 4.20. As an error indicator, we use the quantity η^N_K defined by

$$\eta^N_K = \left(\left|\!\left|\!\left| \epsilon \nabla \tilde{u}^N_{h_N} - \tilde{\boldsymbol{y}}^*_N \right|\!\right|\!\right|^2_{*(K)} + 2D_{F,N,K}(\tilde{u}^N_{h_N}, -\Lambda^* \tilde{\boldsymbol{y}}^*_N) \right)^{\frac{1}{2}}. \tag{4.191}$$

where $D_{F,N,K}(\tilde{u}^N_{h_N}, -\Lambda^*\tilde{\boldsymbol{y}}^*_N)$ is defined by (4.139) but with integration taking place only on elements $K \in \Omega_{ions}$. From (4.51) we have the following upper bounds for the error in energy and combined energy norm

$$\begin{aligned}
&|||\nabla(\tilde{u}^N_{h_N} - \tilde{u}^N)||| \leq \sqrt{2} M_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N) \\
&|||(\tilde{u}^N_{h_N} - \tilde{u}^N, \tilde{\boldsymbol{y}}^*_N - \tilde{\boldsymbol{p}}^*_N)|||_{\mathrm{CEN}} \leq \sqrt{2} M_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N).
\end{aligned}$$

We will denote by $\tilde{u}^N_i$ the finite element approximations of $\tilde{u}^N$ and by $\tilde{\boldsymbol{y}}^*_{N,i}$ the approximations of the flux $\tilde{\boldsymbol{p}}^*_N$ at mesh refinement level $i, i = 0, 1, 2, ..., \bar{p}$, where $\bar{p} = 6$. By $J^N_{h_L} : H^1_0(\Omega) \to \mathbb{R} \cup \{+\infty\}$ we denote the functional defined by

$$J^N_{h_L}(v) := \begin{cases} \displaystyle\int_\Omega \left[\frac{\epsilon}{2}|\nabla v|^2 + \overline{k}^2 \cosh(v + G + u^L_{h_L})\right] dx, \text{ if } \overline{k}^2 \cosh(v + G + u^L_{h_L}) \in L^1(\Omega), \\ +\infty, \text{ if } \overline{k}^2 \cosh(v + G + u^L_{h_L}) \notin L^1(\Omega). \end{cases} \tag{4.192}$$

The unique minimizer of $J^N_{h_L}$ over $H^1_0(\Omega)$ is the solution $\tilde{u}^N$ to the problem (4.134) with $u^L_{h_L}$ in it. The subscript h_L in the notation for the functional $J^N_{h_L}$ corresponds to the mesh refinement level on which the approximation $u^L_{h_L}$ of u^L is computed. Since in this case we take u^L_2 as an approximation of u^L, we are interested in the values of $J^N_2(\tilde{u}^N_i)$ on levels $i = 0, 1, ..., \bar{p}$ in the adaptive solution for $\tilde{u}^N$ (see Table 4.14).

Table 4.14: Example 1. System 1.

System 1: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\|\tilde{u}^N_i\|_0$	$\|\|\|\nabla\tilde{u}^N_i\|\|\|$	$\|\|\|\epsilon\nabla\tilde{u}^N_i - \tilde{\boldsymbol{y}}^*_{N,i}\|\|\|_*$	$\sqrt{2}M_{\oplus,N}(\tilde{u}^N_i, \tilde{\boldsymbol{y}}^*_{N,i})$	$J^N_2(\tilde{u}^N_i)$
0	667 008	927.667	324.330	155.122	192.502	259 017 030.56
1	1 315 573	928.279	320.425	70.1561	108.592	259 013 323.94
2	5 800 985	928.384	321.655	39.3835	61.8732	259 012 091.45
3	9 514 417	928.394	321.675	33.1119	52.9724	259 011 967.30
4	13 957 123	928.399	321.741	29.4864	47.2872	259 011 894.50
5	18 286 791	928.401	321.782	26.7945	43.4640	259 011 849.22
6	22 883 680	928.403	321.807	24.8527	40.5678	259 011 818.93

As for the part u^N we can define the following guaranteed lower and upper bounds on the relative errors. For any $i, j, k, s \in \{0, 1, \ldots, \bar{p}\}$ we have

$$\frac{|||\nabla(\tilde{u}^N_i - \tilde{u}^N)|||}{|||\nabla\tilde{u}^N|||} \leq \frac{\sqrt{2}M_{\oplus,N}(\tilde{u}^N_i, \tilde{\boldsymbol{y}}^*_{N,k})}{\left|\left|\left|\nabla\tilde{u}^N_j\right|\right|\right| - \sqrt{2}M_{\oplus,N}(\tilde{u}^N_j, \tilde{\boldsymbol{y}}^*_{N,s})} =: \mathrm{REN}^{\mathrm{N,Up}}_{i,j,k,s} \tag{4.193}$$

By using the main error identity (4.51), the lower bound (4.54) for the combined energy norm, and the estimates

$$\begin{aligned}\left|\!\left|\!\left|(\tilde{u}_i^N, \tilde{\boldsymbol{y}}_{N,s}^*)\right|\!\right|\!\right|_{\mathrm{CEN}} - \sqrt{2}M_{\oplus,N}\left(\tilde{u}_i^N, \tilde{\boldsymbol{y}}_{N,s}^*\right) &\le \left|\!\left|\!\left|(\tilde{u}^N, \tilde{\boldsymbol{p}}_N^*)\right|\!\right|\!\right|_{\mathrm{CEN}} \\ &\le \left|\!\left|\!\left|(\tilde{u}_i^N, \tilde{\boldsymbol{y}}_{N,s}^*)\right|\!\right|\!\right|_{\mathrm{CEN}} + \sqrt{2}M_{\oplus,N}\left(\tilde{u}_i^N, \tilde{\boldsymbol{y}}_{N,s}^*\right)\end{aligned}$$

for all $i, s = 0, 1, ..., \bar{p}$, we can define the following guaranteed lower and upper bounds for the relative error in combined energy norm:

$$\frac{\left|\!\left|\!\left|\left(\tilde{u}_i^N - \tilde{u}^N, \tilde{\boldsymbol{y}}_{N,i}^* - \tilde{\boldsymbol{p}}_N^*\right)\right|\!\right|\!\right|_{\mathrm{CEN}}}{\left|\!\left|\!\left|(\tilde{u}^N, \tilde{\boldsymbol{p}}_N^*)\right|\!\right|\!\right|_{\mathrm{CEN}}} \le \frac{\sqrt{2}M_{\oplus,N}\left(\tilde{u}_i^N, \tilde{\boldsymbol{y}}_{N,i}^*\right)}{\left|\!\left|\!\left|\left(\tilde{u}_j^N, \tilde{\boldsymbol{y}}_{N,s}^*\right)\right|\!\right|\!\right|_{\mathrm{CEN}} - \sqrt{2}M_{\oplus,N}\left(\tilde{u}_j^N, \tilde{\boldsymbol{y}}_{N,s}^*\right)} := \mathrm{RCEN}_{i,j,s}^{\mathrm{N,Up}},$$

$$\mathrm{RCEN}_{i,j,s}^{\mathrm{N,Low}} := \frac{\frac{1}{\sqrt{2}}\left|\!\left|\!\left|\epsilon\nabla\tilde{u}_i^N - \tilde{\boldsymbol{y}}_{N,i}^*\right|\!\right|\!\right|}{\left|\!\left|\!\left|\left(\tilde{u}_j^N, \tilde{\boldsymbol{y}}_{N,s}^*\right)\right|\!\right|\!\right|_{\mathrm{CEN}} + \sqrt{2}M_{\oplus,N}\left(\tilde{u}_j^N, \tilde{\boldsymbol{y}}_{N,s}^*\right)} \le \frac{\left|\!\left|\!\left|\left(\tilde{u}_i^N - \tilde{u}^N, \tilde{\boldsymbol{y}}_{N,i}^* - \tilde{\boldsymbol{p}}_N^*\right)\right|\!\right|\!\right|_{\mathrm{CEN}}}{\left|\!\left|\!\left|(\tilde{u}^N, \tilde{\boldsymbol{p}}_N^*)\right|\!\right|\!\right|_{\mathrm{CEN}}}.$$

The sharpest values for $\mathrm{REN}_{i,j,k,s}^{\mathrm{N,Up}}$, $\mathrm{RCEN}_{i,j,s}^{\mathrm{N,Up}}$, and $\mathrm{RCEN}_{i,j,s}^{\mathrm{N,Low}}$ at each level i are expected to be obtained when $j = \bar{p}, k = \bar{p}, s = \bar{p}$ (assuming that we do not have another better approximation $\tilde{\boldsymbol{y}}_N^*$ for the flux $\tilde{\boldsymbol{p}}_N^*$). The practical estimation $P_{\mathrm{rel},i}^{\mathrm{N,CEN}}$ for the relative error in combined energy norm is given by

$$\mathrm{P}_{\mathrm{rel},i}^{\mathrm{N,CEN}} := \frac{\left|\!\left|\!\left|\epsilon\nabla\tilde{u}_i^N - \tilde{\boldsymbol{y}}_{N,i}^*\right|\!\right|\!\right|_*}{\sqrt{2}\left|\!\left|\!\left|\nabla\tilde{u}_i^N\right|\!\right|\!\right|} \text{ for all } i = 0, 1, ..., \bar{p}. \tag{4.194}$$

We also introduce a practical upper bound

$$\mathrm{PREN}_{i,j}^{\mathrm{N,Up}} := \frac{\left|\!\left|\!\left|\epsilon\nabla\tilde{u}_i^N - \tilde{\boldsymbol{y}}_{N,j}^*\right|\!\right|\!\right|_*}{\left|\!\left|\!\left|\nabla\tilde{u}_i^N\right|\!\right|\!\right|} \text{ for all } i, j = 0, 1, ..., \bar{p}, \tag{4.195}$$

for the relative error in energy norm which is based on the relation

$$\left|\!\left|\!\left|\nabla(\tilde{u}_i^N - \tilde{u}^N)\right|\!\right|\!\right| \le \left|\!\left|\!\left|(\tilde{u}_i^N - \tilde{u}^N, \tilde{\boldsymbol{y}}_{N,i}^* - \tilde{\boldsymbol{p}}_N^*)\right|\!\right|\!\right|_{\mathrm{CEN}} \approx \left|\!\left|\!\left|\epsilon\nabla\tilde{u}_i^N - \tilde{\boldsymbol{y}}_{N,i}^*\right|\!\right|\!\right|_*$$

and is useful when it is suspected that the guaranteed upper bound for the relative error overestimates the real error. The above introduced bounds on the relative errors are presented in Table 4.15 and Table 4.16.

Finally, according to (4.146) the overall error in the regular component u will be

$$\begin{aligned}&\left|\!\left|\!\left|\nabla(u_h - u)\right|\!\right|\!\right| \le 2M_{\oplus,L}(u_{h_L}^L, \boldsymbol{y}_L^*) + \sqrt{2}M_{\oplus,N}(\tilde{u}_{h_N}^N, \tilde{\boldsymbol{y}}_N^*) \\ &= 2M_{\oplus,L}(u_2^L, \boldsymbol{y}_{L,2}^*) + \sqrt{2}M_{\oplus,N}(\tilde{u}_6^N, \tilde{\boldsymbol{y}}_{N,6}^*) = 2 \times 968.37 + 40.57 = 1977.31\end{aligned}$$

For comparison, the energy norm of the approximate regular component $u_h = u_2^L + \tilde{u}_6^N$ is $\left|\!\left|\!\left|\nabla u_h\right|\!\right|\!\right| = 16276.2$. This means that the relative error in energy norm is no more than approximately $1977.31/16276.2 = 12.15\%$.

Table 4.15: Example 1. System 1.

System 1: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\mathrm{PREN}^{\mathrm{N,Up}}_{i,i}$[%]	$\mathrm{REN}^{\mathrm{N,Up}}_{i,i,i,i}$[%]	$\mathrm{RCEN}^{\mathrm{N,Low}}_{i,i,i}$[%]	$\mathrm{P}^{\mathrm{N,CEN}}_{\mathrm{rel},i}$[%]	$\mathrm{RCEN}^{\mathrm{N,Up}}_{i,i,i}$[%]
0	667 008	47.828	146.02	16.227	33.819	66.164
1	1 315 573	21.894	51.263	8.7551	15.481	31.076
2	5 800 985	12.243	23.817	5.3764	8.6578	15.695
3	9 514 417	10.293	19.714	4.6026	7.2786	13.152
4	13 957 123	9.1646	17.229	4.1455	6.4803	11.579
5	18 286 791	8.3269	15.616	3.7964	5.8880	10.546
6	22 883 680	7.7228	14.424	3.5422	5.4608	9.7758

Table 4.16: Example 1. System 1.

System 1: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\mathrm{REN}^{\mathrm{N,Up}}_{i,\bar{p},\bar{p},\bar{p}}$ [%]	$\mathrm{RCEN}^{\mathrm{N,Low}}_{i,\bar{p},\bar{p}}$ [%]	$\mathrm{RCEN}^{\mathrm{N,Up}}_{i,\bar{p},\bar{p}}$ [%]	$\left\|\left\|\left\|\epsilon\nabla\tilde{u}^N_i - \tilde{\boldsymbol{y}}^*_{N,\bar{p}}\right\|\right\|\right\|_*$	$\sqrt{2}M_{\oplus,N}(\tilde{u}^N_i, \tilde{\boldsymbol{y}}^*_{N,\bar{p}})$
0	667 008	40.974	22.129	46.434	89.292	115.23
1	1 315 573	24.311	10.008	26.194	51.983	68.373
2	5 800 985	16.737	5.6183	14.924	32.512	47.071
3	9 514 417	15.855	4.7236	12.777	29.023	44.591
4	13 957 123	15.269	4.2064	11.406	27.033	42.943
5	18 286 791	14.841	3.8224	10.484	25.734	41.741
6	22 883 680	14.424	3.5422	9.7758	24.852	40.567

Example 2. System 1 (Alexa 488 and Alexa 594), 2-term splitting with additional decomposition into u^L and u^N recomputed with u^L_4

Here, we recompute an approximation u_h of the regular component u from Example 1. This time we take u^L_4 as an approximation of u^L and solve with it for $\tilde{u}^N$. For u^L_4, we have $M_{\oplus,L}(u^L_4, \boldsymbol{y}^*_{L,4}) = 653.047$ and $\mathrm{REN}^{\mathrm{L,Up}}_{4,4,4,4} = 4.2489\%$. The final level is $\bar{p} = 3$.
In Table 4.19, it can be seen that the convergence of $\tilde{u}^N_{h_N}$ to $\tilde{u}^N$ is faster compared to the case when we used a worse approximation for u^L. Finally, according to (4.146) the overall error in the regular component u can be estimated by

$$\begin{aligned}\left|\left|\left|\nabla(u_h - u)\right|\right|\right| &\le 2M_{\oplus,L}(u^L_{h_L}, \boldsymbol{y}^*_L) + \sqrt{2}M_{\oplus,N}(\tilde{u}^N_{h_N}, \tilde{\boldsymbol{y}}^*_N)\\ &= 2M_{\oplus,L}(u^L_4, \boldsymbol{y}^*_{L,4}) + \sqrt{2}M_{\oplus,N}(\tilde{u}^N_3, \tilde{\boldsymbol{y}}^*_{N,3})\\ &= 2 \times 653.047 + 35.784 = 1341.878.\end{aligned}$$

For comparison, the energy norm of the approximate regular component $u_h = u^L_4 + \tilde{u}^N_3$ is $\left|\left|\left|\nabla u_h\right|\right|\right| = 16298.534$. This means that the relative error in energy norm is no more than

Table 4.17: Example 2. System 1 recomputed with u_4^L.

System 1: $\overline{k}_{ions}^2 = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\|\tilde{u}_i^N\|_0$	$\|\|\|\nabla\tilde{u}_i^N\|\|\|$	$\|\|\|\epsilon\nabla\tilde{u}_i^N - \tilde{\boldsymbol{y}}_{N,i}^*\|\|\|_*$	$\sqrt{2}M_{\oplus,N}(\tilde{u}_i^N, \tilde{\boldsymbol{y}}_{N,i}^*)$	$J_4^N(\tilde{u}_i^N)$
0	667 008	880.446	314.647	128.067	142.404	259 007 595.76
1	1 389 691	880.942	309.468	39.1633	60.4019	259 004 495.90
2	5 706 468	880.989	309.166	24.0717	40.2461	259 004 177.33
3	8 606 657	880.992	309.138	20.5852	35.7841	259 004 134.61

Table 4.18: Example 2. System 1 recomputed with u_4^L.

System 1: $\overline{k}_{ions}^2 = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\mathrm{PREN}_{i,i}^{\mathrm{N,Up}}$[%]	$\mathrm{REN}_{i,i,i,i}^{\mathrm{N,Up}}$[%]	$\mathrm{RCEN}_{i,i,i}^{\mathrm{N,Low}}$[%]	$\mathrm{P}_{\mathrm{rel},i}^{\mathrm{N,CEN}}$[%]	$\mathrm{RCEN}_{i,i,i}^{\mathrm{N,Up}}$[%]
0	667 008	40.702	82.676	14.972	28.780	44.499
1	1 389 691	12.655	24.251	5.5423	8.9484	15.943
2	5 706 468	7.7860	14.965	3.5610	5.5055	10.125
3	8 606 657	6.6589	13.090	3.0751	4.7085	8.9065

Table 4.19: Example 2. System 1 recomputed with u_4^L.

System 1: $\overline{k}_{ions}^2 = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\mathrm{REN}_{i,\bar{p},\bar{p},\bar{p}}^{\mathrm{N,Up}}$ [%]	$\mathrm{RCEN}_{i,\bar{p},\bar{p}}^{\mathrm{N,Low}}$ [%]	$\mathrm{RCEN}_{i,\bar{p},\bar{p}}^{\mathrm{N,Up}}$ [%]	$\|\|\|\epsilon\nabla\tilde{u}_i^N - \tilde{\boldsymbol{y}}_{N,\bar{p}}^*\|\|\|_*$	$\sqrt{2}M_{\oplus,N}(\tilde{u}_i^N, \tilde{\boldsymbol{y}}_{N,\bar{p}}^*)$
0	667 008	33.507	19.131	35.443	73.624	91.593
1	1 389 691	16.777	5.8504	15.033	31.708	45.861
2	5 706 468	13.637	3.5959	10.017	22.183	37.279
3	8 606 657	13.090	3.0751	8.9065	20.585	35.784

approximately $1341.878/16298.534 = 8.23\%$.

Example 3: System 1 (Alexa 488 and Alexa 594) recomputed with 2-term splitting without additional decomposition into u^L and u^N

Here we recompute an approximation u_h of the regular component u from Example 1 and Example 2 (all parameters are the same) by solving directly problem (3.50) where the nonhomogeneous Dirichlet boundary condition is given by $g - G = -G$ on $\partial\Omega$. We apply the procedure described in Section 4.3.1 on p. 144 (see also Section 4.1.4 for the derivation of the error estimates in the case of nonhomogeneous Dirichlet boundary condition). We

denote the approximation u_h of u on mesh refinement level i by u_i. Similarly, $\boldsymbol{y}_i^*$ denotes the approximation $\boldsymbol{y}^*$ of $\boldsymbol{p}^* = \epsilon\nabla u$ at mesh refinement level i. The last refinement level is $\bar{p} = 11$. To find $\boldsymbol{y}_i^*$ we apply the flux reconstruction described in Remark 4.22. We define the quantities $\mathrm{PREN}_{i,j}^{\mathrm{Up}}$, $\mathrm{REN}_{i,j,k,s}^{\mathrm{Up}}$, $\mathrm{RCEN}_{i,j,s}^{\mathrm{Low}}$, $\mathrm{P}_{\mathrm{rel},i}^{\mathrm{CEN}}$, $\mathrm{RCEN}_{i,j,s}^{\mathrm{Up}}$ in a similar fashion as for the component $\tilde{u}^N$ in the 2-term splitting with the additional decomposition into u^L and u^N. The rest of the notation is the same as in Section 4.3.1 and Section 4.1.4.

This time we obtain a guaranteed upper bound on the relative error in the regular component u of 2.53% (see 4-th column in Table 4.21) compared to around 8.23% when additional decomposition into u^L and u^N is applied. Moreover, the minorant $M_{\ominus}(u_i, u_{\bar{p}})$ and the majorant $\sqrt{2}M_{\oplus}(u_i, \boldsymbol{y}_i^*)$ provide tight bounds on the primal part of the error $J(u_i) - J(u) = M_{\oplus}^2(u_i, \boldsymbol{p}^*)$ (columns 5 and 7 in Table 4.20).

Table 4.20: Example 3. System 1 recomputed with 2-term splitting without additional decomposition into u^L and u^N. Here, $M_{\ominus}(u_i, u_{\bar{p}})$ is the minorant for the primal part of the error defined by (4.120) and $\sqrt{2}M_{\oplus}(u_i, \boldsymbol{y}_i^*)$ is an upper bound both for the primal part of the error and for the energy norm of the error.

System 1: $\overline{k}_{ions}^2 = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$							
level i	#elements	$\|u_i\|_0$	$\|\|\|\nabla u_i\|\|\|$	$\sqrt{2}M_{\ominus}(u_i, u_{\bar{p}})$	$\|\|\|\epsilon\nabla u_i - \boldsymbol{y}_i^*\|\|\|_*$	$\sqrt{2}M_{\oplus}(u_i, \boldsymbol{y}_i^*)$	$J(u_i)$
0	667 008	64847.173	16446.255	3867.60	4881.16	5067.89	137 987 860.82
1	1 619 690	64853.944	16282.891	1384.92	1585.69	1594.16	131 467 694.44
2	3 624 678	64855.019	16278.101	933.867	1104.89	1107.65	130 944 733.77
3	6 830 130	64855.176	16283.830	681.347	850.948	853.001	130 740 796.37
4	9 861 226	64855.249	16284.985	559.190	729.952	731.163	130 665 026.35
5	12 982 453	64855.305	16285.379	476.068	653.787	654.706	130 622 000.18
6	16 420 992	64855.346	16285.807	409.548	597.282	598.027	130 592 544.59
7	20 636 057	64855.370	16286.127	349.061	550.313	550.920	130 569 601.53
8	25 937 013	64855.382	16286.388	289.230	508.493	508.983	130 550 506.67
9	32 602 138	64855.404	16286.599	226.473	470.277	470.673	130 534 324.62
10	40 972 275	64855.430	16286.773	153.787	434.999	435.316	130 520 504.81
11	51 409 492	64855.446	16286.923	-	402.449	402.703	130 508 679.46

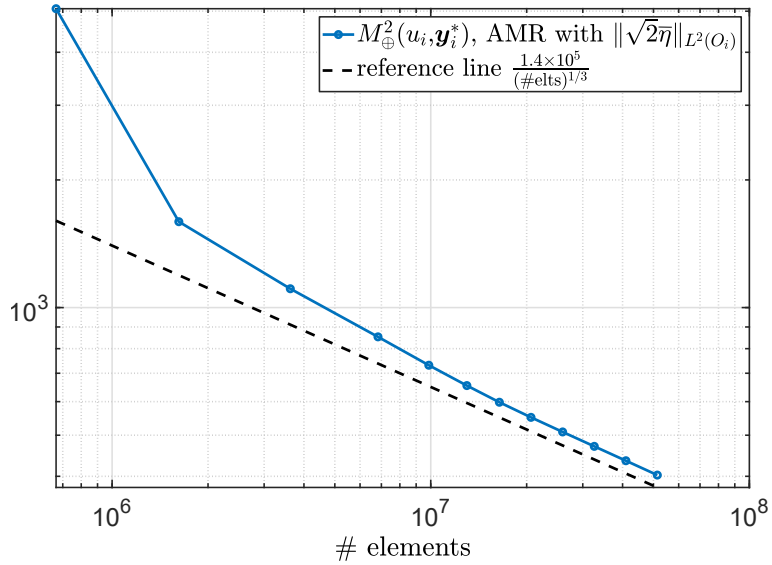


Figure 4.18: Convergence of the majorant $\sqrt{2}M_\oplus(u_i, \boldsymbol{y}_i^*)$ under adaptive mesh refinement with the error indicator $\|\sqrt{2}\overline{\eta}\|_{L^2(O_i)}$, where $\overline{\eta}$ is defined by (4.108) on p. 133 and O_i is the patch of elements around the vertex V_i (see (4.84) on p. 122 for details on the adaptive procedure using the software mmg3d).

Table 4.21: Example 3. System 1 recomputed with 2-term splitting without additional decomposition into u^L and u^N.

System 1: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\text{PREN}^{\text{Up}}_{i,i}$[%]	$\text{REN}^{\text{Up}}_{i,i,i,i}$[%]	$\text{RCEN}^{\text{Low}}_{i,i,i}$[%]	$\text{P}^{\text{CEN}}_{\text{rel},i}$[%]	$\text{RCEN}^{\text{Up}}_{i,i,i}$[%]
0	667 008	29.67	44.539	11.965	20.986	27.088
1	1 619 690	9.738	10.852	4.5437	6.8860	7.4187
2	3 624 678	6.787	7.3014	3.2344	4.7995	5.0486
3	6 830 130	5.225	5.5279	2.5178	3.6951	3.8437
4	9 861 226	4.482	4.7008	2.1711	3.1695	3.2771
5	12 982 453	4.014	4.1886	1.9510	2.8387	2.9246
6	16 420 992	3.667	3.8120	1.7867	2.5933	2.6648
7	20 636 057	3.379	3.5011	1.6495	2.3893	2.4498
8	25 937 013	3.122	3.2260	1.5269	2.2077	2.2592
9	32 602 138	2.887	2.9759	1.4145	2.0417	2.0856
10	40 972 275	2.670	2.7462	1.3104	1.8885	1.9260
11	51 409 492	2.470	2.5352	1.2140	1.7472	1.7791

Example 4: System 1 (Alexa 488 and Alexa 594) recomputed with 3-term splitting without additional decomposition into u^L and u^N

In this example, we solve the PBE for the system consisting of the two chromophores Alexa 488 and Alexa 594 but this time we utilize the 3-term splitting without further decomposition of the regular component u in $u^L + u^N$. The parameters are the same as in the first three examples, i.e., $\epsilon_m = 2$, $\epsilon_s = 80$, $I_s \approx 1.178\, M$, which corresponds to $\overline{k}^2_{ions} = 10\, \mathring{A}^{-2}$ at $T = 298\, K$. In this example, $\Omega_{IEL} = \emptyset$ and thus $\Omega_s \equiv \Omega_{ions}$, $\Omega_m \cup \Omega_{IEL} = \Omega_m$.

Finding the harmonic component u^H

According to the 3-term splitting, we first have to obtain a conforming approximation $\tilde{u}^H$ of u^H by solving problem (3.29). We solve this problem on a sequence of adapted meshes using the error majornat (4.152) and the derived from it error indicator. To reconstruct the numerical flux $\nabla\tilde{u}^H$ and obtain $T(\nabla\tilde{u}^H)$ we can either minimize the majorant in (4.152) (the right-hand side of this inequality) over a subspace of $H(\mathrm{div};\Omega_m)$, like RT_0, or apply some patchwise flux reconstruction technique. We notice that minimization of the majorant over RT_0 defined on the same mesh and applying the patchwise equilibrated flux reconstruction from [30, 31] yields practically the same results. Also, notice that when $\mathrm{div}\left(T\left(\nabla\tilde{u}^H\right)\right) = 0$ the right-hand side of (4.152) coincides with the majorant $M_{\oplus,H}\left(\tilde{u}^H, T(\nabla\tilde{u}^H)\right)$ in (4.154). We have computed $\tilde{u}^H$ on a final mesh with 32 457 251 tetrahedrons. The corresponding value for the majorant in (4.154) is $M_{\oplus,H}\left(\tilde{u}^H, T(\nabla\tilde{u}^H)\right) = 18.410$, where $T(\nabla\tilde{u}^H)$ is obtained by the flux reconstruction in [30, 31] and thus $\mathrm{div}\left(T(\nabla\tilde{u}^H)\right) = 0$. Since in this case

$$\|\nabla\left(\tilde{u}^H - u^H\right)\|_{L^2(\Omega_m)} \leq M_{\oplus,H}\left(\tilde{u}^H, T(\nabla\tilde{u}^H)\right) = 18.410,$$

and $\|\nabla\tilde{u}^H\|_{L^2(\Omega_m)} = 703.092$, we obtain a guaranteed upper bound on the relative error in $H^1(\Omega_m)$ seminorm

$$\frac{\|\nabla\left(\tilde{u}^H - u^H\right)\|_{L^2(\Omega_m)}}{\|\nabla u^H\|_{L^2(\Omega_m)}} \leq \frac{M_{\oplus,H}\left(\tilde{u}^H, T(\nabla\tilde{u}^H)\right)}{\|\nabla\tilde{u}^H\|_{L^2(\Omega_m)} - M_{\oplus,H}\left(\tilde{u}^H, T(\nabla\tilde{u}^H)\right)} = 2.69\,\%.$$

Finding the regular component $\tilde{u}$

Now, we find a conforming approximation $\tilde{u}_h$ of $\tilde{u}$, the exact solution of problem (4.175), by adaptively solving the Galerkin problem (4.178) (with $\tilde{u}_g \equiv 0$). By $\tilde{u}_i$ and $\tilde{\boldsymbol{y}}^*_i$ we denote the finite element approximations at mesh refinement level i to $\tilde{u}$ and $\tilde{\boldsymbol{p}}^*$, respectively. Here $\bar{p} = 18$. In order to find a good approximation $\tilde{\boldsymbol{y}}^*$ of the exact flux $\tilde{\boldsymbol{p}}^*$ having the form $\tilde{\boldsymbol{y}}^* = \boldsymbol{f}_{\tilde{\mathcal{G}}_3} + \tilde{\boldsymbol{y}}^*_0$ with $\tilde{\boldsymbol{y}}^*_0 \in H(\mathrm{div};\Omega)$ and $\mathrm{div}\,\tilde{\boldsymbol{y}}^*_0 = 0$ in $\Omega_m \cup \Omega_{IEL}$ we apply the patchwise flux reconstruction described in Remark 4.28.

We define the quantities $\mathrm{PREN}^{\mathrm{Up}}_{i,j}$, $\mathrm{REN}^{\mathrm{Up}}_{i,j,k,s}$, $\mathrm{RCEN}^{\mathrm{Low}}_{i,j,s}$, $\mathrm{P}^{\mathrm{CEN}}_{\mathrm{rel},i}$, $\mathrm{RCEN}^{\mathrm{Up}}_{i,j,s}$ in a similar fashion as for the component $\tilde{u}^N$ in the 2-term splitting. We should note that the bounds on the error in energy norm obtained by the majorant $\sqrt{2}M_{\oplus}(\tilde{u}_i, \tilde{\boldsymbol{y}}^*_i)$ in Table 4.22 are rather conservative and they could be improved by applying a flux reconstruction involving a higher order

Raviart-Thomas spaces, like RT_1. To obtain an idea of how much the error is overestimated, we can compare the values $\sqrt{2}M_{\oplus}(\tilde{u}_0, \tilde{\boldsymbol{y}}^*_{\bar{p}})$ and $\sqrt{2}M_{\oplus}(\tilde{u}_1, \tilde{\boldsymbol{y}}^*_{\bar{p}})$ of the majorant, evaluated with the last available approximation $\tilde{\boldsymbol{y}}^*_{\bar{p}}$ of the exact flux $\tilde{\boldsymbol{p}}^*$ to the values $\sqrt{2}M_{\oplus}(\tilde{u}_0, \tilde{\boldsymbol{y}}^*_0)$ and $\sqrt{2}M_{\oplus}(\tilde{u}_1, \tilde{\boldsymbol{y}}^*_1)$ evaluated with the current $\tilde{\boldsymbol{y}}^*_0$ and $\tilde{\boldsymbol{y}}^*_1$ at the first two meshes. We see that the overestimation is between around $125.21/70.117 \approx 1.785$ and $76.024/45.312 \approx 1.677$ times. This means that it is safe to assume that the real error $|||\nabla(\tilde{u} - \tilde{u}_{18})|||$ at the last level $\bar{p} = 18$ is no more than approximately $\sqrt{2}M_{\oplus}(\tilde{u}_{\bar{p}}, \tilde{\boldsymbol{y}}^*_{\bar{p}})/1.785 \approx 14.765$. On the other hand, owing to (4.61), we can define a minorant for the primal part $M^2_{\oplus}(\tilde{u}_i, \tilde{\boldsymbol{p}}^*) = \tilde{J}(\tilde{u}_i) - \tilde{J}(\tilde{u})$ of the error, where $\tilde{J} : H^1_0(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\tilde{J}(v) := \begin{cases} \int\limits_{\Omega} \left[\frac{\epsilon}{2} |\nabla v|^2 + \overline{k}^2 \cosh(v) - \boldsymbol{f}_{\tilde{\mathcal{G}}_3} \cdot \nabla v \right] dx, \ \text{ if } \overline{k}^2 \cosh(v) \in L^1(\Omega), \\ +\infty, \ \text{if } \overline{k}^2 \cosh(v) \notin L^1(\Omega), \end{cases} \tag{4.196}$$

For any $w \in H^1_0(\Omega)$ we have

$$\begin{aligned} 2M^2_{\oplus}(\tilde{u}_i, \tilde{\boldsymbol{p}}^*) &= |||\nabla(\tilde{u}_i - \tilde{u})|||^2 + 2D_F(\tilde{u}_i, -\Lambda^* \tilde{\boldsymbol{p}}^*) \\ &= 2\left(\tilde{J}(\tilde{u}_i) - \tilde{J}(\tilde{u})\right) \geq 2\left(\tilde{J}(\tilde{u}_i) - \tilde{J}(w)\right) =: 2M^2_{\ominus}(\tilde{u}_i, w). \end{aligned} \tag{4.197}$$

In practice, we always take for w the last available approximation $\tilde{u}_{\bar{p}}$. Moreover, from (4.49) we know that the nonlinear measure $2D_F(\tilde{u}_i, -\Lambda^* \tilde{\boldsymbol{p}}^*) \eqsim \|\tilde{u}_i - \tilde{u}\|^2_{L^2(\Omega_{ions})}$. Therefore this term converges faster than $|||\nabla(\tilde{u}_i - \tilde{u})|||^2$ (see Figure 4.3 on p. 115) and the minorant $\sqrt{2}M_{\ominus}(\tilde{u}_i, w)$ is approximately a lower bound also for the error in energy norm. From Table 4.24 we see that the primal part of the error $\sqrt{2}M_{\oplus}(\tilde{u}_h, \tilde{\boldsymbol{p}}^*)$ is between 64.949 and 70.117 on the initial mesh, and between 36.827 and 45.312 on the first adapted mesh. With this in mind, we obtain an overall guaranteed bound on the error in energy norm for the regular component u by using (4.181):

$$\begin{aligned} &|||\nabla(u - \tilde{u}_{18})||| \leq \sqrt{\epsilon_m} M_{\oplus,H}\left(\tilde{u}^H, T\left(\nabla \tilde{u}^H\right)\right) + \sqrt{2}M_{\oplus}(\tilde{u}_{18}, \tilde{\boldsymbol{y}}^*_{18}) \\ =&\sqrt{2} \times 18.410 + 26.356 = 52.391. \end{aligned}$$

For comparison, the energy norm of $\tilde{u}_{18}$ is equal to 300.05 (see Table 4.22).

Table 4.22: Example 4. System 1 recomputed with 3-term splitting without additional decomposition into u^L and u^N.

System 1: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\|\tilde{u}_i\|_0$	$\|\|\|\nabla\tilde{u}_i\|\|\|$	$\|\|\|\epsilon\nabla\tilde{u}_i - \tilde{\boldsymbol{y}}^*_i\|\|\|_*$	$\sqrt{2}M_\oplus(\tilde{u}_i, \tilde{\boldsymbol{y}}^*_i)$	$\tilde{J}(\tilde{u}_i)$
0	667 008	58.089	295.42	124.22	125.21	258 884 684.61
1	1 448 301	58.857	298.20	75.479	76.024	258 883 253.51
2	2 345 138	58.909	298.76	62.856	63.370	258 883 072.64
3	3 170 889	58.957	299.09	56.527	56.987	258 882 952.86
4	3 964 923	58.987	299.26	52.422	52.846	258 882 884.40
5	4 805 923	59.008	299.39	49.288	49.683	258 882 836.09
6	5 770 299	59.025	299.48	46.644	47.016	258 882 798.37
7	6 920 812	59.039	299.56	44.234	44.585	258 882 765.86
8	8 330 105	59.049	299.64	41.993	42.322	258 882 736.73
9	10 054 979	59.059	299.70	39.878	40.186	258 882 710.79
10	12 150 021	59.067	299.76	37.882	38.172	258 882 687.45
11	14 681 622	59.075	299.81	36.002	36.274	258 882 666.41
12	17 712 647	59.082	299.86	34.260	34.514	258 882 647.95
13	21 324 443	59.088	299.90	32.645	32.883	258 882 631.77
14	25 599 530	59.095	299.94	31.152	31.376	258 882 617.49
15	30 665 444	59.100	299.97	29.770	29.980	258 882 604.93
16	36 657 462	59.105	300.00	28.484	28.681	258 882 593.86
17	43 733 890	59.109	300.03	27.292	27.477	258 882 584.08
18	52 088 245	59.112	300.05	26.182	26.356	258 882 575.36

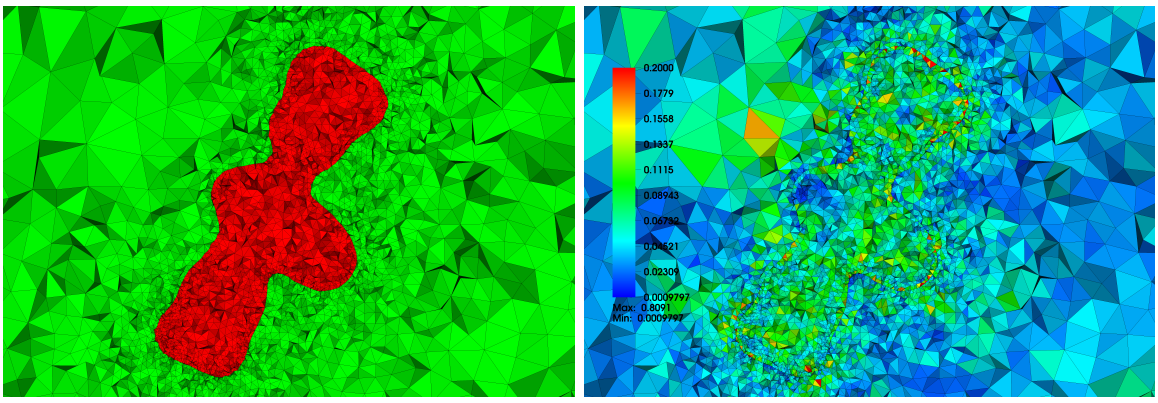


Figure 4.19: On the left: cross section of the mesh with the plane $y = 3\,\mathring{A}$ at level $i = 1$ in the mesh refinement procedure for finding the component $\tilde{u}$ in Example 4. The molecule region Ω_m is marked red (Alexa 594). On the right: error indicator as a piecewise constant function.

Table 4.23: Example 4. System 1 recomputed with 3-term splitting without additional decomposition into u^L and u^N.

System 1: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	#elements	$\mathrm{PREN}^{\mathrm{Up}}_{i,i}$[%]	$\mathrm{REN}^{\mathrm{Up}}_{i,i,i,i}$[%]	$\mathrm{RCEN}^{\mathrm{Low}}_{i,i,i}$[%]	$\mathrm{P}^{\mathrm{CEN}}_{\mathrm{rel},i}$[%]	$\mathrm{RCEN}^{\mathrm{Up}}_{i,i,i}$[%]
0	667 008	42.04	73.56	15.63	29.73	40.23
1	1 448 301	25.31	34.21	10.57	17.89	21.56
2	2 345 138	21.03	26.92	9.057	14.87	17.41
3	3 170 889	18.89	23.53	8.260	13.36	15.40
4	3 964 923	17.51	21.44	7.732	12.38	14.14
5	4 805 923	16.46	19.89	7.321	11.64	13.19
6	5 770 299	15.57	18.62	6.970	11.01	12.40
7	6 920 812	14.76	17.48	6.646	10.44	11.68
8	8 330 105	14.01	16.44	6.341	9.909	11.03
9	10 054 979	13.30	15.48	6.051	9.408	10.42
10	12 150 021	12.63	14.59	5.774	8.936	9.850
11	14 681 622	12.00	13.76	5.512	8.491	9.317
12	17 712 647	11.42	13.00	5.266	8.078	8.827
13	21 324 443	10.88	12.31	5.036	7.696	8.377
14	25 599 530	10.38	11.68	4.822	7.344	7.963
15	30 665 444	9.924	11.10	4.623	7.017	7.583
16	36 657 462	9.494	10.57	4.437	6.713	7.232
17	43 733 890	9.096	10.08	4.263	6.432	6.908
18	52 088 245	8.725	9.630	4.100	6.170	6.608

Table 4.24: Example 4. System 1 recomputed with 3-term splitting without additional decomposition into u^L and u^N.

System 1: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 2$, $\epsilon_s = 80$						
level i	$\mathrm{REN}^{\mathrm{Up}}_{i,\bar{p},\bar{p},\bar{p}}$ [%]	$\mathrm{RCEN}^{\mathrm{Low}}_{i,\bar{p},\bar{p}}$ [%]	$\mathrm{RCEN}^{\mathrm{Up}}_{i,\bar{p},\bar{p}}$ [%]	$\sqrt{2}M_{\ominus}(\tilde{u}_i, \tilde{u}_{\bar{p}})$	$\left\|\left\|\left\|\epsilon\nabla\tilde{u}_i - \tilde{\boldsymbol{y}}^*_{\bar{p}}\right\|\right\|\right\|_*$	$\sqrt{2}M_{\oplus}(\tilde{u}_i, \tilde{\boldsymbol{y}}^*_{\bar{p}})$
0	25.61	19.45	31.39	64.949	69.784	70.117
1	16.55	11.82	19.06	36.827	45.191	45.312
2	15.02	9.843	15.88	31.536	41.005	41.131
3	13.92	8.852	14.28	27.477	37.973	38.103

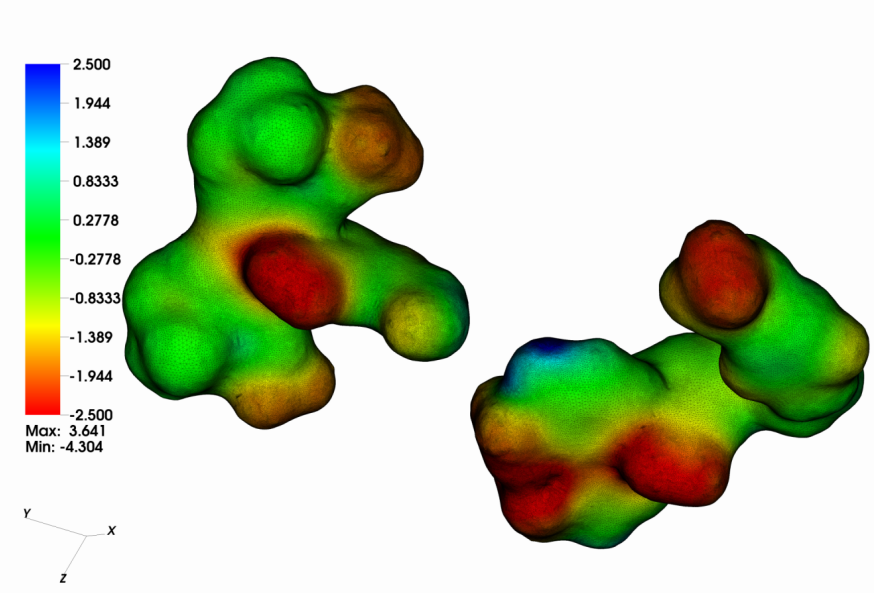


Figure 4.20: Full potential surface map with the 3-term splitting (without additional decomposition into $u^L + u^N$) for the system Alexa 488 and Alexa 594 in units k_BT/e_0. Blue color indicates a positive potential (values$> 2.5k_BT/e_0$) and red color indicates negative potential (values $< -2.5K_BT/e_0$).

Example 5: System 2 (Insulin protein, PDB ID: 1RWE), 3-term splitting without additional decomposition into u^L and u^N

For the third application, we consider the insulin protein with the crystal structure from Protein Data Bank (ID code 1RWE). This is a small protein that functions in the hormonal control of metabolism [184]. Because of its importance in the treatment of diabetes mellitus, it has attracted attention as a target of protein engineering. In recent years, insulin analogues have gained widespread clinical acceptance [42]. Despite such empirical success, the binding of insulin to the insulin receptor still raises many questions. Electrostatic interactions between molecules might contribute significantly to the binding mechanism. The Poisson-Boltzmann equation can be used to calculate the electrostatic surface of insulin, which could help to determine the binding sites. Moreover, the distribution of the electrostatic potential around the protein can be used in simulations of binding dynamics.

Here, the CHARMM-GUI web server was employed to add hydrogen atoms to the system. The total number of atoms (with the added hydrogens) is 1590. The charges for the calculations were taken from the *psf* file created by the CHARMM-GUI. The number of nonzero charges that define the simple Coulomb potential G in (3.11) is 1574 and the total charge of this protein is equal to $-4\,e_0$. In this example, the parameters are $\epsilon_m = 20$, $\epsilon_s = 80$, $I_s \approx 1.178\,M$ which corresponds to $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$ at $T = 298\,K$. Again, there is no ion exclusion layer, i.e., $\Omega_{IEL} = \emptyset$ and $\Omega_s \equiv \Omega_{ions}$. The initial surface mesh is created by NanoShaper [61] with *GridScale=2*.

Finding the harmonic component u^H

As in Example 4, we first have to obtain a conforming approximation $\tilde{u}^H$ of u^H by solving problem (3.29). We solve this problem on a sequence of adapted meshes using the error majornat (4.152) and the derived from it error indicator. To reconstruct the numerical flux $\nabla\tilde{u}^H$ and obtain $T(\nabla\tilde{u}^H)$ we apply the equilibrated patchwise flux reconstruction technique [30, 31] with RT_0 elements. Therefore, we have $\operatorname{div}\left(T\left(\nabla\tilde{u}^H\right)\right) = 0$. We have computed $\tilde{u}^H$ on a final mesh with 31 971 835 tetrahedrons. The corresponding value for the majorant in (4.154) is $M_{\oplus,H}\left(\tilde{u}^H, T\left(\nabla\tilde{u}^H\right)\right) = 9.178$ and for comparison $\|\nabla\tilde{u}^H\|_{L^2(\Omega_m)} = 149.680$. We obtain a guaranteed upper bound on the relative error in the $H^1(\Omega_m)$ seminorm

$$\frac{\|\nabla\left(\tilde{u}^H - u^H\right)\|_{L^2(\Omega_m)}}{\|\nabla u^H\|_{L^2(\Omega_m)}} \leq \frac{M_{\oplus,H}\left(\tilde{u}^H, T\left(\nabla\tilde{u}^H\right)\right)}{\|\nabla\tilde{u}^H\|_{L^2(\Omega_m)} - M_{\oplus,H}\left(\tilde{u}^H, T\left(\nabla\tilde{u}^H\right)\right)} = 6.53\,\%.$$

Finding the regular component $\tilde{u}$

Here, we find a conforming approximation $\tilde{u}_h$ of $\tilde{u}$, the exact solution of problem (4.175), by adaptively solving the Galerkin problem (4.178). Again, by $\tilde{u}_i$ and $\tilde{\boldsymbol{y}}^*_i$ we denote the finite element approximations at mesh refinement level i to $\tilde{u}$ and $\tilde{\boldsymbol{p}}^*$, respectively. Here $\bar{p} = 13$. In order to find a good approximation $\tilde{\boldsymbol{y}}^*$ of the exact flux $\tilde{\boldsymbol{p}}^*$ having the form $\tilde{\boldsymbol{y}}^* = \boldsymbol{f}_{\tilde{\mathcal{G}}_3} + \tilde{\boldsymbol{y}}^*_0$ with $\tilde{\boldsymbol{y}}^*_0 \in H(\operatorname{div};\Omega)$ and $\operatorname{div}\tilde{\boldsymbol{y}}^*_0 = 0$ in $\Omega_m \cup \Omega_{IEL}$ we apply the patchwise flux reconstruction

described in Remark 4.28.

As in Example 4, the bounds on the error in energy norm obtained by the majorant $\sqrt{2}M_{\oplus}(\tilde{u}_i, \tilde{\boldsymbol{y}}_i^*)$ in Table 4.25 could be improved by applying a flux reconstruction involving a higher order Raviart-Thomas spaces, like RT_1. To obtain an idea of how much the error is overestimated, we compare the values $\sqrt{2}M_{\oplus}(\tilde{u}_0, \tilde{\boldsymbol{y}}_{\bar{p}}^*)$ and $\sqrt{2}M_{\oplus}(\tilde{u}_1, \tilde{\boldsymbol{y}}_{\bar{p}}^*)$ of the majorant, evaluated with the last available approximation $\tilde{\boldsymbol{y}}_{\bar{p}}^*$ of the exact flux $\tilde{\boldsymbol{p}}^*$ to the values $\sqrt{2}M_{\oplus}(\tilde{u}_0, \tilde{\boldsymbol{y}}_0^*)$ and $\sqrt{2}M_{\oplus}(\tilde{u}_1, \tilde{\boldsymbol{y}}_1^*)$ evaluated with the current $\tilde{\boldsymbol{y}}_0^*$ and $\tilde{\boldsymbol{y}}_1^*$ at the first two meshes. We see that the overestimation is between around $199.91/163.38 \approx 1.223$ and $134.64/112.01 \approx 1.202$ times. This means that it is safe to assume that the real error $|||\nabla(\tilde{u} - \tilde{u}_{13})|||$ at the last level $\bar{p} = 13$ is no more than approximately $\sqrt{2}M_{\oplus}(\tilde{u}_{\bar{p}}, \tilde{\boldsymbol{y}}_{\bar{p}}^*)/1.223 \approx 46.78$. Moreover, from Table 4.27 we see that the primal part of the error $\sqrt{2}M_{\oplus}(\tilde{u}_h, \tilde{\boldsymbol{p}}^*)$ is between 152.95 and 163.38 on the initial mesh, and between 96.074 and 112.01 on the first adapted mesh. With this in mind, we obtain an overall guaranteed bound on the error in energy norm for the regular component u by using (4.181):

$$\begin{aligned} |||\nabla(u - \tilde{u}_{13})||| &\leq \sqrt{\epsilon_m} M_{\oplus,H}\left(\tilde{u}^H, T\left(\nabla \tilde{u}^H\right)\right) + \sqrt{2}M_{\oplus}(\tilde{u}_{13}, \tilde{\boldsymbol{y}}_{13}^*) \\ &= \sqrt{20} \times 9.178 + 57.210 = 93.436. \end{aligned}$$

For comparison, the energy norm of $\tilde{u}_{13}$ is equal to 559.21 (see Table 4.25).

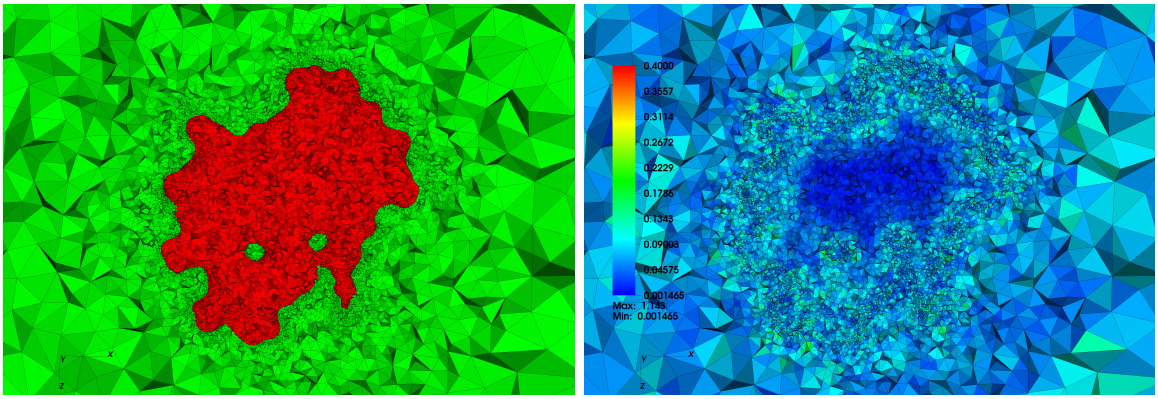


Figure 4.21: Cross section of the mesh with the plane $y = 15\,\mathring{A}$ at level $i = 2$ in the mesh refinement procedure for finding the regular component $\tilde{u}$ in Example 5. On the left, the molecule region Ω_m is marked red. On the right, error indicator as a piecewise constant function.

Table 4.25: Example 5. System 2.

System 2: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 20$, $\epsilon_s = 80$						
level i	#elements	$\|\tilde{u}_i\|_0$	$\|\|\|\nabla\tilde{u}_i\|\|\|$	$\|\|\|\epsilon\nabla\tilde{u}_i - \tilde{\boldsymbol{y}}_i^*\|\|\|_*$	$\sqrt{2}M_\oplus(\tilde{u}_i, \tilde{\boldsymbol{y}}_i^*)$	$\tilde{J}(\tilde{u}_i)$
0	724 737	100.70	541.07	197.54	199.91	5 393 609 145.91
1	1 523 717	101.76	551.97	133.27	134.64	5 393 602 063.00
2	2 325 989	101.92	554.50	113.47	114.69	5 393 600 517.46
3	3 082 380	102.02	555.68	102.78	103.86	5 393 599 760.05
4	3 880 178	102.09	556.42	95.072	96.062	5 393 599 280.10
5	4 785 000	102.14	556.97	88.861	89.780	5 393 598 923.77
6	5 850 441	102.19	557.41	83.527	84.386	5 393 598 635.46
7	7 140 525	102.22	557.79	78.690	79.493	5 393 598 388.08
8	8 713 471	102.25	558.11	74.286	75.037	5 393 598 175.65
9	10 625 629	102.28	558.39	70.215	70.917	5 393 597 989.90
10	12 948 272	102.31	558.63	66.431	67.087	5 393 597 826.84
11	15 764 841	102.33	558.85	62.913	63.526	5 393 597 682.94
12	19 162 081	102.35	559.04	59.670	60.242	5 393 597 557.40
13	23 234 788	102.37	559.21	56.676	57.210	5 393 597 447.87

Table 4.26: Example 5. System 2.

