Doctoral Dissertation
Doctoral Program in Electrical, Electronics and Communications Engineering
(37^{th} cycle)

Fast, Regularized, and High-Order Modelling in Advanced Computational Electromagnetics

Johann Bourhis

* * * * *

Supervisor
Prof. Francesco P. Andriulli

Doctoral Examination Committee:
Prof. Francesco P. Andriulli, Full Professor, Politecnico di Torino, Turin, Italy
Prof. Simon B. Adrian, Full Professor, Universität Rostock, Rostock, Germany (Referee)
Dr. Samuel Nosal, Head of Discipline Hw Simulation, Thales DMS France, Élancourt, France (Referee)
Prof. Roberto D. Graglia, Full Professor, Politecnico di Torino, Turin, Italy
Prof. Kristof Cools, Full Professor, Universiteit Gent, Ghent, Belgium
Prof. Adrien Merlini, Associate Professor, IMT Atlantique, Brest, France

Politecnico di Torino
January 24, 2024

This thesis is licensed under a Creative Commons License, Attribution - Noncommercial-NoDerivative Works 4.0 International: see www.creativecommons.org. The text may be reproduced for non-commercial purposes, provided that credit is given to the original author.

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . . . .

Johann Bourhis

Turin, January 24, 2024

Summary

Electromagnetic simulations are essential across various applications, from antenna design and radar to biomedical imaging and wireless communications, where accurate and fast handling of complex geometries, materials, and wide frequency ranges is often needed. Among the methods developed in computational electromagnetics (CEM), the Boundary Element Method (BEM) offers distinct advantages, especially in reducing unknowns through surface discretization and inherently satisfying radiation conditions at infinity, which is ideal for modelling the scattering from piecewise homogeneous and isotropic media. High-order discretizations further enhance BEM's accuracy, making it suitable for large-scale and complex models. However, BEM faces challenges such as high computational costs, the need for precise singular integral evaluations, and issues with ill-conditioning in complex scenarios. Acceleration techniques, including the fast multipole method (FMM) and low-rank block matrix compression within the framework of hierarchical matrices, have mitigated some of these obstacles by reducing computational and storage costs of matrix assembly and algebraic operations. Moreover, preconditioning strategies, such as quasi-Helmholtz decompositions and Calderón preconditioners, have been developed to address stability and conditioning of the numerical systems, allowing to significantly reduce the number of iterations when resorting to iterative solvers. Limitations remain however, in particular for high-order discretizations and complex material interfaces, for which the extension of these techniques were not fully investigated. This work aims to address these gaps by introducing new quasi-Helmholtz and Calderón preconditioning techniques tailored to high-order elements as well as structures with junctions, providing new advances in BEM's applicability to practical and large-scale problems.

Acknowledgements

A PhD journey can feel like having to solve an ill-conditioned system, but, like numerical problems, it converges more smoothly with the right preconditioners–by which I mean my friends, family, colleagues, and advisors. Without each of you, this solution might never have converged!

To start, I am profoundly grateful to my PhD advisor, Francesco P. Andriulli, for his mentorship, support and encouragements over these past three years. I also want to extend my thanks to Adrien Merlini and Clément Henry, from IMT Atlantique, for their invaluable collaboration and guidance throughout my PhD.

To my colleagues and friends at Politecnico di Torino–Viviana, Damiano, Paolo, Leonardo, Ermanno, Alberto, Simone, Davide, Pierrick, Margaux, Giulio, Tommaso, Matteo, and Giuseppe–thank you for the camaraderie and the memorable times we shared, both in and out of the lab.

A special thanks goes to Damiano Franzò, who worked with me on implementing part of the code we shared during the PhD. That project would not have been as enjoyable without you by my side! You became also one of my most valuable friends during these years and I truly want to thank you for introducing me to your group of friends outside of the office! I extend my thanks to them as well, for their kindness and acceptance for integrating in their group a French guy that was not passionate about learning Italian at all.

I am also deeply grateful to the simulation team at Thales DMS, where I spent half of my PhD experience. To my advisor there, Samuel Nosal, thank you for your inspiring guidance and for the fantastic opportunity to evolve around your team. I also include Gilles Salin, Alexandre Dely and Jean-Paul Martinaud to tell I am very thankful for your time and expertise in helping me apprehend the specifics of computational electromagnetics in industry. To everyone there, additionally including, among others, Guillaume, Bernard, Valentin, Augustin, Carlos, Arnaud and Nicolas, thank you for your warm welcome, unforgettable coffee break discussions, and, of course, the many ping-pong games with some of you! I am looking forward to meeting you again and I hope to have the opportunity to keep collaborating with you in the following years.

To continue, this PhD was funded by the Marie Skłodowska-Curie Actions from the European Commission. I am very grateful to my funders for this wonderful opportunity of being part of that high-level experience and for the incredible financial support they provided.

I would also like to take this opportunity to express my sincere gratitude to the members of my PhD jury, whose constructive feedback and insightful comments greatly contributed to improving this thesis. I deeply appreciate their involvement in this important phase of my academic journey.

I am also thankful to my professors at INSA Rennes and Rennes 1 University, to have provided me with the knowledge and the interest for the field of applied mathematics. I extend these thanks to Agnès Pujols at the French Alternative Energies and Atomic Energy Commission (CEA) for the introduction to the field of computational EM during a very decisive internship I did under her supervision four years ago. Without these advisors and mentors, I would not have

followed the same path, to which I am so proud and grateful for!