System 2: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 20$, $\epsilon_s = 80$						
level i	#elements	$\mathrm{PREN}^{\mathrm{Up}}_{i,i}[\%]$	$\mathrm{REN}^{\mathrm{Up}}_{i,i,i,i}[\%]$	$\mathrm{RCEN}^{\mathrm{Low}}_{i,i,i}[\%]$	$\mathrm{P}^{\mathrm{CEN}}_{\mathrm{rel},i}[\%]$	$\mathrm{RCEN}^{\mathrm{Up}}_{i,i,i}[\%]$
0	724 737	36.50	58.59	14.09	25.81	33.79
1	1 523 717	24.14	32.26	10.16	17.07	20.46
2	2 325 989	20.46	26.07	8.842	14.47	16.91
3	3 082 380	18.49	22.98	8.105	13.07	15.07
4	3 880 178	17.08	20.86	7.562	12.08	13.78
5	4 785 000	15.95	19.21	7.118	11.28	12.76
6	5 850 441	14.98	17.83	6.732	10.59	11.91
7	7 140 525	14.10	16.62	6.378	9.975	11.14
8	8 713 471	13.31	15.53	6.051	9.411	10.45
9	10 625 629	12.57	14.54	5.747	8.891	9.822
10	12 948 272	11.89	13.64	5.461	8.408	9.242
11	15 764 841	11.25	12.82	5.194	7.960	8.709
12	19 162 081	10.67	12.07	4.945	7.547	8.221
13	23 234 788	10.13	11.39	4.713	7.166	7.775

Table 4.27: Example 5. System 2.

System 2: $\overline{k}^2_{ions} = 10\,\mathring{A}^{-2}$, $\epsilon_m = 20$, $\epsilon_s = 80$						
level i	$\mathrm{REN}^{\mathrm{Up}}_{i,\bar{p},\bar{p},\bar{p}}$ [%]	$\mathrm{RCEN}^{\mathrm{Low}}_{i,\bar{p},\bar{p}}$ [%]	$\mathrm{RCEN}^{\mathrm{Up}}_{i,\bar{p},\bar{p}}$ [%]	$\sqrt{2}M_{\ominus}(\tilde{u}_i, \tilde{u}_{\bar{p}})$	$\left\|\left\|\left\|\epsilon\nabla\tilde{u}_i - \tilde{\boldsymbol{y}}^*_{\bar{p}}\right\|\right\|\right\|_*$	$\sqrt{2}M_{\oplus}(\tilde{u}_i, \tilde{\boldsymbol{y}}^*_{\bar{p}})$
0	32.54	16.43	27.17	152.95	162.83	163.38
1	22.31	11.08	18.29	96.074	111.67	112.01
2	19.36	9.438	15.58	78.352	96.842	97.196
3	17.73	8.548	14.11	68.002	88.649	89.017

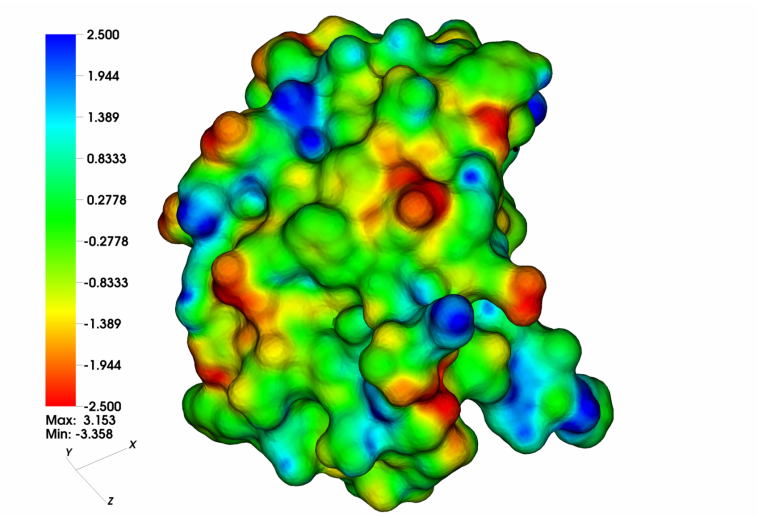


Figure 4.22: Full potential surface map of the insulin protein (PDB ID: 1RWE) in units k_BT/e_0. Blue color indicates a positive potential (values$> 2.5k_BT/e_0$) and red color indicates negative potential (values $< -2.5K_BT/e_0$).

Example 6: System 3 (SecYEG protein membrane channel with ion exclusion layer), 3-term splitting without additional decomposition into u^L and u^N

In this example, the computations are performed on the SecYEG membrane protein channel with an open plug and with a signal sequence in the lateral gate. The channel is located in the plasma membrane of bacteria and provides a lateral exit into the bilayer for membrane proteins, while simultaneously offering a pathway into the aqueous interior for secreted proteins. The molecular mechanisms that determine the functionality of the channel for these two pathways and driving forces of the translocation are not comprehensively understood. Important contributions to both may come from the electrostatic interactions between the SecYEG and translocated peptide.

The pdb file with the atoms positions is taken from a frame in a molecular dynamics

simulation which includes 46 373 atoms, 9 563 of which belong to the SecYEG. Here, the number of nonzero charges that define the simple Coulomb potential G in (3.11) is 7 305 and the total charge of the SecYEG protein channel is equal to $13\, e_0$. With the help of the open-source 3D computer graphics software Blender [18], the original membrane, which is created by NanoShaper [61] with *GridScale=2*, is embedded in a big slab parallel to the XY coordinate plane extending to the boundaries of the computational domain Ω, a cube with dimensions $1000 \times 1000 \times 1000\, \mathring{A}^3$. The slab has a thickness equal to the average thickness of the original membrane, i.e., $38.84\, \mathring{A}$. Ion exclusion layer (Ω_{IEL}) with a tickness of $2\, \mathring{A}$ is added around the membrane and the SecYEG channel. The dielectric coefficient in the channel and the membrane is $\epsilon_m = 4$ and it is varying in the solvent region Ω_s, 80 in the part of the solvent region that is outside the region between the two planes parallel to the XY coordinate plane (dashed lines on Figure 4.23) and it decreases linearly to a value of 4 at $z = z_0 = -4\, \mathring{A}$ in the solvent region. The two planes bounding the region of varying dielectric $\epsilon_s(x)$ are at $z = -17.49\, \mathring{A}$ and $z = 9.49\, \mathring{A}$, respectively. The initial mesh has 10 287 866 tetrahedral elements and a total of $\bar{p} = 3$ mesh refinement steps are made.

Introducing the membrane region into the continuum electrostatic models extends the applicability of the PBE to treating membrane channel proteins. However, incorporating the membrane with the embedded in it membrane channel in finite element calculations is a hard task. Recently, an automated method to generate meshes for implicit membrane solvation models, containing membrane transport proteins, was developed in [128]. This methodology in conjunction with a finite element solution, but without adaptivity, of the PBE for membrane channel proteins was presented in [109]. There, the membrane is again approximated as a solvent-inaccessible planar low-dielectric slab, but no ion exclusion layer around the membrane and the channel is created.

Finding the harmonic component u^H

First, we compute an approximation $\tilde{u}^H$ of u^H. The initial triangulation of the molecular domain Ω_m consists of 4 113 729 tetrahedrons. In Table 4.28 are shown all 5 levels of adaptive mesh refinement and the corresponding values for the norm $\|\nabla \tilde{u}^H\|_{L^2(\Omega_m)}$ as well as for the majorant $M_{\oplus,H}\left(\tilde{u}_i^H, T\left(\nabla \tilde{u}_i^H\right)\right)$ and the approximate upper bound $\frac{M_{\oplus,H}\left(\tilde{u}_i^H, T\left(\nabla \tilde{u}_i^H\right)\right)}{\|\nabla \tilde{u}_i^H\|_{L^2(\Omega_m)}}$ [%] for the relative error. Here, $T\left(\nabla \tilde{u}^H\right) \in RT_0$ is the patchwise reconstruction from [30, 31] of the numerical flux $\nabla \tilde{u}^H$.

Finding the regular component $\tilde{u}$

Here, we find a conforming approximation $\tilde{u}_h$ of $\tilde{u}$, the exact solution of problem (4.175), by adaptively solving the Galerkin problem (4.178) with $T\left(\nabla \tilde{u}_5^H\right)$ from the last refinement level. Again, $\tilde{u}_i$ and $\tilde{\boldsymbol{y}}_i^*$ denote the finite element approximations at mesh refinement level i to $\tilde{u}$ and $\tilde{\boldsymbol{p}}^*$, respectively. Again, to obtain a good approximation of $\tilde{\boldsymbol{p}}^*$ we apply the flux reconstruction from Remark 4.28. On the last refinement level $\bar{p} = 3$ we achieve an

approximate upper bound for the relative error in energy norm for the approximation $\tilde{u}_3$ of $\tilde{u}$ of 16.09 % (see Table 4.29). Here we note that this estimate is rather conservative. This is evident from the fact that at level $i = 0$, the ratio between the upper and lower bounds on the primal part of the error is more than 2 (see columns 4 and 5 in Table 4.29). The upper bound, i.e., the value of the majorant $\sqrt{2}M_{\oplus}(\tilde{u}_i, \tilde{\boldsymbol{y}}_i^*)$ can be improved if we consider a flux reconstruction in a bigger subspace of $H(\mathrm{div};\Omega)$, like RT_1. According to (4.181) we have the following estimate:

$$\begin{aligned} &|||\nabla(u - \tilde{u}_3)||| \leq \sqrt{\epsilon_m} M_{\oplus,H}\left(\tilde{u}^H, T\left(\nabla\tilde{u}^H\right)\right) + \sqrt{2}M_{\oplus}(\tilde{u}_3, \tilde{\boldsymbol{y}}_3^*) \\ =&\sqrt{4} \times 116.98 + 245.72 = 479.68. \end{aligned}$$

For comparison the energy norm of the approximate regular component $\tilde{u}_3$ is $|||\nabla\tilde{u}_3||| = 1526.84$.

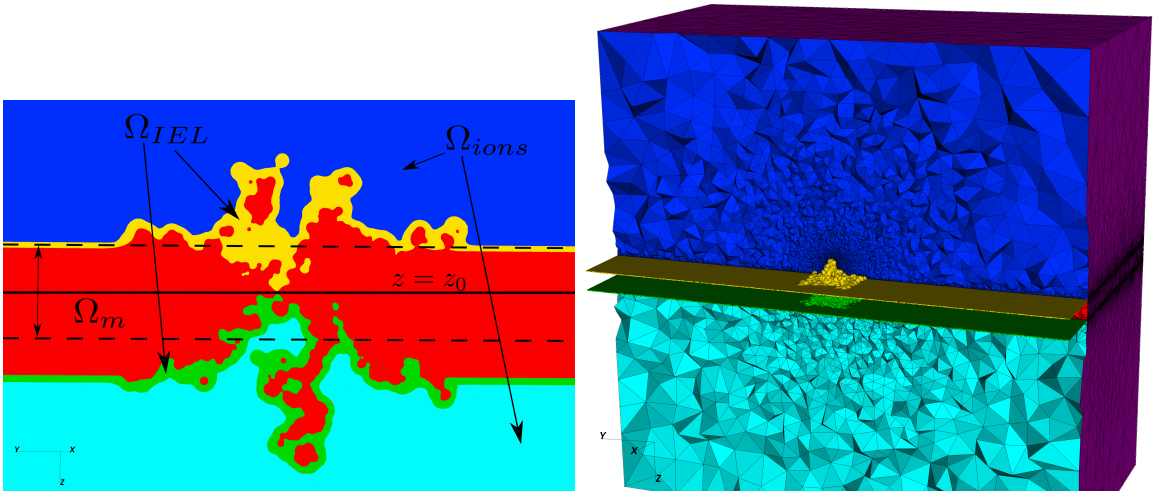


Figure 4.23: On the left: different regions for the SecYEG channel with an ion exclusion layer (IEL). On the right: mesh at refinement level $i = 3$. Molecular region is maerked in red, whereas the ion exclusion layer is in yellow and green respectively above and below the membrane.

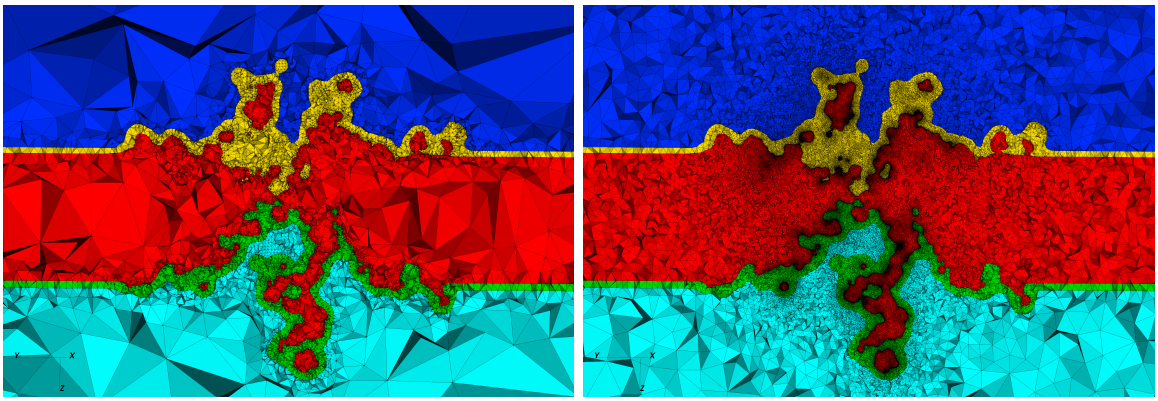


Figure 4.24: On the left: initial mesh. On the right: mesh at refinement level $i = 3$.

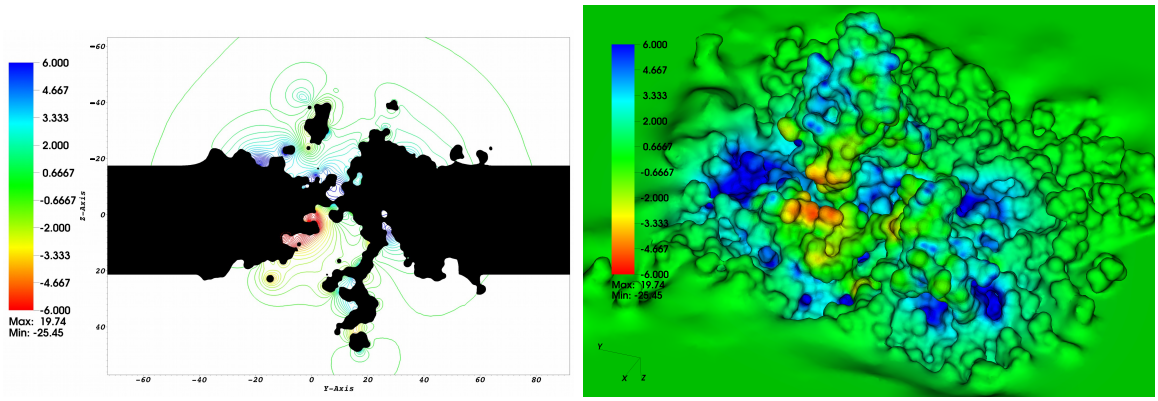


Figure 4.25: On the left: electrostatic potential φ in units $\frac{K_B T}{e_0}$ at mesh refinement level $i = 3$. On the right: surface potential map in units $\frac{K_B T}{e_0}$ at mesh refinement level $i = 3$.

Table 4.28: Example 6. System 3. Here $J_H : H^1_{-G}(\Omega_m) \to \mathbb{R}$ is defined by $J_H(v) := \int_{\Omega_m} \frac{1}{2} |\nabla v|^2 \, dx$

System 3: $I_s = 0.1\,M$, $\overline{k}^2_{ions} = 0.843\,\mathring{A}^{-2}$, $T = 300\,K$, $\epsilon_m = 4$, $\epsilon_s = 80$					
level i	# elements	$\|\nabla \tilde{u}_i^H\|_{L^2(\Omega_m)}$	$M_{\oplus,H}\left(\tilde{u}_i^H, T\left(\nabla \tilde{u}_i^H\right)\right)$	$J_H\left(\tilde{u}_i^H\right)$	$\frac{M_{\oplus,H}\left(\tilde{u}_i^H, T\left(\nabla \tilde{u}_i^H\right)\right)}{\|\nabla \tilde{u}_i^H\|_{L^2(\Omega_m)}}$ [%]
0	4 113 729	1662.3557	544.81	1 381 713.30	32.77
1	9 859 418	1644.8591	264.61	1 352 780.89	16.08
2	19 497 424	1642.8727	185.06	1 349 515.49	11.26
3	33 167 516	1642.0064	150.29	1 348 092.61	9.153
4	48 298 685	1641.4157	130.73	1 347 122.81	7.964
5	65 604 295	1641.0111	116.98	1 346 458.68	7.128

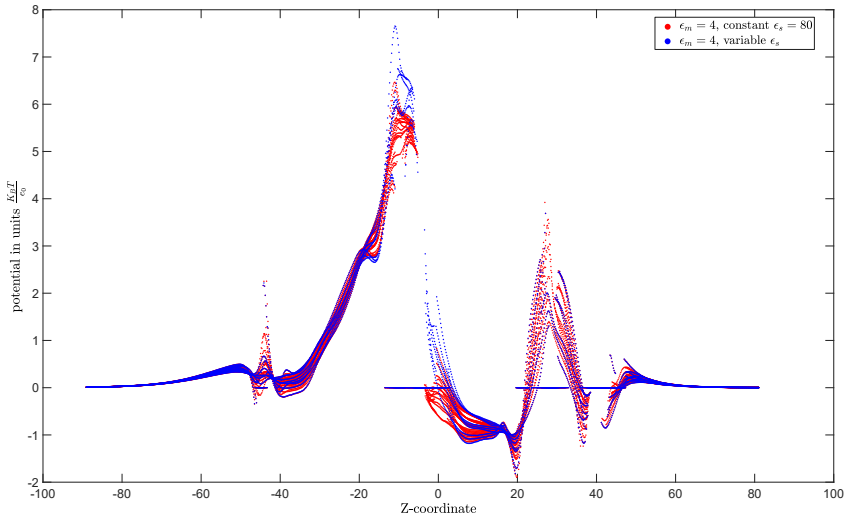


Figure 4.26: Potential in units $k_B T/e_0$ along 9 segments parallel to the Z-axis and passing through 9 uniformly distributed points in a rectangle in the $XY-$plane with a center at $(4.05, 0.55)$ and sides $2.1\,\mathring{A}$ and $2.3\,\mathring{A}$. The rectangle is chosen in such a way that most of the lines pass through the channel. Only the potential outside the channel is plotted, i.e., we have plotted only the regular component u in the 3-term splitting. The zero values of the potential indicate that at these coordinates the particular segment crosses the interior of the SecYEG channel (the region Ω_m). In blue are the values of the potential computed with a variable dielectric coefficient $\epsilon_s(x)$ and in red are the values computed with a constant $\epsilon_s = 80$.

Table 4.29: Example 6. System 3.

System 3: $I_s = 0.1\,M$, $\overline{k}^2_{ions} = 0.843\,\mathring{A}^{-2}$, $T = 300\,K$, $\epsilon_m = 4$, $\epsilon_s = 80$ with $\tilde{u}_5^H$						
level i	# elements	$\|\|\|\nabla\tilde{u}_i\|\|\|$	$\sqrt{2}M_{\ominus}(\tilde{u}_i, \tilde{u}_{\bar{p}})$	$\sqrt{2}M_{\oplus}(\tilde{u}_i, \tilde{\boldsymbol{y}}_i^*)$	$\tilde{J}\,(\tilde{u}_i)$	$\frac{\sqrt{2}M_{\oplus}(\tilde{u}_i, \tilde{\boldsymbol{y}}_i^*)}{\|\|\|\nabla\tilde{u}_i\|\|\|}$ [%]
0	10 287 866	1462.18	446.41	996.39	805 687 862.26	68.14
1	18 713 711	1514.89	191.40	466.35	805 606 537.60	30.78
2	43 525 735	1524.72	80.102	293.41	805 591 428.09	19.24
$\bar{p} = 3$	79 867 368	1526.84	-	245.72	805 588 219.90	16.09

Chapter 5

Goal-oriented error estimates

The purpose of this chapter is to derive goal-oriented error estimates for the electrostatic interaction between molecules and to apply and verify them in practice. In Section 5.1, we start by stating the primal problem that defines the electrostatic potential in a system of biomolecules and introduce the electrostatic interaction between molecules in terms of a goal functional. In order to have a consistent description of the symmetric in nature electrostatic interaction, one needs to consider the Linearized Poisson-Boltzmann equation with a homogeneous Dirichlet boundary condition as the primal problem governing the electrostatic potential.

In Section 5.2, we continue with a brief overview of the dual weighted residual (DWR) method, a goal-oriented error estimation approach that involves the adjoint problem. For regular goal functionals, which are also in L^2, and primal problems with a regular right-hand side, also in L^2, we show the steps in the derivation of an error representation for the goal quantity in terms of the unknown adjoint solution. Based on this error representation, one can obtain effective elementwise and nodewise error indicators (5.17) and (5.18). Since these error indicators involve the unknown solution z of the adjoint problem, we also comment on the error that one makes when using an approximation of z.

Further, in Section 5.3, we proceed with the derivation of goal-oriented error estimates for the electrostatic interaction between molecules. Now, the right-hand side of the primal problem is not a regular functional, i.e., it is not representable by a locally summable function, and it is also not in $H^{-1}(\Omega)$. In the case of the 2-term splitting, it is a surface density, representing the jump in the normal component of the displacement field, and in the case of no splitting applied, it is a linear combination of delta functions. What is more, the goal functional is also not regular and also does not belong to $H^{-1}(\Omega)$: it is a linear combination of pointwise evaluations, which are naturally expressed in terms of delta functions. It is clear that the steps involved in the derivation of the error representation in the case of a regular L^2 goal functional and a primal problem with a regular L^2 right-hand side become nontrivial for the problem that we are interested in. A standard approach in such situations is to regularize the goal functional

by replacing the point evaluations with averaging over small balls $B(x_{i,1}, \rho)$, where ρ is the radius around the points of interest $x_{i,1}$. In our case, the potential is a harmonic function in a neighborhood of $x_{i,1}$, and therefore such an averaging is equivalent to taking point evaluations. In Section 5.3.1 we derive **error estimate 1** for the electrostatic interaction due to the reaction field part of the potential, where the goal functional is regularized by using averaging over balls. Since the reaction field potential is in H^1 and the solution z of the adjoint problem is in H^1, this is the only case in which all the steps in the derivation of the representation of the error in the goal quantity are straight forward. We obtain one more error estimate (**error estimate 3**) which involves regularization of the goal functional and which is applied directly to the full potential. The derivation of the corresponding error representation in terms of the adjoint solution z is based on Proposition 5.10, which is the main result in Section 5.3.3.

In general, regularizing the goal functional by means of averaging over balls, changes the goal functional and requires the computation of integrals of discontinuous functions over balls. To achieve a high enough accuracy, the mesh needs to be a priori adapted around the points of interest and special integration rules have to be used. In order to avoid the conceptual shortcomings of this approach we derive two more representations of the error (**error estimate 2** and **error estimate 4**) in the goal quantity which do not require averaging and directly exploit the original goal functional. In Section 5.3.2 we present **error estimate 2** for the electrostatic interaction due to the reaction field potential in which the goal quantity is a linear combination of pointwise evaluations. The main result in this section is Proposition 5.6 which is the basis for the derivation of the error representation for the goal quantity. Finally, in Section 5.3.4, we obtain **error estimate 4** for the full electrostatic interaction where the goal functional is again formed by a linear combination of pointwise evaluations. In this case, the derivation of the error representation in the goal quantity is based on Proposition 5.12, which is also the main result in the section.

In Section 5.4 we present a large collection of tests performed on problems with analytically known solutions and demonstrate the efficiency and optimality of the derived error estimates. Additionally, we compare our results to the results obtained with the software package MEAD (Macroscopic Electrostatics with Atomic Detail) *version 2.2.8a*, a well-known representative of the solvers utilizing the finite difference method with uniform Cartesian grids. We end this chapter with Section 5.5, where we present two practical biophysical applications related to the computation of Fröster resonance energy transfer (FRET) and the electrostatic interaction between chromophores in their ground state. Again, a comparison with MEAD is performed.

5.1 Electrostatic interaction between two molecules

In Chapter 4 we have derived error estimates that give guaranteed bounds on the error measured in energy norm. These are global estimates that provide a general idea of the

quality of a computed solution. However, in many applications one is interested not in the solution itself over the whole computational domain, but rather in a specific quantity that depends on the solution. Often these quantities of interest have local nature and can be expressed in terms of a linear functional, called a goal functional. For example, if one is interested in the average value of a solution u over a small region $\omega \subset \Omega$, the goal functional is

$$\langle \mathscr{L}, u \rangle = \frac{1}{|\omega|} \int_{\omega} u dx.$$

The methods of estimating the error $\langle \mathscr{L}, u - \tilde{u} \rangle$ between the exact solution u and the approximate solution $\tilde{u}$ in terms of the goal functional often involve the so-called adjoint boundary value problem whose right-hand side is formed by $\mathscr{L}$ (see, e.g., [4,144]). Convergence analysis for adaptive finite element methods based on goal oriented error estimates can be found, for example, in [78, 104, 105].

In this chapter we derive goal-oriented error estimates for the electrostatic interaction between two molecules whose interior regions are denoted by $\Omega_{m,1}$ and $\Omega_{m,2}$, and their ion exclusion layers by $\Omega_{IEL,1}$ and $\Omega_{IEL,2}$ (see Figure 5.1). With this in mind, we can write $\Omega_m = \Omega_{m,1} \cup \Omega_{m,2}$, $\Omega_{IEL} = \Omega_{IEL,1} \cup \Omega_{IEL,2}$, and $\Gamma = \Gamma_1 \cup \Gamma_2$, where Γ_1 and Γ_2 are the Lipschitz boundaries of $\Omega_{m,1}$ and $\Omega_{m,2}$, respectively.

In the biophysical applications that we will consider, $\Omega_{m,1}$ and $\Omega_{m,2}$ are occupied by two chromophores (Alexa 594 and Alexa 488) and Ω_s is occupied by a dielectric, such as water in which moving ions, such as Na^+ and Cl^-, can be present. The presented theory and methods apply for more general configurations with more molecular domains, $\Omega_{m,1}, \ldots, \Omega_{m,l}$, $l \in \mathbb{N}$, and they are not limited to the applications that we present. When the two molecules are very close to each other, the regions $\Omega_{m,1}$ and $\Omega_{m,2}$ cannot be distinguished and are merged into one connected region Ω_m. Similarly, if the distance between $\overline{\Omega_{m,1}}$ and $\overline{\Omega_{m,2}}$ is less than $2R_{ion}$, $\Omega_{IEL,1}$ and $\Omega_{IEL,2}$ are merged into Ω_{IEL}, where R_{ion} is the thickness of the ion exclusion layer.
To model the electrostatic interaction between the two chromophores, we will use the linearized Poisson-Boltzmann equation (3.9) for the potential φ with dimension $[\varphi] = \left[\frac{\text{charge}}{\text{length}}\right]$ in which we will assume that $g = 0$. This ensures linearity of the problem and in particular ensures that the electrostatic interaction does not depend on whether the potential is created by the partial charges of one or the other chromophore.

We use the following notation. For short, by Dye I we denote Alexa 594, occupying $\Omega_{m,1}$, and by Dye II we denote Alexa 488, occupying $\Omega_{m,2}$. Further, by E_{1-2} we denote the electrostatic interaction computed by the formula $E_{1-2} = \sum_{i=1}^{N_2} \varphi_1(x_{i,2}) q_{i,2}$, where φ_1 is the electrostatic potential created by the partial charges of Dye I, N_2 is the number of partial charges in Dye II, $q_{i,2} = z_{i,2} e_0$, $i = 1, 2, \ldots, N_2$ are the charges of Dye II, and $x_{i,2}$, $i = 1, 2, \ldots, N_2$

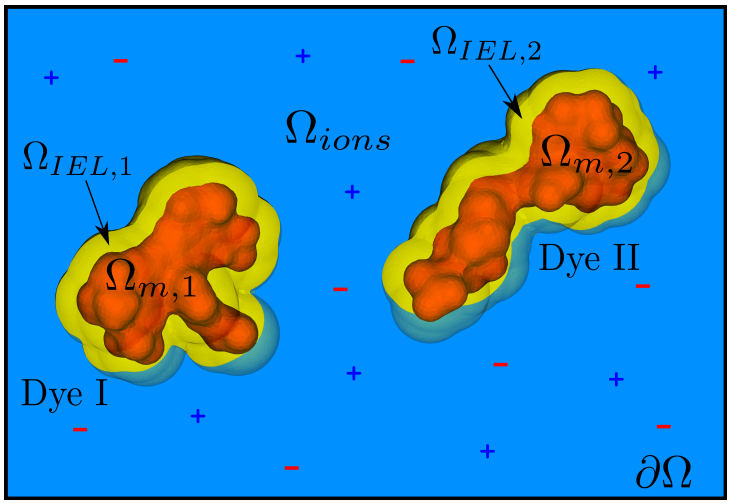


Figure 5.1: Different regions in the case $I_s \neq 0$

are thier coordinates. Similarly, E_{2-1} denotes the electrostatic interaction computed by the formula $E_{2-1} = \sum_{i=1}^{N_1} \varphi_2(x_{i,1}) q_{i,1}$. Here, N_1 is the number of partial charges in Dye I, $q_{i,1} = z_{i,1} e_0$, $i = 1, 2, \ldots, N_1$ are its partial charges, and $x_{i,1}$, $i = 1, 2, \ldots, N_1$ are the corresponding coordinates. In what follows, the primal problem that we consider is the LPBE for the potential φ_2, $[\varphi_2] = \left[\frac{\text{charge}}{\text{length}}\right]$, created by the charges in Dye II (see Figure 5.1)

$$-\nabla \cdot (\epsilon \nabla \varphi_2) + \overline{k}^2 \varphi_2 = 4\pi e_0 \sum_{i=1}^{N_2} z_{i,2} \delta_{x_{i,2}} =: \mathscr{F}_2 \quad \text{in } \Omega, \tag{5.1a}$$

$$[\varphi_2]_\Gamma = 0, \tag{5.1b}$$

$$[\epsilon \nabla \varphi_2 \cdot n]_\Gamma = 0, \tag{5.1c}$$

$$\varphi_2 = 0 \quad \text{on } \partial\Omega \tag{5.1d}$$

and the weak formulation that φ_2 satisfies reads

$$\varphi_2 \in \bigcap_{p < \frac{d}{d-1}} W_0^{1,p}(\Omega), \quad \int_\Omega \epsilon \nabla \varphi_2 \cdot \nabla v dx + \int_\Omega \overline{k}^2 \varphi_2 v dx = \langle \mathscr{F}_2, v \rangle, \ \forall v \in \bigcup_{q > d} W_0^{1,q}(\Omega). \tag{5.2}$$

We remind that since we impose a homogeneous Dirichlet boundary condition on $\partial\Omega$, it does not matter whether the electrostatic interaction is computed with the potential created by the charges of one or the other dye and there is no loss of generality by considering the LPBE for φ_2. Indeed, the weak formulation for φ_1 reads

$$\varphi_1 \in \bigcap_{p < \frac{d}{d-1}} W_0^{1,p}(\Omega), \quad \int_\Omega \epsilon \nabla \varphi_1 \cdot \nabla v dx + \int_\Omega \overline{k}^2 \varphi_1 v dx = \langle \mathscr{F}_1, v \rangle, \ \forall v \in \bigcup_{q > d} W_0^{1,q}(\Omega), \tag{5.3}$$

where $\mathscr{F}_1 := 4\pi e_0 \sum_{i=1}^{N_1} z_{i,1} \delta_{x_{i,1}}$. In Proposition 5.12 we will prove that we can test (5.2) with

φ_1 and (5.3) with φ_2 which implies

$$\langle \mathscr{F}_2, \varphi_1 \rangle = \int_\Omega \epsilon \nabla \varphi_2 \cdot \nabla \varphi_1 dx + \int_\Omega \bar{k}^2 \varphi_2 \varphi_1 dx = \langle \mathscr{F}_1, \varphi_2 \rangle. \tag{5.4}$$

Therefore

$$\begin{aligned} 4\pi \sum_{i=1}^{N_2} q_{i,2} \varphi_1(x_{i,2}) &= 4\pi \sum_{i=1}^{N_1} q_{i,1} \varphi_2(x_{i,1}) \\ \Leftrightarrow E_{1-2} &= E_{2-1} =: E. \end{aligned}$$

However, for the nonlinear PBE we cannot make this conclusion and in general $E_{2-1} \neq E_{1-2}$. We further introduce the following notation. By G_2 and G_1, $[G_i] = \left[\frac{\text{charge}}{\text{length}}\right]$, $i = 1, 2$, we denote the Coulomb potentials created by the charges in Dye II and Dye I, respectively.

$$G_1 := \sum_{i=1}^{N_1} \frac{q_{i,1}}{\epsilon_m |x - x_{i,1}|}, \qquad G_2 := \sum_{i=1}^{N_2} \frac{q_{i,2}}{\epsilon_m |x - x_{i,2}|} \tag{5.5}$$

Similarly, by u_2 and u_1, $[u_i] = \left[\frac{\text{charge}}{\text{length}}\right]$, $i = 1, 2$, we denote the reaction field part of the total potential φ_2 and φ_1, respectively. Finally, by E_{G_2}, E_{u_2}, E_{G_1}, and E_{u_1} we denote the electrostatic interaction corresponding to G_2, u_2, G_1, u_1, respectively, and they are defined as follows:

$$\begin{aligned} E_{G_2} &:= \sum_{i=1}^{N_1} G_2(x_{i,1}) q_{i,1}, & E_{u_2} &: = \sum_{i=1}^{N_1} u_2(x_{i,1}) q_{i,1} \\ E_{G_1} &:= \sum_{i=1}^{N_2} G_1(x_{i,2}) q_{i,2}, & E_{u_1} &: = \sum_{i=1}^{N_2} u_1(x_{i,2}) q_{i,2} \end{aligned} \tag{5.6}$$

Since $\varphi_1 = G_1 + u_1$ and $\varphi_2 = G_2 + u_2$, with the notation above, it holds

$$E_{2-1} = E_{G_2} + E_{u_2} \quad \text{and} \quad E_{1-2} = E_{G_1} + E_{u_1}. \tag{5.7}$$

As we mentioned earlier, we assume that we have partial charges only in Dye II. In order to find E_{2-1} we need to compute the values of the potential φ_2 or u_2 at the positions $x_{i,1}$ of the partial charges in Dye I. More precisely, we are interested in the accurate evaluation of the quantity of interest E_{2-1} or E_{u_2}. Mathematically, the goal functional can be written as

$$\frac{1}{4\pi} \mathscr{F}_1 = \sum_{i=1}^{N_1} q_{i,1} \delta_{x_{i,1}}. \tag{5.8}$$

Therefore, assuming that φ_2 and u_2 are continuous at least in a neighborhood of $\{x_{i,1}\}_{i=1}^{N_1}$, then we can write

$$E_{2-1} = \langle \frac{1}{4\pi} \mathscr{F}_1, \varphi_2 \rangle \quad \text{and} \quad E_{u_2} = \langle \frac{1}{4\pi} \mathscr{F}_1, u_2 \rangle. \tag{5.9}$$

We will see in Section 5.3.2 and Section 5.3.4 that these quantities are indeed well defined since the functions φ_2 and u_2 have higher regularity in a neighborhood of the point evaluations involved in $\frac{1}{4\pi}\mathscr{F}_1$.

According to Theorem 3.4, the equation defining the reaction field potential $u_2 \in H^1_{-G_2}(\Omega)$ in the splitting $\varphi_2 = G_2 + u_2$ is

$$\int_\Omega \epsilon \nabla u_2 \cdot \nabla v dx + \int_\Omega \overline{k}^2 u_2 v dx = \int_{\Omega_s} (\epsilon_m - \epsilon_s) \nabla G_2 \cdot \nabla v dx - \int_\Omega \overline{k}^2 G_2 v dx =: \langle \mathcal{T}_2, v \rangle \tag{5.10}$$

for all $v \in H^1_0(\Omega)$.

5.2 A general goal-oriented error estimate approach

In this section, we show the (general) form of the goal-oriented a posteriori error estimates that involve the adjoint problem by considering primal problems with regular right hand-side and regular goal functional. For this, we use the paradigm of the problem

$$\begin{aligned} &\text{Find } u \in H^1_g(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla u \cdot \nabla v dx = \int_\Omega f v dx =: \langle \mathscr{F}, v \rangle, \quad \text{for all } v \in H^1_0(\Omega) \end{aligned} \tag{5.11}$$

where the right hand side is a regular functional represented by the function $f \in L^2(\Omega)$ (see [138]), $g \in C^{0,1}(\partial\Omega)$, and $\epsilon \geq \epsilon_0 > 0$ is a bounded function. For a natural number k, and a mesh parameter h, by V_h^k we denote the finite element space defined on a mesh $\mathscr{T}_h$ with continuous piecewise polynomial functions of degree k. Further, for a function $g \in H^{1/2}(\partial\Omega)$, by $V_{g,h}^k$ we denote the subset of V_h^k in which all functions have boundary values equal to g. Let u_h denote a Galerkin approximation from the finite element space $V_{g,h}^k \subset H^1_g(\Omega)$ defined on a triangulation $\mathscr{T}_h$ with piecewise polynomial functions of degree k. Let the quantity of interest $\langle \mathscr{Z}, u \rangle$ be also a regular functional represented by the $L^2(\Omega)$ function l. In this case

$$\langle \mathscr{Z}, u \rangle = \int_\Omega l u dx$$

and we want to derive an estimate for the quantity $\langle \mathscr{Z}, u - u_h \rangle$. We introduce the adjoint problem of (5.11), given by

$$\begin{aligned} &\text{Find } z \in H^1_0(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla v \cdot \nabla z dx = \int_\Omega l v dx = \langle \mathscr{Z}, v \rangle \quad \text{for all } v \in H^1_0(\Omega) \end{aligned} \tag{5.12}$$

and a Galerkin approximation z_τ from a finite element space $V_{0,\tau}^{k^*} \subset H^1_0(\Omega)$ with piecewise polynomial functions of degree k^* on a mesh $\mathscr{T}_\tau$ that may or may not coincide with $\mathscr{T}_h$. We

can express the error measured in terms of the goal functional $\mathscr{Z}$ in the following way:

$$\begin{aligned}\langle \mathscr{Z}, u-u_h\rangle &= \int_\Omega \epsilon\nabla(u-u_h)\cdot\nabla z dx \\ =\langle \mathscr{F}, z\rangle &- \int_\Omega \epsilon\nabla z\cdot\nabla u_h dx = (f,z) - (\epsilon\nabla z, \nabla u_h) =: E(u_h, z).\end{aligned} \tag{5.13}$$

Thus $\langle \mathscr{Z}, u-u_h\rangle$ can be computed provided that z is available. In practice, one has to find an approximation z_τ of z. In this case it is better to rewrite $E(u_h, z)$ in the form $E(u_h, z) = E(u_h, z_\tau) + \mathcal{E}(u, u_h, z, z_\tau)$ (see [138]), where

$$\begin{aligned}\mathcal{E}(u, u_h, z, z_\tau) &= \langle \mathscr{F}, z-z_\tau\rangle - (\epsilon\nabla u_h, \nabla(z-z_\tau)) \\ &= \int_\Omega \epsilon\nabla(u-u_h)\cdot\nabla(z-z_\tau)dx\end{aligned} \tag{5.14}$$

If $V_{0,h}^k \equiv V_{0,\tau}^{k^*}$, due to Galerkin orthogonality, it follows that $E(u_h, z_\tau) = 0$ and $\langle \mathscr{Z}, u-u_h\rangle = \mathcal{E}$. In the case $V_{0,h}^k \neq V_{0,\tau}^{k^*}$ we have the approximate equality $\langle \mathscr{Z}, u-u_h\rangle \approx E(u_h, z_\tau)$. Indeed, if we have sufficiently regular solutions, for example, $u, z \in H^2(\Omega)$, and a sequence of regular triangulations $\mathscr{T}_h$ and $\mathscr{T}_\tau$, since $z_\tau \to z$ as $\tau \to 0$, $|||\nabla z_\tau|||$ is bounded and

$$|E(u_h, z_\tau)| = \left|\int_\Omega \epsilon\nabla(u-u_h)\cdot\nabla z_\tau dx\right| \leq |||\nabla(u-u_h)|||\,|||\nabla z_\tau||| = O(h). \tag{5.15}$$

For $\mathcal{E}$ we have

$$\begin{aligned}|\mathcal{E}| &= \left|\int_\Omega \epsilon\nabla(u-u_h)\cdot\nabla(z-z_\tau)dx\right| \\ &\leq |||\nabla(u-u_h)|||\,|||\nabla(z-z_\tau)||| = O(h\tau).\end{aligned} \tag{5.16}$$

Therefore $\mathcal{E}$ has a higher order of convergence than $E(u_h, z_\tau)$ which means that asymptotically, $E(u_h, z_\tau)$ contains the major part of the error. (see [115, 139]) For example, H^2 regularity for the primal and adjoint problem will hold when $\epsilon \in C^{0,1}(\overline{\Omega})$, Ω is an open bounded convex domain (thus with Lipschitz boundary), and there is a function $\tilde{g} \in H^2(\Omega)$ such that the trace of $\tilde{g}$ on $\partial\Omega$ coincides with g (see Theorem 3.2.1.2 in [95] and Remark 5.1). Instead of neglecting the term $\mathcal{E}$, one can also replace the unknown functions $z, \nabla u$, and ∇z by means of postprocessing the available approximations u_h and z_τ. One such approach is to use the postprocessed gradients $G_h(\nabla u_h)$ and $G_\tau(\nabla z_\tau)$ in $\mathcal{E}$ instead of ∇u and ∇z, where

$$G_h : [L^2(\Omega)]^d \to [V_h]^d, \quad G_\tau : [L^2(\Omega)]^d \to [V_\tau]^d$$

are averaging operators and to exploit superconvergence properties of the primal and adjoint problem (see [100, 115, 119, 139]). Another approach (see [138]) is to further rewrite the term

$\mathcal{E}$ in a form that does not depend on the unknown solution z of the adjoint problem and apply superconvergent postprocessing of the function u_h and a regularization of the adjoint flux $\epsilon\nabla z_\tau$. A third approach that exploits Galerkin orthogonality is suggested in [159] and is based on the relation (see [13])

$$\begin{aligned}\langle \mathscr{Z}, u-u_h\rangle &= \langle \mathscr{F}, z-I_h z\rangle - \int_\Omega \epsilon\nabla u_h\cdot\nabla(z-I_h z)dx \\ &= \sum_{K\in\mathscr{T}_h} \underbrace{\{(f,(z-I_h z))_K - (\epsilon\nabla u_h, \nabla(z-I_h z))_K\}}_{=:\eta_K} =: \eta(u_h,z),\end{aligned} \tag{5.17}$$

where u_h is the Galerkin approximation of u in the space $V^k_{g,h}$ and I_h denotes the finite element interpolant in $V^k_{g,h}$. There, the authors suggest a new localization approach based on the partition of unity $\sum_{i=1}^{N_V}\psi_i \equiv 1$ which is introduced in the error identity (5.17):

$$\begin{aligned}&\langle \mathscr{Z}, u-u_h\rangle \\ &= \int_\Omega \left\{ f\left((z-I_h z)\sum_{i=1}^{N_V}\psi_i\right) - \epsilon\nabla u_h\cdot\nabla\left((z-I_h z)\sum_{i=1}^{N_V}\psi_i\right)\right\}dx \\ &= \sum_{i=1}^{N_V} \underbrace{\{(f,(z-I_h z)\psi_i) - (\epsilon\nabla u_h, \nabla\left((z-I_h z)\psi_i\right))\}}_{=:\eta_i^{PU}} =: \eta^{PU}(u_h,z)\end{aligned} \tag{5.18}$$

and the unknown interpolation error $z-I_h z$ is approximated in an appropriate way. Here $\{\psi_i\}_{i=1}^{N_V}$ are the nodal P_1 basis functions and N_V is the number of nodes. It is easy to see that the introduced error, if one uses an approximation z_τ instead of z in (5.17), is again given by the expression $\mathcal{E}(u,u_h,z,z_\tau)$. In this case

$$\begin{aligned}\langle \mathscr{Z}, u-u_h\rangle &= \eta(u_h,z_\tau) + \mathcal{E}(u,u_h,z,z_\tau) \\ &= \eta^{PU}(u_h,z_\tau) + \mathcal{E}(u,u_h,z,z_\tau)\end{aligned} \tag{5.19}$$

and $\eta(u_h,z_\tau) = \eta^{PU}(u_h,z_\tau) = E(u_h,z_\tau)$. The resulting error indicators η_K in (5.17) are element-wise contributions and the error indicators η_i^{PU} in (5.18) are node-wise contributions of the error (see [159]). Here we note that deriving the error identity (5.17) in the general case presented above is straight forward since we have standard weak formulations for u and z that involve H^1 spaces and regular right-hand sides, i.e., right-hand sides being functions in $L^2(\Omega)$. The adjoint solution z of (5.12) is sometimes referred to as the influence function and the quantity $z-I_h z$ in (5.17) is referred to as the sensitivity factor (see, e.g. [13, 144, 159]). The influence function demonstrates how the information from the residual

$$R(u_h)\in H^{-1}(\Omega):\ \langle R(u_h), v\rangle = \int_\Omega (fv - \epsilon\nabla u_h\cdot\nabla v)\,dx,\ \forall v\in H^1_0(\Omega)$$

in the primal problem is propagated to the error $\mathscr{Z}(u-u_h)$ in the quantity of interest. This approach to a posteriori error estimation with respect to a quantity of interest is called the

dual weighted residual method (DWR method) since the residual $R(u_h)$ is weighted with the quantity z or $z - I_h z$ obtained by a solution of the dual problem (5.12). In the next section, we will derive four goal-oriented error estimates for problem (5.1), two variations in the case of 2-term splitting and two variations in the case of no splitting applied. The main difficulty in deriving each of the four goal-oriented error estimates for problem (5.1) will be to prove the error identity (5.13) or equivalently (5.17) and (5.18) in each separate case.

Remark 5.1
If Ω is a bounded convex open subset of $\mathbb{R}^d$ (thus with Lipschitz boundary), $\tilde{g} \in H^2(\Omega)$, and $\epsilon \in C^{0,1}(\overline{\Omega})$, then the solution u of (5.11) is in $H^2(\Omega)$. To see this, let $u = \tilde{g} + u_0$, where u_0 satisfies the homogenized version of (5.11)

$$\begin{aligned} &\text{Find } u_0 \in H_0^1(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla u_0 \cdot \nabla v dx = -\int_\Omega \epsilon \nabla \tilde{g} \cdot \nabla v dx + \int_\Omega f v dx \text{ for all } v \in H_0^1(\Omega). \end{aligned} \tag{5.20}$$

We have that $\epsilon \in C^{0,1}(\overline{\Omega}) \subset W^{1,\infty}(\Omega)$ and $\nabla \tilde{g} \in \left[H^1(\Omega)\right]^d$. If $v \in C_0^\infty(\Omega)$, then by Theorem 2.9, $\nabla \tilde{g} v \in \left[H^1(\Omega)\right]^d$. Moreover, Leibniz' formula applies for the weak derivatives of $\nabla \tilde{g} v$ and we obtain

$$\operatorname{div}(\nabla \tilde{g} v) = \operatorname{div}(\nabla \tilde{g}) + \nabla \tilde{g} \cdot \nabla v. \tag{5.21}$$

Since $\nabla \tilde{g} v$ has compact support in Ω, it follows by Theorem 2.10 that $\nabla \tilde{g} v \in \left[H_0^1(\Omega)\right]^d$. By the divergence theorem applied to the functions $\epsilon \in H^1(\Omega)$ and $\nabla \tilde{g} v \in \left[H_0^1(\Omega)\right]^d$ and by using (5.21) we obtain

$$\begin{aligned} \int_\Omega \nabla \epsilon \cdot \nabla \tilde{g} v dx &= -\int_\Omega \epsilon \operatorname{div}(\nabla \tilde{g} v)\, dx + \underbrace{\int_{\partial\Omega} \gamma_2(\nabla \tilde{g} v) \cdot \boldsymbol{n}_{\partial\Omega} \epsilon ds}_{=0} \\ &= -\int_\Omega \epsilon \operatorname{div}(\nabla \tilde{g})\, v dx - \int_\Omega \epsilon \nabla \tilde{g} \cdot \nabla v dx. \end{aligned} \tag{5.22}$$

Since $\epsilon \in W^{1,\infty}(\Omega)$, by a standard density argument, (5.22) is also valid for all $v \in H_0^1(\Omega)$ and for the first term on the right-hand side of (5.20) we obtain

$$-\int_\Omega \epsilon \nabla \tilde{g} \cdot \nabla v dx = \int_\Omega \underbrace{(\nabla \epsilon \cdot \nabla \tilde{g} + \epsilon \operatorname{div}(\nabla \tilde{g}))}_{\in L^2(\Omega)} v dx. \tag{5.23}$$

Therefore, $u_0 \in H_0^1(\Omega)$ satisfies the equation

$$\int_\Omega \epsilon \nabla u_0 \cdot \nabla v dx = \int_\Omega \tilde{f} v dx \text{ for all } v \in H_0^1(\Omega), \tag{5.24}$$

where $\tilde{f} := \nabla\epsilon \cdot \nabla\tilde{g} + \epsilon \operatorname{div}(\nabla\tilde{g}) + f \in L^2(\Omega)$. *Finally, by Theorem 3.2.1.2 in [95] it follows that* $u_0 \in H^2(\Omega)$ *and consequently* $u = \tilde{g} + u_0 \in H^2(\Omega)$. *If, in addition,* $\nabla\tilde{g}$ *was in* $\left[H^1(\Omega)\right]^d \cap [L^\infty(\Omega)]^d$, *then by Theorem 2.8,* $\epsilon\nabla\tilde{g} \in \left[H^1(\Omega)\right]^d$ *and in particular* $\epsilon\nabla\tilde{g} \in H(\operatorname{div};\Omega)$. *Thus, we directly obtain*

$$-\int_\Omega \epsilon\nabla\tilde{g}\cdot\nabla v dx = \int_\Omega \underbrace{\operatorname{div}(\epsilon\nabla\tilde{g})}_{\in L^2(\Omega)} v dx. \tag{5.25}$$

5.3 Error estimates for the electrostatic interaction

In this section, we present four goal-oriented a posteriori error estimates in which the quantity of interest is the electrostatic interaction $E_{2-1} = E_{1-2}$ between the two dyes (Dye I and Dye II). The first two estimates are applied to the solution u_2 of the regular problem (5.10) and the other two are applied to the whole potential φ_2, the solution of (5.2). The first two estimates are convenient to use when $\frac{\epsilon_m}{\epsilon_s}$ is close to 1 and the other two estimates are especially appropriate in the case $\frac{\epsilon_m}{\epsilon_s} \ll 1$ and $I_s > 0$. We recall that when $\frac{\epsilon_m}{\epsilon_s} \ll 1$ and $I_s > 0$, due to the strong dielectric screening, the reaction field potential u_2 is very close in absolute value to the Coulomb part of the potential G_2 but with opposite sign, i.e $|G_2 + u_2| \ll |u_2|$ and small relative errors in u_2 result in huge relative errors in $\varphi_2 = G_2 + u_2$.

Since computing the electrostatic interaction involves point evaluations of the potential, the goal functional is not regular as it is the case with the standard formulation in (5.12). Instead, it is a linear combination of delta functions. The usual way to treat such goal functionals is to regularize them by means of some averaging over small balls $B(x_i, \rho)$ with a radius ρ around the points of interest x_i. There are two problems with this approach. The first one is that in this way we change the goal functional and we are no more solving the original problem of estimating real point values of the solution. Instead we are estimating the averaged solution at these points with the parameter ρ. The second problem is that in order to perform this averaging we need to be able to integrate expressions of the form $\int_\Omega \chi_{B(x_i,\rho)} v dx$, where v is a finite element function from a finite element space of degree k or k^*. These expressions involve integration of the discontinuous functions $\chi_{B(x_i,\rho)}v$ with support in the balls $B(x_i, \rho)$ and appear when assembling the load vector for the adjoint problem and also when evaluating the goal functional at the approximate solution of the primal problem. To achieve a high enough accuracy, the mesh needs to be a priori adapted around the points of interest and special integration rules are required. Of course, the averaging can also be done by means of mollification with a small parameter ρ. In this case the functions to be integrated are not discontinuous, but in order to achieve good enough accuracy, the mesh should be again a priori adapted to the regions of mollification. However, the first problem remains. Namely, we are again solving a perturbed adjoint problem where the original goal functional (a linear combination of delta functions) is replaced by its mollified version. Moreover, even at the exat solution of the primal problem, the quantity of interest is not exact but rather a mollified

version of it with the parameter ρ.

Two of the presented below goal-oriented error estimates (**error estimate 2** and **error estimate 4**) completely resolve the above mentioned issues. This is made possible by deriving representations of the error in the goal quantity which do not involve averaging and exploit directly the original goal functional. The theoretical justification of these representations of the error is also presented below. The other two error estimates that we present (**error estimate 1** and **error estimate 3**) resolve only the first issue. Namely, the regularization by averaging does not change the original quantity of interest because the solution of the primal problem is a harmonic function in the molecular region. However, the integration problem remains and needs to be carefully addressed.

In the estimates that we will derive for the electrostatic interaction between the two chromophores, we will use one and the same mesh for the primal and adjoint problem, i.e., $\mathscr{T}_\tau \equiv \mathscr{T}_h$. For the primal problem we will use Galerkin approximations with continuous piecewise linear finite elements and for the adjoint problem we will use Galerkin approximations with continuous piecewise polynomials of degree 2, i.e., $k = 1$ and $k^* = 2$. We note that in the error estimates presented in Sections 5.3.1, 5.3.2, 5.3.3, and 5.3.4 other finite element spaces can also be used on possibly different meshes $\mathscr{T}_h$ and $\mathscr{T}_\tau$.

5.3.1 Error estimation 1: 2-term splitting in primal problem and regular goal functional

Testing (5.10) with functions $v \in C_0^\infty(\Omega_m)$ we see that the reaction field potential u_2 is harmonic in Ω_m, and in particular, in $\Omega_{m,1}$. This means that we can write

$$\langle \frac{1}{4\pi}\mathscr{F}_1, u_2\rangle = \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, u_2\rangle,$$

where

$$\langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, u_2\rangle := \sum_{i=1}^{N_1} \frac{q_{i,1}}{|(B(x_{i,1}, \overline{r}_{i,1}))|} \int\limits_{B(x_{i,1},\overline{r}_{i,1})} u_2 dx \tag{5.26}$$

with

$$|B(x_{i,1}, \overline{r}_{i,1})| = \frac{4\pi(\overline{r}_{i,1})^3}{3} \text{ in } d = 3 \text{ and } |B(x_{i,1}, \overline{r}_{i,1})| = \pi(r_{i,1})^2 \text{ in } d = 2.$$

In (5.26), $B(x_{i,1}, \overline{r}_{i,1})$ is the open ball with center at $x_{i,1}$ and a radius $\overline{r}_{i,1}$, where $\overline{r}_{i,1} < r_{i,1}$ and $r_{i,1}$ is the Van der Waals radius of the i-th atom in Dye I. Now, the goal functional $\frac{1}{4\pi}\overline{\mathscr{F}}_1$ is a bounded linear functional over $H_0^1(\Omega)$. In fact, it is representable by a square summable function l throught the formula

$$\langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, v\rangle = \int\limits_\Omega l v dx, \tag{5.27}$$

where l is defined by

$$l(x) := \sum_{i=1}^{N_1} \frac{3}{4\pi(\overline{r}_{i,1})^3} \chi_{B(x_{i,1},\overline{r}_{i,1})}(x). \tag{5.28}$$

Let $u_{2,h}$ denote the Galerkin finite element approximation of u_2 in a finite element space $V^k_{-G_2,h} \subset H^1_{-G_2}(\Omega)$ with $k = 1$ (see Remark 5.4). Our goal is to estimate the quantity $\langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, u_2 - u_{2,h}\rangle$. We introduce the adjoint problem

$$\begin{aligned} &\text{Find } z \in H^1_0(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla v \cdot \nabla z dx + \int_\Omega \overline{k}^2 v z dx = \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, v\rangle, \ \forall v \in H^1_0(\Omega), \end{aligned} \tag{5.29}$$

which is the weak formulation of the problem

$$\begin{aligned} -\nabla \cdot (\epsilon \nabla z) + \overline{k}^2 z &= l \quad \text{in } \Omega, && \text{(5.30a)} \\ [z]_\Gamma &= 0, && \text{(5.30b)} \\ [\epsilon \nabla z \cdot \boldsymbol{n}_\Gamma]_\Gamma &= 0, && \text{(5.30c)} \\ z &= 0 \quad \text{on } \partial\Omega. && \text{(5.30d)} \end{aligned}$$

As in Section 5.2 we obtain the error representation in terms of the adjoint solution $z \in H^1_0(\Omega)$

$$\begin{aligned} \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, u_2 - u_{2,h}\rangle &= \int_\Omega \epsilon \nabla (u_2 - u_{2,h}) \cdot \nabla z dx + \int_\Omega \overline{k}^2 (u_2 - u_{2,h}) z dx \\ &= \langle \mathcal{T}_2, z\rangle - \int_\Omega \epsilon \nabla u_{2,h} \cdot \nabla z dx - \int_\Omega \overline{k}^2 u_{2,h} z dx =: E(u_{2,h}, z). \end{aligned} \tag{5.31}$$

Note that by the Sobolev embedding theorem $z \in H^1_0(\Omega)$ implies $z \in L^6(\Omega)$ for $d = 2, 3$ and that z satisfies the equation

$$\int_\Omega \epsilon \nabla v \cdot \nabla z dx = \int_\Omega \underbrace{\left(-\overline{k}^2 z + l\right)}_{\in L^6(\Omega)} v dx, \ \forall v \in H^1_0(\Omega), \tag{5.32}$$

which by Theorem 2.32 means that $z \in L^\infty(\Omega)$. This means that the term on the right-hand side of (5.32) defines a bounded linear functional over $W^{1,p}_0(\Omega)$ for any $1 \le p < \frac{d}{d-1}$. If we additionally assume that $\Gamma \in C^1$, from Theorem 2.34 it follows that $z \in W^{1,q}_0(\Omega)$ for some $q > d$ and by the Sobolev embedding theorem it follows that z has a continuous representative.

Therefore, the nodal P_1 interpolant in V_h^1 of z is well defined and we can write

$$\begin{aligned} E(u_{2,h},z) &= \langle \mathcal{T}_2, z - I_h z\rangle - \int_\Omega \epsilon \nabla u_{2,h}\cdot\nabla(z-I_h z)dx - \int_\Omega \overline{k}^2 u_{2,h}(z-I_h z)dx \\ &= \sum_{K\in\mathscr{T}_h} \eta_K := \eta(u_{2,h},z) \\ &= \sum_{i=1}^{N_V} \underbrace{\left\{ \langle \mathcal{T}_2, (z-I_h z)\,\psi_i\rangle - (\epsilon\nabla u_{2,h}, \nabla\left((z-I_h z)\psi_i\right)) - \left(\overline{k}^2 u_{2,h}, (z-I_h z)\,\psi_i\right)\right\}}_{=:\eta_i^{PU}} \\ &=: \eta^{PU}(u_{2,h},z), \end{aligned} \tag{5.33}$$

where N_V is the number of vertices (nodes) in the mesh $\mathscr{T}_h$, $\{\psi_i\}_{i=1}^{N_V}$ are the nodal P_1 basis functions, I_h is the nodal interpolation operator in V_h^1, and

$$\begin{aligned} \eta_K(u_{2,h},z) = &\left((\epsilon_m-\epsilon_s)\nabla G_2\chi_{\Omega_s} - \epsilon\nabla u_{2,h}, \nabla(z-I_h z)\right)_K \\ &- \left(\overline{k}^2(G_2+u_{2,h}), (z-I_h z)\right)_K. \end{aligned} \tag{5.34}$$

To apply in practice the error indicators η_i^{PU} or η_K, we use a Galerkin finite element approximation $z_h^{(2)}$ of z from the finite element space $V_{0,\tau}^{k^*}$ with $k^*=2$ and $\tau\equiv h$, i.e., $V_{0,\tau}^{k^*}$ is defined on the same triangulation $\mathscr{T}_h$ as the finite element space $V^1_{-G_2,h}$ for the primal problem. Taking into account (5.7) and (5.9), the approximate electrostatic interaction E_{2-1}^P, computed by means of the *Primal* solution, is given by

$$E_{2-1}^P = E_{G_2} + \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, u_{2,h}\rangle \tag{5.35}$$

and the error estimation for the electrostatic interaction E_{2-1} is

$$\begin{aligned} E_{2-1} - E_{2-1}^P &= E_{G_2} + \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, u_2\rangle - \left(E_{G_2} + \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, u_{2,h}\rangle\right) \\ &= \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, u_2 - u_{2,h}\rangle \\ &= E(u_{2,h},z) = E(u_{2,h},z_h^{(2)}) + \mathcal{E}(u_2,u_{2,h},z,z_h^{(2)}). \end{aligned} \tag{5.36}$$

As in Section 5.2, we can bound the term $E(u_{2,h},z_h^{(2)})$ by

$$\begin{aligned} \left|E(u_{2,h},z_h^{(2)})\right| &= \left|\int_\Omega \epsilon\nabla(u_2-u_{2,h})\cdot\nabla z_h^{(2)}dx + \int_\Omega \overline{k}^2\,(u_2-u_{2,h})\,z_h^{(2)}dx\right| \\ &\le \epsilon_{\max}\|\nabla(u_2-u_{2,h})\|_{L^2(\Omega)}\|\nabla z_h^{(2)}\|_{L^2(\Omega)} + \overline{k}^2_{ions}\|u_2-u_{2,h}\|_{L^2(\Omega)}\|z_h^{(2)}\|_{L^2(\Omega)} \\ &\le C_1\left(\epsilon_{\max}\|\nabla(u_2-u_{2,h})\|_{L^2(\Omega)} + \overline{k}^2_{ions}\|u_2-u_{2,h}\|_{L^2(\Omega)}\right) \end{aligned} \tag{5.37}$$

whereas the error $\mathcal{E}(u_2, u_{2,h}, z, z_h^{(2)})$ that we make when using an approximation $z_h^{(2)}$ instead of z in $E(u_{2,h}, z)$ can be bounded as follows

$$\begin{aligned}
\mathcal{E}(u_2, u_{2,h}, z, z_h^{(2)}) &= \langle \mathcal{T}_2, z - z_h^{(2)} \rangle - \int_\Omega \epsilon \nabla u_{2,h} \cdot \nabla (z - z_h^{(2)}) dx - \int_\Omega \overline{k}^2 u_{2,h} \left(z - z_h^{(2)} \right) dx \\
&= \int_\Omega \epsilon \nabla (u_2 - u_{2,h}) \cdot \nabla (z - z_h^{(2)}) dx + \int_\Omega \overline{k}^2 \left(u_2 - u_{2,h} \right) (z - z_h^{(2)}) dx \\
&\leq \epsilon_{\max} \|\nabla (u_2 - u_{2,h})\|_{L^2(\Omega)} \|\nabla (z - z_h^{(2)})\|_{L^2(\Omega)} + \overline{k}_{ions}^2 \|u_2 - u_{2,h}\|_{L^2(\Omega)} \|z - z_h^{(2)}\|_{L^2(\Omega)} \\
&\leq C_2 \left(\epsilon_{\max} \|\nabla (u_2 - u_{2,h})\|_{L^2(\Omega)} + \overline{k}_{ions}^2 \|u_2 - u_{2,h}\|_{L^2(\Omega)} \right) \|\nabla (z - z_h^{(2)})\|_{L^2(\Omega)}. \qquad (5.38)
\end{aligned}$$

In (5.37) and (5.38) C_1 and C_2 are generic constants. In the derivation of the upper bound (5.37) we have used the fact that $\|\nabla z_h^{(2)}\|_{L^2(\Omega)}$ and $\|z_h^{(2)}\|_{L^2(\Omega)}$ are bounded since $\|\nabla (z - z_h^{(2)})\|_{L^2(\Omega)} \to 0$, which is true for regular triangulations $\mathscr{T}_h$ by virtue of Cea's lemma, even without assumptions on the regularity of z. We have also used Poincareé's inequality to bound the L^2 norms of $z_h^{(2)}$ and $z - z_h^{(2)}$ in (5.37) and (5.38), respectively. From (5.37) and (5.38) it is seen that $\mathcal{E}(u_2, u_{2,h}, z, z_h^{(2)})$ converges faster to zero than $E(u_{2,h}, z_h^{(2)})$ and, therefore, we can assume that the major part of the error in (5.36) is contained in the term $E(u_{2,h}, z_h^{(2)})$. For this reason, we can skip $\mathcal{E}(u_2, u_{2,h}, z, z_h^{(2)})$ in (5.36) and write the approximate equality

$$E_{2-1} - E_{2-1}^P \approx E(u_{2,h}, z_h^{(2)}), \qquad (5.39)$$

which can also be written in the form

$$E_{2-1} \approx E_{2-1}^P + E(u_{2,h}, z_h^{(2)}). \qquad (5.40)$$

The quantity $E_{2-1}^P + E(u_{2,h}, z_h^{(2)})$ is a corrected value for the electrostatic interaction.

Remark 5.2

For the practical application of this error estimation approach where instead of the original goal functional $\frac{1}{4\pi}\mathscr{F}_1$ we use a regularized version of it, $\frac{1}{4\pi}\overline{\mathscr{F}}_1$, special quadrature rules are needed for the evaluation of the terms $\langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, u_{2,h} \rangle$ and $\langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, z_h^{(2)} \rangle$ (see, e.g. [108, 145] for such quadrature rules).

Remark 5.3

Even if the interface Γ is not C^1, the elementwise error indicators $\eta_K(u_{2,h}, z_h^{(2)})$ and the nodewise error indicators $\eta_i^{PU}(u_{2,h}, z_h^{(2)})$ are well defined since $z_h^{(2)}$ is a continuous function and, thus, $I_h z_h^{(2)}$ is well defined.

Remark 5.4

In particular, if we assume that $-G_2$ is exactly representable on $\partial\Omega$ in the finite element space

V_h^k, then we can take $u_{2,h} \in V^k_{-G_2,h} \subset V_h^k$ to be a finite element solution of the corresponding Galerkin problem

$$\begin{aligned} &\text{Find } u_{2,h} \in V^k_{-G_2,h} \text{ such that} \\ &\int_\Omega \epsilon \nabla u_{2,h} \cdot \nabla v dx + \int_\Omega \overline{k}^2 u_{2,h} v dx = \langle \mathcal{T}_2, v \rangle \quad \text{for all } v \in V^k_{0,h}. \end{aligned} \tag{5.41}$$

Else, if $-G_2$ is not exactly representable on $\partial\Omega$ in the finite element space V_h^k, then one can obtain $u_{2,h}$ by solving a Galerkin formulation for the homogenized version of equation (5.10)

$$\begin{aligned} &\text{Find } u_0 \in H^1_0(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla u_0 \cdot \nabla v dx + \int_\Omega \overline{k}^2 u_0 v dx = \langle \mathcal{T}_2, v \rangle + \int_\Omega \epsilon \nabla \tilde{G}_2 \cdot \nabla v dx + \int_\Omega \overline{k}^2 \tilde{G}_2 v dx. \end{aligned} \tag{5.42}$$

One finds $u_{0,h} \in V^k_{0,h}$ by solving the Galerkin formulation

$$\int_\Omega \epsilon \nabla u_{0,h} \cdot \nabla v dx + \int_\Omega \overline{k}^2 u_{0,h} v dx = \langle \mathcal{T}_2, v \rangle + \int_\Omega \epsilon \nabla \tilde{G}_2 \cdot \nabla v dx + \int_\Omega \overline{k}^2 \tilde{G}_2 v dx \tag{5.43}$$

for all $v \in V^k_{0,h}$, where $\tilde{G}_2 : \Omega \to \mathbb{R}$ is in $H^2(\Omega)$ with $\gamma_2(-\tilde{G}_2) = -G_2$ on $\partial\Omega$. Consequently, one defines $u_{2,h} := -\tilde{G}_2 + u_{0,h} \in H^1_{-G_2}(\Omega)$. In this case, if $V^k_{0,h} \equiv V^{k^*}_{0,\tau}$ and $z_\tau \in V^{k^*}_{0,\tau}$ is the Galerkin approximation of the adjoint problem (5.29)*, the term $E(u_{2,h}, z_\tau)$ is still zero.*

Remark 5.5
For the construction of $\tilde{G}_2$ just take $(\psi G_2)_{|\overline{\Omega}} \in C^\infty(\overline{\Omega})$, where $\psi \in C^\infty_0(\mathbb{R}^d)$ is such that it is equal to 1 in a neighborhood of $\partial\Omega$ and with support in $\mathbb{R}^d \setminus \overline{\Omega}_m$. For ψ just mollify the characteristic function of the set $(\partial\Omega)^{+\delta} := \{x \in \mathbb{R}^d : dist(x, \partial\Omega) < \delta\}$ with a mollifier η_ρ for $\rho < \delta/2$, where $\delta < \frac{1}{2} dist(\Gamma, \partial\Omega)$.

5.3.2 Error estimation 2: 2-term splitting in primal problem and irregular goal functional

Here, we solve the same primal problem (5.10). This time the right-hand side of the adjoint problem (5.30) is formed by the original goal functional $\frac{1}{4\pi}\mathscr{F}_1 \notin H^{-1}(\Omega)$ defined in (5.8). In this way we avoid the numerical integration of the discontinuous functions arising when averaging over the balls $B(x_{i,1}, \overline{r}_{i,1})$. The functional $\frac{1}{4\pi}\mathscr{F}_1$ is not bounded over $H^1_0(\Omega)$ and the weak form of the adjoint problem is defined in a similar way to the weak formulation (5.2)

$$\begin{aligned} &\text{Find } z \in \bigcap_{p<\frac{d}{d-1}} W^{1,p}_0(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla v \cdot \nabla z dx + \int_\Omega \overline{k}^2 v z dx = \langle \frac{1}{4\pi}\mathscr{F}_1, v \rangle, \, \forall v \in \bigcup_{q>d} W^{1,q}_0(\Omega) \end{aligned} \tag{5.44}$$

Here we note that the solution z to the adjoint problem (5.44) satisfies the relation $z = \frac{1}{4\pi}\varphi_1$ since φ_1 solves the same linear problem but with a right-hand side $\mathscr{F}_1$ and a homogeneous Dirichlet boundary condition (see (5.3)). Now, the physical meaning of the adjoint problem is clear. From Theorem 3.4 we know that (5.44) possesses a unique solution $z \in \bigcap\limits_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega)$ which has the form $z = \frac{1}{4\pi}\varphi_1 = \frac{1}{4\pi}(G_1 + u_1)$ where u_1 satisfies (5.10) with G_1 instead of G_2. Again, by $u_{2,h}$ we denote the Galerkin finite element approximation of u_2 in the space V_h^k with $k = 1$. In order to derive a similar estimate to (5.31) we will need the following Proposition 5.6.

Proposition 5.6
The following equalities hold true

$$\int_\Omega \epsilon\nabla(u_2 - u_{2,h})\cdot\nabla z dx + \int_\Omega \overline{k}^2(u_2 - u_{2,h})z dx = \langle \frac{1}{4\pi}\mathscr{F}_1, u_2 - u_{2,h}\rangle, \tag{5.45}$$

$$\int_\Omega \epsilon\nabla u_2\cdot\nabla z dx + \int_\Omega \overline{k}^2 u_2 z dx = \langle \mathcal{T}_2, z\rangle = \int_\Omega (\epsilon_m - \epsilon)\nabla G_2\cdot\nabla z dx - \int_\Omega \overline{k}^2 G_2 z dx. \tag{5.46}$$

Remark 5.7
In particular, (5.45) *and* (5.46) *mean that the integrals on the left-hand sides are well defined, as well as the expression* $\langle \frac{1}{4\pi}\mathscr{F}_1, u_2 - u_{2,h}\rangle$ *is well defined. Note, that* (5.45) *means that we can test* (5.44) *with the function* $u_2 - u_{2,h}$ *which, in general, is not in* $\bigcup\limits_{q>d} W_0^{1,q}(\Omega)$ *and* (5.46) *means that we can test* (5.10) *with* z *which is not in* $H_0^1(\Omega)$.

Remark 5.8
If Γ *is of class* C^1, *since* $\partial\Omega \in C^{0,1}$, *then from Theorem 2.34 it follows that* $u_2 \in W^{1,\bar{q}}(\Omega)$ *for some* $\bar{q} > 3$. *More precisely, we can apply Theorem 2.34 to the homogenized version* (5.42) *of* (5.10) *with* $u_2 = -\tilde{G}_2 + u_0$. *By applying the Lax-Milgram Theorem we see that problem* (5.42) *has a unique solution* $u_0 \in H_0^1(\Omega)$. *By the Sobolev embedding theorems, we know that* $H_0^1(\Omega) \hookrightarrow L^6(\Omega)$ *for* $d = 1, 2, 3$ *and thus* $\langle \mathcal{T}_2, v\rangle + \int\limits_\Omega \epsilon\nabla\tilde{G}_2\cdot\nabla v dx + \int\limits_\Omega \overline{k}^2\tilde{G}_2 v dx - \int\limits_\Omega \overline{k}^2 u_0 v dx$ *defines a bounded linear functional over* $W_0^{1,r}$ *for all* $r \in [6/5, +\infty)$. *By Theorem 2.34 there exists* $p > d$ *such that* $-\nabla\cdot\epsilon\nabla : W_0^{1,q} \to W_0^{-1,q} = \left(W_0^{1,q'}\right)^*$ *is a topological isomorphism for all* $q \in (p', p)$, *where* p' *and* q' *denote the Hölder conjugates of* p *and* q, *respectively. In particular, it follows that there exists* $\bar{q}$ *such that* $d < \bar{q} < p$ *and* $\bar{q} \le 6$ *for which* $u_0 \in W_0^{1,\bar{q}}(\Omega)$. *Again by the Sobolev embedding theorems it follows that* $\tilde{G}_2 \in H^2(\Omega) \hookrightarrow W^{1,\bar{q}}(\Omega)$, *and consequently,* $u_2 = -\tilde{G}_2 + u_0 \in W^{1,\bar{q}}_{-G_2}(\Omega)$.
It is also clear that $u_{2,h} \in W^{1,\bar{q}}(\Omega)$ *when* $u_{2,h} = -\tilde{G}_2 + u_{0,h}$ *and* $u_{2,h} \in W^{1,\infty}(\Omega)$ *when* $u_{2,h}$ *is a pure finite element function. Thus* $u_2 - u_{2,h} \in W_0^{1,\bar{q}}(\Omega) \subset \bigcup\limits_{q>d} W_0^{1,q}(\Omega)$ *and therefore* (5.45) *is satisfied.*

In order to avoid some technical details we will prove Proposition 5.6 only for the case $I_s = 0$, i.e., $\overline{k} = 0$ and note that the proof for the case $I_s \neq 0$ is similar.

Proof of Proposition 5.6. Let $B_{m,1}(r) := \bigcup_{i=1}^{N_1} B(x_{i,1}, r)$, where $B(x_{i,1}, r)$ is the open ball with center at $x_{i,1}$ and a radius r. Here, r is chosen so small that $B(x_{i,1}, 2r)$ is strictly contained in Ω_m and $\overline{B(x_{i,1}, 2r)} \cap \overline{B(x_{j,1}, 2r)} = \emptyset$ for all $i \neq j$, $i, j = 1, 2, .., N_1$ (see Figure 5.2). The idea to prove (5.45) is to approximate $u_2 - u_{2,h}$ with functions ψ_n in $W_0^{1,q}(\Omega)$ for some fixed $d < q \leq 6$ such that $\psi_n \to u_2 - u_{2,h}$ in $H^1(\Omega \setminus B_{m,1}(r))$ and $\psi_n \to u_2 - u_{2,h}$ in $W^{1,q}(B_{m,1}(r))$. Similarly, to prove (5.46) we find functions $\psi_n \in H_0^1(\Omega)$ such that $\psi_n \to z$ in $W^{1,p}(B_{m,1}(2r))$ for some fixed $p < \frac{d}{d-1}$ and $\psi_n \to z$ in $H^1(\Omega \setminus B_{m,1}(2r))$. Thus, let us fix q such that $d < q \leq 6$ and let p be its Hölder conjugate, i.e., p is such that $\frac{1}{p} + \frac{1}{q} = 1$. It is clear that p satisfies $p < \frac{d}{d-1}$.

Proof of (5.45)**:**
Let $\overline{w} := u_2 - u_{2,h}$. Since $u_{2,h} \in W^{1,\infty}(\Omega)$ and $u_2 \in H^2_{\mathrm{loc}}(\Omega_m)$ (see Theorem 3.4), by the Sobolev embedding theorems it follows that $\overline{w} \in W^{1,q}(B_{m,1}(2r)) \cap H_0^1(\Omega)$. We can find a sequence $w_n \in C_0^\infty(\Omega)$ such that $\|w_n - \overline{w}\|_{H^1(\Omega)} \xrightarrow[n\to\infty]{} 0$. Now, let χ_1 be a smooth cut-off function such that

$$\chi_1(x) = \begin{cases} 1, & x \in B_{m,1}(r), \\ 0, & x \in \Omega \setminus B_{m,1}(2r). \end{cases} \tag{5.47}$$

and define $\psi_n := w_n + (\overline{w} - w_n)\chi_1$. We have

$$\psi_n = \begin{cases} \overline{w}, & x \in B_{m,1}(r), \\ w_n + (\overline{w} - w_n)\chi_1, & x \in A_1 := B_{m,1}(2r) \setminus B_{m,1}(r), \\ w_n, & x \in \Omega \setminus B_{m,1}(2r). \end{cases} \tag{5.48}$$

By the product rule, applied to the functions $(\overline{w} - w_n) \in H_0^1(\Omega)$ and $\chi_1 \in C_0^\infty(\Omega)$ (see Theorem 2.9) we obtain that $(\overline{w} - w_n)\chi_1 \in H^1(\Omega)$ and therefore $\psi_n \in H^1(\Omega)$. Since $(\overline{w} - w_n)\chi_1$ has a compact support in Ω, by Theorem 2.10 it follows that $(\overline{w} - w_n)\chi_1 \in H_0^1(\Omega)$ and consequently that $\psi_n \in H_0^1(\Omega)$. Moreover, for the weak derivatives we have

$$\frac{\partial \psi_n}{\partial x_i} = \frac{\partial w_n}{\partial x_i} + \frac{\partial}{\partial x_i}\left((\overline{w} - w_n)\chi_1\right) = \frac{\partial w_n}{\partial x_i} + \frac{\partial(\overline{w} - w_n)}{\partial x_i}\chi_1 + (\overline{w} - w_n)\frac{\partial \chi_1}{\partial x_i} \tag{5.49}$$

and thus

$$\frac{\partial \psi_n}{\partial x_i} = \begin{cases} \dfrac{\partial \overline{w}}{\partial x_i}, & x \in B_{m,1}(r), \\ \dfrac{\partial w_n}{\partial x_i} + \dfrac{\partial(\overline{w} - w_n)}{\partial x_i}\chi_1 + (\overline{w} - w_n)\dfrac{\partial \chi_1}{\partial x_i}, & x \in A_1 = B_{m,1}(2r) \setminus B_{m,1}(r), \\ \dfrac{\partial w_n}{\partial x_i}, & x \in \Omega \setminus B_{m,1}(2r). \end{cases} \tag{5.50}$$

From the expression (5.48) for ψ_n and the expression (5.50) for $\frac{\partial \psi_n}{\partial x_i}$, by recalling that $\overline{w} \in W^{1,q}(B_{m,2}(2r))$, we see that $\psi_n \in L^q(\Omega)$ and also $\frac{\partial \psi_n}{\partial x_i} \in L^q(\Omega)$ for all $n \in \mathbb{N}$. Therefore,

we conclude that $\psi_n \in W^{1,q}(\Omega)$ and since all ψ_n have compact support in Ω, by Theorem 2.10, it follows that $\psi_n \in W_0^{1,q}(\Omega)$ for all $n \in \mathbb{N}$. Now, for every $n \in \mathbb{N}$, it holds

$$\int\limits_{B_{m,1}(r)} \epsilon \nabla z \cdot \nabla \psi_n dx + \int\limits_{A_1} \epsilon \nabla z \cdot \nabla \psi_n dx + \int\limits_{\Omega \setminus B_{m,1}(2r)} \epsilon \nabla z \cdot \nabla \psi_n dx = \langle \frac{1}{4\pi} \mathscr{F}_1, \psi_n \rangle \tag{5.51}$$

Obviously, $\|\psi_n - \overline{w}\|_{W^{1,q}(B_{m,1}(r))} = 0$, $\forall n \in \mathbb{N}$ and $\|\psi_n - \overline{w}\|_{H^1(\Omega \setminus B_{m,1}(2r))} \xrightarrow[n\to\infty]{} 0$. On the set A_1, since $\|\chi_1\|_{L^\infty(\Omega)} \leq 1$ and $\|\frac{\partial \chi_1}{\partial x_i}\|_{L^\infty(\Omega)}$ is bounded for all $i = 1, 2, 3$, we have

$$\begin{aligned}
&\left\| \frac{\partial \psi_n}{\partial x_i} - \frac{\partial \overline{w}}{\partial x_i} \right\|_{L^2(A_1)} \\
&\leq \left\| \frac{\partial w_n}{\partial x_i} - \frac{\partial \overline{w}}{\partial x_i} \right\|_{L^2(A_1)} + \left\| \frac{\partial (\overline{w} - w_n)}{\partial x_i} \chi_1 \right\|_{L^2(A_1)} + \left\| (\overline{w} - w_n) \frac{\partial \chi_1}{\partial x_i} \right\|_{L^2(A_1)} \\
&\leq \left\| \frac{\partial w_n}{\partial x_i} - \frac{\partial \overline{w}}{\partial x_i} \right\|_{L^2(A_1)} + \left\| \frac{\partial (\overline{w} - w_n)}{\partial x_i} \right\|_{L^2(A_1)} \|\chi_1\|_{L^\infty(\Omega)} \\
&+ \|(\overline{w} - w_n)\|_{L^2(A_1)} \left\| \frac{\partial \chi_1}{\partial x_i} \right\|_{L^\infty(\Omega)} \to 0
\end{aligned} \tag{5.52}$$

and, therefore, $\|\nabla(\psi_n - \overline{w})\|_{L^2(A_1)} \xrightarrow[n\to\infty]{} 0$. By observing that $\nabla z = \frac{1}{4\pi} \nabla \varphi_1 = \frac{1}{4\pi} (\nabla G_1 + \nabla u_1) \in L^2(\Omega \setminus B_{m,1}(r))$ and by applying Hölder's inequality we obtain

$$\left| \int\limits_{A_1} \epsilon \nabla z \cdot \nabla (\psi_n - \overline{w}) \, dx \right| \leq \epsilon_{\max} \|\nabla z\|_{L^2(A_1)} \|\nabla(\psi_n - \overline{w})\|_{L^2(A_1)} \to 0,$$

$$\left| \int\limits_{\Omega \setminus B_{m,1}(2r)} \epsilon \nabla z \cdot \nabla (\psi_n - \overline{w}) \, dx \right| \leq \epsilon_{\max} \|\nabla z\|_{L^2(\Omega \setminus B_{m,1}(2r))} \|\nabla(\psi_n - \overline{w})\|_{L^2(\Omega \setminus B_{m,1}(2r))} \to 0.$$

Also,

$$\begin{aligned}
\left| \langle \frac{1}{4\pi} \mathscr{F}_1, \psi_n \rangle - \langle \frac{1}{4\pi} \mathscr{F}_1, \overline{w} \rangle \right| &= \left| \sum_{i=1}^{N_1} q_{i,1} \psi_n(x_{i,1}) - \sum_{i=1}^{N_1} q_{i,1} \overline{w}(x_{i,1}) \right| \\
&\leq \sum_{i=1}^{N_1} |q_{i,1}| \, |\psi_n(x_{i,1}) - \overline{w}(x_{i,1})| \leq C_E \|\psi_n - \overline{w}\|_{W^{1,q}(B_{m,1}(r))} \sum_{i=1}^{N_1} |q_{i,1}| \to 0,
\end{aligned}$$

where C_E is the embedding constant in the inequality $\|v\|_{L^\infty(\Omega)} \leq C_E \|v\|_{W^{1,q}(\Omega)}$. Finally, letting $n \to \infty$ in (5.51) gives (5.45).

Proof of (5.46)

We have that $z \in \bigcap\limits_{s < \frac{d}{d-1}} W_0^{1,s}(\Omega)$ and in particular $z \in W_0^{1,p}(\Omega)$. There exist functions $z_n \in C_0^\infty(\Omega)$ such that $\|z_n - z\|_{W^{1,p}(\Omega)} \xrightarrow[n\to\infty]{} 0$. We define the functions

$$\psi_n := z + (z_n - z)\chi_1,$$

where χ_1 is defined by (5.47). We have

$$\psi_n = \begin{cases} z_n, & x \in B_{m,1}(r), \\ z + (z_n - z)\chi_1, & x \in A_1 = B_{m,1}(2r) \setminus B_{m,1}(r), \\ z, & x \in \Omega \setminus B_{m,1}(2r). \end{cases} \tag{5.53}$$

By the product rule (see Theorem 2.9), applied to the functions $(z_n - z) \in W_0^{1,p}(\Omega)$ and $\chi_1 \in C_0^\infty(\Omega)$, it follows that $(z_n - z)\chi_1 \in W^{1,p}(\Omega)$. Moreover,

$$\frac{\partial \psi_n}{\partial x_i} = \frac{\partial z}{\partial x_i} + \frac{\partial (z_n - z)}{\partial x_i}\chi_1 + (z_n - z)\frac{\partial \chi_1}{\partial x_i} \tag{5.54}$$

and therefore,

$$\frac{\partial \psi_n}{\partial x_i} = \begin{cases} \dfrac{\partial z_n}{\partial x_i}, & x \in B_{m,1}(r), \\ \dfrac{\partial z}{\partial x_i} + \dfrac{\partial (z_n - z)}{\partial x_i}\chi_1 + (z_n - z)\dfrac{\partial \chi_1}{\partial x_i}, & x \in A_1 = B_{m,1}(2r) \setminus B_{m,1}(r), \\ \dfrac{\partial z}{\partial x_i}, & x \in \Omega \setminus B_{m,1}(2r). \end{cases} \tag{5.55}$$

Since the support of $(z_n - z)\chi_1$ is a compact set in Ω, by Theorem 2.10, it follows that $(z_n - z)\chi_1 \in W_0^{1,p}(\Omega)$ and consequently, $\psi_n \in W_0^{1,p}(\Omega)$, $\forall n \in \mathbb{N}$. From (5.53) and (5.55), by using the fact that $z = \frac{1}{4\pi}(G_1 + u_1) \in H^1(\Omega \setminus B_{m,1}(r))$, we find that $\psi_n \in H^1(\Omega)$, $\forall n \in \mathbb{N}$ and consequently $\psi_n \in H_0^1(\Omega)$, $\forall n \in \mathbb{N}$ (see Remark 5.9).

Now, for every $n \in \mathbb{N}$ we have

$$\int\limits_{B_{m,1}(r)} \epsilon \nabla u_2 \cdot \nabla \psi_n dx + \int\limits_{A_1} \epsilon \nabla u_2 \cdot \nabla \psi_n dx + \int\limits_{\Omega \setminus B_{m,1}(2r)} \epsilon \nabla u_2 \cdot \nabla \psi_n dx = \langle \mathcal{T}_2, \psi_n \rangle \tag{5.56}$$

Obviously, $\|\nabla(\psi_n - z)\|_{L^p(B_{m,1}(r))} \xrightarrow[n\to\infty]{} 0$, whereas on the set A_1, for all $i = 1, 2, 3$ we have

$$\begin{aligned} \left\| \frac{\partial \psi_n}{\partial x_i} - \frac{\partial z}{\partial x_i} \right\|_{L^p(A_1)} &\leq \left\| \frac{\partial (z_n - z)}{\partial x_i}\chi_1 \right\|_{L^p(A_1)} + \left\| (z_n - z)\frac{\partial \chi_1}{\partial x_i} \right\|_{L^p(A_1)} \\ \leq \left\| \frac{\partial (z_n - z)}{\partial x_i} \right\|_{L^p(A_1)} \|\chi_1\|_{L^\infty(\Omega)} &+ \|(z_n - z)\|_{L^p(A_1)} \left\| \frac{\partial \chi_1}{\partial x_i} \right\|_{L^\infty(\Omega)} \to 0 \end{aligned} \tag{5.57}$$

and therefore $\|\nabla(\psi_n - z)\|_{L^p(A_1)} \xrightarrow[n\to\infty]{} 0$.

By applying Hölder's inequality and using the fact that $\nabla u_2 \in L^q(B_{m,1}(2r))$ (by Theorem 3.4

and the Sobolev embedding theorems) we obtain

$$\left| \int_{B_{m,1}(r)} \epsilon \nabla u_2 \cdot \nabla(\psi_n - z) dx \right| \leq \epsilon_{\max} \|\nabla u_2\|_{L^q(B_{m,1}(r))} \|\nabla(\psi_n - z)\|_{L^p(B_{m,1}(r))} \to 0,$$

$$\left| \int_{A_1} \epsilon \nabla u_2 \cdot \nabla(\psi_n - z) dx \right| \leq \epsilon_{\max} \|\nabla u_2\|_{L^q(A_1)} \|\nabla(\psi_n - z)\|_{L^p(A_1)} \to 0,$$

$$\int_{\Omega \setminus B_{m,1}(2r)} \epsilon \nabla u_2 \cdot \nabla \psi_n dx = \int_{\Omega \setminus B_{m,1}(2r)} \epsilon \nabla u_2 \cdot \nabla z dx, \ \forall n \in \mathbb{N},$$

$$\langle \mathcal{T}_2, \psi_n \rangle = \int_{\Omega_s} (\epsilon_m - \epsilon) \nabla G_2 \cdot \nabla \psi_n dx = \langle \mathcal{T}_2, z \rangle, \ \forall n \in \mathbb{N}$$

Finally, letting $n \to \infty$ in (5.56) we obtain (5.46). □

Remark 5.9
Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. If $z \in W^{1,p}(\Omega) \cap W^{1,q}(\Omega)$ with $1 < p < q < \infty$, then $\gamma_p(z) = \gamma_q(z)$ a.e. on $\partial\Omega$. Indeed, let $z_n \in C^\infty(\overline{\Omega})$ such that $z_n \to z$ in $W^{1,q}(\Omega)$. For this sequence, $z_n\restriction_{\partial\Omega} = \gamma_q(z_n) \to \gamma_q(z)$ in $L^q(\partial\Omega)$. But we also have that $z_n \to z$ in $W^{1,p}(\Omega)$ $(1 \leq p < q)$ and thus $z_n\restriction_{\partial\Omega} = \gamma_p(z_n) \to \gamma_p(z)$ in $L^p(\partial\Omega)$. From here we see that $z_n\restriction_{\partial\Omega} \to \gamma_q(z)$ in $L^p(\partial\Omega)$ and $z_n\restriction_{\partial\Omega} \to \gamma_p(z)$ in $L^p(\partial\Omega)$. This means that $\gamma_p(z) = \gamma_q(z)$ a.e. on $\partial\Omega$.

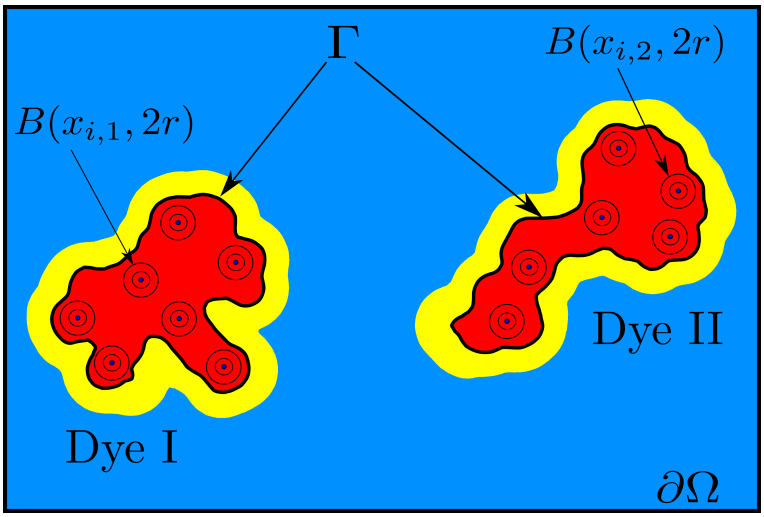


Figure 5.2: Different regions in the case $I_s \neq 0$.

Now, by using first (5.45) and then (5.46) we can derive the error equality for

$\langle \frac{1}{4\pi}\mathscr{F}_1, u_2 - u_{2,h}\rangle$ in terms of the adjoint solution z

$$\begin{aligned}\langle \frac{1}{4\pi}\mathscr{F}_1, u_2 - u_{2,h}\rangle &= \int_\Omega \epsilon\nabla\left(u_2 - u_{2,h}\right)\cdot\nabla z dx + \int_\Omega \overline{k}^2\left(u_2 - u_{2,h}\right) z dx \\ &= \langle \mathcal{T}_2, z\rangle - \int_\Omega \epsilon\nabla u_{2,h}\cdot\nabla z dx - \int_\Omega \overline{k}^2 u_{2,h} z dx =: E(u_{2,h}, z).\end{aligned} \tag{5.58}$$

Note that in order to evaluate the error estimator $E(u_{2,h}, z)$, one has to know the exact solution z of the adjoint problem or an approximation of it. As we already discussed in Section 5.3.1, in practice one finds a good approximation of z in a richer space than the one in which $u_{2,h}$ is found. In particular, we use the Galerkin finite element approximation $z_h^{(2)}$ of z from the space V_h^2 defined on the same mesh $\mathscr{T}_h$ on which V_h^1 is defined. In this case, by using Galerkin orthogonality, similarly to (5.17) (see [13, 159]) we can write

$$\begin{aligned}E(u_{2,h}, z_h^{(2)}) =& \langle \mathcal{T}_2, z_h^{(2)} - I_h z_h^{(2)}\rangle - \int_\Omega \epsilon\nabla u_{2,h}\cdot\nabla(z_h^{(2)} - I_h z_h^{(2)})dx \\ &- \int_\Omega \overline{k}^2 u_{2,h}(z_h^{(2)} - I_h z_h^{(2)})dx \\ =& \sum_{K\in\mathscr{T}_h} \eta_K(u_{2,h}, z_h^{(2)}) := \eta(u_{2,h}, z_h^{(2)}) \\ =& \sum_{i=1}^{N_V} \eta_i^{PU}(u_{2,h}, z_h^{(2)}) =: \eta^{PU}(u_{2,h}, z_h^{(2)}),\end{aligned} \tag{5.59}$$

where N_V is the number of vertices in the mesh $\mathscr{T}_h$, $\{\psi_i\}_{i=1}^{N_V}$ are the nodal P_1 basis functions, I_h is the nodal interpolation operator in V_h^1, and

$$\begin{aligned}\eta_K(u_{2,h}, z_h^{(2)}) =& \left((\epsilon_m - \epsilon_s)\nabla G_2\chi_{\Omega_s} - \epsilon\nabla u_{2,h}, \nabla(z_h^{(2)} - I_h z_h^{(2)})\right)_K \\ &- \left(\overline{k}^2(G_2 + u_{2,h}), (z_h^{(2)} - I_h z_h^{(2)})\right)_K,\end{aligned} \tag{5.60}$$

$$\begin{aligned}\eta_i^{PU}(u_{2,h}, z_h^{(2)}) =& \left((\epsilon_m - \epsilon_s)\nabla G_2\chi_{\Omega_s} - \epsilon\nabla u_{2,h}, \nabla\left((z_h^{(2)} - I_h z_h^{(2)})\psi_i\right)\right) \\ &- \left(\overline{k}^2(G_2 + u_{2,h}), (z_h^{(2)} - I_h z_h^{(2)})\psi_i\right).\end{aligned} \tag{5.61}$$

Taking into account (5.7) and (5.9), the approximate electrostatic interaction E_{2-1}^P is given by

$$E_{2-1}^P = E_{G_2} + \langle \frac{1}{4\pi}\mathscr{F}_1, u_{2,h}\rangle \tag{5.62}$$

and the error estimation for the electrostatic interaction E_{2-1} is

$$\begin{aligned}E_{2-1} - E_{2-1}^P &= E_{G_2} + \langle \frac{1}{4\pi}\mathscr{F}_1, u_2\rangle - \left(E_{G_2} + \langle \frac{1}{4\pi}\mathscr{F}_1, u_{2,h}\rangle\right) \\ &= \langle \frac{1}{4\pi}\mathscr{F}_1, u_2 - u_{2,h}\rangle \\ &= E(u_{2,h}, z) = E(u_{2,h}, z_h^{(2)}) + \mathcal{E}(u_2, u_{2,h}, z, z_h^{(2)}).\end{aligned} \tag{5.63}$$

In view of Section 5.3.1 we expect that the term $\mathcal{E}(u_2, u_{2,h}, z, z_h^{(2)})$ again converges faster than the term $E(u_{2,h}, z_h^{(2)})$. For this reason the following approximate equality holds:

$$E_{2-1} - E_{2-1}^P \approx E(u_{2,h}, z_h^{(2)}), \tag{5.64}$$

which can also be written in the form

$$E_{2-1} \approx E_{2-1}^P + E(u_{2,h}, z_h^{(2)}). \tag{5.65}$$

The quantity $E_{2-1}^P + E(u_{2,h}, z_h^{(2)})$ is a corrected value for the electrostatic interaction and can be used as a better approximation of the interaction E_{2-1}. However, in practice, it is better to use the quantity $E_{1-2}^A := \langle \mathscr{F}_2, z_h^{(2)} \rangle$, computed with the more accurate *Adjoint* solution $z_h^{(2)}$, since it is easier to compute and numerically more stable. Indeed, as we already mentioned, the solution z of the adjoint problem coincides with $\frac{1}{4\pi}\varphi_1$, where φ_1 is the exact potential created by the charges of the first molecule. Therefore, $\langle \mathscr{F}_2, z_h^{(2)} \rangle$ is an approximation to the electrostatic interaction $E_{1-2} = \langle \mathscr{F}_2, z \rangle = \langle \mathscr{F}_2, \frac{1}{4\pi}\varphi_1 \rangle$ which is equal to E_{2-1}. In Section 5.5.1, we demonstrate the improved accuracy of E_{1-2}^A compared to E_{2-1}^P and in Section 5.4 we use E_{1-2}^A as a better approximation to the electrostatic interaction in an application related to FRET.

Alternative approach by 2-term splitting in the adjoint problem

To evaluate approximately $E(u_{2,h}, z)$ we can also apply the 2-term splitting to the adjoint problem and use the approximation $\tilde{z} = \frac{1}{4\pi}G_1 + w_\tau^{k^*} \notin V_\tau^{k^*}$, where $w_\tau^{k^*} \in V_\tau^{k^*}$ is a conforming finite element approximation of the regular part $w = \frac{1}{4\pi}u_1$. Here, the mesh $\mathscr{T}_\tau$ may or may not coincide with $\mathscr{T}_h$ and k^* may or may not be equal to k. We can write

$$E(u_{2,h}, z) = E(u_{2,h}, \tilde{z}) + \mathcal{E}(u_2, u_{2,h}, z, \tilde{z}), \tag{5.66}$$

where

$$E(u_{2,h}, \tilde{z}) = \langle \mathcal{T}_2, \tilde{z} \rangle - \int_\Omega \epsilon \nabla u_{2,h} \cdot \nabla \tilde{z} dx - \int_\Omega \overline{k}^2 u_{2,h} \tilde{z} dx, \tag{5.67}$$

$$\begin{aligned}\mathcal{E}(u_2, u_{2,h}, z, \tilde{z}) &= \langle \mathcal{T}_2, z - \tilde{z} \rangle - \int_\Omega \epsilon \nabla u_{2,h} \cdot \nabla (z - \tilde{z}) dx - \int_\Omega \overline{k}^2 u_{2,h} (z - \tilde{z}) dx \\ &= \int_\Omega \epsilon \nabla (u_2 - u_{2,h}) \cdot \nabla (w - w_\tau^{k^*}) dx + \int_\Omega \overline{k}^2 (u_2 - u_{2,h}) (w - w_\tau^{k^*}) dx.\end{aligned} \tag{5.68}$$

Here we have used the fact that $z - \tilde{z} = w - w_\tau^{k^*}$ $\big(\in H_0^1(\Omega)\big)$. Note that the expression $-\int_\Omega \epsilon \nabla u_{2,h} \cdot \nabla \tilde{z} dx$ is well defined since $\tilde{z} \in L^p(\Omega)$ for all $p < \frac{d}{d-1}$ and $\nabla u_{2,h} \in L^\infty(\Omega)$. The expressions $\langle \mathcal{T}_2, z \rangle$, $\langle \mathcal{T}_2, \tilde{z} \rangle$, $-\int_\Omega \overline{k}^2 u_{2,h} z dx$, and $-\int_\Omega \overline{k}^2 u_{2,h} \tilde{z} dx$ are well defined because $\overline{k}^2$ is zero in Ω_m where z and $\tilde{z}$ have singularities.

Note that even if $V_h^k \equiv V_\tau^{k^*}$ (and the approximation $w_\tau^{k^*}$ is from the space V_h^k), the term $E(u_{2,h}, \tilde{z})$ is in general different from zero. Another thing to keep in mind is that we need to compute the integral $\int\limits_\Omega \frac{\epsilon}{4\pi} \nabla u_{2,h} \cdot \nabla G_1 dx$ which has singularities at the positions $x_{i,1}$, $i = 1, \ldots, N_1$ of the point charges $q_{i,1}$, $i = 1, \ldots, N_1$. In the case of piecewise constant ϵ and a piecewise linear approximation $u_{2,h}$ this integration can be carried out analytically. Otherwise, appropriate quadrature rules have to be used (see, e.g., [108, 145]).

As in (5.15), we can observe by comparing $E(u_{2,h}, \tilde{z})$ and $\mathcal{E}(u_2, u_{2,h}, z, \tilde{z})$ that asymptotically $E(u_{2,h}, \tilde{z})$ contains the major part of the error $\langle \frac{1}{4\pi} \mathscr{F}_1, u_2 - u_{2,h} \rangle$. Indeed,

$$
\begin{aligned}
|E(u_{2,h}, \tilde{z})| &= \left| \langle \mathcal{T}_2, \tilde{z} \rangle - \int\limits_\Omega \epsilon \nabla u_{2,h} \cdot \nabla \tilde{z} dx - \int\limits_\Omega \overline{k}^2 u_{2,h} \tilde{z} dx \right| \\
&\leq \left| \int\limits_\Omega \epsilon \nabla (u_2 - u_{2,h}) \cdot \nabla \tilde{z} dx \right| + \left| \int\limits_\Omega \overline{k}^2 (u_2 - u_{2,h}) \tilde{z} dx \right| \\
&= \left| \int\limits_{B_{m,1}(r)} \epsilon \nabla (u_2 - u_{2,h}) \cdot \nabla \tilde{z} dx + \int\limits_{\Omega \setminus B_{m,1}(r)} \epsilon \nabla (u_2 - u_{2,h}) \cdot \nabla \tilde{z} dx \right| \\
&+ \overline{k}_{ions}^2 \|u_2 - u_{2,h}\|_{L^2(\Omega)} \|\tilde{z}\|_{L^2(\Omega_{ions})} \\
&\leq \epsilon_{\max} \Big(\|\nabla (u_2 - u_{2,h})\|_{L^q(B_{m,1}(r))} \|\nabla \tilde{z}\|_{L^p(B_{m,1}(r))} \\
&\qquad + \|\nabla (u_2 - u_{2,h})\|_{L^2(\Omega \setminus B_{m,1}(r))} \|\nabla \tilde{z}\|_{L^2(\Omega \setminus B_{m,1}(r))} \Big) \\
&+ \overline{k}_{ions}^2 \|u_2 - u_{2,h}\|_{L^2(\Omega)} \|\tilde{z}\|_{L^2(\Omega_{ions})} \\
&\leq \epsilon_{\max} C_1 \left(\|\nabla (u_2 - u_{2,h})\|_{L^q(B_{m,1}(r))} + \|\nabla (u_2 - u_{2,h})\|_{L^2(\Omega \setminus B_{m,1}(r))} \right) \\
&+ \overline{k}_{ions}^2 C_2 \|u_2 - u_{2,h}\|_{L^2(\Omega)} \\
&\leq C_3 \left(\|\nabla (u_2 - u_{2,h})\|_{L^q(B_{m,1}(r))} + \|\nabla (u_2 - u_{2,h})\|_{L^2(\Omega \setminus B_{m,1}(r))} + \|u_2 - u_{2,h}\|_{L^2(\Omega)} \right), \quad (5.69)
\end{aligned}
$$

where C_1, C_2, C_3 are generic constants. In the derivation of the upper bound (5.69) we have used the following facts:
1) that we can test the weak formulation (5.10) for u_2 with $\tilde{z}$ (similar to the proof of (5.46))
2) that u_2 and $u_{2,h}$ are in $W^{1,q}(B_{m,1}(r)$
3) that $\tilde{z} \in L^2(\Omega \setminus B_{m,1}(r))$
4) that $\tilde{z} = \frac{1}{4\pi} G_1 + w_\tau^{k^*}$ and that $\|w_\tau^{k^*}\|_{L^2(\Omega)}$ and $\|\nabla w_\tau^{k^*}\|_{L^2(\Omega}$ are bounded since $w_\tau^{k^*}$ converges to w in $H^1(\Omega)$
5) that $\|\nabla w_\tau^{k^*}\|_{L^p(B_{m,1}(r))} \leq |B_{m,1}(r)|^{\frac{2-p}{2p}} \|\nabla w_\tau^{k^*}\|_{L^2(B_{m,1}(r))}$.

For the term $\mathcal{E}(u_2, u_{2,h}, z, \tilde{z})$ we have the following estimate

$$\begin{aligned}
\mathcal{E}(u_2, u_{2,h}, z, \tilde{z}) &\leq \left| \int_\Omega \epsilon \nabla(u_2 - u_{2,h}) \cdot \nabla(w - w_\tau^{k^*}) dx \right| + \left| \int_\Omega \overline{k}^2 (u_2 - u_{2,h})(w - w_\tau^{k^*}) dx \right| \\
&\leq \left| \int_{B_{m,1}(r)} \epsilon \nabla(u_2 - u_{2,h}) \cdot \nabla(w - w_\tau^{k^*}) dx + \int_{\Omega \setminus B_{m,1}(r)} \epsilon \nabla(u_2 - u_{2,h}) \cdot \nabla(w - w_\tau^{k^*}) dx \right| \\
&+ \overline{k}_{ions}^2 \|u_2 - u_{2,h}\|_{L^2(\Omega)} \|w - w_\tau^{k^*}\|_{L^2(\Omega)} \\
&\leq \epsilon_{\max} \Big(\|\nabla(u_2 - u_{2,h})\|_{L^q(B_{m,1}(r))} \|\nabla(w - w_\tau^{k^*})\|_{L^p(B_{m,1}(r))} \\
&\qquad + \|\nabla(u_2 - u_{2,h})\|_{L^2(\Omega \setminus B_{m,1}(r))} \|\nabla(w - w_\tau^{k^*})\|_{L^2(\Omega \setminus B_{m,1}(r))} \Big) \\
&+ \overline{k}_{ions}^2 C_P \|u_2 - u_{2,h}\|_{L^2(\Omega)} \|\nabla(w - w_\tau^{k^*})\|_{L^2(\Omega)} \\
&\leq C_4 \left(\|\nabla(u_2 - u_{2,h})\|_{L^q(B_{m,1}(r))} + \|\nabla(u_2 - u_{2,h})\|_{L^2(\Omega \setminus B_{m,1}(r))} \right) \|\nabla(w - w_\tau^{k^*})\|_{L^2(\Omega)} \\
&+ \overline{k}_{ions}^2 C_P \|u_2 - u_{2,h}\|_{L^2(\Omega)} \|\nabla(w - w_\tau^{k^*})\|_{L^2(\Omega)} \\
&\leq C_5 \left(\|\nabla(u_2 - u_{2,h})\|_{L^q(B_{m,1}(r))} + \|\nabla(u_2 - u_{2,h})\|_{L^2(\Omega \setminus B_{m,1}(r))} + \|u_2 - u_{2,h}\|_{L^2(\Omega)} \right) \\
&\|\nabla(w - w_\tau^{k^*})\|_{L^2(\Omega)}, \qquad (5.70)
\end{aligned}$$

where C_4 and C_5 are generic constants and C_P is Poincaré's constant in the inequality $\|v\|_{L^2(\Omega)} \leq C_P \|\nabla v\|_{L^2(\Omega)}$ for $v \in H_0^1(\Omega)$. In the derivation of the upper bound (5.70) we have used the fact that $p < \frac{d}{d-1} \leq 2$ and consequently the $L^p(B_{m,1}(r))$ norm of $\nabla(w - w_\tau^{k^*})$ can be bounded from above by the $L^2(B_{m,1}(r))$ norm of $\nabla(w - w_\tau^{k^*})$. By comparing the upper bounds (5.69) and (5.70) we can assume that the term $\mathcal{E}(u_2, u_{2,h}, z, \tilde{z})$ converges faster to zero than $E(u_{2,h}, \tilde{z})$ and thus the major part of the error in (5.66) is asymptotically contained in $E(u_{2,h}, \tilde{z})$ (see [115,139] for similar arguments in the case of high regularity of the solutions of the primal and adjoint problem). It is also seen that if $w_\tau^{k^*} \to w$ in $H^1(\Omega)$, then

$$\langle \frac{1}{4\pi} \mathscr{F}_1, u_2 - u_{2,h} \rangle = E(u_{2,h}, \tilde{z}) + \mathcal{E}(u_2, u_{2,h}, z, \tilde{z}) \to E(u_{2,h}, \tilde{z})$$

Thus, we can write the approximate equality

$$\langle \frac{1}{4\pi} \mathscr{F}_1, u_2 - u_{2,h} \rangle \approx E(u_{2,h}, \tilde{z}),$$

where $E(u_{2,h}, \tilde{z})$ is fully computable. Since u_2, $w \in H^1(\Omega)$, all post processing techniques from [100,115,119,139] and [138] can be applied to approximately evaluate the term $\mathcal{E}(u_2, u_{2,h}, z, \tilde{z})$ although superconvergence of the gradients cannot be guaranteed since our primal and adjoint problems have solutions of low regularity.

5.3.3 Error estimation 3: no splitting in primal problem and regular goal functional

Unfortunately, the 2-term regularization is not always appropriate to use. Typically, in the presence of a strong dielectric screening in the solvent region Ω_s, i.e. $\frac{\epsilon_m}{\epsilon_s} \ll 1$, the analytically known component E_{G_2} and the reaction field component E_{u_2} have an opposite sign but almost the same absolute value. Thus, for the full electrostatic interaction we have

$$|E_{G_2} + E_{u_2}| = |E_{2-1}| \ll |E_{u_2}|\,,$$

which means that a small relative error in the numerically approximated component E_{u_2} results in a large relative error in the full interaction E_{2-1}. It may very well happen that $|E_{2-1}| = 0.001 \times |E_{u_2}|$. This means that a relative error of 1% in $|E_{u_2}|$ ($|e| = 0.01 \times |E_{u_2}|$) results in a relative error of 1000% in E_{2-1}:

$$\frac{|e|}{|E_{2-1}|} = \frac{0.01 \times |E_{u_2}|}{0.001 \times |E_{u_2}|} = 10.$$

In such situations, it is best not to apply the 2-term splitting, but to solve directly equation (5.2), which we recall here

$$\varphi_2 \in \bigcap_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega), \quad \int_\Omega \epsilon \nabla \varphi_2 \cdot \nabla v dx + \int_\Omega \overline{k}^2 \varphi_2 v dx = \langle \mathscr{F}_2, v \rangle, \; \forall v \in \bigcup_{q>d} W_0^{1,q}(\Omega). \tag{5.2}$$

We consider the goal functional $\frac{1}{4\pi}\overline{\mathscr{F}}_1$ defined by (5.27) which is given by a linear combination of averaging operators over the balls $B(x_{i,1}, \overline{r}_{i,1})$. This is justified since the function φ_2 is harmonic in $\Omega_{m,1}$ as the sum of the harmonic in Ω_m functions G_2 and u_2. The electrostatic interaction E_{2-1} can be expressed in terms of the goal functional as follows

$$E_{2-1} = \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_2 \rangle = \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, \varphi_2 \rangle. \tag{5.71}$$

The corresponding adjoint problem is the same as (5.29) and we also recall it here

$$\begin{aligned} &\text{Find } z \in H_0^1(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla v \cdot \nabla z dx + \int_\Omega \overline{k}^2 v z dx = \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, v \rangle, \; \forall v \in H_0^1(\Omega), \end{aligned} \tag{5.29}$$

Let $\varphi_{2,h}$ denote the Galerkin finite element approximation of φ_2 in the space $V_{0,h}^k \subset \bigcap_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega)$ of continuous piecewise polynomial functions of degree k over a mesh $\mathscr{T}_h$

$$\text{Find } \varphi_{2,h} \in V_{0,h}^k, \text{ s.t.} \quad \int_\Omega \epsilon \nabla \varphi_{2,h} \cdot \nabla v dx + \int_\Omega \overline{k}^2 \varphi_{2,h} v dx = \langle \mathscr{F}_2, v \rangle, \; \forall v \in V_{0,h}^k. \tag{5.72}$$

In our numerical experiments, we always take $k = 1$. To derive an error representation involving the adjoint solution z, similar to (5.31), we will need the following Proposition 5.10.

Proposition 5.10
The following equalities hold true

$$\int_\Omega \epsilon\nabla(\varphi_2-\varphi_{2,h})\cdot\nabla z dx+\int_\Omega \overline{k}^2(\varphi_2-\varphi_{2,h})z dx=\langle\frac{1}{4\pi}\overline{\mathscr{F}}_1,\varphi_2-\varphi_{2,h}\rangle, \tag{5.73}$$

$$\int_\Omega \epsilon\nabla\varphi_2\cdot\nabla z dx+\int_\Omega \overline{k}^2\varphi_2 z dx=\langle\mathscr{F}_2,z\rangle. \tag{5.74}$$

Moreover,

$$\langle\frac{1}{4\pi}\overline{\mathscr{F}}_1,\varphi_2\rangle=\langle\mathscr{F}_2,z\rangle. \tag{5.75}$$

In order to avoid some technical details we will prove Proposition 5.10 only for the case $I_s=0$, i.e., $\overline{k}=0$ and note that the proof for the case $I_s\neq 0$ is similar.

Proof. The proof is similar to the proof of Proposition 5.6. Let $B_{m,2}(r):=\bigcup_{i=1}^{N_2}B(x_{i,2},r)$, where $B(x_{i,2},r)$ is the open ball with center at $x_{i,2}$ and a radius r. Again, here, r is chosen so small that $B(x_{i,2},2r)$ is strictly contained in Ω_m and $\overline{B(x_{i,2},2r)}\cap\overline{B(x_{j,2},2r)}=\emptyset$ for all $i\neq j$, $i,j=1,2,\ldots,N_2$ and, additionally, $\overline{B(x_{i,2},2r)}\cap\overline{B(x_{j,1},2r)}=\emptyset$ for all $i=1,2,\ldots,N_2$ and all $j=1,2,\ldots,N_1$ (see Figure 5.2). We also assume that $\overline{B_{m,2}(2r)}\cap\bigcup\limits_{i=1}^{N_1}\overline{B(x_{i,1},\overline{r}_{i,1})}=\emptyset$. The idea to prove (5.73) is to approximate $\varphi_2-\varphi_{2,h}$ with functions ψ_n in $H_0^1(\Omega)$ such that $\psi_n\to\varphi_2-\varphi_{2,h}$ in $H_0^1\left(\Omega\setminus B_{m,2}(2r)\right)$ and $\psi_n\to\varphi_2-\varphi_{2,h}$ in $W^{1,p}(B_{m,2}(2r))$ for some $p<\frac{d}{d-1}$. Similarly, to prove (5.74) we find functions $\psi_n\in W_0^{1,q}(\Omega)$ such that $\psi_n\to z$ in $W^{1,q}(B_{m,2}(r))$ for some $q>d$ and $\psi_n\to z$ in $H^1\left(\Omega\setminus B_{m,2}(r)\right)$. Thus, let us fix q such that $d<q\leq 6$ and let p be its Hölder conjugate, i.e., p is such that $\frac{1}{p}+\frac{1}{q}=1$. It is clear that p satisfies $p<\frac{d}{d-1}$.

Now, φ_2 has singularities at the positions $x_{i,2}$ of the charges $q_{i,2}$ of the second dye. Testing equation (5.29) with

$$v\in C_0^\infty\left(\Omega_m\setminus\bigcup_{i=1}^{N_1}\overline{B(x_{i,1},\overline{r}_{i,1})}\right)$$

we see that (see p. 310 in [70] for interior regularity of linear elliptic problems)

$$z\in H_0^1(\Omega)\cap H^t_{\text{loc}}\left(\Omega_m\setminus\bigcup_{i=1}^{N_1}\overline{B(x_{i,1},\overline{r}_{i,1})}\right)\text{ for all intiger } t\geq 2.$$

From the Sobolev embedding theorem, it follows that $z\in W^{1,q}\left(\Omega_m\setminus\bigcup\limits_{i=1}^{N_1}\overline{B(x_{i,1},\overline{r}_{i,1})}\right)$ and consequently $z\in W^{1,q}\left(B_{m,2}(2r)\right)$ which is also true for $I_s>0$, i.e., $\overline{k}^2_{ions}>0$.

Proof of (5.73)
Since $\overline{w}:=\varphi_2-\varphi_{2,h}\in W_0^{1,p}(\Omega)$, there exist functions $w_n\in C_0^\infty(\Omega)$ such that $\|w_n-$

$\overline{w}\|_{W^{1,p}(\Omega)} \xrightarrow[n\to\infty]{} 0$. We define the functions $\psi_n := \overline{w} + (w_n - \overline{w})\chi_2$, where χ_2 is a smooth cut-off function such that

$$\chi_2(x) = \begin{cases} 1, & x \in B_{m,2}(r), \\ 0, & x \in \Omega \setminus B_{m,2}(2r). \end{cases} \tag{5.76}$$

We have

$$\psi_n(x) = \begin{cases} w_n, & x \in B_{m,2}(r), \\ \overline{w} + (w_n - \overline{w})\chi_2, & x \in A_2 := B_{m,2}(2r) \setminus B_{m,2}(r), \\ \overline{w}, & x \in \Omega \setminus B_{m,2}(2r). \end{cases} \tag{5.77}$$

All functions ψ_n are clearly in $W_0^{1,p}(\Omega)$, as the sum of such functions. Indeed, by the product rule (see Theorem 2.9) applied to the functions $w_n - \overline{w} \in W_0^{1,p}(\Omega)$ and $\chi_2 \in C_0^\infty(\Omega)$, it follows that $(w_n - \overline{w})\chi_2 \in W^{1,p}(\Omega)$. Since $(w_n - \overline{w})\chi_2$ has a compact support in Ω, by applying Theorem 2.10, we see that $(w_n - \overline{w})\chi_2 \in W_0^{1,p}(\Omega)$. Moreover, the weak derivatives of ψ_n are given by

$$\frac{\partial \psi_n}{\partial x_i} = \frac{\partial \overline{w}}{\partial x_i} + \frac{\partial}{\partial x_i}\left((w_n - \overline{w})\chi_2\right) = \frac{\partial \overline{w}}{\partial x_i} + \frac{\partial (w_n - \overline{w})}{\partial x_i}\chi_2 + (w_n - \overline{w})\frac{\partial \chi_2}{\partial x_i},\ i = 1,2,3 \tag{5.78}$$

and therefore

$$\frac{\partial \psi_n}{\partial x_i} = \begin{cases} \dfrac{\partial w_n}{\partial x_i}, & x \in B_{m,2}(r), \\ \dfrac{\partial \overline{w}}{\partial x_i} + \dfrac{\partial (w_n - \overline{w})}{\partial x_i}\chi_2 + (w_n - \overline{w})\dfrac{\partial \chi_2}{\partial x_i}, & x \in A_2 = B_{m,2}(2r) \setminus B_{m,2}(r), \\ \dfrac{\partial \overline{w}}{\partial x_i}, & x \in \Omega \setminus B_{m,2}(2r). \end{cases} \tag{5.79}$$

From (5.77) and (5.79), by observing that $\varphi_2 = G_2 + u_2 \in H^1\left(\Omega \setminus B_{m,2}(r)\right)$, we see that over each subdomain, ψ_n, $\frac{\partial \psi_n}{\partial x_i} \in L^2$ and thus $\psi_n \in H^1(\Omega)$. Because $\psi_n \in W_0^{1,p}(\Omega)$ it also follows that $\psi_n \in H_0^1(\Omega)$ (see Remark 5.9). Now, for all $n \in \mathbb{N}$ we have

$$\int\limits_{B_{m,2}(r)} \epsilon \nabla z \cdot \nabla \psi_n dx + \int\limits_{A_2} \epsilon \nabla z \cdot \nabla \psi_n dx + \int\limits_{\Omega \setminus B_{m,2}(2r)} \epsilon \nabla z \cdot \nabla \psi_n dx = \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, \psi_n \rangle \tag{5.80}$$

Obviously, $\|\nabla(\psi_n - \overline{w})\|_{L^p(B_{m,2}(r))} \xrightarrow[n\to\infty]{} 0$, whereas on the set A_2, for all $i = 1,2,3$ we have

$$\begin{aligned} \left\| \frac{\partial \psi_n}{\partial x_i} - \frac{\partial \overline{w}}{\partial x_i} \right\|_{L^p(A_2)} &\leq \left\| \frac{\partial \left(w_n - \overline{w}\right)}{\partial x_i}\chi_2 \right\|_{L^p(A_2)} + \left\| (w_n - \overline{w})\frac{\partial \chi_2}{\partial x_i} \right\|_{L^p(A_2)} \\ \leq \left\| \frac{\partial \left(w_n - \overline{w}\right)}{\partial x_i} \right\|_{L^p(A_2)} &\|\chi_2\|_{L^\infty(\Omega)} + \|(w_n - \overline{w})\|_{L^p(A_2)} \left\| \frac{\partial \chi_2}{\partial x_i} \right\|_{L^\infty(\Omega)} \to 0 \end{aligned} \tag{5.81}$$

and therefore $\|\nabla(\psi_n - \overline{w})\|_{L^p(A_2)} \xrightarrow[n\to\infty]{} 0$.

By applying Hölder's inequality and using the fact that $\nabla z \in L^q\left(B_{m,2}(2r)\right)$ we obtain

$$\left| \int\limits_{B_{m,2}(r)} \epsilon \nabla z \cdot \nabla(\psi_n - \overline{w}) dx \right| \leq \epsilon_{\max} \|\nabla z\|_{L^q(B_{m,2}(r))} \|\nabla(\psi_n - \overline{w})\|_{L^p(B_{m,2}(r))} \xrightarrow[n\to\infty]{} 0,$$

$$\left| \int\limits_{A_2} \epsilon \nabla z \cdot \nabla(\psi_n - \overline{w}) dx \right| \leq \epsilon_{\max} \|\nabla z\|_{L^q(A_2)} \|\nabla(\psi_n - \overline{w})\|_{L^p(A_2)} \xrightarrow[n\to\infty]{} 0.$$

We also have that for all $n \in \mathbb{N}$,

$$\int\limits_{\Omega \setminus B_{m,2}(2r)} \epsilon \nabla z \cdot \nabla \psi_n dx = \int\limits_{\Omega \setminus B_{m,2}(2r)} \epsilon \nabla z \cdot \nabla \overline{w} dx,$$

where the integral on the right is well defined because $z \in H_0^1(\Omega)$ and $\psi_n \in H_0^1(\Omega)$. For the right-hand side of (5.80), we obviously have

$$\langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, \psi_n - \overline{w} \rangle = \int\limits_{\Omega} l(\psi_n - \overline{w}) dx = \int\limits_{\bigcup\limits_{i=1}^{N_1} \overline{B(x_{i,1}, \overline{r}_{i,1})}} l(\psi_n - \overline{w}) dx = 0 \text{ for all } n \in \mathbb{N}.$$

Finally, letting $n \to \infty$ in (5.80) we obtain (5.73).

Proof of (5.74)
Since $z \in H_0^1(\Omega)$, we can find functions $z_n \in C_0^\infty(\Omega)$, $n = 1, 2 \ldots$ such that $\|z_n - z\|_{H^1(\Omega)} \xrightarrow[n\to\infty]{} 0$. We define the functions $\psi_n := z_n + (z - z_n)\chi_2$, $n \in \mathbb{N}$. We have

$$\psi_n(x) = \begin{cases} z, & x \in B_{m,2}(r), \\ z_n + (z - z_n)\chi_2, & x \in A_2 := B_{m,2}(2r) \setminus B_{m,2}(r), \\ z_n, & x \in \Omega \setminus B_{m,2}(2r). \end{cases} \tag{5.82}$$

Obviously, the function $\psi_n \in H_0^1(\Omega)$ for all $n \in \mathbb{N}$ as the sum of such functions. Indeed, by the product rule (see Theorem 2.9) applied to the functions $z - z_n \in H_0^1(\Omega)$ and $\chi_2 \in C_0^\infty(\Omega)$, it follows that $(z - z_n)\chi_2 \in H^1(\Omega)$. Since $(z - z_n)\chi_2$ has a compact support in Ω, by applying Theorem 2.10, we see that $(z - z_n)\chi_2 \in H_0^1(\Omega)$. Moreover, the weak derivatives of ψ_n are given by

$$\frac{\partial \psi_n}{\partial x_i} = \frac{\partial z_n}{\partial x_i} + \frac{\partial (z - z_n)}{\partial x_i}\chi_2 + (z - z_n)\frac{\partial \chi_2}{\partial x_i}, \; i = 1, 2, 3 \tag{5.83}$$

and, therefore,

$$\frac{\partial \psi_n}{\partial x_i} = \begin{cases} \dfrac{\partial z}{\partial x_i}, & x \in B_{m,2}(r), \\ \dfrac{\partial z_n}{\partial x_i} + \dfrac{\partial (z - z_n)}{\partial x_i}\chi_2 + (z - z_n)\dfrac{\partial \chi_2}{\partial x_i}, & x \in A_2 = B_{m,2}(2r) \setminus B_{m,2}(r), \\ \dfrac{\partial z_n}{\partial x_i}, & x \in \Omega \setminus B_{m,2}(2r). \end{cases} \tag{5.84}$$

From (5.82) and (5.84), by using the fact that $z \in W^{1,q}\left(B_{m,2}(2r)\right)$, we can see that ψ_n, $\frac{\partial \psi_n}{\partial x_i} \in L^q(\Omega)$, $i = 1,2,3$ for all $n = 1,2,\ldots$ and, therefore, $\psi_n \in W^{1,q}(\Omega)$. Since supp $\psi_n \subset\subset \Omega$, again by Theorem 2.10, it follows that $\psi_n \in W_0^{1,q}(\Omega)$. Now, for all $n \in \mathbb{N}$ it holds that

$$\int\limits_{B_{m,2}(r)} \epsilon \nabla\varphi_2 \cdot \nabla\psi_n dx + \int\limits_{A_2} \epsilon \nabla\varphi_2 \cdot \nabla\psi_n dx + \int\limits_{\Omega\setminus B_{m,2}(2r)} \epsilon \nabla\varphi_2 \cdot \nabla\psi_n dx = \langle \mathscr{F}_2, \psi_n \rangle. \tag{5.85}$$

On the set A_2, for $i = 1,2,3$, since $\|z_n - z\|_{H^1(\Omega)} \to 0$ and χ_2, $\frac{\partial \chi_2}{\partial x_i}$, $i = 1,2,3$ are essentially bounded, we have

$$\begin{aligned} &\left\| \frac{\partial \psi_n}{\partial x_i} - \frac{\partial z}{\partial x_i} \right\|_{L^2(A_2)} \\ &\leq \left\| \frac{\partial z_n}{\partial x_i} - \frac{\partial z}{\partial x_i} \right\|_{L^2(A_2)} + \left\| \frac{\partial \left(z - z_n\right)}{\partial x_i} \chi_2 \right\|_{L^2(A_2)} + \left\| (z - z_n) \frac{\partial \chi_2}{\partial x_i} \right\|_{L^2(A_2)} \to 0. \end{aligned} \tag{5.86}$$

By using the fact that $\varphi_2 = G_2 + u_2 \in H^1\left(\Omega \setminus B_{m,2}(r)\right)$ and by applying Hölder's inequality we obtain

$$\left| \int\limits_{A_2} \epsilon \nabla\varphi_2 \cdot \nabla(\psi_n - z) dx \right| \leq \epsilon_{\max} \|\nabla\varphi_2\|_{L^2(A_2)} \|\nabla(\psi_n - z)\|_{L^2(A_2)} \to 0$$

$$\left| \int\limits_{\Omega\setminus B_{m,2}(2r)} \epsilon \nabla\varphi_2 \cdot \nabla(\psi_n - z) dx \right| \leq \epsilon_{\max} \|\nabla\varphi_2\|_{L^2(\Omega\setminus B_{m,2}(2r))} \|\nabla(\psi_n - z)\|_{L^2(\Omega\setminus B_{m,2}(2r))} \to 0$$

$$\langle \mathscr{F}_2, \psi_n \rangle = \langle \mathscr{F}_2, z \rangle \text{ for all } n \in \mathbb{N}.$$

By letting $n \to \infty$ in (5.85), we obtain (5.74). To show (5.75), we only need to verify that

$$\int\limits_{\Omega} \epsilon \nabla\varphi_2 \cdot \nabla z dx + \int\limits_{\Omega} \overline{k}^2 \varphi_2 z dx = \langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, \varphi_2 \rangle.$$

The proof of this equality is the same as the proof of (5.73) where instead of $\overline{w} = \varphi_2 - \varphi_{2,h}$ we take $\overline{w} = \varphi_2$. □

Now, by using first (5.73) and then (5.74) we can derive the error representation for $\langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, \varphi_2 - \varphi_{2,h} \rangle$ in terms of the adjoint solution $z \in H_0^1(\Omega)$

$$\begin{aligned} \langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, \varphi_2 - \varphi_{2,h} \rangle &= \int\limits_{\Omega} \epsilon \nabla \left(\varphi_2 - \varphi_{2,h}\right) \cdot \nabla z dx + \int\limits_{\Omega} \overline{k}^2 \left(\varphi_2 - \varphi_{2,h}\right) z dx \\ &= \langle \mathscr{F}_2, z \rangle - \int\limits_{\Omega} \epsilon \nabla\varphi_{2,h} \cdot \nabla z dx - \int\limits_{\Omega} \overline{k}^2 \varphi_{2,h} z dx =: E(\varphi_{2,h}, z). \end{aligned} \tag{5.87}$$

We recall the observation we made about z in Section 5.3.1 under the assumption that $\Gamma \in C^1$. Namely, there exists $q > d$ such that $z \in W_0^{1,q}(\Omega)$ which, by the Sobolev embedding theorem,

implies that z has a continuous representative. Therefore, the nodal P_1 interpolant in V_h^1 of z is well defined and we can write

$$\begin{aligned} E(\varphi_{2,h}, z) &= \langle \mathscr{F}_2, z - I_h z\rangle - \int_\Omega \epsilon \nabla \varphi_{2,h} \cdot \nabla (z - I_h z) dx - \int_\Omega \overline{k}^2 \varphi_{2,h} (z - I_h z) dx \\ &= \sum_{K \in \mathscr{T}_h} \eta_K(\varphi_{2,h}, z) =: \eta(\varphi_{2,h}, z) \\ &= \sum_{i=1}^{N_V} \underbrace{\left\{ \langle \mathscr{F}_2, (z - I_h z)\, \psi_i\rangle - (\epsilon \nabla \varphi_{2,h}, \nabla ((z - I_h z)\psi_i)) - \left(\overline{k}^2 \varphi_{2,h}, (z - I_h z)\, \psi_i\right)\right\}}_{=:\eta_i^{PU}} \\ &=: \eta^{PU}(\varphi_{2,h}, z), \end{aligned} \tag{5.88}$$

where N_V is the number of nodes in the mesh $\mathscr{T}_h$, I_h is the nodal interpolation operator in V_h^1, and

$$\begin{aligned} \eta_K(\varphi_{2,h}, z) = \langle \mathscr{F}_2, (z - I_h z)\chi_K\rangle - (\epsilon \nabla \varphi_{2,h}, \nabla (z - I_h z))_K \\ - \left(\overline{k}^2 \varphi_{2,h}, (z - I_h z)\right)_K. \end{aligned} \tag{5.89}$$

Note that for an element $K \in \mathscr{T}_h$, the term $\langle \mathscr{F}_2, (z - I_h z)\chi_K\rangle$ can be nonzero only if K contains a partial charge from Dye II in it. Similarly, the term $\langle \mathscr{F}_2, (z - I_h z)\, \psi_i\rangle$ in the definition of η_i^{PU} can be nonzero only if the support of ψ_i contains a partial charge from Dye II. To apply in practice the error indicators η_i^{PU} or η_K, we use a Galerkin finite element approximation $z_h^{(2)}$ of z from the finite element space $V_{0,\tau}^{k^*}$ with $k^* = 2$ and $\tau \equiv h$, i.e., $V_{0,\tau}^{k^*}$ is defined on the same triangulation $\mathscr{T}_h$ as the finite element space $V_{0,h}^1$ for the primal problem. By taking into account (5.9), the approximate electrostatic interaction E_{2-1}^P is given by

$$E_{2-1}^P = \langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, \varphi_{2,h}\rangle \tag{5.90}$$

and the error estimation for the electrostatic interaction E_{2-1} is

$$\begin{aligned} E_{2-1} - E_{2-1}^P &= \langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, \varphi_2 - \varphi_{2,h}\rangle \\ &= E(\varphi_{2,h}, z) = E(\varphi_{2,h}, z_h^{(2)}) + \mathcal{E}(\varphi_2, \varphi_{2,h}, z, z_h^{(2)}). \end{aligned} \tag{5.91}$$

In view of Section 5.3.1, assuming that the term $\mathcal{E}(\varphi_2, \varphi_{2,h}, z, z_h^{(2)})$ converges faster to zero than the term $E(\varphi_{2,h}, z_h^{(2)})$, we can write the approximate equality

$$E_{2-1} - E_{2-1}^P \approx E(\varphi_{2,h}, z_h^{(2)}). \tag{5.92}$$

From the first two lines in (5.87) it can be seen that we also have an approximate version of (5.75), given by the relation

$$E(\varphi_{2,h}, z_h^{(2)}) = \langle \mathscr{F}_2, z_h^{(2)}\rangle - \langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, \varphi_{2,h}\rangle. \tag{5.93}$$

In other words, the practical estimate for the error $E(\varphi_{2,h}, z_h^{(2)})$ is equal to the approximate electrostatic interaction $E_{1-2}^A = \langle \mathscr{F}_2, z_h^{(2)} \rangle$, computed with the more accurate Lagrange $P_{k^*} (= P_2)$ finite element, minus the approximate electrostatic interaction $E_{2-1}^P = \langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, \varphi_{2,h} \rangle$, computed with Lagrange $P_k (= P_1)$ finite element. In practical computations (see Section 5.4, Section 5.5.1, and Section 5.5.2), we use the corrected value for the electrostatic interaction, which is given by E_{1-2}^A:

$$E_{2-1} \approx E_{2-1}^P + E(\varphi_{2,h}, z_h^{(2)}) = E_{2-1}^P + \left(\langle \mathscr{F}_2, z_h^{(2)} \rangle - \langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, \varphi_{2,h} \rangle \right) = E_{1-2}^A. \tag{5.94}$$

Remark 5.11
Similarly to the first error estimation approach, for the practical application of this third error estimate where instead of the original goal functional $\frac{1}{4\pi}\mathscr{F}_1$ we use a regularized version of it, $\frac{1}{4\pi}\overline{\mathscr{F}}_1$, special quadrature rules are needed for the evaluation of the terms $\langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, \varphi_{2,h} \rangle$ and $\langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, z_h^{(2)} \rangle$ (see, e.g. [108, 145] for such quadrature rules).

5.3.4 Error estimation 4: no splitting in primal problem and irregular goal functional

Finally, we present a goal error estimation approach for the electrostatic interaction that involves the primal problem (5.2), where the right-hand side is $\mathscr{F}_2$, and an adjoint problem in which the right-hand side is formed by the irregular goal functional $\frac{1}{4\pi}\mathscr{F}_1 \notin H^{-1}(\Omega)$. With this approach, since we use directly the functional $\frac{1}{4\pi}\mathscr{F}_1$, we avoid the problem with the integration of discontinuous functions in the averaging procedure in Section 5.3.3. We recall the primal problem defining the electrostatic potential φ_2

$$\varphi_2 \in \bigcap_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega), \quad \int_\Omega \epsilon \nabla \varphi_2 \cdot \nabla v dx + \int_\Omega \overline{k}^2 \varphi_2 v dx = \langle \mathscr{F}_2, v \rangle, \; \forall v \in \bigcup_{q>d} W_0^{1,q}(\Omega). \tag{5.2}$$

The weak form of the adjoint problem is given by (5.44) and we recall it here

$$\begin{aligned} &\text{Find } z \in \bigcap_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla v \cdot \nabla z dx + \int_\Omega \overline{k}^2 v z dx = \langle \frac{1}{4\pi} \mathscr{F}_1, v \rangle, \; \forall v \in \bigcup_{q>d} W_0^{1,q}(\Omega). \end{aligned} \tag{5.44}$$

We also recall that the solution z to the adjoint problem (5.44) satisfies the relation $z = \frac{1}{4\pi}\varphi_1$ since φ_1 solves the same linear problem but with a right-hand side $\mathscr{F}_1$ and a homogeneous Dirichlet boundary condition (see (5.3)). In other words, the primal problem defines the potential created by the charges of Dye 2 and the adjoint problem defines the potential created by the charges of Dye I. As it is physically expected, the electrostatic interaction in both cases should be the same, i.e, $E_{2-1} = E_{1-2}$ (see Proposition 5.12). From Theorem 3.4 we know that (5.44) possesses a unique solution $z \in \bigcap_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega)$ which has the form

$z = \frac{1}{4\pi}\varphi_1 = \frac{1}{4\pi}(G_1 + u_1)$ where u_1 satisfies (5.10) with G_1 instead of G_2. By $\varphi_{2,h}$ we denote the Galerkin finite element approximation of φ_2 in the space $V^k_{0,h}$ with $k = 1$ defined by (5.72). In order to derive an error representation for the goal quantity $E_{2-1} = \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_2\rangle$ in terms of the adjoint solution z, we will need the following Proposition 5.12.

Proposition 5.12
The following equalities hold true

$$\int_\Omega \epsilon\nabla(\varphi_2 - \varphi_{2,h})\cdot\nabla z dx + \int_\Omega \overline{k}^2(\varphi_2 - \varphi_{2,h})z dx = \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_2 - \varphi_{2,h}\rangle, \tag{5.95}$$

$$\int_\Omega \epsilon\nabla\varphi_2\cdot\nabla z dx + \int_\Omega \overline{k}^2\varphi_2 z dx = \langle \mathscr{F}_2, z\rangle. \tag{5.96}$$

Moreover,

$$\langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_2\rangle = \langle \mathscr{F}_2, z\rangle = \langle \frac{1}{4\pi}\mathscr{F}_2, \varphi_1\rangle. \tag{5.97}$$

In order to avoid some technical details we will prove Proposition 5.12 only for the case $I_s = 0$, i.e., $\overline{k} = 0$ and note that the proof for the case $I_s \neq 0$ is similar.

Remark 5.13
In particular, (5.95) *and* (5.96) *mean that the integrals on the left hand-sides are well defined, as well as the expressions* $\langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_2 - \varphi_{2,h}\rangle$ *and* $\langle \mathscr{F}_2, z\rangle$ *are well defined. Note, that* (5.95) *means that we can test* (5.44) *with the function* $\varphi_2 - \varphi_{2,h}$ *which is not in* $\bigcup_{q>d} W^{1,q}_0(\Omega)$ *and* (5.96) *means that we can test* (5.2) *with* z *which is also not in* $\bigcup_{q>d} W^{1,q}_0(\Omega)$.

Proof of Proposition 5.12. We start with (5.95). The proof of this equality we split in two steps. The first step is to show that we can test the equation (5.44) also with functions $\overline{v} \in H^1_0(\Omega)\cap W^{1,q}(B_{m,1}(2r))$. Here, $\overline{v} \in W^{1,q}(B_{m,1}(2r))$ means that the restriction of $\overline{v}$ and its first weak derivatives $\frac{\partial\overline{v}}{\partial x_i}$ to $B_{m,1}(2r)$ are in $L^q(B_{m,1}(2r))$. In the second step, we use the functions $\psi_n = \overline{w} + (w_n - \overline{w})\chi_2 \in H^1_0(\Omega)\cap W^{1,q}(B_{m,1}(2r))$ from the proof of (5.73) and we show that by taking the limit $n\to\infty$, we arrive at (5.95). Before we continue, let us fix q such that $d < q \leq 6$ and let p be its Hölder conjugate, defined by $\frac{1}{p}+\frac{1}{q} = 1$. Obviously, $p < \frac{d}{d-1}$.

Step 1. Let $\overline{v} \in H^1_0(\Omega)\cap W^{1,q}(B_{m,1}(2r))$. We will approximate $\overline{v}$ with functions $\theta_n \in W^{1,q}_0(\Omega)$ such that $\theta_n \to \overline{v}$ in $W^{1,q}(B_{m,1}(r))$ and $\theta_n \to \overline{v}$ in $H^1(\Omega\setminus B_{m,1}(r))$. Since $\overline{v} \in H^1_0(\Omega)$, there exist functions $v_n \in C^\infty_0(\Omega)$ such that $v_n \to \overline{v}$ in $H^1(\Omega)$. We define the functions $\theta_n := v_n + (\overline{v} - v_n)\chi_1$, where χ_1 is defined by (5.47). We have

$$\theta_n(x) = \begin{cases} \overline{v}, & x \in B_{m,1}(r), \\ v_n + (\overline{v} - v_n)\chi_1, & x \in A_1 = B_{m,1}(2r)\setminus B_{m,1}(r), \\ v_n, & x \in \Omega\setminus B_{m,1}(2r). \end{cases} \tag{5.98}$$

Obviously, the function $\theta_n \in H_0^1(\Omega)$ for all $n \in \mathbb{N}$ as the sum of such functions. Indeed, by the product rule (see Theorem 2.9) applied to the functions $\overline{v} - v_n \in H_0^1(\Omega)$ and $\chi_1 \in C_0^\infty(\Omega)$, we obtain that $(\overline{v} - v_n)\chi_1 \in H^1(\Omega)$. Since $(\overline{v} - v_n)\chi_1$ has a compact support in Ω, by applying Theorem 2.10, we see that $(\overline{v} - v_n)\chi_1 \in H_0^1(\Omega)$. Moreover, the weak derivatives of θ_n are given by

$$\frac{\partial \theta_n}{\partial x_i} = \frac{\partial v_n}{\partial x_i} + \frac{\partial(\overline{v} - v_n)}{\partial x_i}\chi_1 + (\overline{v} - v_n)\frac{\partial \chi_1}{\partial x_i},\ i = 1, 2, 3 \tag{5.99}$$

and, therefore,

$$\frac{\partial \theta_n}{\partial x_i} = \begin{cases} \dfrac{\partial \overline{v}}{\partial x_i}, & x \in B_{m,1}(r), \\ \dfrac{\partial v_n}{\partial x_i} + \dfrac{\partial(\overline{v} - v_n)}{\partial x_i}\chi_1 + (\overline{v} - v_n)\dfrac{\partial \chi_1}{\partial x_i}, & x \in A_1 = B_{m,1}(2r) \setminus B_{m,1}(r), \\ \dfrac{\partial v_n}{\partial x_i}, & x \in \Omega \setminus B_{m,1}(2r). \end{cases} \tag{5.100}$$

From the expressions (5.98) and (5.100) for θ_n and $\frac{\partial \theta_n}{\partial x_i}$, $i = 1, 2, 3$, respectively, we see that θ_n, $\frac{\partial \theta_n}{\partial x_i} \in L^q(\Omega)$, $i = 1, 2, 3$, and therefore $\theta_n \in W^{1,q}(\Omega)$. Since the support of θ_n is a compact set in Ω, by Theorem 2.10, it follows that $\theta_n \in W_0^{1,q}(\Omega)$ for all $n \in \mathbb{N}$. Now, for all $n \in \mathbb{N}$ we have

$$\int\limits_{B_{m,1}(r)} \epsilon \nabla \theta_n \cdot \nabla z dx + \int\limits_{A_1} \epsilon \nabla \theta_n \cdot \nabla z dx + \int\limits_{\Omega \setminus B_{m,1}(2r)} \epsilon \nabla \theta_n \cdot \nabla z dx = \langle \frac{1}{4\pi}\mathscr{F}_1, \theta_n \rangle. \tag{5.101}$$

On the set A_1, for $i = 1, 2, 3$, since $\|v_n - \overline{v}\|_{H^1(\Omega)} \to 0$ and χ_1, $\frac{\partial \chi_1}{\partial x_i}$, $i = 1, 2, 3$ are essentially bounded, we have

$$\begin{aligned} &\left\| \frac{\partial \theta_n}{\partial x_i} - \frac{\partial \overline{v}}{\partial x_i} \right\|_{L^2(A_1)} \\ &\leq \left\| \frac{\partial v_n}{\partial x_i} - \frac{\partial \overline{v}}{\partial x_i} \right\|_{L^2(A_1)} + \left\| \frac{\partial (\overline{v} - v_n)}{\partial x_i}\chi_1 \right\|_{L^2(A_1)} + \left\| (\overline{v} - v_n)\frac{\partial \chi_1}{\partial x_i} \right\|_{L^2(A_1)} \to 0 \end{aligned} \tag{5.102}$$

and, therefore, $\|\nabla(\theta_n - \overline{v})\|_{L^2(A_1)} \xrightarrow[n\to\infty]{} 0$. Obviously, we also have that

$$\|\nabla(\theta_n - \overline{v})\|_{L^2(\Omega \setminus B_{m,1}(2r))} \xrightarrow[n\to\infty]{} 0.$$

By applying Hölder's inequality and using the fact that $\nabla z = \frac{1}{4\pi}\left(\nabla G_1 + \nabla u_1\right) \in L^2\left(\Omega \setminus B_{m,1}(r)\right)$ we obtain

$$\left| \int\limits_{A_1} \epsilon \nabla z \cdot \nabla(\theta_n - \overline{v}) dx \right| \leq \epsilon_{\max} \|\nabla z\|_{L^2(A_1)} \|\nabla(\theta_n - \overline{v})\|_{L^2(A_1)} \to 0,$$

$$\left| \int\limits_{\Omega \setminus B_{m,1}(2r)} \epsilon \nabla z \cdot \nabla(\theta_n - \overline{v}) dx \right| \leq \epsilon_{\max} \|\nabla z\|_{L^2(\Omega \setminus B_{m,1}(2r))} \|\nabla(\theta_n - \overline{v})\|_{L^2(\Omega \setminus B_{m,1}(2r))} \to 0.$$

We also have that for all $n \in \mathbb{N}$

$$\int\limits_{B_{m,1}(r)} \epsilon \nabla z \cdot \nabla \theta_n dx = \int\limits_{B_{m,1}(r)} \epsilon \nabla z \cdot \nabla \overline{v} dx,$$

$$\langle \frac{1}{4\pi} \mathscr{F}_1, \theta_n \rangle = \sum_{i=1}^{N_1} q_{i,1} \overline{v}(x_{i,1}) =: \langle \frac{1}{4\pi} \mathscr{F}_1, \overline{v} \rangle,$$

where the integral $\int\limits_{B_{m,1}(r)} \epsilon \nabla z \cdot \nabla \overline{v} dx$ is well defined because $z \in \bigcap\limits_{s<\frac{d}{d-1}} W_0^{1,s}(\Omega)$ and $\overline{v} \in H_0^1(\Omega) \cap W^{1,q}\left(B_{m,1}(2r)\right)$, and $\langle \frac{1}{4\pi} \mathscr{F}_1, \overline{v} \rangle$ is well defined because by the Sobolev embedding theorem $\overline{v} \in W^{1,q}\left(B_{m,1}(2r)\right) \hookrightarrow C^{0,\lambda}\left(\overline{B}_{m,1}(2r)\right)$, $0 < \lambda \leq 1 - d/q$ (see Theorem 2.11). Finally, by letting $n \to \infty$ in (5.101) we obtain that (5.44) holds true when tested with $\overline{v} \in H_0^1(\Omega) \cap W^{1,q}\left(B_{m,1}(2r)\right)$.

Step 2. Now, we take the functions $\psi_n = \overline{w} + (w_n - \overline{w})\chi_2 \in H_0^1(\Omega)$ from the proof of (5.73). By observing the facts that G_2 is smooth in $B_{m,1}(2r)$ and that by the Sobolev embedding theorem $u_2 \in H^2_{\text{loc}}(\Omega_m)$ implies $u_2 \in W^{1,q}\left(B_{m,1}(2r)\right)$, we see that $\varphi_2 = G_2 + u_2 \in W^{1,q}\left(B_{m,1}(2r)\right)$. Now, since $\overline{w} = \varphi_2 - \varphi_{2,h} \in W^{1,q}\left(B_{m,1}(2r)\right)$, we see that $\psi_n = \overline{w} + (w_n - \overline{w})\chi_2 \in H_0^1(\Omega) \cap W^{1,q}\left(B_{m,1}(2r)\right)$ for all $n \in \mathbb{N}$. By the first step, we have that for all $n \in \mathbb{N}$ it holds

$$\begin{aligned} &\int\limits_{B_{m,1}(r)} \epsilon \nabla z \cdot \nabla \psi_n dx + \int\limits_{A_1} \epsilon \nabla z \cdot \nabla \psi_n dx + \int\limits_{\Omega \setminus (B_{m,1}(2r) \cup B_{m,2}(2r))} \epsilon \nabla z \cdot \nabla \psi_n dx \\ &+ \int\limits_{A_2} \epsilon \nabla z \cdot \nabla \psi_n dx + \int\limits_{B_{m,2}(r)} \epsilon \nabla z \cdot \nabla \psi_n dx = \langle \frac{1}{4\pi} \mathscr{F}_1, \psi_n \rangle \end{aligned} \tag{5.103}$$

On the domains of integration for the first three integrals in (5.103), $\nabla \psi_n = \nabla \overline{w}$. From the proof of (5.73) we have that

$$\|\nabla(\psi_n - \overline{w})\|_{L^p(B_{m,2}(r))} \xrightarrow[n\to\infty]{} 0 \quad \text{and} \quad \|\nabla(\psi_n - \overline{w})\|_{L^p(A_2)} \xrightarrow[n\to\infty]{} 0.$$

By applying Hölder's inequality and using the fact that $\nabla z = \frac{1}{4\pi}\left(\nabla G_1 + \nabla u_1\right) \in L^q\left(B_{m,2}(2r)\right)$ (this is due to the fact that $u_1 \in H^2_{\text{loc}}(\Omega_m)$ and the Sobolev embedding theorem) we obtain

$$\left| \int\limits_{B_{m,2}(r)} \epsilon \nabla z \cdot \nabla(\psi_n - \overline{w}) dx \right| \leq \epsilon_{\max} \|\nabla z\|_{L^q(B_{m,2}(r))} \|\nabla(\psi_n - \overline{w})\|_{L^p(B_{m,2}(r))} \to 0,$$

$$\left| \int\limits_{A_2} \epsilon \nabla z \cdot \nabla(\psi_n - \overline{w}) dx \right| \leq \epsilon_{\max} \|\nabla z\|_{L^q(A_2)} \|\nabla(\psi_n - \overline{w})\|_{L^p(A_2)} \to 0.$$

$$\langle \mathscr{L}, \psi_n \rangle = \sum_{i=1}^{N_1} q_{i,1} \overline{w}(x_{i,1}) =: \langle \frac{1}{4\pi} \mathscr{F}_1, \overline{w} \rangle, \ \forall n \in \mathbb{N}.$$

Finally, letting $n \to \infty$ in (5.103), we obtain (5.95).
The proofs of (5.96) and (5.97) are similar to the proof of (5.95) and we skip them. □

Now, by using first (5.95) and then (5.96) we can derive the error equality for $\langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_2 - \varphi_{2,h}\rangle$ in terms of the adjoint solution z

$$\begin{aligned}\langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_2 - \varphi_{2,h}\rangle &= \int_\Omega \epsilon\nabla\left(\varphi_2 - \varphi_{2,h}\right)\cdot\nabla z dx + \int_\Omega \overline{k}^2\left(\varphi_2 - \varphi_{2,h}\right) z dx \\ &= \langle \mathscr{F}_2, z\rangle - \int_\Omega \epsilon\nabla\varphi_{2,h}\cdot\nabla z dx - \int_\Omega \overline{k}^2\varphi_{2,h} z dx =: E(\varphi_{2,h}, z).\end{aligned} \tag{5.104}$$

Note that in order to evaluate the error estimator $E(\varphi_{2,h}, z)$, one has to know the exact solution z of the adjoint problem or an approximation of it. As we have already discussed, in practice one finds a good approximation of z in a richer space than the one in which $\varphi_{2,h}$ is found. In particular, we use the Galerkin finite element approximation $z_h^{(2)}$ of z from the space V_h^2 defined on the same mesh $\mathscr{T}_h$ on which V_h^1 is defined. In this case, by using Galerkin orthogonality, similarly to (5.17) (see [13, 159]), we can write

$$\begin{aligned}E(\varphi_{2,h}, z_h^{(2)}) =& \langle \mathscr{F}_2, z_h^{(2)} - I_h z_h^{(2)}\rangle - \int_\Omega \epsilon\nabla\varphi_{2,h}\cdot\nabla(z_h^{(2)} - I_h z_h^{(2)})dx \\ &- \int_\Omega \overline{k}^2\varphi_{2,h}(z_h^{(2)} - I_h z_h^{(2)})dx \\ =& \sum_{K\in\mathscr{T}_h}\eta_K(\varphi_{2,h}, z_h^{(2)}) =: \eta(\varphi_{2,h}, z_h^{(2)}) \\ =& \sum_{i=1}^{N_V}\eta_i^{PU}(\varphi_{2,h}, z_h^{(2)}) =: \eta^{PU}(\varphi_{2,h}, z_h^{(2)}),\end{aligned} \tag{5.105}$$

where N_V is the number of nodes in the mesh $\mathscr{T}_h$, $\{\psi_i\}_{i=1}^{N_V}$ are the nodal P_1 basis functions, I_h is the nodal interpolation operator in V_h^1, and

$$\begin{aligned}\eta_K(\varphi_{2,h}, z_h^{(2)}) =& \langle \mathscr{F}_2, (z_h^{(2)} - I_h z_h^{(2)})\chi_K\rangle - \left(\epsilon\nabla\varphi_{2,h}, \nabla(z_h^{(2)} - I_h z_h^{(2)})\right)_K \\ &- \left(\overline{k}^2\varphi_{2,h}, (z_h^{(2)} - I_h z_h^{(2)})\right)_K,\end{aligned} \tag{5.106}$$

$$\begin{aligned}\eta_i^{PU}(\varphi_{2,h}, z_h^{(2)}) =& \langle \mathscr{F}_2, \left(z_h^{(2)} - I_h z_h^{(2)}\right)\psi_i\rangle - \left(\epsilon\nabla\varphi_{2,h}, \nabla\left((z_h^{(2)} - I_h z_h^{(2)})\psi_i\right)\right) \\ &- \left(\overline{k}^2\varphi_{2,h}, \left(z_h^{(2)} - I_h z_h^{(2)}\right)\psi_i\right).\end{aligned} \tag{5.107}$$

Recall that for an element $K \in \mathscr{T}_h$, the term $\langle \mathscr{F}_2, (z_h^{(2)} - I_h z_h^{(2)})\chi_K\rangle$ can be nonzero only if K contains a partial charge from Dye II in it. Similarly, the term $\langle \mathscr{F}_2, \left(z_h^{(2)} - I_h z_h^{(2)}\right)\psi_i\rangle$ in the definition of η_i^{PU} can be nonzero only if the support of ψ_i contains a partial charge from Dye II. The approximate electrostatic interaction E_{2-1}^P is given by

$$E_{2-1}^P = \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_{2,h}\rangle \tag{5.108}$$

and the error estimation for the electrostatic interaction E_{2-1} is

$$\begin{aligned} E_{2-1} - E_{2-1}^P &= \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_2 - \varphi_{2,h}\rangle \\ &= E(\varphi_{2,h}, z) = E(\varphi_{2,h}, z_h^{(2)}) + \mathcal{E}(\varphi_2, \varphi_{2,h}, z, z_h^{(2)}), \end{aligned} \tag{5.109}$$

where

$$\begin{aligned} \mathcal{E}(\varphi_2, \varphi_{2,h}, z, z_h^{(2)}) &= \langle \mathscr{F}_2, z - z_h^{(2)}\rangle - \int_\Omega \epsilon \nabla\varphi_{2,h}\cdot\nabla(z - z_h^{(2)})dx - \int_\Omega \overline{k}^2\varphi_{2,h}(z - z_h^{(2)})dx \\ &= \int_\Omega \epsilon\nabla(\varphi_2 - \varphi_{2,h})\cdot\nabla(z - z_h^{(2)})dx + \int_\Omega \overline{k}^2(\varphi_2 - \varphi_{2,h})(z - z_h^{(2)})dx. \end{aligned} \tag{5.110}$$

We expect that the term $\mathcal{E}(\varphi_2, \varphi_{2,h}, z, z_h^{(2)})$ again converges faster than the term $E(\varphi_{2,h}, z_h^{(2)})$, as we can assume in Section 5.3.1, and in practice we use the approximate equality

$$E_{2-1} - E_{2-1}^P \approx E(\varphi_{2,h}, z_h^{(2)}) = \langle \mathscr{F}_2, z_h^{(2)}\rangle - \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_{2,h}\rangle, \tag{5.111}$$

which can also be written in the form

$$E_{2-1} \approx E_{2-1}^P + \left(\langle \mathscr{F}_2, z_h^{(2)}\rangle - \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_{2,h}\rangle\right) = \langle \mathscr{F}_2, z_h^{(2)}\rangle. \tag{5.112}$$

The quantity $E_{2-1}^P + E(\varphi_{2,h}, z_h^{(2)}) = E_{1-2}^A$ is the "corrected" value for the electrostatic interaction and it is used in the biophysical applications as a better approximation of the interaction E_{2-1} (see Section 5.4, Section 5.5.1, and Section 5.5.2).

Remark 5.14
We have proved Proposition 5.6, Proposition 5.10, and Proposition 5.12 in the case when all charges $\{q_{i,1}\}_{i=1}^{N_1}$ and $\{q_{i,2}\}_{i=1}^{N_2}$ are placed inside the molecular region Ω_m. However, in the proofs we only use the assumptions that ϵ_m is constant, that $\epsilon_s \in C^{0,1}(\overline{\Omega}_s)$, and that the charges are at a positive distance from the interface Γ. These assumptions guarantee the H^2_{loc} regularity of the reaction field part of the potential in the regions Ω_m and Ω_s (see Theorem 3.4). Therefore, if we further assume ϵ_s to be constant (ϵ_s will play the role of the constant ϵ_m in the definition of the Green's function G), the same proofs with slight technical modifications would work in the case when $I_s = 0$ and either the charges $\{q_{i,1}\}_{i=1}^{N_1}$ or $\{q_{i,2}\}_{i=1}^{N_2}$ or both $\{q_{i,1}\}_{i=1}^{N_1}$ and $\{q_{i,2}\}_{i=1}^{N_2}$ are in the solvent region Ω_s.

*When $I_s > 0$ ($\overline{k}^2_{ions} > 0$), let us assume without loss of generality that $\{q_{i,1}\}_{i=1}^{N_1} \subset \Omega_s$. The regularity of the primal problem, defining the potential u_2, in the case of **error estimate 1** and **error estimate 2**, or φ_2, in the case of **error estimate 3** and **error estimate 4**, is analyzed as usual through the 2-term splitting (Theorem 3.4) where*

$$G_2 := \sum_{i=1}^{N_2} \frac{q_{i,2}}{\epsilon_m\,|x - x_{i,2}|}.$$

For the analysis of the adjoint problem in the case of ***error estimate 2*** *and* ***error estimate 4****, assume additionally that ϵ_s is constant (the adjoint problem in the case of* ***error estimate 1*** *and* ***error estimate 3*** *has a regular L^∞ right-hand side and hence there are no difficulties). Observe that $\frac{1}{|x-x_{i,1}|^t} \in L^1(\Omega)$ if and only if $t < d$. Thus, if $d = 3$, then $\frac{1}{\epsilon_s|x-x_{i,1}|} \in L^2(\Omega_s)$. Therefore the function G_1 defined by*

$$G_1 := \sum_{i=1}^{N_1} \frac{q_{i,1}}{\epsilon_s\,|x - x_{i,1}|}$$

is also in $L^2(\Omega_s)$. It is clear that $G_1 \in \bigcap_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega)$, and similarly to (3.15)*, G_1 satisfies*

$$\int_\Omega \epsilon_s \nabla G_1 \cdot \nabla v dx = \langle \mathscr{F}_1, v\rangle,\ \forall v \in \bigcup_{q>d} W_0^{1,q}(\Omega), \tag{5.113}$$

where $\mathscr{F}_1 = 4\pi \sum_{i=1}^{N_1} q_{i,1}\delta_{x_{i,1}}$.
The adjoint problem defining the potential $z = \frac{1}{4\pi}\varphi_1$ is given by

$$\begin{aligned} &\text{Find } z \in \bigcap_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla v \cdot \nabla z dx + \int_\Omega \overline{k}^2 v z dx = \langle \frac{1}{4\pi}\mathscr{F}_1, v\rangle,\ \forall v \in \bigcup_{q>d} W_0^{1,q}(\Omega). \end{aligned} \tag{5.44}$$

We can perform the splitting of φ_1 into G_1 and u_1, where φ_1 solves (5.44) *but with a right-hand side $\mathscr{F}_1$, and $u_1 \in H^1_{-G_1}(\Omega) \subset \bigcap_{p<\frac{d}{d-1}} W^{1,p}_{-G_1}(\Omega)$ satisfies*

$$\int_\Omega \epsilon \nabla u_1 \cdot \nabla v dx + \int_\Omega \overline{k}^2 u_1 v dx = \int_{\Omega_m} (\epsilon_s - \epsilon_m)\nabla G_1 \cdot \nabla v dx - \int_\Omega \overline{k}^2 G_1 v dx,\ \forall v \in H^1_0(\Omega). \tag{5.114}$$

Indeed, the rigth-hand side of (5.114) *defines a functional in H^{-1} (in the first integral on the right-hand side, G_1 is smooth, whereas in the second intergal, G_1 has singularities in Ω_s) and hence by the Lax-Milgram Theorem* (5.114) *has a unique solution $u_1 \in H^1_{-G_1}(\Omega)$. The uniqueness of φ_1 under the assumption that $\Gamma \in C^1$ is done in a similar way to the proof in Theorem 3.4. Finally, by testing with functions $v \in C_0^\infty(\Omega_m), v \in C_0^\infty(\Omega_s)$ and using the fact that $G_1 \in L^2(\Omega)$, we obtain that $u_1 \in H^1_{-G_1}(\Omega) \cap H^2_{loc}(\Omega_m) \cap H^2_{loc}(\Omega_s)$ (see p. 309 in [70] for interior H^2 regularity of linear elliptic problems).*

5.3.5 Summary of all four error estimates

Before we move on to the applications, we summarize the four error estimates that we derived in Sections 5.3.1, 5.3.2, 5.3.3, 5.3.4 and briefly discuss their features and applicability.

For the relevant notation below, we refer to the respective sections. We recall that since we solve the linearized Poisson-Boltzmann equation and we impose homogeneous Dirichlet boundary conditions on the potential, the computed electrostatic interaction does not depend on whether the charges are in Dye II or in Dye I, i.e., $E_{2-1} = E_{1-2}$. Thus, without loss of generality, we assume that there are only charges in Dye II and we are interested in computing the electrostatic potential at the positions of the charges in Dye I. However, in the case of the nonlinear PBE, the symmetry of the electrostatic interaction is in general lost and $E_{2-1} \neq E_{1-2}$ even with homogeneous Dirichlet boundary conditions.

Two of the derived error estimates, namely **error estimate 1** and **error estimate 3**, exploit averaging over balls $B(x_{i,1}, \overline{r}_{i,1})$ in the goal functional. If the finite element mesh is not aligned with the balls $B(x_{i,1}, \overline{r}_{i,1})$, the rigorous application of these two estimates necessitates the use of special quadrature rules that can perform the integration of discontinuous functions with high enough accuracy (see [108, 145]). As an alternative to these two error estimates, we present **error estimate 2** and **error estimate 4**, where the goal functional is a linear combination of pointwise evaluations and thus the integration problem is eliminated.

Error estimate 1 and error estimate 2

These two approaches are applicable when only the reaction field potential is needed or when the dielectric screening in the solvent region Ω_s is weak. The latter is typically the case when the ratio ϵ_m/ϵ_s is close to one and the ionic strength I_s is zero. The primal problem in **error estimate 1** and **error estimate 2** defines the reaction field potential $u_2 \in H^1_{-G_2}(\Omega)$ in the two term splitting $\varphi_2 = G_2 + u_2$ and according to Theorem 3.4 is given by

$$\begin{aligned} &\int_\Omega \epsilon \nabla u_2 \cdot \nabla v dx + \int_\Omega \overline{k}^2 u_2 v dx \\ &= \int_{\Omega_s} (\epsilon_m - \epsilon_s) \nabla G_2 \cdot \nabla v dx - \int_\Omega \overline{k}^2 G_2 v dx = \langle \mathcal{T}_2, v \rangle \text{ for all } v \in H^1_0(\Omega). \end{aligned} \tag{5.10}$$

Equation (5.10) is the weak formulation of the linear interface problem

$$-\nabla \cdot (\epsilon \nabla u_2) + \overline{k}^2 u_2 = -\overline{k}^2 G_2 \quad \text{in } \Omega_m \cup \Omega_s, \tag{5.115a}$$

$$[u_2]_\Gamma = 0, \tag{5.115b}$$

$$[\epsilon \nabla u_2 \cdot \boldsymbol{n}_\Gamma]_\Gamma = -[\epsilon \nabla G_2 \cdot \boldsymbol{n}_\Gamma]_\Gamma, \tag{5.115c}$$

$$u_2 = -G_2 \quad \text{on } \partial\Omega. \tag{5.115d}$$

The right-hand side of the adjoint problem in **error estimate 1** is formed by the regular goal functional $\frac{1}{4\pi}\overline{\mathscr{F}}_1 \in H^{-1}(\Omega) \cap L^\infty(\Omega)$ defined by

$$\langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, v \rangle = \int_\Omega l v dx, \text{ with } l(x) = \sum_{i=1}^{N_1} \frac{\chi_{B(x_{i,1}, \overline{r}_{i,1})}(x)}{|B(x_{i,1}, \overline{r}_{i,1})|} \tag{5.116}$$

and the weak form of the adjoint problem is

$$\begin{aligned} &\text{Find } z \in H_0^1(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla v \cdot \nabla z dx + \int_\Omega \overline{k}^2 v z dx = \langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, v \rangle, \ \forall v \in H_0^1(\Omega). \end{aligned} \tag{5.29}$$

Since u_2 is harmonic in Ω_m, for the reaction field part E_{u_2} of the interaction it holds

$$E_{u_2} = \sum_{i=1}^{N_1} q_{i,1} u_2(x_{i,1}) = \langle \frac{1}{4\pi} \mathscr{F}_1, u_2 \rangle = \langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, u_2 \rangle.$$

On the other hand, the right-hand side of the adjoint problem in **error estimate 2** is formed by the goal functional $\frac{1}{4\pi}\mathscr{F}_1 \notin H^{-1}(\Omega)$ and the weak formulation in this case is

$$\begin{aligned} &\text{Find } z \in \bigcap_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla v \cdot \nabla z dx + \int_\Omega \overline{k}^2 v z dx = \langle \frac{1}{4\pi} \mathscr{F}_1, v \rangle, \ \forall v \in \bigcup_{q>d} W_0^{1,q}(\Omega). \end{aligned} \tag{5.44}$$

In all four error estimates, the corresponding approximate solution $u_{2,h}$ and $\varphi_{2,h}$ of the primal problem is found by solving a Galerkin formulation in the space V_h^1 of continuous piecewise linear functions over a mesh $\mathscr{T}_h$ whereas the approximate solution $z_h^{(2)}$ of the adjoint problem is found by a Galerkin formulation in the space V_h^2 of continuous piecewise quadratic functions over the same mesh $\mathscr{T}_h$.

In **error estimate 1**, the approximate electrostatic interaction E_{2-1}^P is given by

$$E_{2-1}^P = E_{G_2} + \langle \frac{1}{4\pi} \overline{\mathscr{F}}_1, u_{2,h} \rangle, \tag{5.35}$$

whereas in **error estimate 2**, the approximate electrostatic interaction E_{2-1}^P is given by

$$E_{2-1}^P = E_{G_2} + \langle \frac{1}{4\pi} \mathscr{F}_1, u_{2,h} \rangle. \tag{5.62}$$

The error estimator for the electrostatic interaction E_{2-1} is $E(u_{2,h}, z_h^{(2)})$, i.e.,

$$E_{2-1} - E_{2-1}^P = E(u_{2,h}, z) \approx E(u_{2,h}, z_h^{(2)}),$$

where the quantity $E(u_{2,h}, z_h^{(2)})$ is given in both cases by

$$E(u_{2,h}, z_h^{(2)}) = \langle \mathcal{T}_2, z_h^{(2)} \rangle - \int_\Omega \epsilon \nabla u_{2,h} \cdot \nabla z_h^{(2)} dx - \int_\Omega \overline{k}^2 u_{2,h} z_h^{(2)} dx.$$

The corrected value for E_{2-1} is $E_{2-1}^P + E(u_{2,h}, z_h^{(2)})$. However, in the case of **error estimate 2** it is better to use the quantity $E_{1-2}^A = \langle \mathscr{F}_2, z_h^{(2)} \rangle$ since it is easier to compute and numerically

more stable (note that $E_{1-2} = \langle \mathscr{F}_2, z \rangle$). As error indicators, we use either the elementwise indicators η_K or the nodewise indicators η_i^{PU}. In both error estimation approaches we have

$$\begin{aligned} E(u_{2,h}, z_h^{(2)}) =& \langle \mathcal{T}_2, z_h^{(2)} - I_h z_h^{(2)} \rangle - \int_\Omega \epsilon \nabla u_{2,h} \cdot \nabla (z_h^{(2)} - I_h z_h^{(2)}) dx \\ & - \int_\Omega \overline{k}^2 u_{2,h} (z_h^{(2)} - I_h z_h^{(2)}) dx \\ =& \sum_{K \in \mathscr{T}_h} \eta_K(u_{2,h}, z_h^{(2)}) = \sum_{i=1}^{N_V} \eta_i^{PU}(u_{2,h}, z_h^{(2)}), \end{aligned} \tag{5.117}$$

where N_V is the number of nodes in the mesh $\mathscr{T}_h$, $\{\psi_i\}_{i=1}^{N_V}$ are the nodal P_1 basis functions, I_h is the nodal interpolation operator in V_h^1, and

$$\begin{aligned} \eta_K(u_{2,h}, z_h^{(2)}) =& \left((\epsilon_m - \epsilon_s) \nabla G_2 \chi_{\Omega_s} - \epsilon \nabla u_{2,h}, \nabla (z_h^{(2)} - I_h z_h^{(2)}) \right)_K \\ & - \left(\overline{k}^2 (G_2 + u_{2,h}), (z_h^{(2)} - I_h z_h^{(2)}) \right)_K, \end{aligned} \tag{5.118}$$

$$\begin{aligned} \eta_i^{PU}(u_{2,h}, z_h^{(2)}) =& \left((\epsilon_m - \epsilon_s) \nabla G_2 \chi_{\Omega_s} - \epsilon \nabla u_{2,h}, \nabla \left(\left(z_h^{(2)} - I_h z_h^{(2)} \right) \psi_i \right) \right) \\ & - \left(\overline{k}^2 (G_2 + u_{2,h}), (z_h^{(2)} - I_h z_h^{(2)}) \psi_i \right). \end{aligned} \tag{5.119}$$

Error estimate 3 and error estimate 4

The next two approaches, **error estimate 3** and **error estimate 4**, are appropriate to use in the presence of strong dielectric screening when one needs the the full potential and not only the reaction field part of it (however, we show in Section 5.4.2 and Section 5.4.3 that even in the presence of strong dielectric screening, **Solver 1** and **Solver 2** perform equally well when one has to sompute electrostatic interactions). This time we do not perform the 2-term splitting and the primal problem in these two cases defines the full potential φ_2

$$\varphi_2 \in \bigcap_{p < \frac{d}{d-1}} W_0^{1,p}(\Omega), \quad \int_\Omega \epsilon \nabla \varphi_2 \cdot \nabla v dx + \int_\Omega \overline{k}^2 \varphi_2 v dx = \langle \mathscr{F}_2, v \rangle, \, \forall v \in \bigcup_{q > d} W_0^{1,q}(\Omega). \tag{5.2}$$

In **error estimate 3**, the goal functional $\frac{1}{4\pi}\overline{\mathscr{F}}_1$ and the corresponding adjoint problem are the same as in **error estimate 1**. This time, the goal quantity is the full electrostatic interaction $E_{2-1} = \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, \varphi_2 \rangle$ and the approximate interaction is given by

$$E_{2-1}^P = \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, \varphi_{2,h} \rangle.$$

As an alternative to **error estimate 3**, **error estimate 4** does not exploit averaging over the balls $B(x_{i,1}, \overline{r}_{i,1})$ and thus avoids the cumbersome numerical integration of discontinuous functions. In this case, the goal functional is $\frac{1}{4\pi}\mathscr{F}_1$ and the corresponding adjoint problem

is the same as in **error estimate 2**. The goal quantity is $E_{2-1} = \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_2 \rangle$ and the approximate interaction is

$$E_{2-1}^P = \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_{2,h} \rangle.$$

The error estimator for the electrostatic interaction E_{2-1} is $E(\varphi_{2,h}, z_h^{(2)})$, i.e.,

$$E_{2-1} - E_{2-1}^P = E(\varphi_{2,h}, z) \approx E(\varphi_{2,h}, z_h^{(2)}),$$

where the quantity $E(\varphi_{2,h}, z_h^{(2)})$ is given in both cases by

$$E(\varphi_{2,h}, z_h^{(2)}) = \langle \mathscr{F}_2, z_h^{(2)} \rangle - \int_\Omega \epsilon \nabla \varphi_{2,h} \cdot \nabla z_h^{(2)} dx - \int_\Omega \overline{k}^2 \varphi_{2,h} z_h^{(2)} dx.$$

In each case, we can further rewrite the quantity $E(\varphi_{2,h}, z_h^{(2)})$ as

$$E(\varphi_{2,h}, z_h^{(2)}) = \begin{cases} \langle \mathscr{F}_2, z_h^{(2)} \rangle - \langle \frac{1}{4\pi}\overline{\mathscr{F}}_1, \varphi_{2,h} \rangle, & \text{for } \textbf{error estimate 3}, \\ \langle \mathscr{F}_2, z_h^{(2)} \rangle - \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_{2,h} \rangle, & \text{for } \textbf{error estimate 4}. \end{cases} \tag{5.120}$$

The corrected value for E_{2-1}, which we use in practice, is $E_{2-1}^P + E(\varphi_{2,h}, z_h^{(2)}) = \langle \mathscr{F}_2, z_h^{(2)} \rangle = E_{1-2}^A$ for both error estimates. Recall that E_{1-2}^A is computed with the higher accurate adjoint solution $z_h^{(2)} \in V_h^2$. As error indicators, we use either the elementwise indicators η_K or the nodewise indicators η_i^{PU}. In both error estimation approaches we have

$$\begin{aligned} E(\varphi_{2,h}, z_h^{(2)}) =& \langle \mathscr{F}_2, z_h^{(2)} - I_h z_h^{(2)} \rangle - \int_\Omega \epsilon \nabla \varphi_{2,h} \cdot \nabla (z_h^{(2)} - I_h z_h^{(2)}) dx \\ & - \int_\Omega \overline{k}^2 \varphi_{2,h} (z_h^{(2)} - I_h z_h^{(2)}) dx \\ =& \sum_{K \in \mathscr{T}_h} \eta_K(\varphi_{2,h}, z_h^{(2)}) = \sum_{i=1}^{N_V} \eta_i^{PU}(\varphi_{2,h}, z_h^{(2)}), \end{aligned} \tag{5.121}$$

where

$$\begin{aligned} \eta_K(\varphi_{2,h}, z_h^{(2)}) =& \langle \mathscr{F}_2, \left(z_h^{(2)} - I_h z_h^{(2)} \right) \chi_K \rangle - \left(\epsilon \nabla \varphi_{2,h}, \nabla \left(z_h^{(2)} - I_h z_h^{(2)} \right) \right)_K \\ & - \left(\overline{k}^2 \varphi_{2,h}, \left(z_h^{(2)} - I_h z_h^{(2)} \right) \right)_K, \end{aligned} \tag{5.122}$$

$$\begin{aligned} \eta_i^{PU}(\varphi_{2,h}, z_h^{(2)}) =& \langle \mathscr{F}_2, \left(z_h^{(2)} - I_h z_h^{(2)} \right) \psi_i \rangle - \left(\epsilon \nabla \varphi_{2,h}, \nabla \left((z_h^{(2)} - I_h z_h^{(2)}) \psi_i \right) \right) \\ & - \left(\overline{k}^2 \varphi_{2,h}, \left(z_h^{(2)} - I_h z_h^{(2)} \right) \psi_i \right). \end{aligned} \tag{5.123}$$

Remark 5.15

We note that in all four error estimates it is not necessary to use the spaces V_h^1 and V_h^2 for the primal and adjoint problem, respectively. Instead, one can use conforming spaces V_h^k and

$V_h^{k^*}$ of polynomial degrees $k < k^*$. It is even possible to take an approximation of z in the same space V_h^k as for the primal problem and then use some reconstruction to reinterpret it as a function in the higher accurate space $V_h^{k^*}$. For such reconstructions we refer to [159] and the references therein. Finally, the underlying meshes for the two discrete spaces V_h^k and $V_h^{k^*}$ can be also different.

Adaptive algorithm

In general, adaptive algorithms can be represented with the following diagram (Figure 5.3): Instead of the mesh parameter h, we will use a subindex $l = 1, 2, \ldots$. to denote the refinement

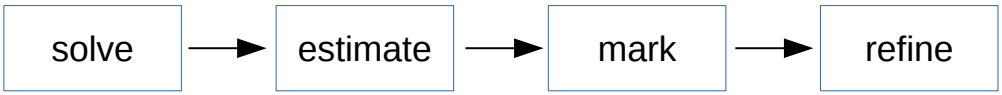


Figure 5.3: Schematic representation of an adaptive refinement algorithm.

level. The number of elements in the mesh $\mathscr{T}_l$ at refinement level l will be denoted by N_E^l and the number of vertices by N_V^l. Based on our experience, we have not noticed significant difference between the two error indicators η_K and η_i^{PU} and thus, we will present results obtained only with the indicators η_K. The adaptive finite element solver based on the four error estimates above can be summarized as follows.

Solve

In this step, a Galerkin finite element approximation $u_{2,l}$ or $\varphi_{2,l}$ from the space V_l^1, defined on the current mesh $\mathscr{T}_l$, is found by solving (5.41) or (5.72), respectively.

Estimate

Next, we find an approximation $z_l^{(2)} \in V_{0,h}^2$ for the adjoint problem by solving a Galerkin formulation of (5.29) for **error estimate 1** and **error estimate 3** or (5.44) for **error estimate 2** and **error estimate 4**. Once we have $z_l^{(2)}$, we compute the error indicators $\eta_K(u_{2,l}, z_l^{(2)})$ in the case of **error estimate 1** and **error estimate 2** or $\eta_K(\varphi_{2,l}, z_l^{(2)})$ in the case of **error estimate 3** and **error estimate 4**.

Mark

Based on the values of η_K, a set of elements $\mathscr{M}_l \subset \mathscr{T}_l$ is marked for refinement. This set can be chosen through different marking strategies.

(i) One such marking strategy takes the average of all indicators $\frac{\sum_{K\in\mathscr{T}_l} |\eta_K|}{N_E}$ and compares

it to each $|\eta_K|$. If $|\eta_K|$ is larger than this average, the element K is marked for refinement.

(ii) Another marking strategy ranks the values $|\eta_K|$ in descending order and for a parameter $\theta \in (0,1)$ it chooses the first θN_E of them.

(iii) A third, very popular method, is the so-called bulk criterion or greedy algorithm, which for a given bulk parameter $\theta \in (0,1)$ selects the smallest set of elements $\mathscr{M}_l \subset \mathscr{T}_l$ such that $\sum_{K\in\mathscr{M}_l} |\eta_K| \geq \theta\overline{\eta}$, where $\overline{\eta} = \sum_{K\in\mathscr{T}_l} |\eta_K|$.

Depending on the marking strategy, the sets of selected elements for refinement can be also different.

If the mesh generator performs local refinement using information different from that of elements being marked or not, one has to define an algorithm to deliver this information based on the element error indicators η_K. We will illustrate this in the following for the mesh generator mmg3d [62], which requires as an input for refinement a new local mesh size h_i^{new} at each vertex V_i, $i = 1, 2, \ldots, N_V$. Then mmg3d tries to generate a mesh in which the local mesh size is as close as possible to h_i^{new}. If h_i^{old} denotes the old (current) mesh size, defined as the arithmetic mean of the lengths of all edges connected to vertex V_i, then the new local mesh size h_i^{new} is computed from the indicators η_K and the old mesh size h_i^{old} by the formula

$$h_i^{\text{new}} = h_i^{\text{old}} \max\left\{ \frac{1}{\max\{\frac{\eta_{O_i}}{AM\{\eta_{O_i}\}}, 1\}}, r \right\}, \tag{5.124}$$

where $O_i := \bigcup\{K \in \mathscr{T}_l : V_i \in K\}$ (all elements K are compact sets), $\eta_{O_i} = \left| \sum_{K\subset O_i} \eta_K \right|$ are patchwise (nodewise) error indicators (very similar to the error indicators η_i^{PU} that were constructed with partition of unity), and $AM\{\eta_{O_i}\}$ is the arithmetic mean of all η_{O_i}, i.e.,

$$AM\{\eta_{O_i}\} := \frac{\sum_{i=1}^{N_V} \eta_{O_i}}{N_V}.$$

In (5.124), $r \in (0,1)$ is a parameter which guarantees that the minimum local mesh size h_i^{new} satisfies $h_i^{\text{new}} \geq r h_i^{\text{old}}$ herewith limiting the decrease of the mesh size. In our computations, we use $r = 0.35$.

Refine

Finally, a new mesh $\mathscr{T}_{l+1}$ is obtained through a refinement based on the marked elements $\mathscr{M}_l$ on the previous step.

Below, we give a pseudocode of the adaptive solver based on **error estimate 4** in which any marking strategy can be used. The pseudocodes for the adaptive solvers based on the other three error estimates are similar.

Algorithm 1 Adaptive solver based on **error estimate 4**

1: $l \leftarrow 0$, initialize $\mathscr{T}_0$, V_0^1, V_0^2
2: **do**
3: $\quad \varphi_{2,l} \leftarrow \text{Solve}^{\text{prim}}(\mathscr{T}_l, V_l^1)$
4: $\quad z_l^{(2)} \leftarrow \text{Solve}^{\text{adj}}(\mathscr{T}_l, V_l^2)$
5: $\quad \{\eta_K\}_{K\in\mathscr{T}_l} \leftarrow \text{Estimate}(\varphi_{2,l}, z_l^{(2)}, \mathscr{T}_l)$
6: $\quad \mathscr{M}_l \leftarrow \text{Mark}(\{\eta_K\}_{K\in\mathscr{T}_l})$
7: $\quad \mathscr{T}_{l+1} \leftarrow \text{Refine}(\mathscr{T}_l, \mathscr{M}_l)$
8: $\quad l \leftarrow l+1$
9: **while** convergence or maximum number of refinement levels reached

Above, the procedure $\varphi_{2,l} \leftarrow \text{Solve}^{\text{prim}}(\mathscr{T}_l, V_l^1)$ finds a Galerkin approximation of φ_2 based on (5.72), whereas $z_l^{(2)} \leftarrow \text{Solve}^{\text{adj}}(V_l^2)$ finds a Galerkin approximation of z based on (5.29). We prefer to use **error estimate 2** and **error estimate 4** since no special quadrature rule is needed. In all presented examples and applications, we have used the conjugate gradient (CG) method with a simple Jacoby preconditioner in the procedures $Solve^{prim}$ and $Solve^{adj}$. The error tolerance used in the stopping criteria of the CG method is set very low in order to ensure that the iterative error is negligible compared to the discretization error. This allows us to demonstrate the efficiency of the goal-oriented error estimates developed above, independently from the choice of a particular iterative solver.

5.4 Verification of the error estimates

In this section, we present 3 tests in which the exact solution is known and we compare the results obtained by applying the derived error estimates. We will abbreviate by **Solver 1**, **Solver 2**, **Solver 3**, or **Solver 4** the adaptive algorithm based on **error estimate 1**, **error estimate 2**, **error estimate 3**, or **error estimate 4**, respectively. We recall that the number of DOFs in the primal problem is equal to the number of mesh vertices in the mesh $\mathscr{T}_h$, whereas the number of DOFs in the adjoint problem is approximately 8 times larger in 3D. We will often compare our results to the results obtained with the software package MEAD (Macroscopic Electrostatics with Atomic Detail) *version 2.2.8a* and thus we give a brief overview of it. MEAD is a collection of several applications for the purpose of modeling electrostatics in molecules where the electrostatic potential is determined by approximately solving the LPBE with the finite difference method on uniform Cartesian grids. To improve the accuracy without massively increasing the computational costs, MEAD utilizes a so-called focusing scheme where a sequence of several computations on progressively finer grids are done. The first and coarsest grid has n_1 grid points in each coordinate direction and covers a

relatively big region Ω in $\mathbb{R}^3$ with a coarse grid spacing h_1. The boundary condition for this first grid is specified by the analytical solution of a simplified problem. For example, in the case of a molecule with a dielectric coefficient $\epsilon = \epsilon_m$ embedded in a solvent with $\epsilon = \epsilon_s \geq \epsilon_m$, the potential g on the boundary $\partial\Omega$ is given by a simple Coulomb potential with a dielectric coefficient equal to ϵ_s everywhere, i.e.,

$$g = \sum_{i=1}^{N_m} \frac{q_i}{\epsilon_s \left| x - x_i \right|}, \tag{5.125}$$

where q_i, $i = 1, 2, \ldots, N_m$ are the partial charges of the molecule. Next, a computation on a finer grid, with a smaller extent, is done, where the values for the potential on the boundary are taken from the solution on the previous grid. Depending on the size of the molecular system and the application, three or more focused grids can be used, where the grids are usually focused on the areas of interest. Each focused grid level l in MEAD is specified by its centering, number of grid points per direction n_l, and grid spacing h_l $\mathring{A}$. The centering can be specified by one of the following two key words ON_ORIGIN and ON_GEOM_CENT or by specifying a center of interest by its three coordinates. For example, a grid configuration with three focused grids, all centered at the geometrical center of the molecular system, looks like

$$\begin{array}{llll} \text{ON\_GEOM\_CENT} & n_1 & h_1 \\ \text{ON\_GEOM\_CENT} & n_2 & h_2 \\ \text{ON\_GEOM\_CENT} & n_3 & h_3, \end{array}$$

where n_l, $l = 1, 2, \ldots$ must be odd numbers. Typical values for n_l, h_l, $l = 1, 2, 3$ for one of the applications that we will be interested in are $n_1 = 97$, $h_1 = 3$, $n_2 = 161$, $h_2 = 1$, $n_3 = 161$, $h_3 = 0.25$. If all focused grids are centered at ON_GEOM_CENT, for short we write

$$\text{ON\_GEOM\_CENT} \quad n_1 - h_1,\, n_2 - h_2,\, n_3 - h_3.$$

To compute the electrostatic interaction with MEAD we have used either *multiflex* or *potential*. In the case of *multiflex*, to find the electrostatic interaction, MEAD performs two calculations for each focused grid - one for the electrostatic potential created by the charges of the first molecule and one for the electrostatic potential created by the charges of the second molecule. The final value for the electrostatic interaction is the average of the computed values for E_{2-1} and E_{1-2}. In the case of *potential*, MEAD outputs the potential at a priori specified by the user coordinates. Thus, we first compute the potential created by the charges $\{q_{i,1}\}_{i=1}^{N_1}$ of the first molecule and specify the positions $\{x_{i,2}\}_{i=1}^{N_2}$ of the charges of the second molecule at which the potential has to be returned. Next, we run *potential* with charges in the second molecule and specify the positions of the charges of the first molecule at which the potential has to be written out. Finally, by using the returned potentials at the positions $\{x_{i,1}\}_{i=1}^{N_1}$ and $\{x_{i,2}\}_{i=1}^{N_2}$ and the charges $\{q_{i,1}\}_{i=1}^{N_1}$ and $\{q_{i,2}\}_{i=1}^{N_2}$, respectively, we compute the approximations for E_{2-1} and E_{1-2}, respectively, and take their average. Both approaches produce the same

values for the electrostatic interaction. Since *multiflex* performs many calculations that are not relevant to the electrostatic interaction, we mostly use the procedure *potential*.

5.4.1 Uniform dielectric

In the first test, the dielectric coefficient is set to 1 in the whole $\mathbb{R}^3$ and there are no ions present in the solvent domain, i.e., $\overline{k} = 0$ and $\Omega_{IEL} = \emptyset$. We present the results for a total of 139 different configurations. Since $\epsilon_m = \epsilon_s = 1$ in Ω, it means that the effect of the interface Γ and its approximation by a piecewise triangular surface is eliminated. All volume meshes are generated in FreeFem++ [98] with TetGen [170] and then adaptively refined with mmg3d [62]. Notice that the Dirichlet boundary condition specified by MEAD is given by (5.125), which in this case ($\epsilon_m = \epsilon_s$) coincides with the exact potential on the boundary of the coarsest focused grid used by MEAD.

120 FRET frames with electroneutral dyes and atomic transition charges

The first 120 configurations correspond to 120 frames from a molecular dynamics (MD) simulation on the Alexa 594 and Alexa 488 dyes attached to a polyproline with a length of six amino residues. This MD simulation is related to the calculation of the Fröster resonance energy transfer (FRET) and thus the partial charges in each dye correspond to the so-called atomic transition charges. We will give more information on FRET in Section 5.5.1. The number of nonzero charges in each dye is $N_1 = 74$ and $N_2 = 52$, respectively, and the total charge in each dye is zero. The computational domain Ω is a cube with a very big, relative to the distances between the charges in the system, side length A. In particular, for this test, we use side lengths A between $130\,000\,\mathring{A}$ and $250\,000\,\mathring{A}$ depending on the dimensions of the smallest box that contains all $N_1 + N_2$ charges in the system. The boundary condition for the potential is set to zero on $\partial\Omega$ and due to the big size of Ω it approximates the exact potential on $\partial\Omega$ with a very high accuracy.

For this test, we used as initial meshes the ones that we used for the case of different dielectric coefficient in Ω_m and Ω_s (see Section 5.5.1). This means that the surfaces of the two dyes are actually triangulated and we have assigned $\epsilon_m = \epsilon_s = 1$ in both Ω_m and Ω_s. Since these meshes have uniform edge length (equal to the average edge length of the triangulated molecular surface Γ) for the elements inside the molecular region Ω_m, it means that they have unnecessarily many elements for this test, where $\epsilon_m = \epsilon_s$. However, in the next tests we will use much coarser meshes with much less elements and will demonstrate that this does not affect the adaptive algorithm and achieved accuracy.

We compare the obtained solution by using **Solver 4** and the exact solution, given by

$$E_{2-1} = \sum_{i=1}^{N_1} q_{i,1} G_2(x_{i,1}) = \sum_{i=1}^{N_2} q_{i,2} G_1(x_{i,2}) = E_{1-2}. \tag{5.126}$$

We note that the relative error in the solution of **Solver 2** is of the order $10^{-10}\,\%$ and thus we do not present any plots for it. The reason for this small error is that **Solver 2** finds the reaction field potential u_2, which in this case is zero. More precisely, since the inhomogeneous Dirichlet boundary condition $u_2 = -G_2$ (in the 2-term splitting of the potential $\varphi_2 = G_2 + u_2$) is prescribed on a surface which is at around 130 000 to 250 000 $\mathring{A}$ from the molecular system, the numerical approximation $u_{h,2}$ is also very close to zero. In fact, for the numerical approximation $\tilde{E}_{u_2}$ of the electrostatic interaction due to the reaction field potential we have $\frac{\tilde{E}_{u_2}}{E_{G_2}} \approx 10^{-12}$, where this ratio for the exact reaction field interaction E_{u_2} is, of course, zero. On the other hand, the relative error for **Solver 4** at refinement level 6 and 7 is of the order $10^{-2}\,\%$ to $10^{-3}\,\%$ as it can be seen on Figure 5.4. The reason for the higher relative error is that now **Solver 4** solves for the full potential φ_2 without exploiting the 2-term splitting and thus all singularities around the fixed partial charges of Dye II have to be approximated numerically.

On Figure 5.4, we also present the relative error for the software package MEAD with two different focused grids. The grid specification of the first one is ON_GEOM_CENT 97-3, 161-1, 161-0.25 which means that there are three focused grids, all centered at the geometrical center of the molecular system (Dye I and Dye II). The coarsest one has 97 grid points in each coordinate direction with a uniform grid spacing of 3 $\mathring{A}$. The second, finer grid, has 161 grid points per direction with a grid spacing of 1 $\mathring{A}$, and the third, finest grid, has 161 grid points per direction with a grid spacing of 0.25 $\mathring{A}$. The second grid configuration for MEAD is ON_GEOM_CENT 285-3, 285-1, 285-0.25. Clearly, for the first focused grid, the computational domain is a cube with edge length equal to $96 \times 3 = 288\,\mathring{A}$ and for the second focused grid, it is a cube with edge length of $284 \times 3 = 852\,\mathring{A}$. It is seen that in almost all 120 frames, **Solver 4** gives a better solution than MEAD even at the initial mesh refinement level (MRL). After only one refinement, the average relative error in **Solver 4** decreases approximately 6 times from 0.18% to 0.03% which is already 22.8 times less than the average error in MEAD with the better grid.

To compare the number of degrees of freedom (DOFs) on each refinement level with the corresponding relative error, in Table 5.1 we present the average number of DOFs in the primal and adjoint problems for each MRL and the corresponding average relative errors. We recall that the adjoint problem is solved in the space V_h^2 of continuous piecewise quadratic polynomials, and hence, in dimension three, the finite element space V_h^2 has approximately eight times more DOFs than the space V_h^1. For comparison, the number of DOFs for the first grid configuration in MEAD is 912 673, 4 173 281, 4 173 281, and for the second grid configuration is 23 149 125, 23 149 125, 23 149 125. Here we recall that on each focused grid MEAD has to solve two finite difference systems - one for the potential created by the charges of the first molecule and one for the potential created by the charges of the second molecule. This means that for the case of three focused grids, MEAD has to solve a total of six finite difference systems. The final electrostatic interaction $\overline{E}_M$ is the average of the

computed values for E_{2-1} and E_{1-2}.

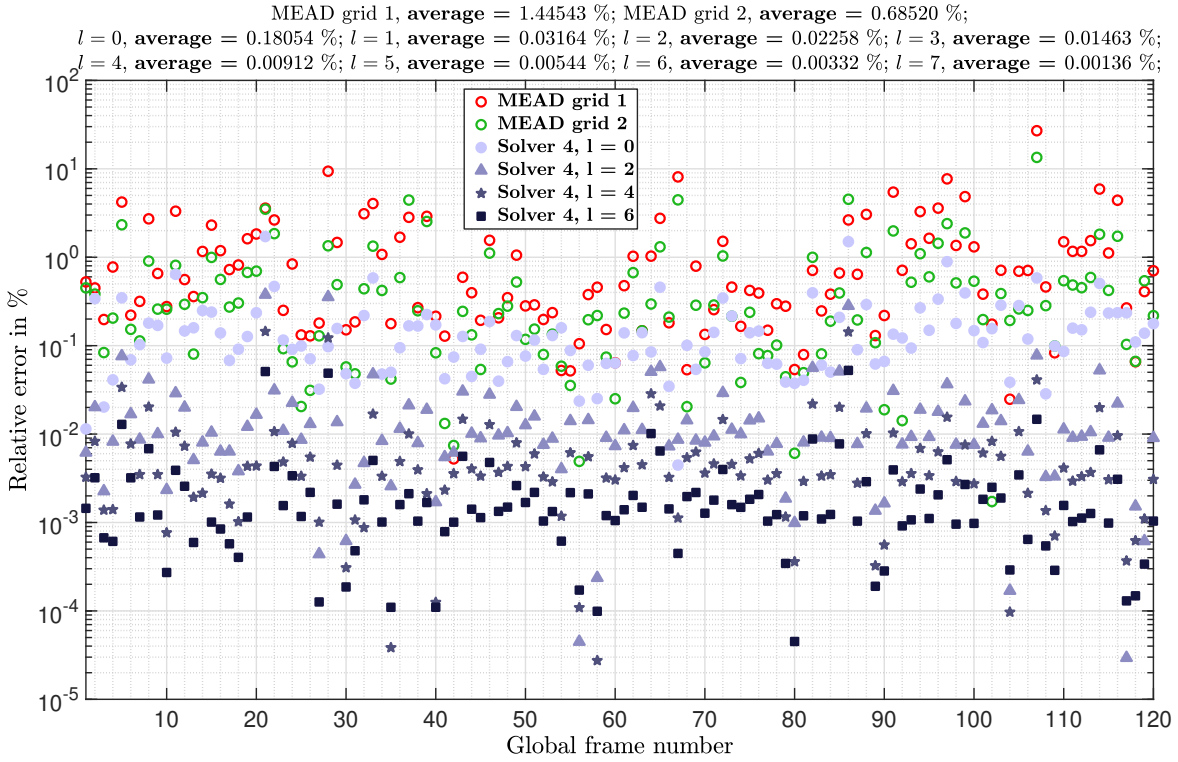


Figure 5.4: Neutral dyes with transition state atomic charges. Relative errors for MEAD with grid 1 and grid 2 compared to the relative errors $\frac{|E_{2-1}-E^A_{1-2}|}{|E_{2-1}|}[\%]$ for **Solver 4** on different refinement levels l, where $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ and $E_{1-2} = E_{2-1} = E$. Grid 1 in MEAD with specifications ON_GEOM_CENT 97-3, 161-1, 161-0.25 and grid 2 with specifications ON_GEOM_CENT 285-3, 285-1, 285-0.25. Some frames have less refinement levels since the maximum number of refinement steps has been set lower.

Recall that for **error estimate 4** we have the following error representation

$$\begin{aligned} E_{2-1} - E^P_{2-1} &= \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_2 - \varphi_{2,h} \rangle \\ &= E(\varphi_{2,h}, z) = E(\varphi_{2,h}, z_h^{(2)}) + \mathcal{E}(\varphi_2, \varphi_{2,h}, z, z_h^{(2)}), \end{aligned} \tag{5.109}$$

where $E^P_{2-1} + E(\varphi_{2,h}, z_h^{(2)}) = E^A_{1-2}$ and

$$E(\varphi_{2,h}, z) = \int_\Omega \epsilon \nabla (\varphi_2 - \varphi_{2,h}) \cdot \nabla z dx + \int_\Omega \overline{k}^2 (\varphi_2 - \varphi_{2,h}) z dx,$$

$$\mathcal{E}(\varphi_2, \varphi_{2,h}, z, z_h^{(2)}) = \int_\Omega \epsilon \nabla(\varphi_2 - \varphi_{2,h}) \cdot \nabla(z - z_h^{(2)}) dx + \int_\Omega \overline{k}^2 (\varphi_2 - \varphi_{2,h})(z - z_h^{(2)}) dx.$$

Table 5.1: **Solver 4.** 120 FRET frames with atomic transition charges and neutral dyes, $\overline{k} = 0$, $\epsilon_m = \epsilon_s = 1$.

level l	aver. # DOFs primal	aver. # DOFs adjoint	aver. $\frac{\left\|E_{2-1} - E^P_{2-1}\right\|}{\left\|E_{2-1}\right\|}$ [%]	aver. $\frac{\left\|E_{2-1} - E^A_{1-2}\right\|}{\left\|E_{2-1}\right\|}$ [%]
0	54 156	429 196	4.62501	0.18054
1	73 913	584 813	1.40682	0.03163
2	104 776	826 317	1.15969	0.02257
3	153 669	1 208 861	0.94636	0.01462
4	236 377	1 857 374	0.72293	0.00912
5	379 346	2 979 865	0.55447	0.00544
6	627 629	4 931 084	0.42831	0.00331
7	1 100 814	8 651 503	0.24057	0.00136

If we assume for a moment that φ_2 and z have higher regularity, say in the Sobolev space $H^3(\Omega)$, that we have a regular family of triangulations $\{\mathscr{T}_h\}_{h\to 0}$, and that the adjoint problem in the Aubin-Nitsche technique is H^2-regular, then one can show that the following relations hold (see, e.g., [32, 68]):

$$\begin{aligned} \|\varphi_2 - \varphi_{2,h}\|_{L^2(\Omega)} &= O(h^2), \qquad \|\nabla(\varphi_2 - \varphi_{2,h})\|_{L^2(\Omega)} &= O(h), \\ \|z - I_h z\|_{L^2(\Omega)} &= O(h^2), \qquad \|\nabla(z - I_h z)\|_{L^2(\Omega)} &= O(h), \\ \|z - z_h^{(2)}\|_{L^2(\Omega)} &= O(h^3), \qquad \|\nabla\left(z - z_h^{(2)}\right)\|_{L^2(\Omega)} &= O(h^2). \end{aligned}$$

Therefore, by using Galerkin orthogonality and the Cauchy-Schwarz inequality we find

$$\begin{aligned} E(\varphi_{2,h}, z) &= \int_\Omega \epsilon \nabla(\varphi_2 - \varphi_{2,h}) \cdot \nabla(z - I_h z)\, dx + \int_\Omega \overline{k}^2 (\varphi_2 - \varphi_{2,h})(z - I_h z)\, dx \\ &\leq \epsilon_{\max} \|\nabla(\varphi_2 - \varphi_{2,h})\|_{L^2(\Omega)} \|\nabla(z - I_h z)\|_{L^2(\Omega)} \\ &\qquad + k_{\max} \|(\varphi_2 - \varphi_{2,h})\|_{L^2(\Omega)} \|z - I_h z\|_{L^2(\Omega)} \leq C_1 h^2 \end{aligned} \tag{5.127}$$

and in a similar way we find

$$\mathcal{E}(\varphi_2, \varphi_{2,h}, z, z_h^{(2)}) \leq C_2 h^3, \tag{5.128}$$

where I_h is the nodal interpolation operator in the space V_h^1 and C_1, $C_2 > 0$ are constants independent of h. This means that under uniform mesh refinement in 3D we would have

$$\left|E_{2-1} - E^P_{2-1}\right| = |E(\varphi_{2,h}, z)| \leq C_1 h^2 \approx \frac{C_1}{(\#DOFs)^{2/3}}, \tag{5.129}$$

$$\left|E_{2-1} - E^A_{1-2}\right| = \left|\mathcal{E}(\varphi_2, \varphi_{2,h}, z, z_h^{(2)})\right| \leq C_2 h^3 \approx \frac{C_2}{\#DOFs}. \tag{5.130}$$

In the case of adaptive mesh refinement, it is more appropriate to measure the approximation error in terms of the number of degrees of freedom $\#DOFs$ instead of the maximum mesh

size h. From Figure 5.5 it seems that the same optimal convergence order, $O((\#DOFs)^{-2/3})$ and $O((\#DOFs)^{-1})$, hold for the errors $\left|E_{2-1} - E^P_{2-1}\right|$ and $\left|E_{2-1} - E^A_{1-2}\right|$, respectively, in the case where φ_2 and z are much less regular (both in $\bigcap_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega)$). On Figure 5.6

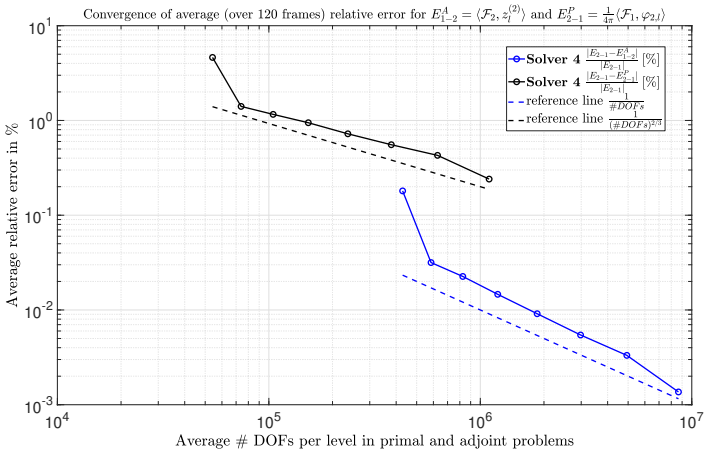


Figure 5.5: Neutral dyes with transition state atomic charges. Convergence of average (over 120 frames) relative errors of **Solver 4** for the quantities $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ and $E^P_{2-1} = \frac{1}{4\pi}\langle \mathscr{F}_1, \varphi_{2,l} \rangle$. The average number of DOFs per MRL and the average relative errors for the primal and adjoint problems used in this plot are given in Table 5.1.

we show the convergence of the relative errors in the primal and adjoint problems for two particular frames.

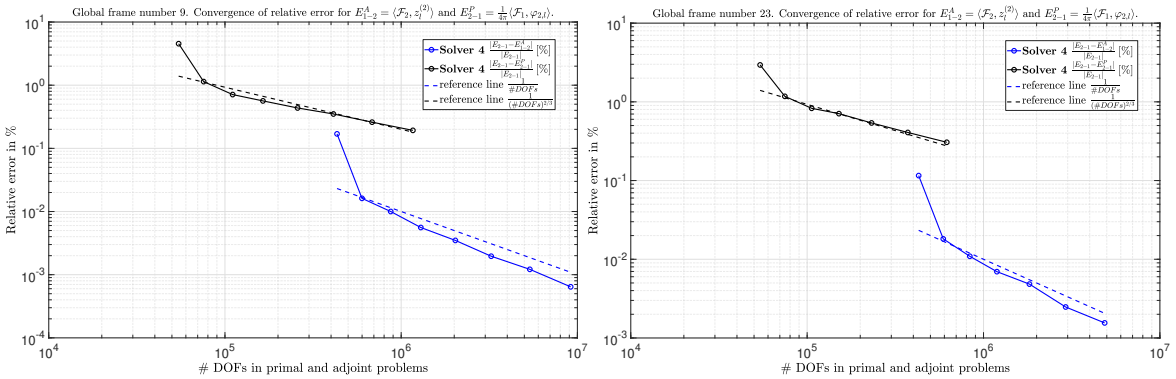


Figure 5.6: **Solver 4.** Neutral dyes with transition state atomic charges. Convergence of relative error in the primal and adjoint problems for frames with global numbers 9 and 23.

19 frames with electroneutral dyes and atomic ground state charges

In the next 19 frames the partial charges in the two dyes correspond to their ground state charge density. The first 10 of these frames are randomly chosen from the same MD simulation from which the previous 120 frames were chosen, and thus the dyes are attached

to a polyproline with a length of 6 amino residues. The other 9 frames are from another MD simulation, where the two dyes are attached to a polyproline with a length of 20 amino residues. The number of nonzero charges in Dye I and Dye II is 79 and 73, respectively. Again, the two dyes are electroneutral, and thus the total charge sum in each dye is zero. The computational domain Ω is a cube with a side length $A = 4000 \times a$, where a is the edge length of the smallest cube, with edges parallel to the coordinate axes, that contains the two dyes. For the first 10 frames $a \approx 40\,\mathring{A}$ and $A \approx 160\,000\,\mathring{A}$, whereas for the other 9 frames a takes values of up to $100\,\mathring{A}$, corresponding to $A \approx 400\,000\,\mathring{A}$. Since the mesh is very coarse towards the boundary of Ω, increasing A even to over $1\,000\,000\,\mathring{A}$ is at the cost of a few more tetrahedrons. Again, as in the previous test, the meshes used here have much more elements than needed for this kind of numerical experiment, where $\epsilon_m = \epsilon_s$. This is due to the fact that we reuse the meshes that were generated for the case of different dielectric coefficients ϵ_m and ϵ_s, where it is important to resolve the interface Γ with a good precision. Nevertheless, we will present two more tests, where the initial meshes have approximately 13 times less elements and the edge length of the cube Ω reaches more than $5 \times 10^6\,\mathring{A}$.

The results that we have obtained with **Solver 4** are also compared to the results obtained with the application *potential* in the software package MEAD. We have observed that there is an accuracy issue with the default stopping criteria in the iterative solver of MEAD, which is Successive over-relaxation (SOR). For this reason, we have run MEAD with 6 different configurations for the focused grids and different stopping criteria for the SOR method. Below we list the six configurations that we have tested in MEAD.

- **Configuration 1**: ON_GEOM_CENT: 97-3, 161-1, 161-0.25,
  Default SOR accuracy: `maxiter`$=10 \times \frac{3}{2}(\texttt{grid\_dim} - 1)$, `maxrmsdiff`$= \frac{2\times10^{-5}}{\texttt{grid\_dim}-1}$;
- **Configuration 2**: ON_GEOM_CENT: 97-3, 161-1, On the geometric center of the dye without charges: 161-0.25,
  Default SOR accuracy: `maxiter`$=10 \times \frac{3}{2}(\texttt{grid\_dim} - 1)$, `maxrmsdiff`$= \frac{2\times10^{-5}}{\texttt{grid\_dim}-1}$;
- **Configuration 3**: ON_GEOM_CENT: 97-3, 161-1, On the geometric center of the dye without charges: 401-0.1,
  `maxiter`$=6000$, `maxrmsdiff`$= \frac{2\times10^{-8}}{\texttt{grid\_dim}-1}$;
- **Configuration 4**:
  Frames 1-10: ON_GEOM_CENT: 101-3, 201-1, 321-0.25 ,
  Frames 11:19: ON_GEOM_CENT: 101-3, 201-1, 401-0.25,
  Default SOR accuracy: `maxiter`$=10 \times \frac{3}{2}(\texttt{grid\_dim} - 1)$, `maxrmsdiff`$= \frac{2\times10^{-5}}{\texttt{grid\_dim}-1}$;
- **Configuration 5**:
  Frames 1-10: ON_GEOM_CENT: 101-3, 201-1, 321-0.25,
  Frames 11-19: ON_GEOM_CENT: 101-3, 201-1, 401-0.25,
  `maxiter`$=3000$, `maxrmsdiff`$= \frac{2\times10^{-8}}{\texttt{grid\_dim}-1}$;

- **Configuration 6**:
  Frames 1-10: ON_GEOM_CENT: 101-3, 201-1, 321-0.25,
  Frames 11-19: ON_GEOM_CENT: 101-3, 201-1, 401-0.25,
  `maxiter`=12000, `maxrmsdiff`= $\frac{2\times 10^{-8}}{\texttt{grid\_dim}-1}$;

In MEAD, the variable `minits` specifies the minimum number of iterations, which by default is set to $3(\texttt{grid\_dim}-1)/2$, where `grid_dim` is the number of grid points in each coordinate direction. Further, by `maxrmsdiff` is denoted the MEAD variable specifying the tolerance for the maximum root mean square of the difference (`rmsdiff`) between two consecutive solutions φ^k and φ^{k+1} in the SOR method, and by `maxiter` is denoted the MEAD variable specifying the maximum number of iterations. By default, `maxrmsdiff` is set to $2\times 10^{-5}/(\texttt{grid\_dim}-1)$ and `maxiter` is set to $10\times$ `minits`. The SOR procedure in MEAD stops when either the maximum number of iterations is reached or the current iteration number is greater than or equal to `minits` and simultaneously `rmsdiff` is less than or equal to `maxrmsdiff`. Note that in configuration 2 and 3, the third focused grid is centered at the geometric center of the dye without charges. More precisely, we make two calculations with *potential*: one with charges in Dye I and one with charges in Dye II. When the charges are in Dye I, we specify the positions $\{x_{i,2}\}_{i=1}^{N_2}$ of the charges in Dye II at which the computed potential has to be written out by *potential* and hence the last focused grid is centered at Dye II in attempt to better resolve the potential at these positions $\{x_{i,2}\}_{i=1}^{N_2}$. Similarly, when the charges are in Dye II, we specify the positions $\{x_{i,1}\}_{i=1}^{N_1}$ of the charges of Dye I at which the computed potential has to be written out by *potential* and the last focused grid is centered at Dye I. As we have already explained, the final value for the electrostatic interaction is the average of the computed values for E_{2-1} and E_{1-2}. Additionally, in configurations 4, 5, and 6, the last focused grid for frames 11 through 19 has a higher number of grid points per coordinate direction since the length of the polyproline connecting the two dyes is larger for these frames, and thus the minimal box that contains the molecular system is also larger.

In Table 5.2 are presented the average (over the 19 frames) number of DOFs per MRL for the primal and adjoint problems as well as the average (over all 19 frames) relative error in percents for the primal and adjoint problems. On Figure 5.8 are shown the relative errors for all 19 frames for three MEAD configurations and three MRLs for **Solver 4**, whereas on Figure 5.7 is shown the convergence of the average relative errors with respect to the average number of DOFs per MRL (with the data from Table 5.2). For comparison, in Table 5.3 we present the average (over all 19 frames) relative errors for MEAD together with the number of DOFs for each focused grid, where $E = E_{2-1} = E_{1-2}$ is the exact electrostatic interaction and $\overline{E}_M$ is the average of the two computed by MEAD approximations for E_{2-1} and E_{1-2}. It is clear that the default stopping criteria in SOR (configurations 1, 2 , 4) is not quite appropriate. Another conclusion that can be made from the results presented in Table 5.3 is that focusing the last and finest grid on the dye of interest does not improve the quality of the

computed electrostatic potential. As a consequence, the quality of the computed electrostatic interaction is also not improved (see the averages for configurations 2 and 3).

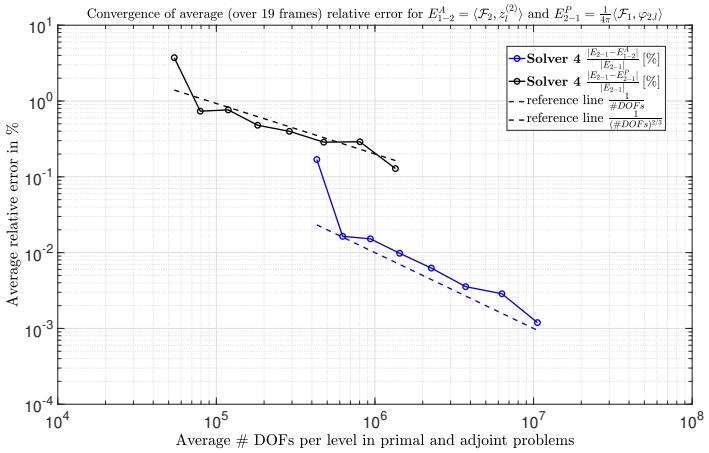


Figure 5.7: Neutral dyes with ground state atomic charges. Convergence of average (over 19 frames) relative errors of **Solver 4** for the quantities $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ and $E^P_{2-1} = \frac{1}{4\pi}\langle \mathscr{F}_1, \varphi_{2,l} \rangle$. The average number of DOFs per MRL and the average relative errors for the primal and adjoint problems used in this plot are given in Table 5.2.

Table 5.2: **Solver 4.** 19 frames with atomic ground state charges and neutral dyes, $\overline{k} = 0$, $\epsilon_m = \epsilon_s = 1$.

level l	aver. # DOFs primal	aver. # DOFs adjoint	aver. $\frac{\lvert E_{2-1} - E^P_{2-1}\rvert}{\lvert E_{2-1}\rvert}$ [%]	aver. $\frac{\lvert E_{2-1} - E^A_{1-2}\rvert}{\lvert E_{2-1}\rvert}$ [%]
0	54 268	430 026	3.73208	0.16928
1	79 173	625 903	0.73420	0.01639
2	118 824	935 871	0.76450	0.01519
3	182 132	1 430 898	0.47930	0.00980
4	288 720	2 266 601	0.39852	0.00625
5	475 450	3 733 080	0.28664	0.00355
6	803 911	6 315 438	0.29003	0.00286
7	1 346 508	10 582 506	0.12850	0.00119

Table 5.3: **MEAD.** 19 frames with atomic ground state charges and neutral dyes, $\overline{k} = 0$, $\epsilon_m = \epsilon_s = 1$. Here, $E = E_{2-1} = E_{1-2}$ and $\overline{E}_M$ is the computed with MEAD value.

configuration	# DOFs in focused grid 1	# DOFs in focused grid 2	# DOFs in focused grid 3	aver. $\frac{\lvert E - \overline{E}_M\rvert}{\lvert E\rvert}$ [%]
1	912 673	4 173 281	4 173 281	1.51675
2	912 673	4 173 281	4 173 281	4.58068
3	912 673	4 173 281	64 481 201	4.15532
4	1 030 301	8 120 601	33 076 161 or 64 481 201	1.38888
5	1 030 301	8 120 601	33 076 161 or 64 481 201	0.28533
6	1 030 301	8 120 601	33 076 161 or 64 481 201	0.29212

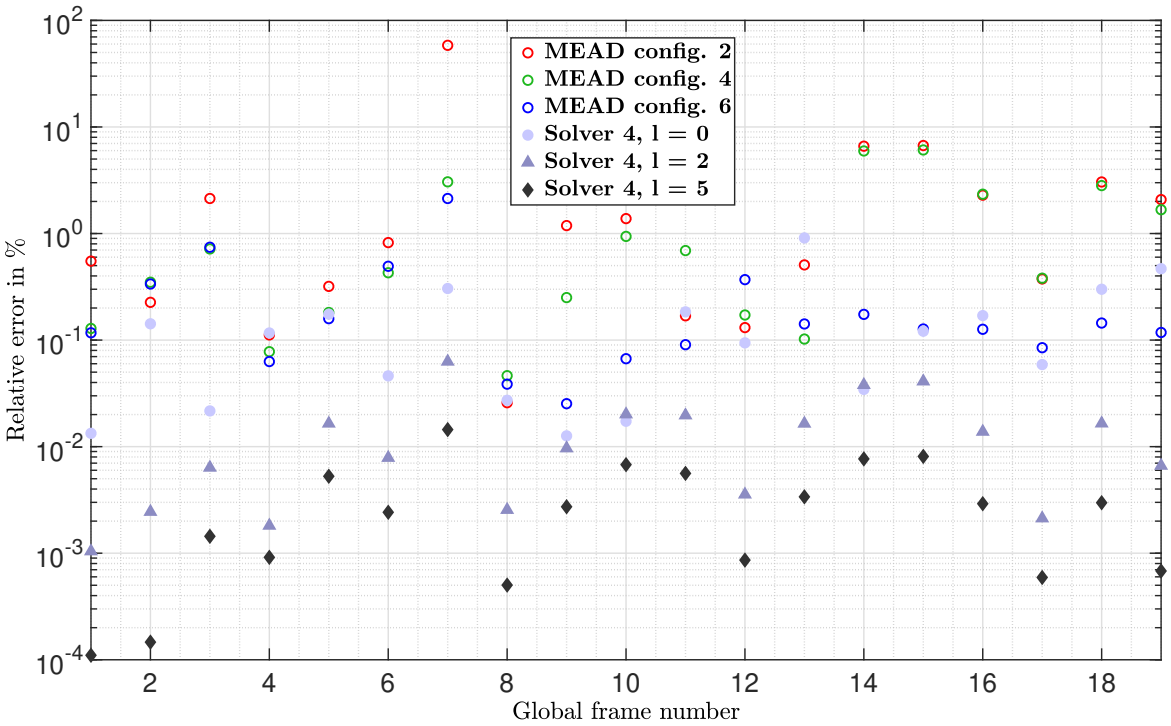


Figure 5.8: Neutral dyes with ground state atomic charges. Relative errors for MEAD with three configurations versus relative errors $\frac{|E_{2-1}-E^A_{1-2}|}{|E_{2-1}|}[\%]$ for the interaction computed with the adjoint solution in **Solver 4** for three different mesh refinement levels l.

19 frames with electroneutral dyes and atomic ground state charges. Very coarse initial meshes.

Now, we repeat the previous test, but this time we use much coarser initial meshes which means much less elements. Moreover, the edge length of the cube Ω is between 3.7×10^6 $\mathring{A}$ and 6×10^6 $\mathring{A}$ and so the homogeneous Dirichlet boundary condition on the potential is prescribed with a very high accuracy. In Table 5.4 we present the average number of DOFs per MRL as well as the corresponding average relative error in the primal and adjoint problems. We can see that a relative error of 0.23 % in the electrostatic interaction is achieved on average with 131 868 DOFs in the adjoint problem. For comparison, MEAD achieves an average error of 0.29 % with configurations 5 and 6 at a significant computational cost: 33 076 161 DOFs for frames 1-10 and 64 481 201 DOFs for frames 11-19 in the last focused grid (see Table 5.3). On Figure 5.9 is shown the convergence of the average relative error versus the average number of DOFs per MRL in primal and adjoint problems (with the data from Table 5.4).

Table 5.4: **Solver 4.** Very coarse initial meshes. 19 frames with atomic ground state charges and neutral dyes, $\overline{k} = 0$, $\epsilon_m = \epsilon_s = 1$.

level l	aver. # DOFs primal	aver. # DOFs adjoint	aver. $\frac{\lvert E_{2-1} - E^P_{2-1} \rvert}{\lvert E_{2-1} \rvert}$ [%]	aver. $\frac{\lvert E_{2-1} - E^A_{1-2} \rvert}{\lvert E_{2-1} \rvert}$ [%]
0	3 951	29 739	15.50291	3.17777
1	6 411	49 223	7.47266	0.55509
2	10 684	82 785	3.07105	0.39941
3	16 965	131 868	1.79431	0.22948
4	26 283	204 637	1.89854	0.15297
5	41 136	320 745	1.17146	0.10046
6	66 581	520 204	0.82917	0.06163

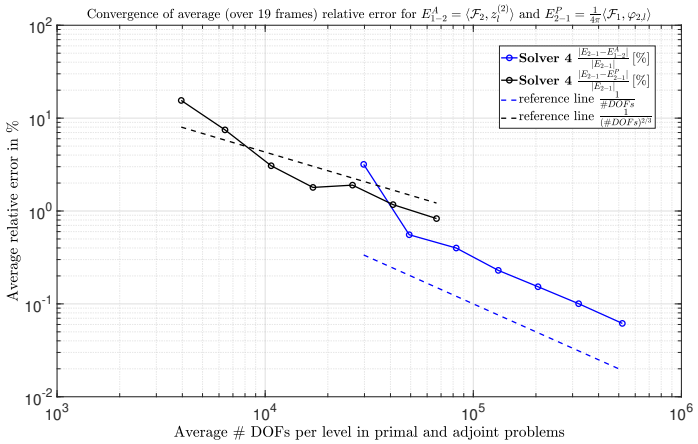


Figure 5.9: Neutral dyes with ground state atomic charges. Very coarse initial meshes. Convergence of average (over 19 frames) relative errors of **Solver 4** for the quantities $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ and $E^P_{2-1} = \frac{1}{4\pi}\langle \mathscr{F}_1, \varphi_{2,l} \rangle$. The average number of DOFs per MRL and the average relative errors for the primal and adjoint problems used in this plot are given in Table 5.4.

19 frames with charged dyes and atomic ground state charges. Very coarse initial meshes.

In this test, we use the same 19 frames (positions of the charges of the two dyes) as in the previous two tests, but with different set of charges for both dyes. In particular, both dyes are charged with a total charge of $-2\,e_0$ in each dye. Again, the number of nonzero charges in Dye I and Dye II is 79 and 73, respectively. In Table 5.5 are presented the average relative error for the primal and adjoint problems as well as the respective average number of DOFs per MRL. On Figure 5.10 is shown the convergence of the average per MRL relative error versus the respective average number of DOFs (with the data from Table 5.5). We note that after MRL 7 the error for the (more accurately solved) adjoint problem starts to stagnate. This is due to the approximate boundary condition on $\partial\Omega$. For a higher accuracy than around $0.002\,\%$ in the case of charged dyes, a larger domain Ω is required.

Table 5.5: **Solver 4.** Very coarse initial meshes. 19 frames with atomic ground state charges and charged dyes, $\overline{k} = 0$, $\epsilon_m = \epsilon_s = 1$.

level l	aver. # DOFs primal	aver. # DOFs adjoint	aver. $\frac{\lvert E_{2-1} - E^P_{2-1}\rvert}{\lvert E_{2-1}\rvert}$ [%]	aver. $\frac{\lvert E_{2-1} - E^A_{1-2}\rvert}{\lvert E_{2-1}\rvert}$ [%]
0	3 951	29 739	21.89265	3.02298
1	5 455	41 623	7.64000	0.36758
2	10 009	77 657	2.67448	0.06875
3	17 576	137 170	1.94356	0.03040
4	30 091	235 446	1.37424	0.01756
5	50 773	397 871	0.98067	0.00914
6	84 089	659 384	0.73550	0.00564
7	140 790	1 104 774	0.55714	0.00383

For comparison, in this case with charged dyes, the average (over 19 frames) relative error with MEAD for configurations 5 and 6 (see p. 227 for the description of different configurations) is 0.081 % and 0.115 %, respectively. The corresponding number of DOFs is between 33×10^6 and 64×10^6 depending on the size of the smallest box that contains the two dyes. On the other hand, the number of DOFs with **Solver 4** for a comparable average relative error is less than 78×10^4 (see Table 5.5). This is approximately a factor of 850 times less compared to MEAD with the above configurations.

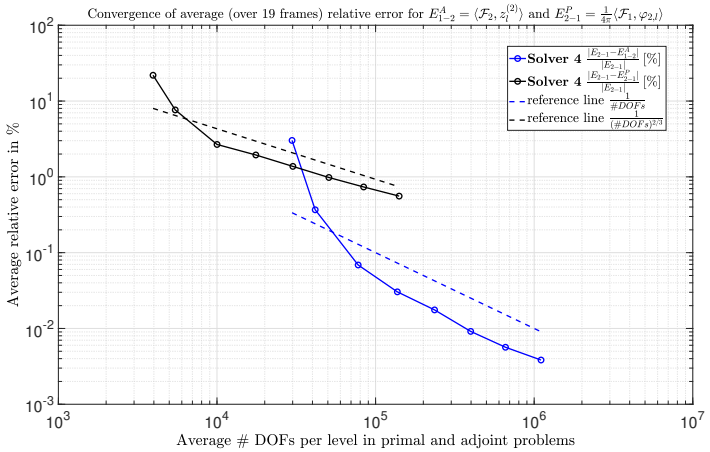


Figure 5.10: Charged dyes with ground state atomic charges and a total charge of $-2\,e_0$ in each dye. Very coarse initial meshes. Convergence of average (over 19 frames) relative errors of **Solver 4** for the quantities $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ and $E^P_{2-1} = \frac{1}{4\pi}\langle \mathscr{F}_1, \varphi_{2,l} \rangle$. The average number of DOFs per MRL and the average relative errors for the primal and adjoint problems used in this plot are given in Table 5.5.

5.4.2 Born ion model with $I_s = 0$

Solver 4

In this test we check the performance of **Solver 4** in the presence of a sharp interface and we also compare it to MEAD. The set up is as follows: an ion, with a Born radius R equal to 1, 2, and 3 $\mathring{A}$ with $\epsilon_m = 4$ and charge $q_1 = 1\,e_0$, is placed in a dielectric medium with $\epsilon_s = 80$ and zero ionic strength (see Figure 5.11). We compute the electrostatic interaction between the ion and a test charge $q_2 = -1\,e_0$ placed at distances varying from $1\,\mathring{A}$ to $74\,\mathring{A}$. In other words, we compute the potential at the coordinates of the test charge and multiply it by $-1\,e_0$. In this case, it is easy to check that the electrostatic potential φ is given by

$$\varphi(r) = \begin{cases} \dfrac{q_1}{\epsilon_m r} + \dfrac{q_1\,(\epsilon_m - \epsilon_s)}{\epsilon_m \epsilon_s R}, & r \le R, \\ \dfrac{q_1}{\epsilon_s r}, & r \ge R, \end{cases} \tag{5.131}$$

where r is the distance to the charge q_1 at the center of the Born ion.

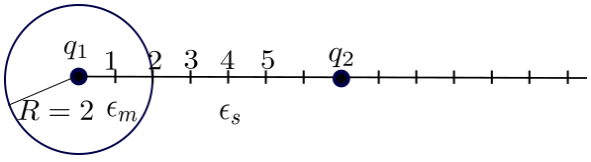
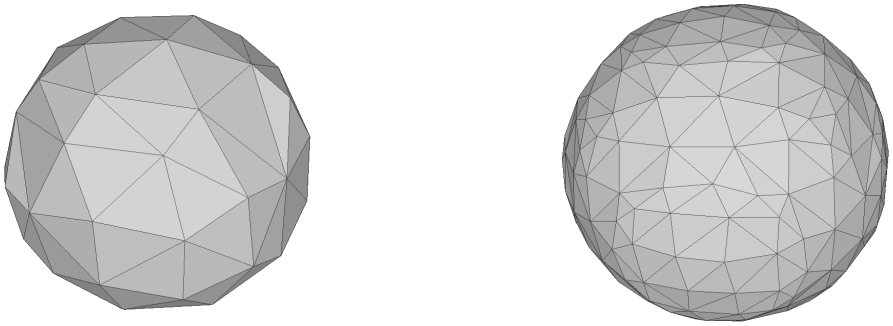


Figure 5.11: Born ion model with $I_s = 0$.

The computational domain in **Solver 4** is a cube with an edge length of approximately $1\,350\,000\,\mathring{A}$ and a homogeneous Dirichlet boundary condition is prescribed at its boundary.

We have done computations with two levels of geometric approximation of the surface of the Born ion sphere. The triangulated surface meshes are generated with NanoShaper [61], where the quality of the geometric approximation is controlled with the parameter *GridScale*. A *GridScale* parameter equal to 2 means that in the construction of the surface mesh is used a cubical grid with spacing of $h = 0.5\,\mathring{A}$ and *GridScale* $= 4$ means that the spacing in the cubical grid is $h = 0.25\,\mathring{A}$ (see Figure 5.12). A grid scale of 2 produces an average edge length of the triangulated surface mesh of around $0.45\,\mathring{A}$, whereas *GridScale* $= 4$ results in an average edge length on the molecular surface of around $0.225\,\mathring{A}$.



(a) *GridScale*=2. (b) *GridScale*=4.

Figure 5.12: Triangulated with NanoShaper surfaces of a sphere with radius $R = 1\,\mathring{A}$.

This means that on Figure 5.14 one should compare the solution produced by MEAD on focused grids with a spacing of the final grid $h_3 = 0.25\,\mathring{A}$ to the solution produced by **Solver 4** corresponding to a surface mesh generated with *GridScale* $= 4$ in NanoShaper. Since the finite difference solution strongly depends on the relative position and orientation of the grid and molecular structure, the values for the relative errors of MEAD, used in Figure 5.14, are averages of the values sampled in different directions from the Born ion's center. More precisely, the value $\overline{e}_s$ for the relative error at a distance $s\,\mathring{A}$ from the Born ion's center is obtained by averaging over 10 different directions $\vec{d}_i$, $i = 1, 2, \ldots, 10$, where $\vec{d}_1 = (1, 0, 0)$, $\vec{d}_2 = (-1, 0, 0)$, $\vec{d}_3 = (0, 1, 0)$, $\vec{d}_4 = (0, -1, 0)$, $\vec{d}_5 = (0, 0, 1)$, $\vec{d}_6 = (0, 0, -1)$, $\vec{d}_7 = (1, 1, 1)$, $\vec{d}_8 = (-1, -1, -1)$, $\vec{d}_9 = (1, 0.5, 0.2)$, $\vec{d}_{10} = (-1, -0.5, -0.2)$. The precise definition of $\overline{e}_s$ is

$$\overline{e}_s := \frac{\sum_{i=1}^{10} e\left(\vec{c} + s\frac{\vec{d}_i}{|\vec{d}_i|_2}\right)}{10},$$

where $\vec{c}$ is the position of the Born ion's center and $e(\vec{x})$ is the error at the point $\vec{x}$. In the examples below, we have chosen $\vec{c} = (0.05, 0.03, -0.025)$ so that the finite difference grid is not centered exactly at the Born ion's center.

Notice, that the default stopping criteria in this MEAD version (2.2.8 a) is not appropriate: the relative error for both used grids, ON_ORIGIN 97-3, 161-1, 161-0.25 and ON_ORIGIN

285-1, 285-0.5, 285-0.1, approaches values of 20 % and 50 %, respectively, at 35 $\mathring{A}$ from the Born ion (the red and yellow curves on Figure 5.14). The corresponding relative errors at a distance of 74 $\mathring{A}$ reach 30 % and 80 %, respectively. On the other hand, the relative errors of MEAD with the improved stopping criteria do not exceed 0.1 %. Here we note that MEAD uses single precision arithmetics. The transition to double precision arithmetics could possibly improve the results even further. In comparison, the relative errors for **Solver 4** do not exceed 0.022 %. However, at the interface Γ between the interior region of the Born ion sphere and the exterior dielectric medium, the finite difference solution given by MEAD is of very poor quality with relative errors reaching more than 60 % and more than 20 % for the grids ON_ORIGIN 97-3, 161-1, 161-0.25 and ON_ORIGIN 285-1, 285-0.5, 285-0.1, respectively, for both default SOR method and the one with improved stopping criteria. The relative errors around the interfaces in the solution produced by the adaptive finite element **Solver 4** do not exceed more than 2.2 % and more than 0.6 % with *GridScale*=2 and *GridScale=4*, respectively. It is seen that the solution produced by **Solver 4** for a relatively coarse surface representation of the Born sphere ($GridScale = 2$) has a better quality even than the MEAD solution corresponding to the grid with spacing in the last focused grid of 0.1 $\mathring{A}$ and with the improved stopping criteria.

Further, on Figure 5.15 and Figure 5.16 is shown the convergence of the relative errors at all mesh refinement levels $l = 0, 1, \ldots, 5$ for *GridScale*=2 and *GridScale*=4, respectively. Additionally, in Table 5.6 and Table 5.7, we present the average number of DOFs in the adjoint problem for $R = 1, 2, 3\,\mathring{A}$ and for each refinement level with *GridScale*=2 and *GridScale*=4, respectively. The average is taken over all distances from 5 $\mathring{A}$ to 74 $\mathring{A}$ from the Born ion's center. At each of these distances, the corresponding number of DOFs is very close to the average. The number of DOFs corresponding to the distances $1, 2, 3, 4\,\mathring{A}$ are much less, and thus we do not include them in the averaging. We note that we have used initial meshes with a uniform mesh size inside the Born ion sphere related to the average edge length on the triangulated sphere. This means that the initial meshes have a large number of elements, especially for spheres with larger radius R. The computational cost can be further reduced, without sacrificing the convergence orders and accuracy, if one uses coarser meshes with nonuniform mesh size inside the triangulated sphere, as we have demonstrated in the previous two examples.

Remark 5.16
Recall that the primal problem gives the potential φ_2 created by the charges of the second molecule. In this example, there is no second molecule, only a test charge which we have chosen to be the one that creates the potential φ_2. Thus, the right-hand side of the primal problem is given by

$$\mathscr{F}_2 = 4\pi q_2 \delta_{x_2},$$

where x_2 is the position of the test charge q_2. The right hand side of the adjoint problem is

given by

$$\frac{1}{4\pi}\mathscr{F}_1 = q_1\delta_{x_1},$$

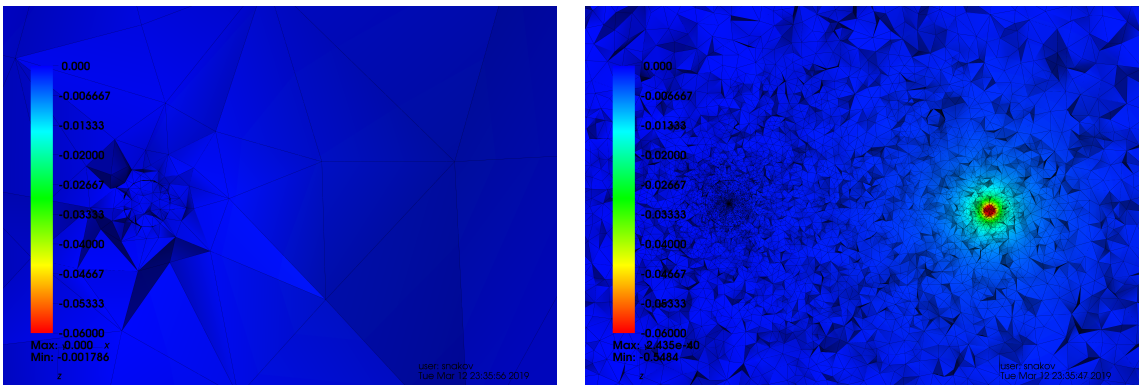
where $x_1 = (0,0,0)$ *is the position of the charge* q_1 *in the Born ion. The solution* z *of the adjoint problem is equal to* $\frac{1}{4\pi}\varphi_1$ *and the more accurate approximation for the electrostatic interaction that we use is* $\langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_{2,h}\rangle + E(\varphi_{2,h}, z_h^{(2)}) = \langle \mathscr{F}_2, z_h^{(2)}\rangle = E^A_{1-2}$. *Notice also that the test charge is placed in the solvent region* Ω_s. *In this case, the derived representations for the error in the goal quantity are still valid based on Remark 5.14.*

Table 5.6: **Solver 4.** *GridScale*=2. Born ion model, $\overline{k} = 0$, $\epsilon_m = 4$, $\epsilon_s = 80$. Averages of number of DOFs and errors for adjoint problem for distances from 5 $\mathring{A}$ to 74 $\mathring{A}$. Uniform initial mesh size inside the Born ion sphere related to the average edge length on the triangulated surface.

	$R = 1\,\mathring{A}$		$R = 2\,\mathring{A}$		$R = 3\,\mathring{A}$	
level	aver. # DOFs adjoint	aver. error [%]	aver. # DOFs adjoint	aver. error [%]	aver. # DOFs adjoint	aver. error [%]
0	7 509	6.26998	18 201	6.55494	63 310	1.03680
1	10 341	0.82301	23 192	0.62322	85 871	0.08286
2	19 556	0.22892	41 040	0.11699	146 781	0.02227
3	36 953	0.08525	74 919	0.03849	282 397	0.01115
4	71 184	0.04124	144 749	0.02071	538 558	0.00796
5	135 774	0.01962	275 398	0.01114	995 427	0.00672

Table 5.7: **Solver 4.** *GridScale*=4. Born ion model, $\overline{k} = 0$, $\epsilon_m = 4$, $\epsilon_s = 80$. Averages of number of DOFs and errors for adjoint problem for distances from 5 $\mathring{A}$ to 74 $\mathring{A}$. Uniform initial mesh size inside the Born ion sphere related to the average edge length on the triangulated surface.

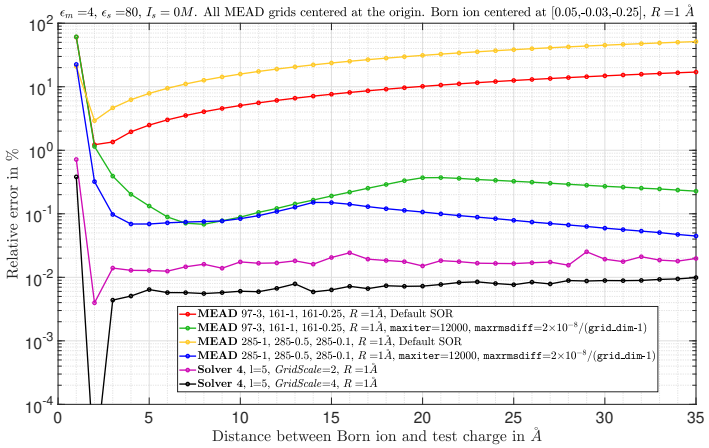
	$R = 1\,\mathring{A}$		$R = 2\,\mathring{A}$		$R = 3\,\mathring{A}$	
level	aver. # DOFs adjoint	aver. error [%]	aver. # DOFs adjoint	aver. error [%]	aver. # DOFs adjoint	aver. error [%]
0	18 543	3.69832	121 019	0.68096	229 342	2.71617
1	23 061	0.52291	151 693	0.06063	245 086	0.25489
2	41 278	0.11451	263 636	0.01419	312 620	0.03238
3	77 046	0.03822	513 567	0.00834	557 970	0.00800
4	154 979	0.01903	1 000 098	0.00663	1 056 868	0.00622
5	301 694	0.01044	1 860 927	0.00597	2 011 042	0.00555



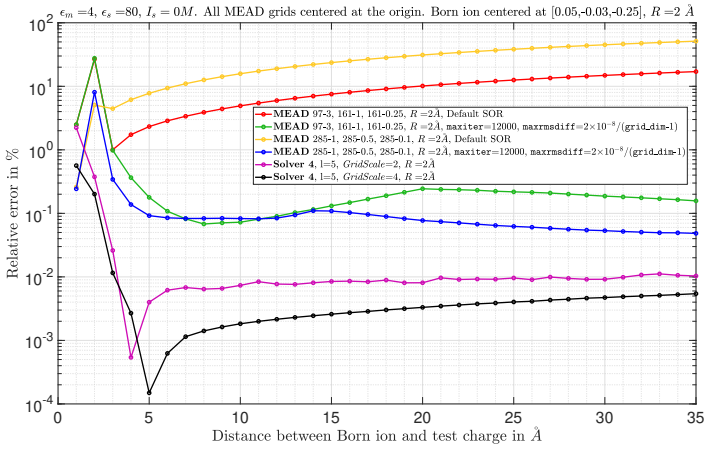
(a) Initial mesh, $l = 0$.

(b) Final mesh, $l = 9$.

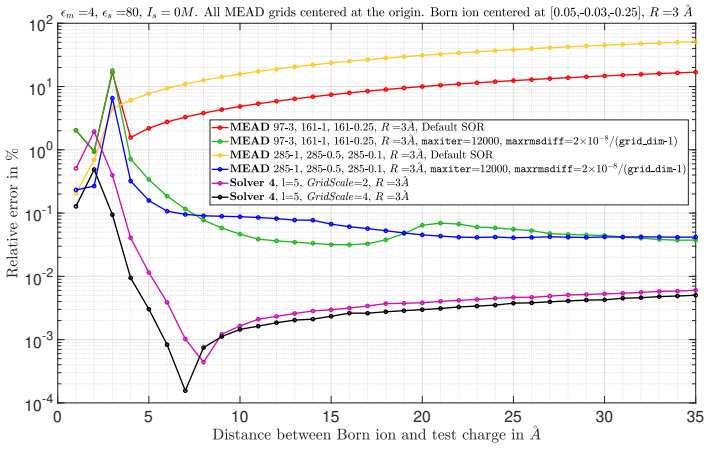
Figure 5.13: **Solver 4.** *GridScale*=2. Born ion model, $R = 1\,\mathring{A}$, $I_s = 0M$, $\epsilon_m = 4$, $\epsilon_s = 80$. Distance between q_2 and q_1 is $10\,\mathring{A}$. Full potential $\varphi_{2,l}$ (obtained by solving the primal problem) at initial and final meshes in units $e_0/\mathring{A}$. Pictures generated with VisIt [52]



(a) $R = 1\,\mathring{A}$.

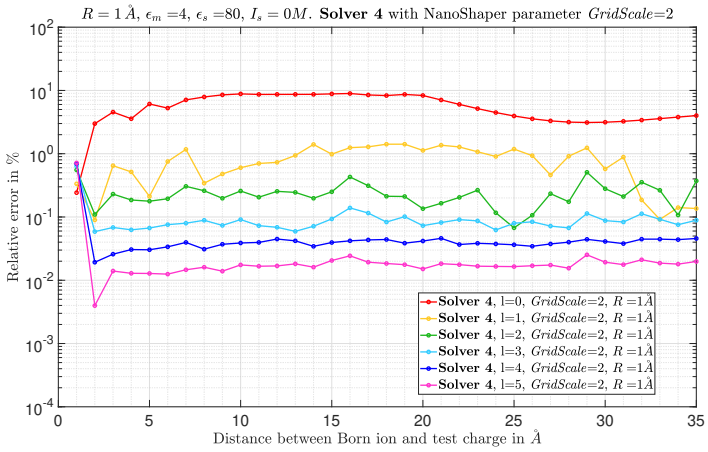


(b) $R = 2\,\mathring{A}$.

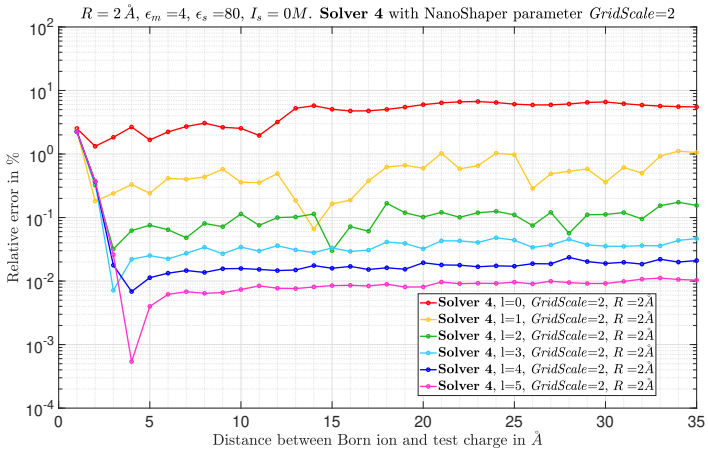


(c) $R = 3\,\mathring{A}$.

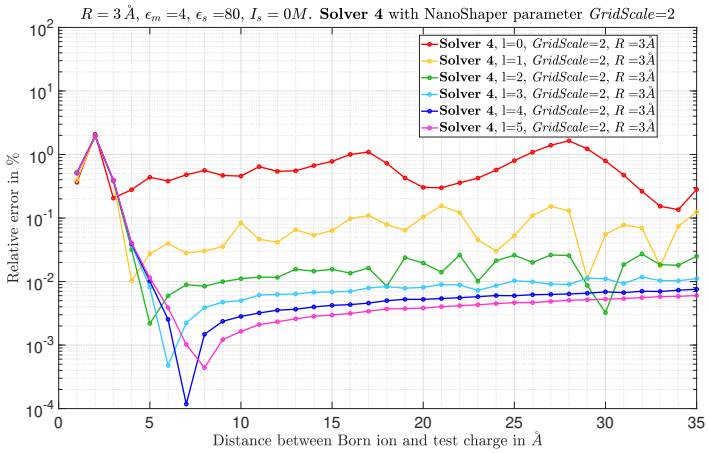
Figure 5.14: Born ion model, $I_s = 0M$. Relative errors of **Solver 4** in the computed electrostatic interaction $E^A_{1-2} = \langle \mathscr{F}_2, z^{(2)}_l \rangle$ at mesh refinement level $l = 5$ versus MEAD relative errors for four configurations.



(a) $R = 1\,\mathring{A}$.

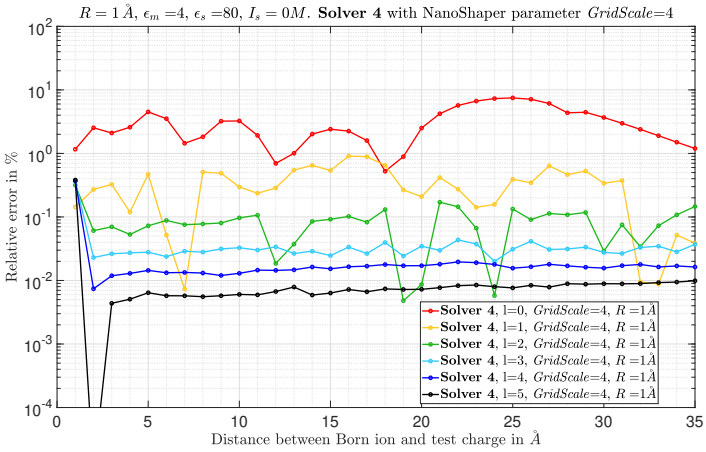


(b) $R = 2\,\mathring{A}$.

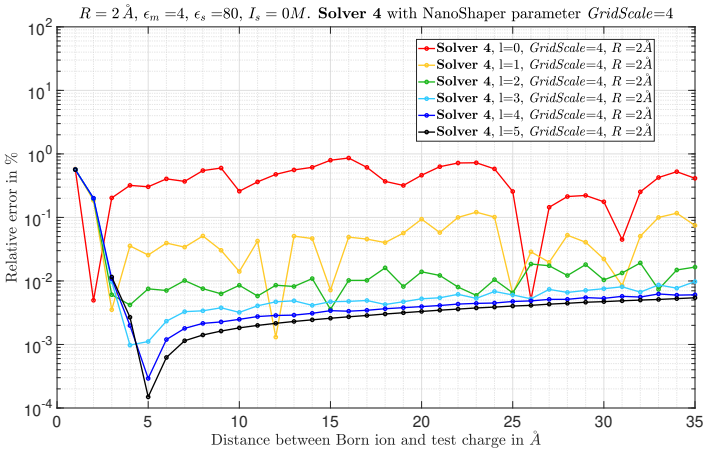


(c) $R = 3\,\mathring{A}$.

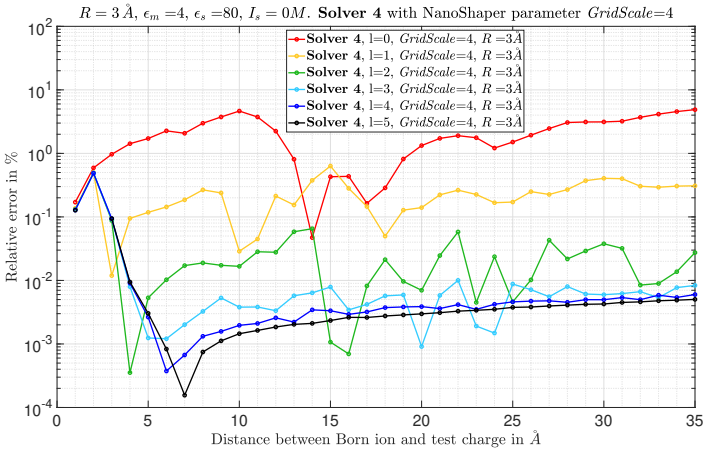
Figure 5.15: Born ion model, $I_s = 0M$. Relative errors of **Solver 4** in the computed electrostatic interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ at mesh refinement levels $l = 0, 1, 2, \ldots, 5$. Born ion sphere triangulated using NanoShaper with parameter *GridScale*=2.



(a) $R = 1$ Å.



(b) $R = 2$ Å.



(c) $R = 3$ Å.

Figure 5.16: Born ion model, $I_s = 0M$. Relative errors of **Solver 4** in the computed electrostatic interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ at mesh refinement levels $l = 0, 1, 2, \ldots, 5$. Born ion sphere triangulated using NanoShaper with parameter *GridScale*=4.

Solver 2

Now we demonstrate the performance of **Solver 2** in the case $\epsilon_m = 4$, $\epsilon_s = 80$ and note that the case of a smaller ratio ϵ_m/ϵ_s is easier to handle. Here, we only present the results for $R = 2\,\mathring{A}$. The computational domain is the same as with **Solver 4**, i.e., a cube with an edge length of approximately $1\,350\,000\,\mathring{A}$ on whose boundary is prescribed a homogeneous Dirichlet boundary condition. We have done computations with two levels of geometric approximation of the surface of the Born ion sphere. Recall that the solution of the primal problem gives the reaction field part u_2 of the full potential $\varphi_2 = G_2 + u_2$:

$$\int_\Omega \epsilon \nabla u_2 \cdot \nabla v dx + \int_\Omega \overline{k}^2 u_2 v dx = \int_{\Omega_s} (\epsilon_m - \epsilon_s) \nabla G_2 \cdot \nabla v dx - \int_\Omega \overline{k}^2 G_2 v dx = \langle \mathcal{T}_2, v \rangle \tag{5.10}$$

for all $v \in H_0^1(\Omega)$, whereas the solution z of the adjoint problem gives the potential φ_1 scaled by $\frac{1}{4\pi}$:

$$\begin{aligned} &\text{Find } z \in \bigcap_{p<\frac{d}{d-1}} W_0^{1,p}(\Omega) \text{ such that} \\ &\int_\Omega \epsilon \nabla v \cdot \nabla z dx + \int_\Omega \overline{k}^2 v z dx = \langle \frac{1}{4\pi} \mathscr{F}_1, v \rangle, \; \forall v \in \bigcup_{q>d} W_0^{1,q}(\Omega). \end{aligned} \tag{5.44}$$

In order to demonstrate more clearly the difference between the refined meshes obtained by applying the 2-term splitting and those obtained without it (compare Figure 5.13 to Figure 5.20), we relabel the charge in the Born ion to be $q_2 = 1\,e_0$ and the test charge to be $q_1 = -1\,e_0$ (see Figure 5.17). In this way, the primal problem defines the reaction field potential u_2 created by the charge q_2 in the Born ion. Recall that based on Remark 5.14, the weak formulations for the primal and adjoint problems are well defined even if all charges are in the solution domain Ω_s and also the representations for the error in the goal quantity, obtained in Section 5.3.1, Section 5.3.2, Section 5.3.3, and Section 5.3.4, ramain valid. Then, as usual, we have

$$G_2 = \frac{q_2}{\epsilon_m |x - x_2|}, \qquad \mathscr{F}_2 = 4\pi q_2 \delta_{x_2}, \qquad \text{and} \qquad \frac{1}{4\pi}\mathscr{F}_1 = q_1 \delta_{x_1},$$

where $x_2 = (0,0,0)$ is the position of q_2 and x_1 is the position of the test charge.

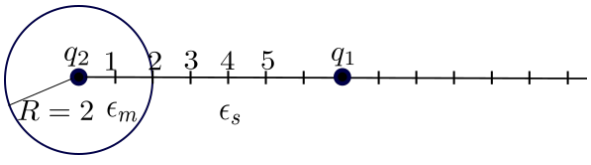
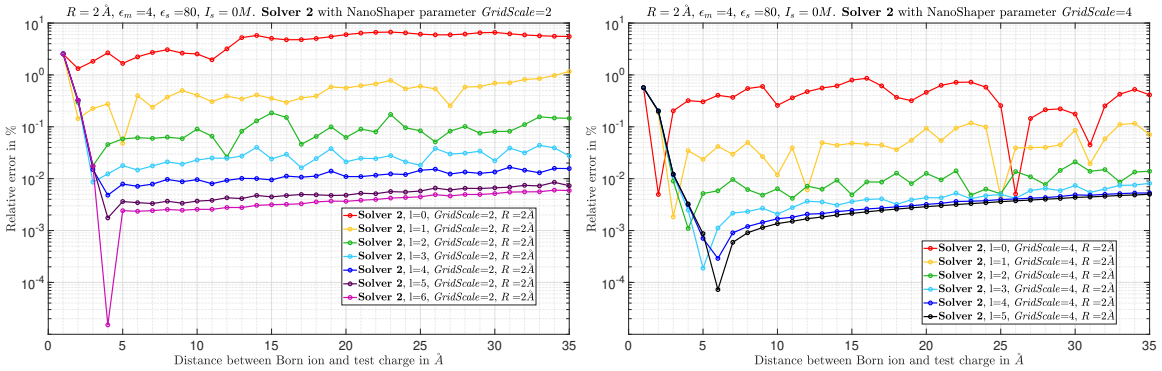


Figure 5.17: Born ion model with $I_s = 0$.

On Figure 5.18 we present the convergence of the relative error in the quantity $\langle \mathscr{F}_2, z_l^{(2)} \rangle$ (the approximate value for the interaction that we use in practice) for distances from 1 $\mathring{A}$

to 35 $\mathring{A}$ from the charge q_2. It is seen that for *GridScale*=2 the relative error at the last MRL is smaller than the relative error with *GridScale*=2 at the last MRL for **Solver 4** (compare with Figure 5.15b)). This might be due to the fact that with **Solver 2** only one singularity has to be approximated, the one in the solution z of the adjoint problem, whereas with **Solver 4** both solutions of the primal and adjoint problem have singularities which have to be approximated. As a result, the distribution of the nodes in the mesh $\mathscr{T}_h$ has to be balanced between both singularities (compare Figure 5.20 to Figure 5.13). We should note that the convergence of the error slows down after several MRL due to the geometric approximation of the Born ion sphere and of the boundary conditions.



(a) *GridScale*=2. (b) *GridScale*=4.

Figure 5.18: Born ion model, $I_s = 0M$. Relative errors of **Solver 2** in the computed electrostatic interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ at all mesh refinement levels. Born ion sphere triangulated using NanoShaper with parameter *GridScale*=2 and *GridScale*=4.

On Table 5.8 we present the number of DOFs and relative errors for a distance of 15 $\mathring{A}$ between the charges q_2 and q_1 when *GridScale*=2. From (5.131) we know that

$$E_{u_2} = \frac{q_1 q_2(\epsilon_m - \epsilon_s)}{\epsilon_m \epsilon_s r}, \qquad E_{G_2} = \frac{q_1 q_2}{\epsilon_m r}, \qquad E_{2-1} = \frac{q_1 q_2}{\epsilon_s r}$$

Since in this example the ratio $\epsilon_m/\epsilon_s = 1/20 \ll 1$, we see that $E_{u_2}/E_{2-1} = \frac{\epsilon_m - \epsilon_s}{\epsilon_m} = -19$. This means that the relative error in the full electrostatic interaction E^P_{2-1} is a factor of 19 larger than the relative error in the approximation $u_{2,h}$ of the reaction field potential u_2 (see columns 3 and 4 in Table 5.8). If the electrostatic interaction E_{u_2} is needed with high accuracy, then as an approximation to E_{u_2} one should use

$$E_{u_2} \approx \langle \mathscr{F}_2, z_h^{(2)} \rangle - E_{G_2} = E^A_{1-2} - E_{G_2}. \tag{5.132}$$

The relative errors $\frac{|E_{u_2} - (E^A_{1-2} - E_{G_2})|}{|E_{u_2}|}$ in percents are given in Table 5.9. Now, since the quantity $E_{2-1} = \langle \mathscr{F}_2, z \rangle$ is 19 times smaller than E_{u_2} in absolute value, the relative error in the approximation $E^A_{1-2} - E_{G_2}$ of E_{u_2} is 19 times smaller than the relative error in E^A_{1-2}. In other words, with **Solver 2** we can obtain approximations with very high accuracy for both

the full electrostatic interaction $E_{2-1} = E_{1-2}$ and E_{u_2}, even if $\epsilon_m/\epsilon_s \ll 1$. Note that the same trick can be used with **Solver 4**, i.e., we can subtract E_{G_2} from E^A_{1-2} to obtain a good approximation of E_{u_2}.

Table 5.8: **Solver 2.** *GridScale*=2. Born ion model, $R = 2\,\mathring{A}$, $I_s = 0M$, $\epsilon_m = 4$, $\epsilon_s = 80$. Distance between q_2 and q_1 is 15 $\mathring{A}$. Number of DOFs for primal and adjoint problems as well as relative errors in the quantities $\langle \frac{1}{4\pi}\mathscr{F}_1, u_{2,l}\rangle$, E^A_{1-2}, and E^P_{2-1}, where $E^A_{1-2} = \langle \mathscr{F}_2, z^{(2)}_l\rangle$ and $E^P_{2-1} = E_{G_2} + \langle \frac{1}{4\pi}\mathscr{F}_1, u_{2,l}\rangle$. Exact values: $E_{u_2} = 0.01583333333\, e_0^2\mathring{A}^{-1}$, $E_{G_2} = -0.01666666667\, e_0^2\mathring{A}^{-1}$, $E_{2-1} = -0.00083333333\, e_0^2\mathring{A}^{-1}$.

level	# DOFs primal	$\frac{\lvert E_{u_2} - \langle \frac{1}{4\pi}\mathscr{F}_1, u_{2,l}\rangle\rvert}{\lvert E_{u_2}\rvert}$ [%]	$\frac{\lvert E_{2-1} - E^P_{2-1}\rvert}{\lvert E_{2-1}\rvert}$ [%]	# DOFs adjoint	$\frac{\lvert E_{2-1} - E^A_{1-2}\rvert}{\lvert E_{2-1}\rvert}$ [%]
0	2 437	33.1168	629.220	18 201	5.06380
1	3 083	11.1502	211.855	23 261	0.35202
2	5 522	3.29127	62.5342	42 623	0.18523
3	10 075	1.89948	36.0901	78 728	0.02398
4	21 522	1.08123	20.5434	169 752	0.00947
5	46 644	0.65646	12.4728	369 433	0.00445
6	97 881	0.39601	7.52423	775 895	0.00314
7	209 838	0.23719	4.50663	1 665 167	0.00254

Table 5.9: **Solver 2.** *GridScale*=2. Born ion model, $R = 2\,\mathring{A}$, $I_s = 0M$, $\epsilon_m = 4$, $\epsilon_s = 80$. Distance between q_2 and q_1 is 15 $\mathring{A}$. Relative error in the approximation $E^A_{1-2} - E_{G_2}$ of the quantity E_{u_2} at all MRLs.

level	# DOFs adjoint	$\frac{\lvert E_{u_2} - \left(E^A_{1-2} - E_{G_2}\right)\rvert}{\lvert E_{u_2}\rvert}$ [%]
0	18 201	0.26651
1	23 261	0.01852
2	42 623	0.00974
3	78 728	0.00126
4	169 752	0.00049
5	369 433	0.00023
6	775 895	0.00016
7	1 665 167	0.00013

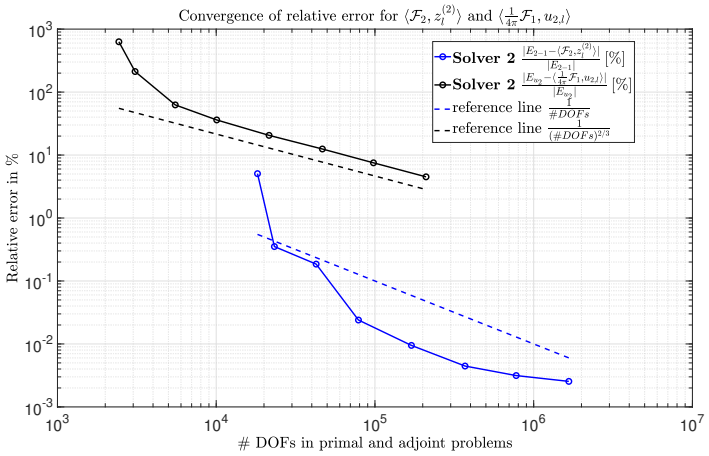
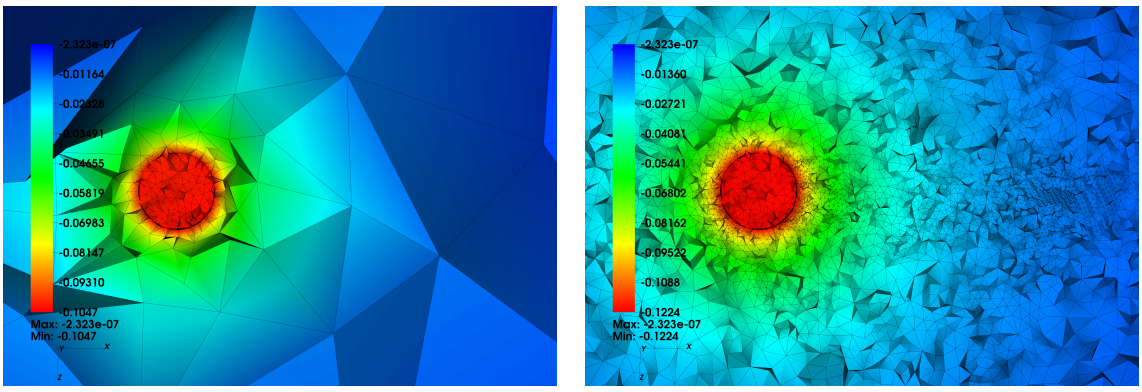


Figure 5.19: **Solver 2.** *GridScale*=2. Born ion model, $R = 2\,\mathring{A}$, $I_s = 0M$, $\epsilon_m = 4$, $\epsilon_s = 80$. Distance between q_2 and q_1 is $15\,\mathring{A}$. Convergence of relative error in the quantities $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ and $\frac{1}{4\pi}\langle \mathscr{F}_1, u_{2,l} \rangle$. Data from Table 5.8. Note that the error in $\langle \mathscr{F}_2, z_l^{(2)} \rangle$ stagnates at around 10^6 DOFs due to the error in the geometric approximation of the Born ion sphere and approximate boundary condition on $\partial\Omega$.



(a) Initial mesh, $l = 0$.

(b) Final mesh, $l = 7$.

Figure 5.20: **Solver 2.** *GridScale*=2. Born ion model, $R = 2\,\mathring{A}$, $I_s = 0M$, $\epsilon_m = 4$, $\epsilon_s = 80$. Distance between q_2 and q_1 is $15\,\mathring{A}$. Reaction field potential $u_{2,l}$ (obtained by solving the primal problem) at initial and final meshes in units $e_0/\mathring{A}$. Pictures generated with VisIt [52]

5.4.3 Born ion model with $I_s > 0$ and an ion exclusion layer

Solver 4

In this test we check the performance of **Solver 4** in the presence of sharp interface, an ion exclusion layer, and positive ionic strength I_s. We again compare our results to MEAD, as a representative of a finite difference solver with uniform Cartesian grids. The set up is as follows: an ion, with a Born radius R equal to 1, 2, and 3 $\mathring{A}$ with $\epsilon_m = 4$ and charge $q_1 = 1\, e_0$, is placed in a dielectric medium with $\epsilon_s = 80$ and ionic strength $I_s = 0.3\, M$, corresponding to $\overline{k}^2_{ions} = 2.52916\, \mathring{A}^{-2}$ at $T = 300$ K. Additionally, an IEL with a thickness of a probe ion's radius $R_{ion} = 2\, \mathring{A}$ is added around the Born ion (see Figure 5.21). We compute the electrostatic interaction between the ion and a test charge $-1\, e_0$ placed at distances varying from $1\, \mathring{A}$ to $35\, \mathring{A}$. In other words, we compute the potential at the coordinates of the test charge and multiply it by $-1\, e_0$. In this case, it is easy to check that the electrostatic potential φ is given by

$$\varphi(r) = \begin{cases} \dfrac{q_1}{\epsilon_m r} + \dfrac{q_1\,(\epsilon_m - \epsilon_s)}{\epsilon_m \epsilon_s R} - \dfrac{q_1 \frac{\overline{k}}{\sqrt{\epsilon_s}}}{\left(1 + a\frac{\overline{k}}{\sqrt{\epsilon_s}}\right)\epsilon_s}, & r \le R, \\ \dfrac{q_1}{\epsilon_s r} - \dfrac{q_1 \frac{\overline{k}}{\sqrt{\epsilon_s}}}{\left(1 + a\frac{\overline{k}}{\sqrt{\epsilon_s}}\right)\epsilon_s}, & R \le r \le a, \\ \dfrac{q_1 \exp\left(a\frac{\overline{k}}{\sqrt{\epsilon_s}}\right)}{\left(1 + a\frac{\overline{k}}{\sqrt{\epsilon_s}}\right)\epsilon_s} \dfrac{\exp\left(-r\frac{\overline{k}}{\sqrt{\epsilon_s}}\right)}{r}, & r \ge a, \end{cases} \tag{5.133}$$

where r is the distance to the charge q_1 at the center of the Born ion and $a = R + R_{ion}$. The computational domain in **Solver 4** is a cube with an edge length of approximately

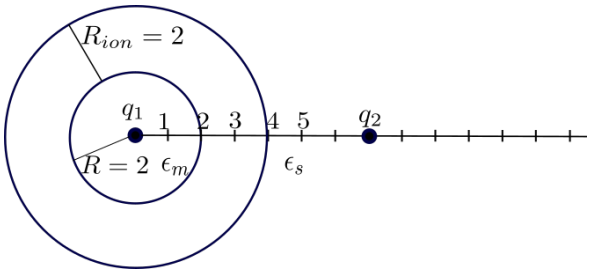


Figure 5.21: Born ion model with ion exclusion layer and $I_s > 0$.

$240\,000\, \mathring{A}$ and a homogeneous Dirichlet boundary condition is prescribed at its boundary. We have again performed all computations with two levels of geometric approximation of the surface of the Born ion sphere and of the IEL surface: with *GridScale*=2 and *GridScale*=4.

On Figure 5.22 are shown the relative errors for $R = 1,\, 2,\, 3\, \mathring{A}$ for both **Solver 4** and MEAD with four different configurations. Again, as in the previous test, each value for the relative

error in MEAD is an average over 10 different directions. This time the default stopping criteria in the SOR method in MEAD is more or less adequate and the iterative error is not higher than the discretization error, except in the regions where the distance to the test charge is in the ranges $(R+1)$ to $15\,\mathring{A}$ and $(R+1)$ to $20\,\mathring{A}$ for the grids ON_ORIGIN 97-3, 161-1, 161-0.25 and ON_ORIGIN 285-1, 285-0.5, 285-0.1, respectively. Notice that the last focused grid of the first grid configuration in MEAD extends exactly to a distance of 20 $\mathring{A}$ from the origin in each coordinate direction and the last focused grid for the second grid configuration extends to 14.2 $\mathring{A}$. The error at the interface Γ with the first (coarser) grid configuration in MEAD reaches 70 % for $R = 1\,\mathring{A}$ and drops to less than 30 % for $R = 3\,\mathring{A}$. Decreasing the grid spacing to $0.1\,\mathring{A}$ in the last focused grid in the second grid configuration reduces the relative error at the Born ion's surface around 3 times to values of 25 % at $R = 1\,\mathring{A}$ and 9 % at $R = 3\,\mathring{A}$. For comparison, the relative error with **Solver 4** and *GridScale*=2 does not exceed 2.23 % in the region around the interface Γ for all $R = 1, 2, 3\,\mathring{A}$, and it drops 4 times when using surface meshes with 2 times smaller average edge length (*GridScale*=4). The relative errors for **Solver 4** with *GridScale*=2 stay at a constant level less than 0.2 % everywhere and for all $R = 1, 2, 3\,\mathring{A}$. The improved geometric approximation with *GridScale*=4 results in 4 times smaller errors for all $R = 1, 2, 3\,\mathring{A}$ and all distances of the test charge to the Born ion's center. Here we note that the accuracy of **Solver 4** seems to be limited only by the geometric quality of the Born ion sphere and almost not influenced by the distance between the Born ion and the boundary $\partial\Omega$, where the homogeneous Dirichlet boundary condition is prescribed. Recall that *GridScale*=4 means that in the construction of triangulated surface mesh in NanoShaper is used a cubic grid with a spacing $h = 0.25\,\mathring{A}$. It is clear that the geometric approximation of the molecular surface with *GridScale*=4 corresponds to a uniform finite difference grid with spacing $h = 0.25\,\mathring{A}$ and hence the results obtained with MEAD for the first (coarser) mesh should be compared to the results obtained with **Solver 4** for *GridScale*=4 (see Figure 5.22).

Further, on Figure 5.23 and Figure 5.24 is shown the convergence of the relative errors at all mesh refinement levels $l = 0, 1, \ldots, 5$ for *GridScale*=2 and *GridScale*=4, respectively. Additionally, in Table 5.10 and Table 5.11, we present the average number of DOFs in the adjoint problem for $R = 1, 2, 3\,\mathring{A}$ and for each refinement level with *GridScale*=2 and *GridScale*=4, respectively. The average is taken over all distances from $5\,\mathring{A}$ to $35\,\mathring{A}$ from the Born ion's center. At each of these distances, the corresponding number of DOFs is close to the average.

Remark 5.17

Similarly to Remark 5.16, recall that the primal problem gives the potential φ_2 created by the charges of the second molecule. As it is the case in the previous example, there is no second molecule, only a test charge which is chosen to be the one that creates the potential φ_2. Then, the right-hand side of the primal problem is given by

$$\mathscr{F}_2 = 4\pi q_2 \delta_{x_2},$$

where x_2 is the position of the test charge q_2. The right hand side of the adjoint problem is given by

$$\frac{1}{4\pi}\mathscr{F}_1 = q_1 \delta_{x_1},$$

where $x_1 = (0,0,0)$ is the position of the charge q_1 in the Born ion. The solution z of the adjoint problem is equal to $\frac{1}{4\pi}\varphi_1$ and the corrected approximation for the electrostatic interaction that we use is $\langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_{2,h}, \rangle + E(\varphi_{2,h}, z_h^{(2)}) = \langle \mathscr{F}_2, z_h^{(2)} \rangle$.

Table 5.10: **Solver 4.** *GridScale*=2. Born ion model with IEL, $I_s = 0.3M$, $\overline{k}_{ions} = \sqrt{2.52916}\,\mathring{A}^{-1}$, $\epsilon_m = 4$, $\epsilon_s = 80$. Averages of number of DOFs and errors for adjoint problem for distances from $5\,\mathring{A}$ to $35\,\mathring{A}$. Almost uniform initial mesh size inside the Born ion sphere and IEL.

	$R = 1\,\mathring{A}$		$R = 2\,\mathring{A}$		$R = 3\,\mathring{A}$	
level	aver. # DOFs adjoint	aver. error [%]	aver. # DOFs adjoint	aver. error [%]	aver. # DOFs adjoint	aver. error [%]
0	67 337	10.6634	128 199	1.47893	198 495	3.09911
1	78 786	1.46270	151 767	0.20302	229 093	0.44182
2	101 142	0.23882	208 175	0.18200	290 703	0.17553
3	153 626	0.19039	328 007	0.18174	435 816	0.16043
4	259 281	0.19348	561 477	0.18009	711 183	0.15786
5	447 791	0.19416	975 413	0.17963	1 228 265	0.15936

Solver 2

Here, we demonstrate the performance of **Solver 2** in the case of $I_s > 0$ and a relatively high jump in the dielectric coefficient ϵ across the interface Γ. The set up is the same as before: $I_s = 0.3M$, $\epsilon_m = 4$, $\epsilon_s = 80$, $T = 300$ K, $\overline{k}_{ions} = \sqrt{2.52916}\,\mathring{A}^{-1}$ and we only present the results for $R = 2\,\mathring{A}$. We again relabel the charges and denote by $q_2 = 1\,e_0$ the charge in the Born ion, and by $q_1 = -1\,e_0$ the test charge. The primal problem defines the reaction field potential u_2 due to the charge q_2 in the Born ion, whereas the adjoint problem defines the scaled by $\frac{1}{4\pi}$ potential created by the test charge q_1. On Figure 5.25 is shown the convergence of the relative errors in the quantity E^A_{1-2} for all mesh refinement levels with *GridScale*=2 and *GridScale*=4. Decreasing two times the average edge length on the triangulated Born ion and IEL surfaces, improves the geometric approximation and results in a decrease of

Table 5.11: **Solver 4.** *GridScale*=4. Born ion model with IEL, $I_s = 0.3M$, $\overline{k}_{ions} = \sqrt{2.52916}\mathring{A}^{-1}$, $\epsilon_m = 4$, $\epsilon_s = 80$. Averages of number of DOFs and errors for adjoint problem for distances from $5\,\mathring{A}$ to $35\,\mathring{A}$. Almost uniform initial mesh size inside the Born ion sphere and IEL.

	$R = 1\,\mathring{A}$		$R = 2\,\mathring{A}$		$R = 3\,\mathring{A}$	
level	aver. # DOFs adjoint	aver. error [%]	aver. # DOFs adjoint	aver. error [%]	aver. # DOFs adjoint	aver. error [%]
0	92 813	1.62254	199 442	3.16998	316 748	1.28987
1	108 763	0.34620	224 197	0.38920	355 793	0.14383
2	141 467	0.08001	277 619	0.07653	445 032	0.04494
3	214 166	0.04943	406 363	0.04475	670 251	0.03946
4	356 606	0.05000	662 737	0.04484	1 037 000	0.04061
5	631 951	0.04973	1 133 363	0.04482	1 744 445	0.04050

the error by a factor of 4 at the last refinement level, where the maximum accuracy for the respective geometry approximation is obtained.

This time, since $I_s > 0$, the screening of the electrostatic potential is even stronger than in the case $I_s = 0$. As a result, $E_{u_2}/E_{2-1} = 240.964336395$ and a relative error of 1 % in the approximation $\langle \frac{1}{4\pi}\mathscr{F}_1, u_{2,l}\rangle$ of E_{u_2} becomes a realtive error of 240.964336395 % in the quantity $E_{G_2} + \langle \frac{1}{4\pi}\mathscr{F}_1, u_{2,l}\rangle$ approximating the full interaction E_{2-1} (see Table 5.12). If the interaction E_{u_2} is needed, then one should use the approximation given by (5.132). This time a relative error of 1 % in the quantity $E^A_{1-2} = \langle \mathscr{F}_2, z^{(2)}_l\rangle$ results in a relative error of $1/240.964336395\,\%$ in the approximation $E^A_{1-2} - E_{G_2}$ to E_{u_2} (see Table 5.13).

Table 5.12: **Solver 2**, *GridScale*=4. Born ion model, $R = 2\,\mathring{A}$, $I_s = 0.3M$, $\epsilon_m = 4$, $\epsilon_s = 80$. Distance between q_2 and q_1 is 15 $\mathring{A}$. Number of DOFs for primal and adjoint problems as well as relative errors in the quantities $\langle \frac{1}{4\pi}\mathscr{F}_1, u_{2,l}\rangle$, E^A_{1-2}, and E^P_{2-1}, where $E^A_{1-2} = \langle \mathscr{F}_2, z^{(2)}_l\rangle$ and $E^P_{2-1} = E_{G_2} + \langle \frac{1}{4\pi}\mathscr{F}_1, u_{2,l}\rangle$. Exact values: $E_{u_2} = 0.01659778599\, e_0^2\mathring{A}^{-1}$, $E_{G_2} = -0.01666666667\, e_0^2\mathring{A}^{-1}$, $E_{2-1} = -6.888067437 \times 10^{-5}\, e_0^2\mathring{A}^{-1}$.

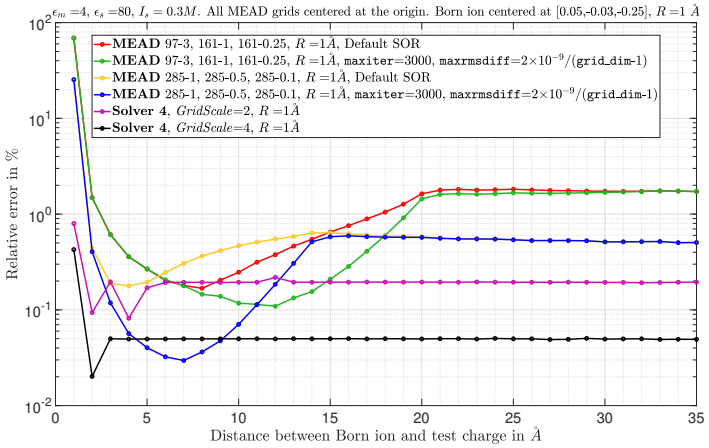
level	# DOFs primal	$\frac{\lvert E_{u_2} - \langle \frac{1}{4\pi}\mathscr{F}_1, u_{2,l}\rangle\rvert}{\lvert E_{u_2}\rvert}$ [%]	$\frac{\lvert E_{2-1} - E^P_{2-1}\rvert}{\lvert E_{2-1}\rvert}$ [%]	# DOFs adjoint	$\frac{\lvert E_{2-1} - E^A_{1-2}\rvert}{\lvert E_{2-1}\rvert}$ [%]
0	25 398	1.34194	323.360	199 442	0.07305
1	28 094	0.25676	61.8705	220 717	0.19776
2	34 085	0.03821	9.20855	268 172	0.09099
3	49 764	0.01674	4.03422	392 598	0.04373
4	84 578	0.00493	1.18946	669 104	0.04410
5	148 492	0.00746	1.79961	1 176 079	0.04339
6	287 232	0.00620	1.49610	2 276 800	0.04337
7	572 645	0.00378	0.91316	4 542 599	0.04336
8	1 115 642	0.00247	0.59569	8 845 074	0.04336

Table 5.13: **Solver 2**, *GridScale*=4. Born ion model, $R = 2\,\mathring{A}$, $I_s = 0.3M$, $\epsilon_m = 4$, $\epsilon_s = 80$. Distance between q_2 and q_1 is 15 $\mathring{A}$. Relative error in the approximation $E^A_{1-2} - E_{G_2}$ of the quantity E_{u_2} at all MRLs.

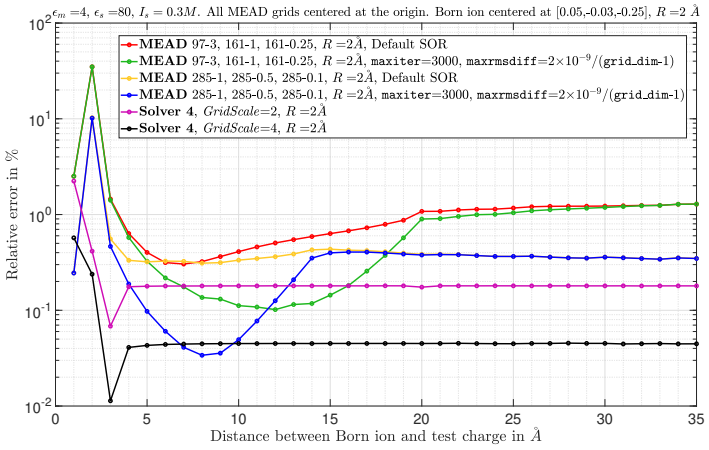
level	# DOFs adjoint	$\frac{\lvert E_{u_2} - \left(E^A_{1-2} - E_{G_2}\right)\rvert}{\lvert E_{u_2}\rvert}$ [%]
0	199 442	0.00030312
1	220 717	0.00082073
2	268 172	0.00037766
3	392 598	0.00018152
4	669 104	0.00018305
5	1 176 079	0.00018012
6	2 276 800	0.00018004
7	4 542 599	0.00018001
8	8 845 074	0.00018000

Remark 5.18

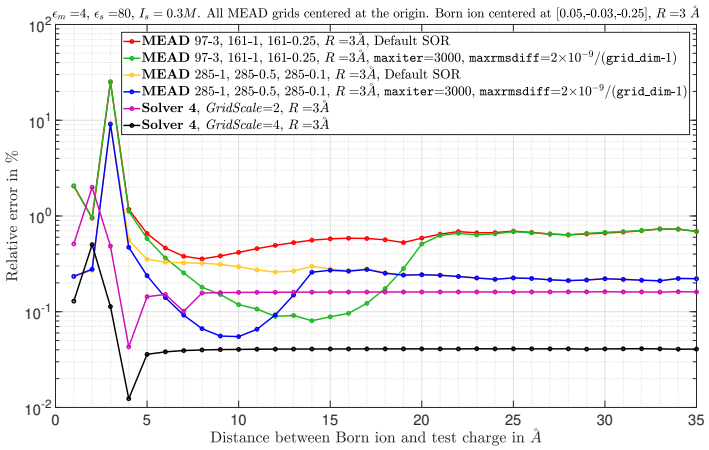
Note that if the electrostatic potential is needed at one point of interest x_0 that is not necessarily in the molecular region Ω_m, one can just place a unit fictitious (test) charge at this position and apply one of the error estimates developed in this chapter to compute a fictitious electrostatic interaction. The mesh will be refined in such a way that the electrostatic interaction, which is the same as the potential at x_0, is computed with high accuracy.



(a) $R = 1\,\mathring{A}$.

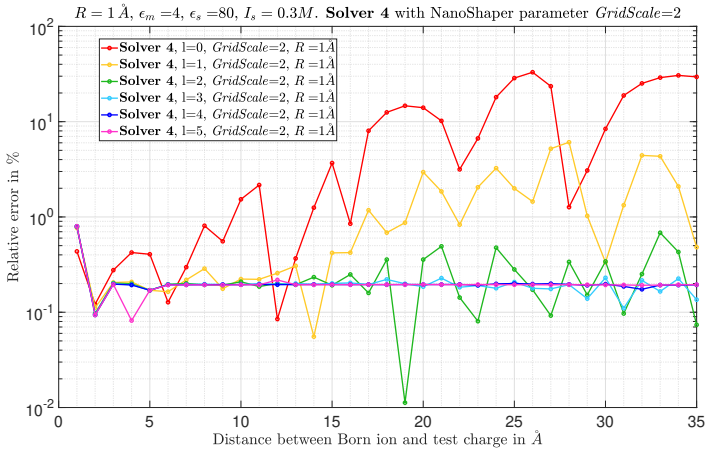


(b) $R = 2\,\mathring{A}$.

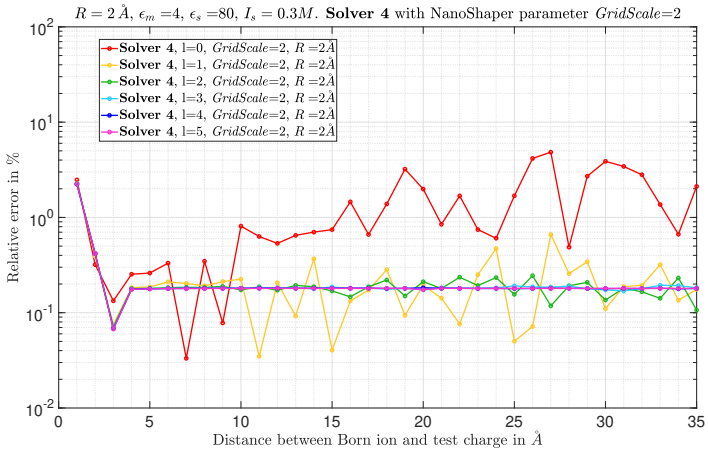


(c) $R = 3\,\mathring{A}$.

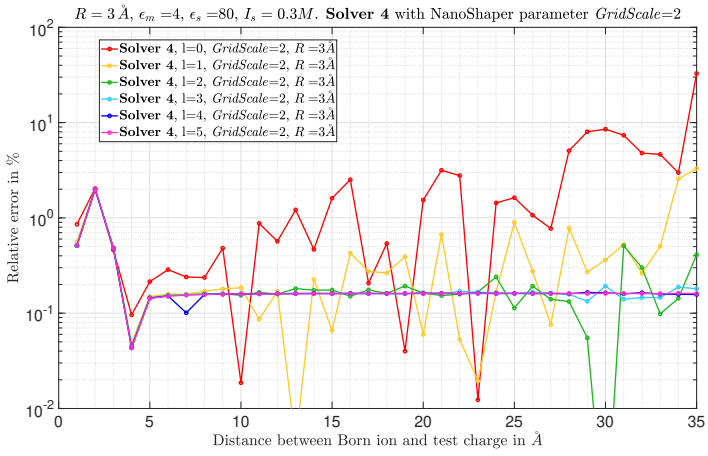
Figure 5.22: Born ion model with IEL, $I_s = 0.3M$. Relative errors of **Solver 4** in the computed electrostatic interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ at mesh refinement level $l = 5$ versus MEAD relative errors for four configurations.



(a) $R = 1\,\mathring{A}$.

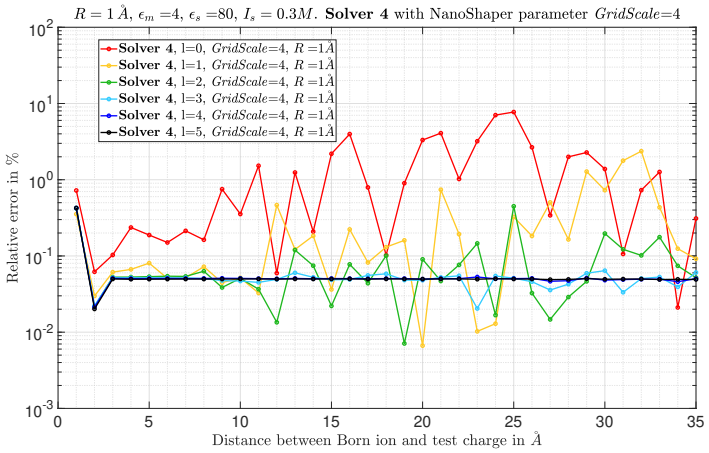


(b) $R = 2\,\mathring{A}$.

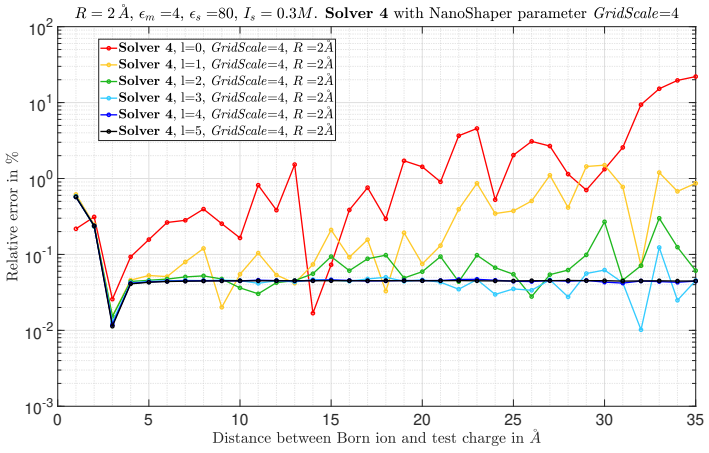


(c) $R = 3\,\mathring{A}$.

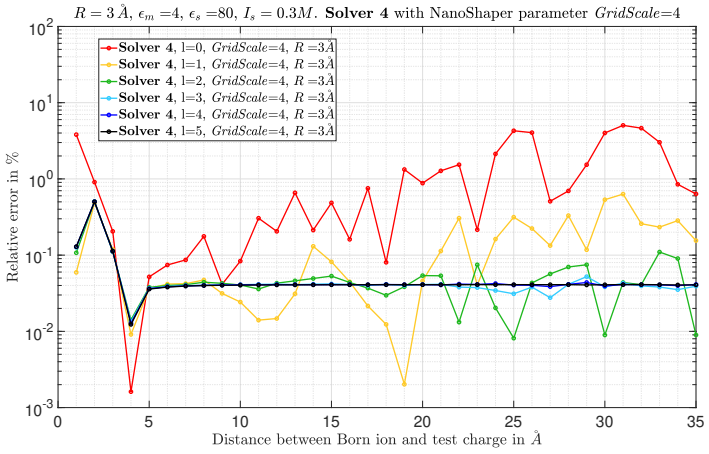
Figure 5.23: **Solver 4**. Born ion model with IEL, $I_s = 0.3M$. Relative errors in the computed electrostatic interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ at mesh refinement levels $l = 0, 1, 2, \ldots, 5$. Born ion sphere triangulated using NanoShaper with parameter *GridScale*=2.



(a) $R = 1\,\mathring{A}$.

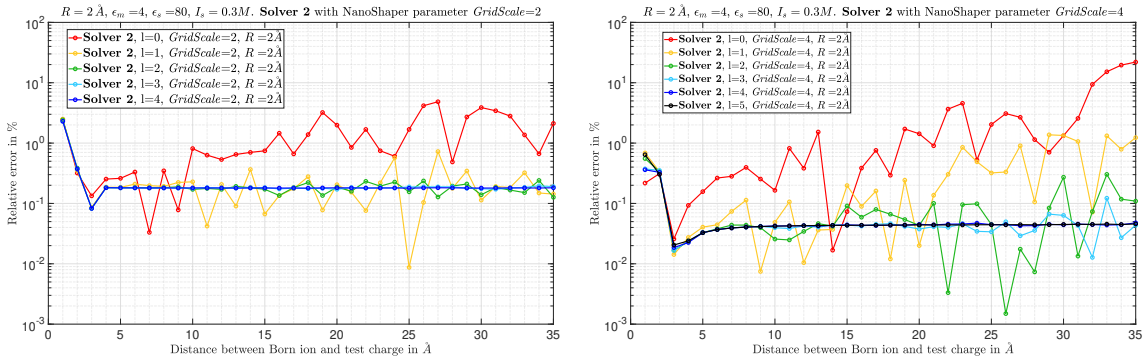


(b) $R = 2\,\mathring{A}$.



(c) $R = 3\,\mathring{A}$.

Figure 5.24: **Solver 4**. Born ion model with IEL, $I_s = 0.3M$. Relative errors in the computed electrostatic interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ at mesh refinement levels $l = 0, 1, 2, \ldots, 5$. Born ion sphere triangulated using NanoShaper with parameter *GridScale*=4.



(a) *GridScale*=2. (b) *GridScale*=4.

Figure 5.25: **Solver 2**. Born ion model with IEL, $R = 2\,\mathring{A}$, $I_s = 0.3M$, $\epsilon_m = 4$, $\epsilon_s = 80$. Relative errors in the computed electrostatic interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ at all mesh refinement levels.

5.5 Applications

We will consider electrostatic interactions between the chromophores Alexa 488 and Alexa 594, attached to the all-trans polyproline helices of 6 and 20 residues (POLY6 and POLY20), placed in aqueous solution with 300 mM NaCl (see Figure 5.26). This system allows the introduction of different types of interactions that provide us with an opportunity to compare the performance of the solvers on different physical models. We compute the coupling between the dyes for the time frames of all-atom molecular dynamics (MD) simulations, performed with NAMD package (for more details, see [172]). For each polyproline length, 10 trajectories are generated, each with a time window of 200 ns. Standard CHARMM force field (*version 35b3*) is used for the polyproline. The force field parameters of the Alexa chromophores are created by an analogy approach from that of similar chemical groups in the CHARMM force field. The water molecules in the MD simulations are parameterized, using a TIP3P model.

The calculation of the interaction follows the usual procedure, starting with computing of electrostatic potential from the Poisson or lineraized Poisson-Boltzmann equations. The obtained potential is then used to calculate the electrostatic coupling $E_{2-1} = E_{1-2}$. We present two different physical models where the electrostatic interaction is required. The first one is related to FRET and the second one describes the interaction energy between chromophores in their ground state.

Obviously, the interaction energy also depends on the definition of the solute surface (see, e.g [5]). However, the study of this problem is beyond the scope of this thesis. For all experiments that we present in the rest of this chapter, the molecular surface of the chromophores is defined as the solvent excluded surface (SES), formed by the contact points of the Van der Waals surface and a solvent probe sphere that is rolled over it (see [94, 158, 165]). In 1983 Connolly gave an analytic description of the solvent excluded surface and therefore it is also known as Connolly surface (see [53] for the piecewise analytic definition of this surface). The surface meshes of the two dyes are generated with NanoShaper [61] using *GridScale*=2 and a probe sphere radius of 1 $\mathring{A}$. The initial volume meshes are generated with TetGen [170] and after this they are adaptively refined with mmg3d [62]. We compare our results to the results obtained with MEAD in which the molecular surface is also represented by the SES. Thus, in all MEAD calculations, we use the same probe sphere radius of 1 $\mathring{A}$.

5.5.1 Application to FRET

The first model is related to the Fröster resonance energy transfer (FRET) between the two dyes. FRET is an important mechanism for the estimation of intermolecular distances in fluorescent labeled proteins. In experiments, FRET is measured between a donor and an acceptor dye, linked to a protein. A nonradiative relaxation process transmits the electronic

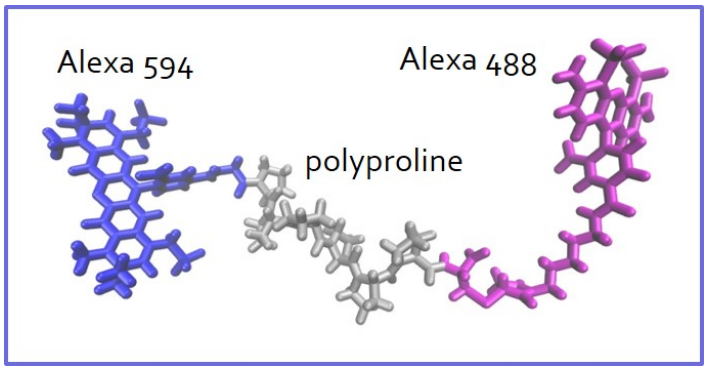


Figure 5.26: Alexa chromophores attached to the all-trans polyproline helice of 6 residues.

excitation from the excited donor to the acceptor chromophore, which is initially in the ground state. In FRET only water (and ions) electronic polarization is involved in the screening of the chromophores' interaction. This determines an optical dielectric constant of the medium ϵ_s equal to 2. The efficiency of the transfer depends on the electrostatic coupling between the dyes, which can be calculated with the Poisson equation (see [172]). Atomic transition charges are obtained by fitting the ab initio electrostatic potential of the transition density on a 3D grid around the chromophores using the program CHELP-BOW (for more details, see [172]). These charges are the same as the ones used in the first example with uniform dielectric ϵ equal to 1. Both dyes are electroneutral and the number of nonzero charges in Dye I and Dye II is $N_1 = 74$ and $N_2 = 52$, respectively. Here we present the results for 120 frames from the MD simulation for POLY6. The computational domain Ω is a cube with an edge length between 130 000 $\mathring{A}$ and 250 000 $\mathring{A}$ depending on the dimension of the smallest box that contains both dyes. Again, a homogeneous Dirichlet boundary condition is prescribed on $\partial\Omega$.

On Figure 5.27 is shown the relative distance between the coupling $\overline{E}_M$, obtained with MEAD with three different configurations, and the coupling $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$, obtained with **Solver 2** using the adjoint solution $z_l^{(2)}$ at MRL $l = 5$. Additionally, in Table 5.14 are given the avergae (over all 120 frames) number of DOFs in the primal and adjoint problems. Recall that since the primal problem is solved with the finite element method for the Lagrange P_1 element, the number of DOFs in the primal problem for each MRL l is equal to the number of vertices in the mesh $\mathscr{T}_l$. Similarly to the case of uniform dielectric coefficient $\epsilon = 1$, one to two refinements of the mesh are enough for convergence (see Table 5.1 on p. 225).

In the first two configurations, grid 1 and grid 2, MEAD is run with its default stopping criteria in the SOR method (see Figure 5.27). The average relative distance for grid 1 in MEAD is 2.229 %, whereas with grid 2 it drops to 1.227 % despite the fact that all focused grids have the same spacings $h_1 = 3\,\mathring{A}$, $h_2 = 1\,\mathring{A}$, $h_3 = 0.25\,\mathring{A}$. The only difference between the grids is the region they span. From Figure 5.27 it is seen that the drop in the average

distance is mostly due to the frame with number 28, where the relative distance decreases from 60 % to 2 %. A similar drop in the relative error between $\overline{E}_M$ and the exact coupling $E_{2-1} = E_{1-2} = E_{G_2} = E_{G_1}$ in the case of uniform dielectric $\epsilon = 1$ is observed when the span of the focused grids is increased (see Figure 5.4 for $\epsilon = 1$ with grid 1 and grid 2). In the case of different dielectric coefficient in Ω_m and Ω_s, the increase of the span of the coarsest grid also improves the accuracy in the approximate boundary condition used by MEAD.

On Figure 5.27, we also show the relative distance between the interaction computed with **Solver 2** and the interaction computed with MEAD using grid 2 but with improved stopping criteria in SOR: `maxrmsdiff` $= 2 \times 10^{-9}/($`grid_dim` $- 1)$ and `maxiter`=3000.

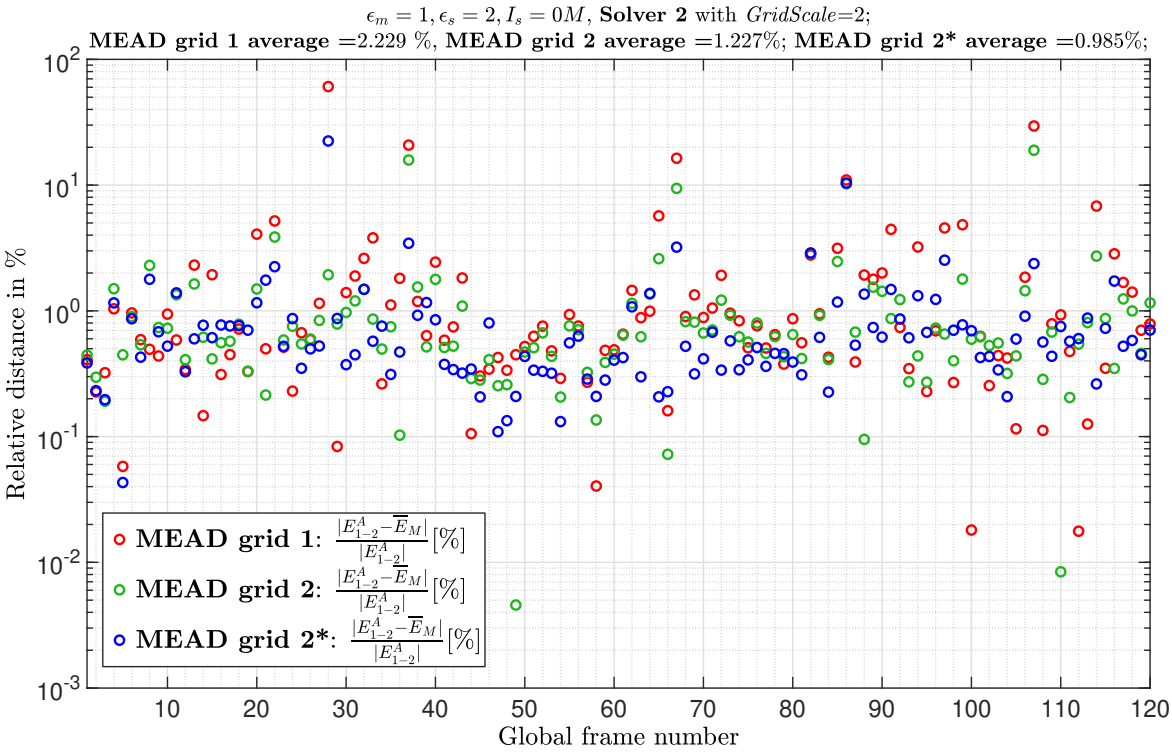


Figure 5.27: FRET application. Neutral dyes with transition state atomic charges. Relative distance $\frac{|E^A_{1-2} - \overline{E}_M|}{|E^A_{1-2}|}$ between MEAD coupling $\overline{E}_M$ and the coupling $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ obtained with **Solver 2** at MRL $l = 5$ with *GridScale*=2. Grid 1 in MEAD with specifications ON_GEOM_CENT 97-3, 161-1, 161-0.25 and grid 2 with specifications ON_GEOM_CENT 285-3, 285-1, 285-0.25. Grid 2* is the same as grid 2, but the stopping criteria in SOR is improved: `maxrmsdiff` $= 2 \times 10^{-9}/($`grid_dim` $- 1)$ and `maxiter`=3000.

Table 5.14: **Solver 2**, *GridScale*=2, $\epsilon_m = 1$, $\epsilon_s = 2$, $I_s = 0$. Neutral dyes with transition state atomic charges. Average (over all 120 frames) number of DOFs in the primal and adjoint problems for each MRL $l = 0, 1, \ldots, 5$. All initial meshes have a uniform mesh size in the molecular region, related to the average edge length on the molecular surface Γ.

level	aver. # DOFs primal	aver. # DOFs adjoint
0	54 156	429 196
1	74 124	586 695
2	106 702	842 292
3	157 736	1 242 095
4	242 896	1 910 434
5	391 030	3 074 173

5.5.2 Application to interaction between chromophores in ground state

The second model that we consider describes interactions between chromophores, being in their ground state and in equilibrium with the solvent. In this model the induced polarization of water molecules is caused by both electronic polarization and orientation of the molecular dipoles. Averaging over all degrees of freedom yields a dielectric constant $\epsilon_s = 80$. This results in a relatively high jump in the dielectric coefficient across the interface Γ, since the dielectric permittivity ϵ_m of proteins and chromophores is low. In this application, we take $\epsilon_m = 4$.

To study the effect of different levels of approximation in the physical model describing the electrostatic potential we use MD frames for POLY6 and POLY20 labeled with Alexa chromophores with their ground-state charges. For every frame we compute the coupling between the dyes, using 5 different levels of approximation:

(1) NO-INTERFACE (no interfaces): chromophores are represented by the fixed point charges at the places of their atoms, immersed in the solution with dielectric constant $\epsilon = 80$ everywhere; since the reaction field potential is zero, the interaction is given by the analytically known Coulomb coupling $E_{G_2} = E_{G_1}$;

(2) SURF-DYES (dyes-solution interface): chromophores with point charges and $\epsilon_m = 4$ are separated from the solution ($\epsilon_s = 80$) by their molecular surface Γ (see Figure 5.28a));

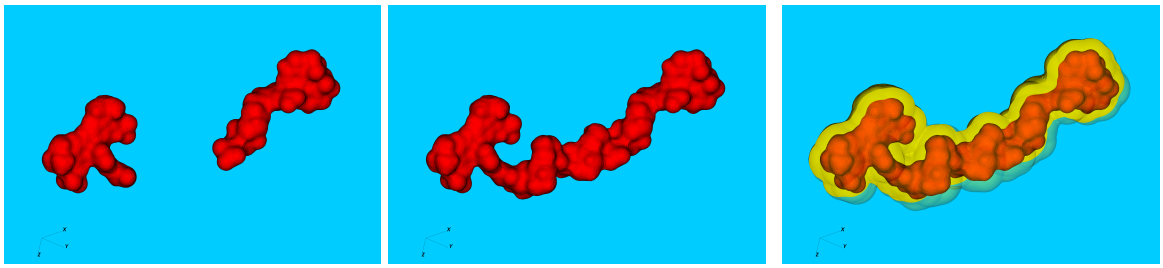
(3) SURF-DYES-POLYPRO (dyes-solution and polyproline-solution interfaces): chromophores are represented like in the previous model; polyproline ($\epsilon_m = 4$) is separated from the solution with its molecular surface (see Figure 5.28b));

(4) SURF-DYES-POLYPRO-IONS (dyes-solution and polyproline-solution interfaces with an IEL and presence of ions): the previous model is complemented by introduction of

ions in the solution. The ionic strength that is used for the experiments is equal to $0.3M$, which at $T = 300$ K results in $\overline{k} = \sqrt{2.52916}\,\mathring{A}^{-1}$ (see Figure 5.28c));

(5) SURF-DYES-POLYPRO-IONS-NO-IEL: the previous model but without an ion exclusion layer.

Furthermore, the strength of the electrostatic interaction is influenced by the total charge in each chromophore. We perform calculations with **Solver 4** for charged dyes with a total charge sum equal to $-2\,e_0$ and for neutral dyes with a total charge sum equal to $0\,e_0$. The computational domain Ω is a cube with a side length between 160 000 $\mathring{A}$ and 400 000 $\mathring{A}$ depending on the size of the smallest box containing the molecular system. The obtained results are compared to the ones obtained with MEAD. The ground-state charges of the chromophores are the same as in the examples with uniform dielectric coefficient $\epsilon = 1$ and are assigned, based on the CHARMM force field (version v35b3). The number of nonzero charges in the dyes is $N_1 = 79$ and $N_2 = 73$. We present results for 19 frames: 10 frames with POLY6 and 9 frames with POLY20.



(a) SURF-DYES (b) SURF-DYES-POLYPRO (c) SURF-DYES-POLYPRO-IONS

Figure 5.28: Different levels of approximation in the physical model describing the electrostatic potential in the system Alexa 594-POLY6-Alexa 488.

Charged dyes with a total charge sum of $-2\,e_0$

On Figure 5.29 is shown the relative distance between the interaction $\overline{E}_M$, obtained with MEAD for three of the six configurations that we used in the examples with uniform dielectric coefficient $\epsilon = 1$ (see p. 227), and the interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$, obtained with **Solver 4** at MRL $l = 4$. The average (over $3 \times 19 = 57$ frames) distances for all 6 configurations are as follows:

- configuration 1: 7.89 %
- configuration 2: 7.13 %
- configuration 3: 1.83 %

- configuration 4: 7.11 %
- configuration 5: 1.27 %
- configuration 6: 1.23 %

It is seen that configurations 3, 5, and 6 give the best match with the results obtained with **Solver 4**. Those are also the configurations with improved stopping criteria in the SOR procedure. Additionally, in Table 5.15 is given the average number of DOFs in the primal and adjoint problem for each MRL l in the case of the approximation SURF-DYES-POLYPRO. All initial meshes have a uniform mesh size in Ω_m which is related to the average edge length on the triangulated surface of the molecules. If the mesh size is nonuniform in Ω_m, then the number of mesh vertices decrease around two times for the longer polyproline POLY20.

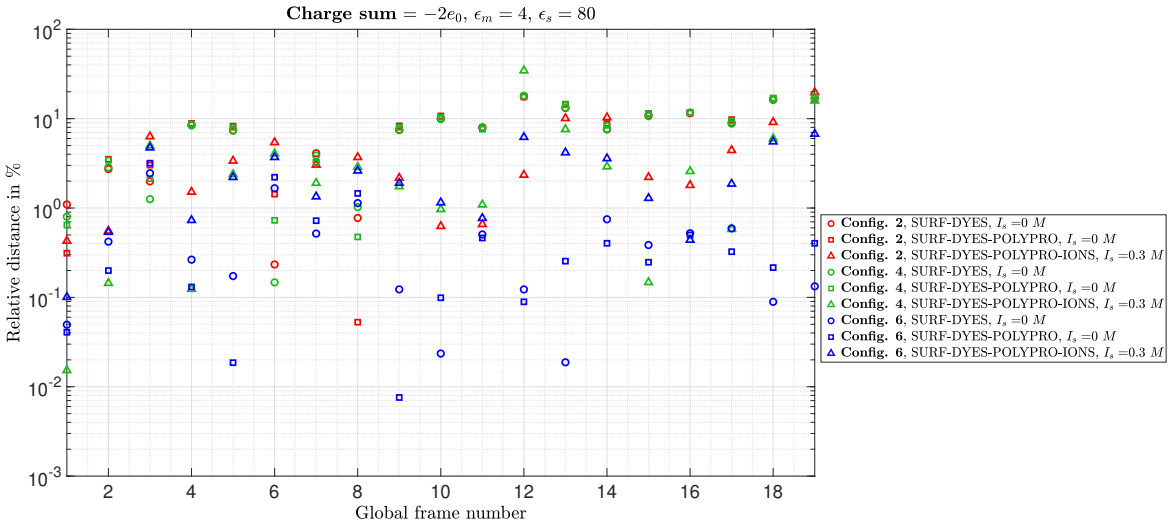


Figure 5.29: Charged dyes with ground state charges and total charge sum of $-2\,e_0$. Relative distance $\frac{|E^A_{1-2} - \overline{E}_M|}{|E^A_{1-2}|}[\%]$ between the interaction $\overline{E}_M$ obtained with MEAD and the interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ obtained with **Solver 4** at MRL $l = 4$. MEAD interaction computed with configurations 2, 4, and 6 for the grids and stopping criteria in the SOR method from p. 227.

Further, on Figure 5.30 is shown the dependence of the coupling between the two dyes on the approximation level used in the physical model describing the electrostatic potential in the system. The first three approximations give very similar couplings for most of the 19 frames. It is evident that the analytically known Coulomb interaction $E_{G_2} = E_{G_1}$, provides a very good approximation to the interaction when $I_s = 0$, especially for the frames 11 to 19, where the two dyes are sufficiently far away from each other (frames 11-19 correspond to POLY20). When $I_s > 0$, the screening of the electrostatic potential by the ions in the solution is much stronger and hence the absolute value of the interaction is several orders of magnitude smaller. Of course, the effect of the screening is even stronger when there is a larger region, occupied by the solvent, between the dyes (frames 11-19). With this said, even though the ion exclusion

Table 5.15: **Solver 4**, *GridScale*=2, $\epsilon_m = 4$, $\epsilon_s = 80$, $I_s = 0.3M$. Charged dyes with ground state atomic charges and a charge sum of $-2\,e_0$ in each dye. Average number of DOFs in the primal and adjoint problems for each MRL $l = 0, 1, \ldots, 5$ for the approximation SURF-DYES-POLYPRO. The average for POLY6 is taken over frames 1-10 and for POLY20 over frames 11-19. All initial meshes with uniform mesh size inside the molecular region, related to the average edge length on the molecular surface.

	aver. # DOFs primal		aver. # DOFs adjoint	
level	POLY6	POLY20	POLY6	POLY20
0	74 324	125 829	589 576	998 602
1	100 227	151 289	793 537	1 199 203
2	156 028	228 415	1 231 934	1 807 599
3	254 192	418 277	2 001 100	3 300 607
4	422 054	743 386	3 319 367	5 859 052
5	711 480	1 297 965	5 592 983	10 217 478

layer has a thickness of only 2 $\mathring{A}$, removing this layer is enough to cause a further decrease in the interaction approximately by a factor of 2 (SURF-DYES-POLYPRO-IONS-NO-IEL).

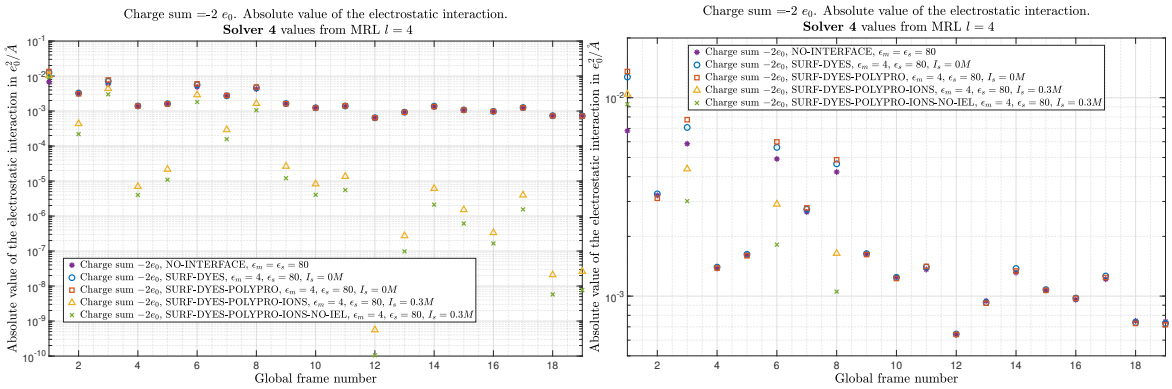


Figure 5.30: Charged dyes with ground state charges and total charge sum $-2\,e_0$. Absolute value of the coupling $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ between the two dyes for the 5 levels of approximation made in the physical model. Coupling obtained with **Solver 4** at MRL $l = 4$.

Neutral dyes (charge sum of $0\,e_0$)

Here, we use the meshes from the previous example and only the set of charges $\{q_{i,1}\}_{i=1}^{N_1}$ and $\{q_{i,2}\}_{i=1}^{N_2}$ are different. On Figure 5.32 is shown the relative distance between the interaction $\overline{E}_M$, obtained with MEAD for three of the six configurations that we use in the examples with uniform dielectric coefficient $\epsilon = 1$ (see p. 227), and the interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$, obtained with **Solver 4** at MRL $l = 4$. The average (over $3 \times 19 = 57$ frames) distances for each of the 6 configurations and the average (over $3 \times 19 - 1 = 56$ frames) without including the case SURF-DYES-POLYPRO-IONS in frame 6 are given in Table 5.16. Again,

Table 5.16: Neutral dyes with ground state atomic charges. Average relative distance between couplings obtained with MEAD for 6 different configurations and **Solver 4**.

MEAD configuration	average over 57 frames [%]	average over 56 frames [%]
1	50.06	39.02
2	40.67	31.24
3	27.76	25.24
4	38.18	26.84
5	27.12	15.74
6	24.66	13.24

configurations 3, 5, and 6 give the best match with the results obtained with **Solver 4**. In this case the differences between the solutions obtained with MEAD and **Solver 4** are significant. Before further commenting on the discrepancies with MEAD, on Figure 5.33 we show the ratios of the couplings obtained with MEAD to the couplings obtained with **Solver 4**. It is evident that there are accuracy issues in MEAD since for several frames, the couplings obtained for different configurations differ by a factor of 1.5 to 2 and some even have a different sign (see frames 10 and 15). We give more details for several problematic frames in Table 5.18, Table 5.19, and Table 5.20. There, we compare the values for E_{1-2} and E_{2-1} obtained by MEAD and **Solver 4**. It is seen that with configurations 1, 2, 4, corresponding to a default stopping criteria in SOR, often times the values for E_{1-2} and E_{2-1} have different signs. On the other hand, with configurations 5, 6, and 7, corresponding to improved stopping criteria in SOR, MEAD produces approximations of E_{1-2} and E_{2-1} which are closer in value and have the same sign. We recall that the grid specifications in configurations 5, 6, and 7 are the same and the only difference is the number of maximum iterations allowed in the SOR procedure: for configuration 5 `maxiter`=3000, for configuration 6 `maxiter`=12000, and for configuration 7 `maxiter`=4000. As in the examples with a uniform dielectric coefficient $\epsilon = 1$, it is seen that focusing on the dye of interest, i.e., focusing on the region at which the potential is needed, does not improve the quality of the computed interaction. Looking at Table 5.18, Table 5.19, and Table 5.20 is clear that the variation in the iteration number in the SOR procedure has a strong impact on the computed interaction. On the other hand, the difference in the two values obtained by solving the adjoint (with P_2 elements) and the primal (with P_1 elements) problems with **Solver 4** is negligible.

We identify several reasons for the discrepancies in the interactions computed with MEAD. The two most important ones are the ineffective stopping criteria in SOR and the use of single precision arithmetic. If the number of iterations in SOR is too low, then the iterative error dominates the discretization one. Increasing too much the number of iterations in SOR is also not recommended since in this case round off errors, especially with single precision arithmetic, start accumulating and the quality of the solution starts deteriorating.

Another cause of errors, that is common for all solvers based on finite difference schemes with uniform Cartesian grids, is the fact that the position of the interface Γ, as well as the positions and magnitudes of the charges in the molecular region Ω_m are not taken into account when computing the electrostatic interaction, i.e., the grid is not refined accordingly. As an illustration, consider the example of a Born ion with a charge of 1 e_0 positioned at the origin of the coordinate system and let φ be the potential it creates. Let $\{q_i\}_{i=1}^N$ be test charges and let $\{r_i\}_{i=1}^N$ be the respective distances from the Born ion's center. Assume that the average error in the computed potential φ_h at a distance r from the Born ion's center is $\overline{e}$. Since the problem is spherically symmetric and all test charges are treated equal (the solver is not aware of their presence), on average, the error in the computed electrostatic interaction $\tilde{E}$ will be $\sum_{i=1}^{N} q_i \overline{e}(r_i)$. Now, it is easy to analyze the overall error in E when taking different combinations of test charges q_i and distances r_i from the center. For example, if $q_i = 1\, e_0$ for all $i = 1, 2, \ldots N$, and $r_i = r, i = 1, 2, \ldots, N$, then the error in the interaction is $E - \tilde{E} = N\overline{e}$. For MEAD with improved stopping criteria and a grid with specifications ON_ORIGIN 97-3 161-1 161-0.25, we have calculated the average relative errors for the Born ion with a radius $R = 14, 16, 18\,\mathring{A}$ over samples of 400 points distributed on the spheres with radii $R - 2 + 0.5i\,\mathring{A}$, $i = 0, 1, \ldots, 8$ (see Figure 5.31). The absolute values of the relative errors for the spheres with radii different from the radius R of the Born ion vary anywhere between 0.13 % to 1 % (for example, see Table 5.17). The average error $\overline{e}$ is higher for spheres close to the interface Γ and reaches around 5 % for the spheres coinciding with the Born ion sphere. This means that for N unit charges, the relative error in the interaction can easily reach $N\%$. Of course, if the sum of all charges q_i is 0 and all of them are at the same distance r from the center, then on average the error will be zero. However, if not all of them are at the same distance from the center, then the average error can become high again.

The fact that the test charges are not "visible" is true even for adaptive finite element solvers that rely on energy norm error estimates. In contrast, the idea of the adaptive FE solvers based on the goal-oriented error estimates that we have presented is to take into account not only the charges creating the potential φ, but also the presence of the test charges so that for a given number of DOFs the error $E - \tilde{E}$ is as small as possible.

Finally, on Figure 5.34 we show the dependence of the electrostatic interaction on the different levels of approximation in the model. This time, the simple Coulomb interaction $E_{G_2} = E_{G_1}$ differs significantly from the interactions computed with the other approximations in the model.

Table 5.17: Born ion with $R = 16\,\mathring{A}$, $\epsilon_m = 4$, $\epsilon_s = 80$, $I_s = 0M$. Average relative errors $\left(\sum_{i=1}^{400} \frac{E(r)-\tilde{E}_i(r)}{E(r)}\right)/400\,[\%]$ for $r = R-2+0.5i$, $i = 0, 1, \ldots, 9$. Here $E(r) = E_{2-1}(r) = E_{1-2}(r)$ is the exact interaction between the Born ion and a test charge at a distance r and $\tilde{E}_i(r)$ is the computed with MEAD interaction at the i-th sample point.
config. (1.1): ON_ORIGIN 97-3 161-1 161-0.25, `maxiter`=4000, `maxrmsfidd`=$\frac{2\times10^{-8}}{\texttt{grid\_dim}-1}$;
config. (2.1): ON_ORIGIN 285-1 285-0.5 285-0.15, `maxiter`=4000, `maxrmsfidd`=$\frac{2\times10^{-8}}{\texttt{grid\_dim}-1}$;
config. (2.2): ON_ORIGIN 285-1 285-0.5 285-0.15, `maxiter`=12000, `maxrmsfidd`=$\frac{2\times10^{-8}}{\texttt{grid\_dim}-1}$;

configuration	$R-2$	$R-1.5$	$R-1$	$R-0.5$	R	$R+0.5$	$R+1$	$R+1.5$	$R+2$
config. (1.1)	0.282	0.348	0.438	0.577	-4.554	-0.209	-0.214	-0.219	-0.224
config. (2.1)	0.381	0.486	0.641	0.899	-1.785	0.775	0.802	0.829	0.856
config. (2.2)	0.147	0.191	0.254	0.351	-2.686	-0.152	-0.154	-0.155	-0.156

Table 5.18: Frame 15, SURF-DYES approximation. Neutral dyes with ground state charges, $\epsilon_m = 4$, $\epsilon_s = 80$, $I_s = 0M$. Check of symmetry in the electrostatic interactions E^M_{1-2} and E^M_{2-1} obtained with MEAD. In the last row are the respective values $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ (adjoint problem), $E^P_{2-1} = \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_{2,l} \rangle$ (less accurately solved primal problem), and $\overline{E} = \frac{E^A_{1-2}+E^P_{2-1}}{2}$ obtained with **Solver 4** at MRL $l = 7$. Configuration 7 with specifications: ON_GEOM_CENT 101-3, 201-1, 401-0.25, `maxiter`=4000, `maxrmsfidd`=$\frac{2\times10^{-8}}{\texttt{grid\_dim}-1}$.

configuration	E^M_{1-2}	E^M_{2-1}	$\overline{E}_M = \frac{E^M_{1-2}+E^M_{2-1}}{2}$
MEAD, config. 1	-9.665545e-08	-2.167925e-07	-1.567239e-07
MEAD, config. 2	8.639515e-08	-2.671485e-07	-9.037667e-08
MEAD, config. 3	1.606346e-07	-9.567210e-08	3.248125e-08
MEAD, config. 4	1.201878e-07	-1.873185e-07	-3.356532e-08
MEAD, config. 5	1.044903e-07	4.926980e-08	7.688007e-08
MEAD, config. 6	1.030285e-07	9.154090e-08	9.728472e-08
MEAD, config. 7	1.048365e-07	6.787570e-08	8.635612e-08
Solver 4, $l = 7$	E^A_{1-2} =2.036060e-07	E^P_{2-1} =1.979671e-07	$\overline{E}$ =2.007865e-07

Table 5.19: Frame 10, SURF-DYES approximation. Neutral dyes with ground state charges, $\epsilon_m = 4$, $\epsilon_s = 80$, $I_s = 0M$. Check of symmetry in the electrostatic interactions E^M_{1-2} and E^M_{2-1} obtained with MEAD. In the last row are the respective values $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ (adjoint problem), $E^P_{2-1} = \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_{2,l} \rangle$ (less accurately solved primal problem), and $\overline{E} = \frac{E^A_{1-2}+E^P_{2-1}}{2}$ obtained with **Solver 4** at MRL $l = 7$. Configuration 7 with specifications: ON_GEOM_CENT 101-3, 201-1, 321-0.25, `maxiter`=4000, `maxrmsfidd`=$\frac{2\times10^{-8}}{\texttt{grid\_dim}-1}$.

configuration	E^M_{1-2}	E^M_{2-1}	$\overline{E}_M = \frac{E^M_{1-2}+E^M_{2-1}}{2}$
MEAD, config. 1	2.601868e-08	-2.639895e-07	-1.189854e-07
MEAD, config. 2	-2.957653e-07	-3.436340e-07	-3.196996e-07
MEAD, config. 3	-2.472066e-07	-4.216845e-07	-3.344455e-07
MEAD, config. 4	-1.646121e-07	-4.858850e-08	-1.066003e-07
MEAD, config. 5	1.359186e-07	2.225950e-07	1.792568e-07
MEAD, config. 6	1.362875e-07	1.996230e-07	1.679552e-07
MEAD, config. 7	1.356246e-07	2.072520e-07	1.714383e-07
Solver 4, $l = 5$	E^A_{1-2} =2.782821e-07	E^P_{2-1} = 2.789001e-07	$\overline{E}$ =2.785911e-07

Table 5.20: Frame 16, SURF-DYES-POLYPRO approximation. Neutral dyes with ground state charges, $\epsilon_m = 4$, $\epsilon_s = 80$, $I_s = 0M$. Check of symmetry in the electrostatic interactions E^M_{1-2} and E^M_{2-1} obtained with MEAD. In the last row are the respective values $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ (adjoint problem), $E^P_{2-1} = \langle \frac{1}{4\pi}\mathscr{F}_1, \varphi_{2,l} \rangle$ (less accurately solved primal problem), and $\overline{E} = \frac{E^A_{1-2}+E^P_{2-1}}{2}$ obtained with **Solver 4** at MRL $l = 7$. Configuration 7 with specifications: ON_GEOM_CENT 101-3, 201-1, 401-0.25, `maxiter`=4000, `maxrmsfidd`=$\frac{2\times10^{-8}}{\texttt{grid\_dim}-1}$.

configuration	E^M_{1-2}	E^M_{2-1}	$\overline{E}_M = \frac{E^M_{1-2}+E^M_{2-1}}{2}$
MEAD, config. 1	-1.047835e-07	-2.442948e-07	-1.745391e-07
MEAD, config. 2	2.258915e-07	-2.009694e-07	1.246102e-08
MEAD, config. 3	2.269249e-07	-1.088780e-07	5.902342e-08
MEAD, config. 4	2.327025e-07	-1.595722e-07	3.656515e-08
MEAD, config. 5	2.370433e-07	1.065088e-07	1.717760e-07
MEAD, config. 6	2.749463e-07	2.678155e-07	2.713809e-07
MEAD, config. 7	2.472049e-07	1.676093e-07	2.074071e-07
Solver 4, $l = 6$	E^A_{1-2} =2.938374e-07	E^P_{2-1} =2.920518e-07	$\overline{E}$ =2.929446e-07

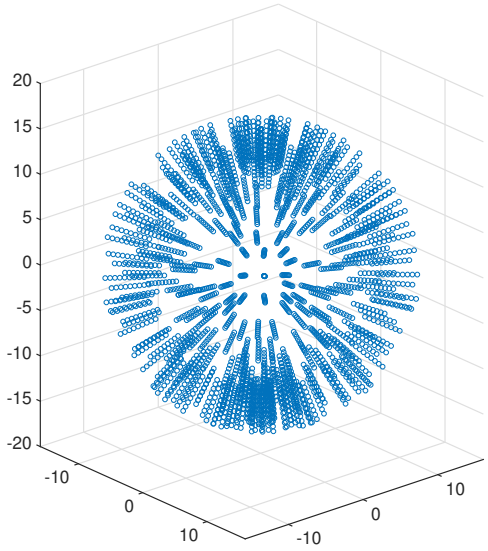


Figure 5.31: Nine concentric spheres with radii $R - 2 + 0.5i\,\mathring{A}$, $i = 0, 1, \ldots, 8$ for $R = 16\,\mathring{A}$. On each sphere there are 400 points over which an average value of the relative error is computed.

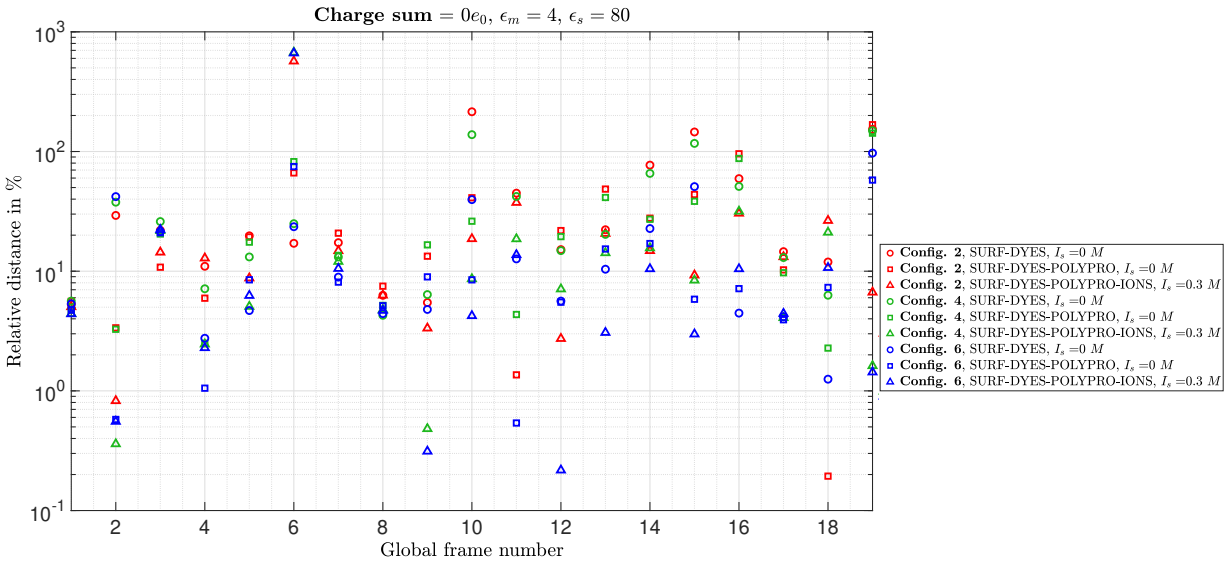


Figure 5.32: Neutral dyes with ground state charges. Relative distance $\frac{|E^A_{1-2} - \overline{E}_M|}{|E^A_{1-2}|}[\%]$ between the interaction $\overline{E}_M$ obtained with MEAD and the interaction $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ obtained with **Solver 4** at MRL $l = 4$. MEAD interaction computed with configurations 2, 4, and 6 for the grids and stopping criteria in the SOR method from p. 227.

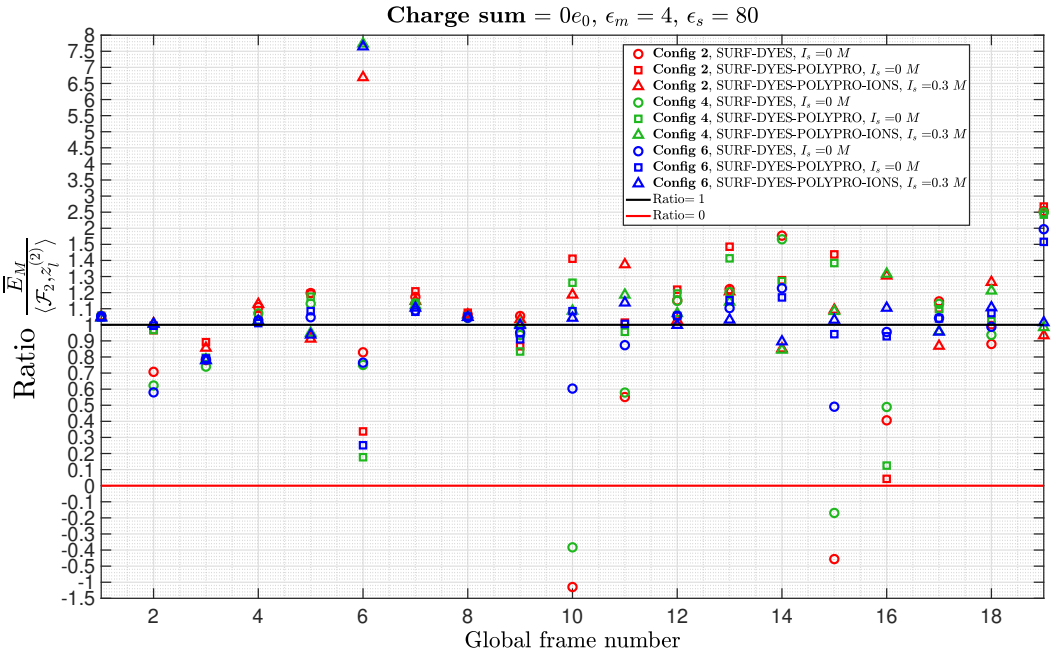


Figure 5.33: Neutral dyes with ground state charges. Ratio $\frac{\overline{E}_M}{E^A_{1-2}}$ of the coupling $\overline{E}_M$ obtained with MEAD to the coupling $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ obtained with **Solver 4** at refinement level 4.

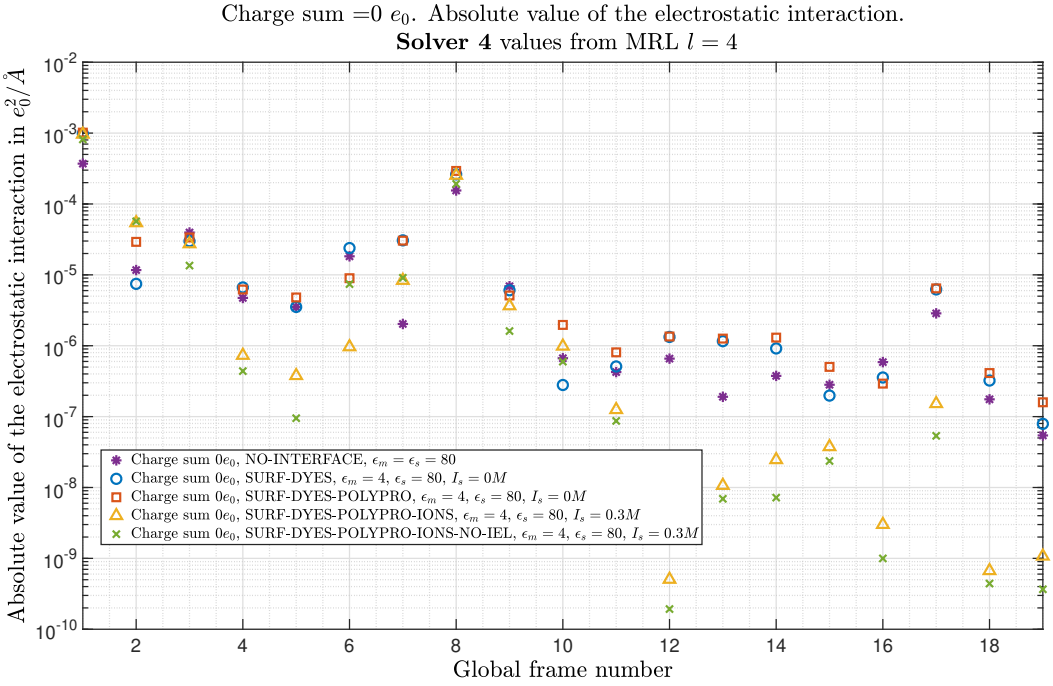


Figure 5.34: Neutral dyes with ground state charges. Absolute value of the coupling $E^A_{1-2} = \langle \mathscr{F}_2, z_l^{(2)} \rangle$ between the two dyes for the 5 levels of approximation made in the physical model. Coupling obtained with **Solver 4** at MRL $l = 4$.

Chapter 6

Conclusion

This thesis had three main goals, which can be summarized as follows:

- present a rigorous solution theory for the linearized and nonlinear Poisson-Boltzmann equation;
- derive a posteriori error estimates in terms of global energy norms of the solution as well as in terms of a specific quantity of interest;
- implement adaptive FE solvers based on the aforementioned estimates and apply these solvers to the treatment of problems of practical interest.

With respect to the solution theory, we have successfully proven the existence and uniqueness of a solution to the linearized PBE and existence of a solution to the nonlinear PBE in suitable function spaces for problems with measure right-hand side. This is achieved by employing a 2- and 3-term splittings of the full potential. Moreover, we have shown that the unique solution of the LPBE can be obtained by considering a standard weak formulation, involving H^1 spaces, for the regular component of the solution in both splittings. Likewise, a particular solution of the nonlinear PBE can be obtained by considering a weak formulation for the regular component of the potential in both splittings.

The fact that the regular component of the solution satisfies a weak formulation involving H^1 spaces means that this component can be numerically approximated by means of well studied methods, such as the FE method. Besides, it also means that the duality theory for convex variational problems is applicable. This allowed us to derive functional type a posteriori error estimates for the regular component of the solution to the nonlinear PBE. The advantage of this approach, based on the duality theory, is that it requires only the structure of the problem and therefore no global or local mesh-dependent constants enter the estimate. This is in contrast to other methods, e.g., residual based a posteriori error estimates, which depend on the particular triangulation. Therefore functional type estimates do not only give an error indicator but also a fully computable and guaranteed bound on

the error. To the best of our knowledge, this approach to the a posteriori error estimation has not been previously applied to problems of continuum electrostatics. Despite the high popularity of the PBE in the biophysics community, it seems that the use of adaptive FE methods in conjunction with a posteriori error estimation techniques is limited mainly to the developments in [48, 102, 103], where residual type error estimates are derived.

Besides functional type error estimates, we have also considered goal oriented ones. The motivation for such kind of estimates came while we were involved in an interdisciplinary project, where FRET (Fröster resonance energy transfer) rates had to be calculated. The efficiency of the transfer depends on the electrostatic coupling between the two dyes exhibiting FRET, and the calculation of the coupling only involves the values of the potential at the positions of the fixed charges in one of the dyes. Therefore, it was clear that functional type error estimates controlling the global quality of the solution, are not efficient in such situations.

The third main goal of this work was to design adaptive FE solvers that utilize the above-mentioned functional and goal oriented error estimates and to utilize them in real biophysical applications. We have successfully implemented in FreeFem++ [98] several such solvers that can solve with high accuracy the PBE for complex molecular structures. The solver based on the functional type error estimate can also return guaranteed bound on the relative error in energy norms. The solute surface meshes are generated with NanoShaper [61] and then a tetrahedral volume mesh is constructed with the help of TetGen [170]. The adaptive mesh refinement is driven either by the functional type error indicators or by the goal oriented ones, where the remeshing is done with the help of mmg3d [62]. What is more, we have verified the reliability and efficiency of these solvers on a series of problems with analytically known solutions. Here, we note that other surface mesh generators, volume mesh generators, and mesh refinement softwares can also be used.

As a possible future work we can consider the following research directions:

- Using higher order Raviart-Thomas or Brezzi-Douglas-Marini finite elements for the patchwise equilibrated flux reconstruction of the dual variable in the functional type error estimates;
- Proving the convergence of the adaptive FE algorithms based on the derived a posteriori error estimates;
- Utilizing isoparametric or isogeometric finite elements to represent more accurately the curved solute-solvent interface;
- Implementation of the adaptive FE solvers in C++ and making them available to the biophysics community.

List of Notation

Ω - Computational domain

Ω_m - Molecular/solute domain

Ω_{IEL} - Ion exclusion layer

Ω_{ions} - Ionic domain

Ω_s - Solution domain

$\partial\Omega$ - Boundary of Ω

Γ - Solute-solvent interface/boundary of Ω_m

$\boldsymbol{n}_{\partial\Omega}$ - Unit outward normal vector to $\partial\Omega$

$\boldsymbol{n}_{\Gamma}$ - Unit outward normal vector to Γ

ϵ - Dielectric coefficient

ϵ_m - Dielectric coefficient in the molecular domain

ϵ_s - Dielectric coefficient in the solution domain

$\overline{k}$ - Inverse Debye length coefficient (piecewise constant) for the dimensionless PBE and LPBE

$\overline{k}_{ions}$ - Inverse Debye length in Ω_{ions} for the dimensionless PBE and LPBE

k - Coefficient in the general semilinear problem

ϕ - Dimensionless electrostatic potential

φ - Electrostatic potential with dimension $\frac{charge}{length}$

g - Dirichlet boundary condition for the dimensionless potential ϕ

g - Dirichlet boundary condition for the potential φ with dimension $\frac{charge}{length}$

G - Coulomb part of the potential. In Chapter 3 and Chapter 4 is dimensionless and in Chapter 5 has dimension $\frac{charge}{length}$

u - Regular component of the potential in the 2- and 3-term splittings. In the 2-term splitting is also called reaction field potential. In Chapter 3 and Chapter 4 is dimensionless and in Chapter 5 has dimension $\frac{charge}{length}$.

u^H - Harmonic component of the potential in the 3-term splitting. In Chapter 3 and Chapter 4 is dimensionless and in Chapter 5 has dimension $\frac{charge}{length}$.

$\mathcal{G}_2$ - Functional on the right-hand side of the weak formulation for the regular component in the 2-term splitting for the PBE.

$\mathcal{G}_3$ - Functional on the right-hand side of the weak formulation for the regular component in the 3-term splitting for the PBE.

$H^1_{\gamma_2(g)}(\Omega)$ - The set of all functions in $H^1(\Omega)$ that are equal to $\gamma_2(g)$ on $\partial\Omega$ in the sense of traces. Here $g \in H^1(\Omega)$

$H^1_g(\Omega)$ - The set of all functions in $H^1(\Omega)$ whose trace is equal to g on $\partial\Omega$. Here $g \in H^{\frac{1}{2}}(\partial\Omega)$.

φ_2 - Potential generated by the charges in molecule II/exact solution of the primal problem (Chapter 5)

φ_1 -Potential generated by the charges in molecule I (Chapter 5)

G_2, u_2 - Coulomb and reaction field part of the potential in the 2-term splitting of φ_2 (Chapter 5)

G_1, u_1 - Coulomb and reaction field part of the potential in the 2-term splitting of φ_1 (Chapter 5)

E_{G_2}, E_{u_2} - Coulomb and reaction field part of the electrostatic interaction with the 2-term splitting of φ_2 (Chapter 5)

E_{G_1}, E_{u_1} - Coulomb and reaction field part of the electrostatic interaction with the 2-term splitting of φ_1 (Chapter 5)

E_{2-1} - Electrostatic interaction between dye/molecule II and dye/molecule I, computed with the potential created by the charges in dye II, evaluated at the positions of the charges in dye I and multiplied by the charges in dye I (Chapter 5)

E_{1-2} - Electrostatic interaction between dye I and dye II, computed with the potential created by the charges in dye I, evaluated at the positions of the charges in dye II and multiplied by the charges in dye II (Chapter 5)

$\varphi_{2,h}$ - Approximate solution of the primal problem with P_1 elements (Chapter 5)

$z_h^{(2)}$ - Approximate solution of the adjoint probem with P_2 elements (Chapter 5)

E_{2-1}^P - Approximate electrostatic interaction between dye II and dye I, computed with the primal problem (Chapter 5)

E_{1-2}^A - Approximate electrostatic interaction between dye I and dye II, computed with the adjoint problem (Chapter 5)

$\frac{1}{4\pi}\mathscr{F}_1$ - Goal functional/functional on the right-hand side of the adjoint problem (Chapter 5)

$\mathscr{F}_2$ - Functional on the right-hand side of the primal problem (Chapter 5)

List of Figures

List of Tables

Bibliography

[1] A collection of molecular surface meshes. `https://www.rocq.inria.fr/gamma/gamma/download/affichage.php?dir=MOLECULE&name=water_mol&last_page=6`. Accessed: 2017-08-18.

[2] R. Adams and J. Fournier. *Sobolev Spaces*, volume 140 of *Pure and Applied Mathematics*. Elsevier, 2003.

[3] M. Ainsworth and J. Oden. A unified approach to a posteriori error estimation using element residual methods. 65(1):23–50, 1993.

[4] M. Ainsworth and J. T. Oden. *A Posteriori Error Estimation in Finite Element Analysis*. John Wiley & Sons, Inc, 2000.

[5] O. Andreussi, N. G. Hörmann, F. Nattino, G. Fisicaro, S. Goedecker, and N. Marzari. Solvent-aware interfaces in continuum solvation. *J. Chem. Theory Comput.*, 15(3):1996–2009, 2019.

[6] I. Babŭska and T. Strouboulis. The finite element method and its reliability. 2001.

[7] N. Baker, D. Sept, S. Joseph, M. Holst, and J. McCammon. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18):10037–10041, 2001.

[8] R. Bank and M. Holst. A new paradigm for parallel adaptive meshing algorithms. *SIAM Journal on Scientific Computing*, 22(4):1411–1443, 2000.

[9] V. Barbu and T. Precupanu. *Convexity and Optimization in Banach Spaces*. Springer Monographs in Mathematics. Springer, 2012.

[10] D. Bartolucci, F. Leoni, L. Orsina, and A. C. Ponce. Semilinear equations with exponential nonlinearity and measure data. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 22(6):799–815, 2005.

[11] D. Bashford. An object-oriented programming suite for electrostatic effects in biological molecules. In *Proceedings of the Scientific Computing in Object-Oriented Parallel Environments*, ISCOPE '97, pages 233–240, London, UK, UK, 1997. Springer-Verlag.

[12] D. Bashford. Macroscopic electrostatic models for protonation states in proteins. *Frontiers in Bioscience*, 9(2):1082–1099, 2004.

[13] R. Becker and R. Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numer.*, 10:1–102, 2001.

[14] P. Bénilan, L. Boccardo, T. Gallouët, R. Gariepy, M. Pierre, and J. Vazquez. An L^1-theory of existence and uniqueness of solutions of nonlinear elliptic equations. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze*, pages 241–273, 1995.

[15] P. Bénilan and H. Brezis. Nonlinear problems related to the Thomas-Fermi equation. *J. Evol. Equ.*, 3(4):673–770, 2003. Dedicated to Philippe Bénilan.

[16] A. Bensoussan, L. Boccardo, and F. Murat. On a nonlinear partial differential equation having natural growth terms and unbounded solution. *Annales de l'I.H.P*, 5(4):347–364, 1988.

[17] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures*, volume 5 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam-New York, 1978.

[18] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Blender Institute, Amsterdam, 2017.

[19] L. Boccardo, A. Dall'Aglio, and L. Orsina. Existence and regularity results for some elliptic equations with degenerate coercivity. *Atti Sem. Mat. Fis. Univ. Modena*, 46:51–81, 1998.

[20] L. Boccardo, S. Segura de León, and C. Trombetti. Bounded and unbounded solutions for a class of quasi-linear elliptic problems with a quadratic gradient term. *J. Math. Pures Appl*, 80(9):919–940, 2001.

[21] L. Boccardo and T. Gallouët. Non-linear elliptic and parabolic equations involving measure data. *Journal of Functional Analysis*, 87:149–169, 1989.

[22] L. Boccardo, T. Gallouët, and L. Orsina. Existence and uniqueness of entropy solutions for nonlinear elliptic equations with measure data. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 13(5):539–551, 1996.

[23] L. Boccardo, T. Gallouët, and L. Orsina. Existence and nonexistence of solutions for some nonlinear elliptic equations. *J. Anal. Math.*, 73:203–223, 1997.

[24] L. Boccardo and F. Murat. A property of nonlinear elliptic equations when the right-hand side is a measure. *Potential Anal.*, 3(3):257–263, 1994.

[25] L. Boccardo, F. Murat, and J. Puel. Existence of bounded solutions for non linear elliptic unilateral problems. *Annali di Matematica Pura ed Applicata*, 152(1):183–196, 1988.

[26] L. Boccardo, F. Murat, and J. Puel. L^∞ estimate for some nonlinear elliptic partial differential equations and application to an existence result. *Siam J. Math. Anal.*, 23(2):326–323, 1992.

[27] A. Boschitsch and M. Fenley. A fast and robust Poisson-Boltzmann solver based on adaptive cartesian grids. *Journal of Chemical Theory and Computation*, 7(5):1524–1540, 2011. PMID: 21984876.

[28] D. Braess. *Finite elements*. Cambridge University Press, Cambridge, third edition, 2007. Theory, fast solvers, and applications in elasticity theory, Translated from the German by Larry L. Schumaker.

[29] D. Braess, V. Pillwein, and J. Schöberl. Equilibrated residual error estimates are p-robust. *Comput. Methods Appl. Mech. Engrg.*, 198(13-14):1189–1197, 2009.

[30] D. Braess and J. Schöberl. Equilibrated residual error estimator for Maxwell's equations. *RICAM report*, 2006.

[31] D. Braess and J. Schöberl. Equilibrated residual error estimator for edge elements. *Math. Comp.*, 77(262):651–672, 2008.

[32] S. Brenner and L. Scott. *The Mathematical Theory of Finite Element Methods*. Texts in applied mathematics. Springer, 2008.

[33] H. Brezis. Nonlinear elliptic equations involving measures. In *Contributions to nonlinear partial differential equations (Madrid, 1981)*, volume 89 of *Res. Notes in Math.*, pages 82–89. Pitman, Boston, MA, 1983.

[34] H. Brezis and F. Browder. Strongly nonlinear elliptic boundary value problems. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze*, pages 587–603, 1978.

[35] H. Brezis and F. Browder. Sur une propriété des espaces de Sobolev. *C. R. Acad. Sc. Paris*, 287:113–115, 1978.

[36] H. Brezis, M. Marcus, and A. C. Ponce. Nonlinear elliptic equations with measures revisited. In *Mathematical aspects of nonlinear dispersive equations*, volume 163 of *Ann. of Math. Stud.*, pages 55–109. Princeton Univ. Press, Princeton, NJ, 2007.

[37] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer,

B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614, 2009.

[38] F. Browder. Nonlinear boundary value problems. *Bull. Amer. Math. Soc.*, 69:864–876, 1963.

[39] F. Browder. Variational boundary value problems for quasi-linear elliptic equations, ii. *Proceedings of the National Academy of Sciences of the United States of America*, 50(4):592–598, 1963.

[40] F. Browder. Variational boundary value problems for quasi-linear elliptic equations of arbitrary order. *Proc Natl Acad Sci U S A*, 50(1):31–37, 1963.

[41] F. Browder. Nonlinear boundary value problems. ii. *Trans. Amer. Math. Soc.*, 117:530–550, 1965.

[42] J. Buse. Insulin analogues. *Curr. Opin. Endocrinol. Diabetes*, 8:95–100, 2001.

[43] T. Can, C.-I Chen, and Y.-fang Wang. Efficient molecular surface generation using level-set methods. *Journal of molecular graphics & modelling*, 25 4:442–54, 2006.

[44] C. Carstensen, M. Feischl, M. Page, and D. Praetorius. Axioms of adaptivity. *Computers & Mathematics with Applications*, 67(6):1195 – 1253, 2014.

[45] Carstensen, C. Quasi-interpolation and a posteriori error analysis in finite element methods. *ESAIM: M2AN*, 33(6):1187–1202, 1999.

[46] D. Chapman. A contribution to the theory of electrocapillarity. *Phil. Mag.*, 25:475–481, 1913.

[47] J. Chaudhry, S. Bond, and L. Olson. Finite element approximation to a finite-size modified Poisson-Boltzmann equation. *J. Sci. Comput.*, 47(3):347–364, 2011.

[48] L. Chen, M. J. Holst, and J. Xu. Adaptive finite element modeling techniques for the Poisson-Boltzmann equation. *Siam J. Numer. Anal.*, 45(6):2298–2320, 2007.

[49] M. Chen, B. Tu, and B. Lu. Triangulated manifold meshing method preserving molecular surface topology. *J. Mol. Graph. Model.*, 38:411–418, 2012.

[50] Y.-Z. Chen and L.-C. Wu. *Second order elliptic equations and elliptic systems*, volume 174 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1998. Translated from the 1991 Chinese original by Bei Hu.

[51] I. Chern, J. Liu, and W. Wang. Accurate evaluation of electrostatics for macromolecules in solution. *Methods and Applications of Analysis*, 10:309–328, 2003.

[52] H. Childs, E. Brugger, B. Whitlock, J. Meredith, S. Ahern, D. Pugmire, K. Biagas, M. Miller, C. Harrison, G. H. Weber, H. Krishnan, T. Fogal, A. Sanderson, C. Garth, E. Wes Bethel, D. Camp, O. Rübel, M. Durant, J. M. Favre, and P. Navrátil. VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data. In *High Performance Visualization–Enabling Extreme-Scale Scientific Insight*, pages 357–372. Oct 2012.

[53] M. L. Connolly. Analytical molecular surface calculation. *Journal of Applied Crystallography*, 16(5):548–558, 1983.

[54] C. Cortis and R. Friesner. Numerical solution of the Poisson-Boltzmann equation using tetrahedral finite-element meshes. *Journal of Computational Chemistry*, 18(13):1591–1608, 1997.

[55] B. Dacorogna. *Direct Methods in the Calculus of Variations*. Springer, 2008.

[56] G. Dal Maso, F. Murat, L. Orsina, and A. Prignet. Definition and existence of renormalized solutions of elliptic equations with general measure data. *C. R. Acad. Sci. Paris Sér. I Math.*, 325(5):481–486, 1997.

[57] Gianni Dal Maso, François Murat, Luigi Orsina, and Alain Prignet. Renormalized solutions of elliptic equations with general measure data. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 28(4):741–808, 1999.

[58] C. Dapogny, C. Dobrzynski, and P. Frey. Three-dimensional adaptive domain remeshing, implicit domain meshing, and applications to free and moving boundary problems. Technical report, March 2013.

[59] R. Dautray and J. L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology. Volume 6*. Springer-Verlag Berlin Heidelberg GmbH, 2000.

[60] P. Debye and E. Hückel. Zur theorie der elektrolyte. *Phys. Zeitschr.*, 24:185–206, 1923.

[61] S. Decherchi and W. Rocchia. A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale. *PLOS ONE*, 8(4):1–15, 04 2013.

[62] C. Dobrzynski. MMG3D: User Guide. Technical Report RT-0422, INRIA, March 2012.

[63] J. Droniou, T. Gallouët, and R. Herbin. A finite volume scheme for a noncoercive elliptic equation with measure data. *SIAM J. Numer. Anal.*, 41(6):1997–2031, 2003.

[64] Y. Eidelman, V. Milman, and A. Tsolomitis. *Functional analysis*, volume 66 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2004. An introduction.

[65] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. North-Holland Publishing Company, 1976.

[66] J. Elschner, J. Rehberg, and G. Schmidt. Optimal regularity for elliptic transmission problems including C^1 interfaces. *Interfaces and Free Boundaries*, 9(2):233–252, 2007.

[67] T. Engel. Computer processing of chemical structure information. In T. Engel and J. Gasteiger, editors, *Chemoinformatics. Basic Concepts and Methods*, chapter 3, pages 103–106. Wiley-VCH, 2018.

[68] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[69] A. Ern and M. Vohralík. Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. *SIAM Journal on Scientific Computing*, 35(4):A1761–A1791, 2013.

[70] L. Evans. *Partial Differential Equations*. American Mathematical Society, 1997.

[71] L. Evans and R. Gariepy. *Measure Theory and Fine Properties of Functions*. Chapman and Hall/CRC, 2015.

[72] L. Faria, O. Miyagaki, D. Motreanu, and M. Tanaka. Existence results for nonlinear elliptic equations with Leray-Lions operator and dependence on the gradient. *Nonlinear Analysis.*, 96:154–166, 2014.

[73] M. Feischl, T. Führer, N. Heuer, M. Karkulik, and D. Praetorius. Adaptive boundary element methods. *Arch. Comput. Methods Eng.*, 22(3):309–389, 2015.

[74] M. Feischl, T. Führer, M. Karkulik, M. Melenk, and D. Praetorius. Quasi-optimal convergence rates for adaptive boundary element methods with data approximation. Part II: Hyper-singular integral equation. *Electron. Trans. Numer. Anal.*, 44:153–176, 2015.

[75] M. Feischl, A. Gantner, G.and Haberl, D. Praetorius, and T. Führer. Adaptive boundary element methods for optimal convergence of point errors. *Numer. Math.*, 132(3):541–567, 2016.

[76] M. Feischl, G. Gantner, A. Haberl, and D. Praetorius. Optimal convergence for adaptive IGA boundary element methods for weakly-singular integral equations. *Numer. Math.*, 136(1):147–182, 2017.

[77] M. Feischl, M. Karkulik, J. M. Melenk, and D. Praetorius. Quasi-optimal convergence rate for an adaptive boundary element method. *SIAM J. Numer. Anal.*, 51(2):1327–1348, 2013.

[78] M. Feischl, D. Praetorius, and K. van der Zee. An abstract analysis of optimal goal-oriented adaptivity. *SIAM J. Numer. Anal.*, 54(3):1423–1448, 2016.

[79] Michael Feischl, Thomas Führer, Michael Karkulik, Jens Markus Melenk, and Dirk Praetorius. Quasi-optimal convergence rates for adaptive boundary element methods with data approximation, part I: weakly-singular integral equation. *Calcolo*, 51(4):531–562, 2014.

[80] M. Fenley, A. Boschitsch, and H.-X. Zhou. Fast boundary element method for the linear Poisson-Boltzmann equation. *The Journal of Physical Chemistry B*, 106(10):2741–2754, 2002.

[81] F. Fogolari, A. Brigo, and H. Molinari. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.*, 15:377–392, 2002.

[82] G. Folland. *Real Analysis. Modern Techniques and Their Applications*. John Wiley & Sons, Inc, 1999.

[83] B. Froese and T. Salvador. Higher-order adaptive finite difference methods for fully nonlinear elliptic equations. *J. Sci. Comput.*, 75(3):1282–1306, 2018.

[84] E. Gagliardo. Caratterizzazioni delle tracce sulla frontiera relative ad alcune classi di funzioni in n variabili. *Rendiconti del Seminario Matematico della Università di Padova*, 27:284–305, 1957.

[85] T. Gallouët and R. Herbin. Existence of a solution to a coupled elliptic system. *Appl. Math. Lett.*, 7(2):49–55, 1994.

[86] T. Gallouët and R. Herbin. Convergence of linear finite elements for diffusion equations with measure data. *C. R. Math. Acad. Sci. Paris*, 338(1):81–84, 2004.

[87] G. Gantner, A. Haberl, D. Praetorius, and B. Stiftner. Rate optimal adaptive FEM with inexact solver for nonlinear operators. *IMA Journal of Numerical Analysis*, 38(4):1797–1831, 09 2017.

[88] T. Gantumur. Adaptive boundary element methods with convergence rates. *Numer. Math.*, 124(3):471–516, 2013.

[89] C. Geuzaine and J.-F. Remacle. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79(11):1309–1331, 2009.

[90] D. Gilbarg and N. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[91] M. Gilson and B. Honig. Calculation of electrostatic potentials in an enzyme active site. *Nature*, 330(6143):84–86, 11 1987.

[92] M. K. Gilson, K. A. Sharp, and B. H. Honig. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *Journal of Computational Chemistry*, 9(4):327–335, 1988.

[93] G. Gouy. Constitution of the electric charge at the surface of an electrolyte. *J. Phys.*, 9:457–468, 1910.

[94] J. Greer and B. L. Bush. Macromolecular shape and surface maps by solvent exclusion. *Proc Natl Acad Sci USA*, 75(1):303–307, 1978.

[95] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Pitman Advanced Publishing Program, 1985.

[96] Brezis H. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2011.

[97] R. Harris, A. Boschitsch, and M. Fenley. Numerical difficulties computing electrostatic potentials near interfaces with the Poisson-Boltzmann equation. *Journal of Chemical Theory and Computation*, 13(8):3945–3951, 2017. PMID: 28640608.

[98] F. Hecht. New development in FreeFem++. *J. Numer. Math.*, 20(3-4):251–265, 2012.

[99] P. Hess. A strongly nonlinear elliptic boundary value problem. *Journal of Mathematical Analysis and Applications*, 43:241–249, 1973.

[100] I. Hlaváček and M. Křížek. On a superconvergent finite element scheme for elliptic systems. I. Dirichlet boundary condition. *Apl. Mat.*, 32(2):131–154, 1987.

[101] M. Holst. *The Poisson-Boltzmann Equation: Analysis and Multilevel Numerical Solution*. Applied Mathematics and CRPC, California Institute of Technology. 1994.

[102] M. Holst, N. Baker, and F. Wang. Adaptive multilevel finite element solution of the Poisson-Boltzmann equation I. Algorithms and examples. *Journal of Computational Chemistry*, 21(15):1319–1342, 2000.

[103] M. Holst, J. McCammon, Z. Yu, Y. C. Zhou, and Y. Zhu. Adaptive finite element modeling techniques for the Poisson-Boltzmann equation. *Commun. Comput. Phys.*, 11:179–214, 2012.

[104] M. Holst and S. Pollock. Convergence of goal-oriented adaptive finite element methods for nonsymmetric problems. *Numer. Methods Partial Differential Equations*, 32(2):479–509, 2016.

[105] M. Holst, S. Pollock, and Y. Zhu. Convergence of goal-oriented adaptive finite element methods for semilinear problems. *Comput. Vis. Sci.*, 17(1):43–63, 2015.

[106] Christian Remling (https://mathoverflow.net/users/48839/christian remling). MathOverflow. URL:https://mathoverflow.net/q/227908 (version: 2017-04-13).

[107] B. Hu. *Blow-up theories for semilinear parabolic equations*, volume 2018 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011.

[108] S. Hubrich, P. Di Stolfo, L. Kudela, S. Kollmannsberger, E. Rank, A. Schröder, and A. Düster. Numerical integration of discontinuous functions: moment fitting and smart octree. *Comput. Mech.*, 60(5):863–881, 2017.

[109] N. Ji, T. Liu, J. Xu, L. Q. Shen, and B. Lu. A finite element solution of lateral periodic Poisson-Boltzmann model for membrane channel proteins. *International Journal of Molecular Sciences*, 19(3), 2018.

[110] B. Kawohl and M. Lucia. Best constants in some exponential Sobolev inequalities. *Indiana University Mathematics Journal*, 57(4):1907–1928, 2008.

[111] S. Kesavan. *Topics in Functional Analysis and Applications*. New Age International (P) Limited, 1989.

[112] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and Their Applications*. SIAM, 2000.

[113] J. Kirkwood. Theory of solutions of molecules containing widely separated charges with special applications to zwitterions. *J. Chem. Phys.*, 7:351–361, 1934.

[114] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins: Structure, Function, and Bioinformatics*, 1(1):47–59, 1 1986.

[115] S. Korotov, P. Neittaanmäki, and S. Repin. A posteriori error estimation of goal-oriented quantities by the superconvergence patch recovery. *J. Numer. Math.*, 11(1):33–59, 2003.

[116] J. Kraus, S. Nakov, and S. Repin. Reliable computer simulation methods for electrostatic biomolecular models based on the Poisson-Boltzmann equation. Preprint on arXiv:1805.11441, 2018.

[117] J. Kraus, S. Nakov, and S. Repin. Reliable numerical solution of a class of nonlinear elliptic problems generated by the Poisson-Boltzmann equation. *Computational Methods in Applied Mathematics*, forthcoming.

[118] A. J. Kurdila and M. Zabarankin. *Convex Functional Analysis*. Birkhäuser Verlag, 2005.

[119] M. Křížek and P. Neittaanmäki. Superconvergence phenomenon in the finite element method arising from averaging gradients. *Numer. Math.*, 45(1):105–116, 1984.

[120] Boccardo L. and Brezis H. Some remarks on a class of elliptic equations with degenerate coercivity. *Bollettino dell'Unione Matematica Italiana*, 6-B(3):521–530, 2003.

[121] M. Landes. Uniqueness and stability of strongly nonlinear elliptic boundary value problems. *Journal of Differential Equations*, 67:122–143, 1987.

[122] G. Leioni. *A First Course in Sobolev Spaces.* American Mathematical Society, 2009.

[123] Bo Li. Minimization of electrostatic free energy and the Poisson-Boltzmann equation for molecular solvation with implicit solvent. *Siam J. Math. Anal.*, 40(6):2536–2566, 2009.

[124] J. Li, S. Wijeratne, X. Qiu, and C.-H. Kiang. DNA under force: Mechanics, electrostatics, and hydration. *Nanomaterials*, 5(1):246–267, 2015.

[125] J. L. Lions and E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications. Volume I.* Springer-Verlag Berlin Heidelberg New York, 1972.

[126] J. Lipfert, S. Doniach, R. Das, and D. Herschlag. Understanding nucleic acid-ion interactions. *Annu Rev Biochem.*, 83:813–841, 2014.

[127] S. Liu. Multiple solutions for elliptic resonant problems. *Proceedings of the Royal Society of Edinburgh*, 138:1281–1289, 2008.

[128] T. Liu, S. Bai, B. Tu, M. Chen, and B. Lu. Membrane-channel protein system mesh construction for finite element simulations. *Computational and Mathematical Biophysics*, 3(1):128–139, 2005.

[129] T. Liu, M. Chen, and B. Lu. Efficient and qualified mesh generation for Gaussian molecular surface using adaptive partition and piecewise polynomial approximation. *SIAM J. Sci. Comput*, 40:507–527, 2018.

[130] B. Lu, X. Cheng, J. Huang, and J. McCammon. Order N algorithm for computation of electrostatic interactions in biomolecular systems. *Proceedings of the National Academy of Sciences*, 103(51):19314–19319, 2006.

[131] B. Lu, X. Cheng, J. Huang, and J. McCammon. An adaptive fast multipole boundary element method for Poisson-Boltzmann electrostatics. *Journal of chemical theory and computation*, 5(6):1692–1699, 2009.

[132] B. Lu and J. McCammon. Improved boundary element methods for Poisson-Boltzmann electrostatic potential and force calculations. *Journal of Chemical Theory and Computation*, 3(3):1134–1142, 2007. PMID: 26627432.

[133] B. Lu, Y. Zhou, M. Holst, and J. McCammon. Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications. *Commun. Comput. Phys.*, 3(5):973–1009, 2008.

[134] J. Madura, J. Briggs, R. Wade, M. Davis, B. Luty, A. Ilin, J. Antosiewicz, M. Gilson, B. Bagheri, L. Scott, and J. McCammon. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Computer Physics Communications*, 91(1):57 – 95, 1995.

[135] S. Mikhlin. *Constants in Some Inequalities of Analysis*. Chichester ; Wiley. Translated by Reinhard Lehmann, 1986.

[136] M. Mirzadeh, M. Theillard, and F. Gibou. A second-order discretization of the nonlinear Poisson-Boltzmann equation over irregular geometries using non-graded adaptive Cartesian grids. *J. Comput. Phys.*, 230(5):2125–2140, 2011.

[137] P. Monk. *Finite Element Methods for Maxwell's Equations*. Numerical Mathematics and Scientific Computing. Clarendon Press Oxford, 2003.

[138] P. Neittaanmäki, S. Repin, and P. Turchyn. New a posteriori error indicator in terms of linear functionals for linear elliptic problems. *Russian J. Numer. Anal. Math. Modelling*, 23(1):77–87, 2008.

[139] P. Neittaanmäki and Repin S. *Reliable Methods for Computer Simulation: Error Control and Posteriori Estimates*. Elsevier, 2004.

[140] A. Nicholls, K. Sharp, and B. Honig. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, 11(4):281–296, 1991.

[141] C. Niedermeier and K. Schulten. Molecular dynamics simulations in heterogeneous dielectrica and Debye-Huckel media-application to the protein bovine pancreatic trypsin inhibitor. *Molecular Simulation*, 8:361–387, 1992.

[142] A. Oberman and I. Zwiers. Adaptive finite difference methods for nonlinear elliptic and parabolic partial differential equations with free boundaries. *J. Sci. Comput.*, 68(1):231–251, 2016.

[143] H. Oberoi and N. Allewell. Multigrid solution of the nonlinear Poisson-Boltzmann equation and calculation of titration curves. *Biophysical Journal*, 65:48–55, 1993.

[144] J. Oden and S. Prudhomme. Goal-oriented error estimation and adaptivity for the finite element. *Computers & Mathematics with Applications*, 41:735–756, 2001.

[145] M. Olshanskii and D. Safin. Numerical integration over implicitly defined domains for higher order unfitted finite element methods. *Lobachevskii J. Math.*, 37(5):582–596, 2016.

[146] L. Orsina and A. Prignet. Strong stability results for solutions of elliptic equations with power-like lower order terms and measure data. *J. Funct. Anal.*, 189(2):549–566, 2002.

[147] W. Orttung. Direct solution of the Poisson equation for biomolecules of arbitrary shape, polarizability density, and charge distribution. *Annals of the New York Academy of Sciences*, 303(1):22–37, 1977.

[148] F. De Paiva. Multiple solutions for elliptic problems with asymmetric nonlinearity. *Journal of Mathematical Analysis and Applications*, 292:317–327, 2004.

[149] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.*, 5:286–292 (1960), 1960.

[150] A. C. Ponce. *Elliptic PDEs, measures and capacities*, volume 23 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2016. From the Poisson equations to nonlinear Thomas-Fermi problems.

[151] A. Porretta and S. Segura de León. Nonlinear elliptic equations having a gradient term with natural growth. *J. Math Pures Appl.*, 85:465–492, 2006.

[152] J. Pousin and J. Rappaz. Consistency, stability, a priori and a posteriori errors for Petrov-Galerkin methods applied to nonlinear problems. *Numerische Mathematik*, 69(2):213–231, 1994.

[153] A. Prignet. Remarks on existence and uniqueness of solutions of elliptic problems with right-hand side measures. *Rend. Mat. Appl. (7)*, 15(3):321–337, 1995.

[154] A. Rashin and K. Namboodiri. A simple method for the calculation of hydration enthalpies of polar molecules with arbitrary shapes. *The Journal of Physical Chemistry*, 91(23):6003–6012, 1987.

[155] S. Repin. A posteriori error estimation for variational problems with uniformly convex functionals. *Math. Comp*, 69:481–500, 2000.

[156] S. Repin. *A Posteriori Estimates for Partial Differential Equations*. Walter de Gruyter‘, 2008.

[157] S. Repin. On measures of errors for nonlinear variational problems. *Russian J. Numer. Anal. Math. Modelling*, 27(6):577–584, 2012.

[158] F. M. Richards. Areas, volumes, packing, and protein structure. *Annual Review of Biophysics and Bioengineering*, 6(1):151–176, 1977.

[159] T. Richter and T. Wick. Variational localizations of the dual weighted residual estimator. *J. Comput. Appl. Math.*, 279:192–208, 2015.

[160] R. T. Rockafellar. Level sets and continuity of conjugate convex functions. *Trans. Amer. Math. Soc.*, 123:46–63, 1966.

[161] R. T. Rockafellar. Integrals which are convex functionals. *Pacific J. Math.*, 24:525–539, 1968.

[162] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[163] N. Rogers and M. Sternberg. Electrostatic interactions in globular proteins: Different dielectric models applied to the packing of α-helices. *Journal of Molecular Biology*, 174(3):527 – 542, 1984.

[164] I. Sakalli, J. Schöberl, and E. W. Knapp. mfes: A robust molecular finite element solver for electrostatic energy computations. *J. Chem. Theory Comput.*, 10:5095–5112, 2014.

[165] M. F. Sanner, A. J. Olson, and J.-C. Spehner. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers*, 38(3):305–320, 1996.

[166] Joachim Schöberl. NETGEN an advancing front 2D/3D-mesh generator based on abstract rules. *Computing and Visualization in Science*, 1(1):41–52, Jul 1997.

[167] J. Serrin. Pathological solutions of elliptic differential equations. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze*, Ser. 3, 18(3):385–387, 1964.

[168] K. Sharp and B. Honig. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem*, 94:7684–7692, 1990.

[169] R. Showalter. *Hilbert Space Methods for Partial Differential Equations*. Courier Corporation, 2010.

[170] H. Si. TetGen, a Delaunay-based quality tetrahedral mesh generator. *ACM Transactions on Mathematical Software (TOMS)*, 41(11), 2015.

[171] K. Siebert. A convergence proof for adaptive finite elements without lower bound. *IMA Journal of Numerical Analysis*, 31(3):947–970, 05 2010.

[172] E. Sobakinskaya, M. Schmidt am Busch, and T. Renger. Theory of FRET "spectroscopic ruler" for short distances: Application to polyproline. *J. Phys. Chem. B*, 112:54–67, 2018.

[173] I. Stakgold and M. Holst. *Green's Functions and Boundary Value Problems*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 2011.

[174] G. Stampacchia. Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus. *Annales de l'institut Fourier*, 15(1):189–257, 1965.

[175] M. Sternberg, F. Hayes, A. Russell, P. Thomas, and A. Fersht. Prediction of electrostatic effects of engineering of protein charges. *Nature*, 330(6143):86–88, 11 1987.

[176] Z.-J. Tan and S.-J. Chen. Predicting electrostatic forces in RNA folding. *Methods Enzymol.*, 469:465–487, 2009.

[177] T. Tao. *An Introduction to Measure Theory.* Graduate Studies in Mathematics. American Mathematical Society, 2011.

[178] C. Trombetti. Non-uniformly elliptic equations with natural growth in the gradient. *Potential Analysis*, 18:391–404, 2003.

[179] N. Trudinger. On imbeddings into Orlicz spaces and some applications. *J. Math. Mech.*, 17:473–483, 1967.

[180] R. Verfürth. A posteriori error estimates for nonlinear problems. Finite element discretizations of elliptic equations. *Mathematics of Computation*, 62(206):445–475, 1994.

[181] R. Verfürth. A posteriori error estimation and adaptive mesh-refinement techniques. *Journal of Computational and Applied Mathematics*, 50(1):67 – 83, 1994.

[182] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques.* Chichester : Wiley, 1996.

[183] Y. N. Vorobjev, J. A. Grant, and H. A. Scheraga. A combined iterative and boundary-element approach for solution of the nonlinear Poisson-Boltzmann equation. *Journal of the American Chemical Society*, 114(9):3189–3196, 1992.

[184] Z. Wan et al. Enhancing the activity of insulin at the receptor interface: crystal structure and photo-cross-linking of A8 analogues. *Biochemistry*, 43:16119–16133, 2004.

[185] J. Warwicker and H. Watson. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *Journal of Molecular Biology*, 157(4):671 – 679, 1982.

[186] J. Webb. On the Dirichlet problem for strongly non-linear elliptic operators in unbounded domains. *J. London Math. Soc.*, 10(2):163–170, 1975.

[187] J. Webb. Boundary value problems for strongly nonlinear elliptic equations. *J. London Math. Soc.*, 21(2):123–132, 1980.

[188] D. Xie. New solution decomposition and minimization schemes for Poisson-Boltzmann equation in calculation of biomolecular electrostatics. *Journal of Computational Physics*, 275:294 – 309, 2014.

[189] Z. Xie and C. Chen. The interpolated coefficient FEM and its application in computing the multiple solutions of semilinear elliptic problems. *International Journal of Numerical Analysis and Modeling*, 25:549–576, 2005.

[190] Z. Xie, C. Chen, and Y. Xu. An improved search-extension method for computing multiple solutions of semilinear pdes. *IMA Journal of Numerical Analysis*, 25:549–576, 2005.

[191] D. Xu and Y. Zhang. Generating triangulated macromolecular surfaces by Euclidean distance transform. *PLOS ONE*, 4(12):1–11, 12 2009.

[192] Z. Yu, M. Holst, Y. Cheng, and J. McCammon. Feature-preserving adaptive mesh generation for molecular shape modeling and simulation. *Journal of molecular graphics & modelling*, 26(8):1370—1380, June 2008.

[193] R. Zauhar and R. Morgan. A new method for computing the macromolecular electric potential. *Journal of Molecular Biology*, 186(4):815 – 820, 1985.

[194] R. Zauhar and R. Morgan. The rigorous computation of the molecular electric potential. *Journal of Computational Chemistry*, 9(2):171–187, 1988.

[195] Z. Zhou, P. Payne, M. Vasquez, N. Kuhn, and M. Levitt. Finite-difference solution of the Poisson-Boltzmann equation: Complete elimination of self-energy. *Journal of Computational Chemistry*, 17(11):1344–1351, 1996.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe. Die vorliegende Dissertation ist mit dem elektronisch übermittelten Textdokument identisch.

Linz, June 2019

_____

Svetoslav Nakov

Curriculum Vitae

Name: Svetoslav Nakov

Nationality: Bulgaria

Date of Birth: 21 February, 1989

Place of Birth: Sofia, Bulgaria

Education:

2003–2008	High School of Mathematics and Natural Sciences, Veliko Tarnovo, Bulgaria
2008–2012	Bachelor Study in Applied Mathematics, University of Sofia "St. Kliment Ohridski"
2012–2014	Master Study in Computational Mathematics and Mathematical Modeling, University of Sofia "St. Kliment Ohridski"
2014–2019	PhD Study in Computational Mathematics, Johannes Kepler University, Linz

Research Stays:

Dec. 2016–Mar. 2017	University of Duisburg-Essen, visiting Prof. Johannes Kraus
Mar. 2017–Jun. 2017	Pennsylvania State University, visiting Prof. Ludmil Zikatanov

Work and Teaching:

Oct. 2014–Sep. 2018	Radon Institute for Computational and Applied Mathematics,

	research scientist
Feb. 2014–Jul. 2014	University of Sofia "St. Kliment Ohridski" graduate teaching assistant for a course in Applications of Mathematics for Modeling Real Problems
Apr. 2013–Sep. 2014	Bulgarian Academy of Sciences, Institute of Mechanics, mathematician
Feb. 2012–Jul. 2013	University of Sofia "St. Kliment Ohridski" graduate teaching assistant for a course in Numerical Linear Algebra

Scholarships and Awards:

2009–2014	European Students Award and Scholarship as part of the project "Students Scholarships and Awards" co-funded by the European Union and the Bulgarian Ministry of Education, Youth
2013–2014	Scholarship for high grade point average (GPA) in the Master program awarded by University of Sofia "St. Kliment Ohridski"
2009–2012	Scholarship for high grade point average (GPA) in the Bachelor program awarded by Sofia University "St. Kliment Ohridski"
14–16 May 2010	Silver Medal from the National Mathematics Olympiad for University Students, V. Tarnovo, Bulgaria
27 Mar. 2012	Gold Medal from the 3-rd National Armwrestling Competition for University Students, +95 kg Right Hand Division, Blagoevgrad, Bulgaria

Special Interests: Armwrestling, Volleyball, Fitness