I would also like to thank Kilian, who was both a friend and classmate before I started my PhD. He introduced me to this opportunity and became one of my greatest supports during these years.

Finally, I owe a deep gratitude to my friends and family for helping me "disconnect" whenever I needed. I will try to not forget any of you in these acknowledgments: to my friends from high-school–Arthur, Tanguy, Kévin, Paul, Vyria, Lucie, Lysiane, Lili, Chloé, Nicolas, Manon, Suzon, Maëlis–my friends from engineering school–Corentin, Erwan, Gurvan, Kilian, Lucas, Martin–and my friends from Québec–Julien, Pauline, Loïc, Lucille, Arnaud, Juliette, Victor, Alizée, Yannis, Anna, Romane. Of course, I am very grateful to my parents and my sister, for their love and their dedicated support; but they already know it very well.. I hope :).

Thank you all!

Contents

List of Tables x

List of Figures xi

0 Introduction 1

1 From Maxwell's Equations to Surface Integral Equations 3
- 1.1 Notations . . . 3
- 1.2 Modelling of Electromagnetic Problems . . . 4
  - 1.2.1 Time-Domain Maxwell's Equations . . . 4
  - 1.2.2 Time-Harmonic Maxwell's Equations . . . 4
  - 1.2.3 Jump Conditions . . . 5
  - 1.2.4 Radiation Conditions at Infinity . . . 5
- 1.3 Derivation of the Vector Helmholtz Equation . . . 6
  - 1.3.1 Scattered Field Solving . . . 6
  - 1.3.2 Expressing the Maxwell's Equations in the Whole Space . . . 7
  - 1.3.3 Jump Formulae . . . 8
  - 1.3.4 Vector Helmholtz Equation . . . 9
  - 1.3.5 Functional Spaces . . . 12
- 1.4 Stratton-Chu Representation Formula . . . 13
  - 1.4.1 Green's Kernel for the Helmholtz Equation . . . 13
  - 1.4.2 Expressing the Electromagnetic Field in Terms of Surface Potentials . . . 15
  - 1.4.3 Tangent Trace of the Surface Potentials . . . 16
- 1.5 Surface Integral Equations and Variational Problems for Dielectric and Metallic Boundaries . . . 18
  - 1.5.1 Case of a Dielectric Interface . . . 18
  - 1.5.2 Case of a PEC Boundary . . . 18
- 1.6 Calderón Identities . . . 19

2 High-Order Boundary Element Method 21
- 2.1 Mesh Discretization of Surface Boundaries . . . 21
  - 2.1.1 Reference Triangle . . . 21
  - 2.1.2 Local-to-Global Parametrization . . . 22
  - 2.1.3 Unit Normal and Jacobian Matrices . . . 22
  - 2.1.4 Example with the Lagrange interpolation . . . 24
  - 2.1.5 Integration and Differentiation on Surface Boundaries . . . 25
- 2.2 Boundary element spaces . . . 28
  - 2.2.1 Definitions . . . 28

2.2.2 Quasi-Helmholtz Decompositions of the Raviart-Thomas and Rotated Raviart-Thomas Spaces . . . 30
2.2.3 Interpolatory Basis Functions . . . 31
2.3 Discretization of the Variational Problem . . . 36
2.3.1 Standard Discretization of the EFIE . . . 36
2.3.2 The case of Thin PEC Layers . . . 38
2.3.3 The Case of PEC Junctions . . . 38
2.4 Numerical Integration . . . 40
2.5 Discussion on the Discretization of the Other Variational Problems . . . 41
2.5.1 Discretization of the MFIE and CFIE . . . 41
2.5.2 Discretization of PMCHWT . . . 42

3 Numerical solvers 43
3.1 Spectral Decompositions, Pseudoinversion and Orthogonal Projections . . . 43
3.2 Matrix Factorizations and Direct Solvers . . . 47
3.3 Iterative Solvers . . . 48
3.3.1 Stationary methods . . . 48
3.3.2 Multigrid methods . . . 49
3.3.3 Krylov methods . . . 51
3.4 Fast Matrix-Vector Products . . . 53
3.5 Discussion and Objectives of the Next Chapters . . . 54

4 Quasi-Helmholtz Projectors for Open Boundaries with Junctions 57
4.1 Introduction . . . 57
4.2 RWG connectivity and dual graph at low-order . . . 58
4.3 Loop and Star Decomposition with Junctions . . . 59
4.4 Quasi-Helmholtz Projectors . . . 60
4.4.1 Definition . . . 60
4.4.2 Study of orthogonality and completeness . . . 61
4.5 Spectral Analysis of Graph Laplacians with Junctions . . . 62
4.6 Numerical Results . . . 64
4.6.1 Pseudoinversion of the Graph Laplacian . . . 64
4.6.2 Efficiency of the Quasi-Helmholtz Preconditioning for Structures with Junctions . . . 66
4.7 Conclusion . . . 68

5 High-Order Quasi-Helmholtz Projectors 69
5.1 Introduction . . . 69
5.2 Definition . . . 69
5.3 Uniqueness of the Projectors . . . 71
5.4 Stabilization of the EFIE . . . 74
5.5 Implementation Details . . . 74
5.6 Numerical Results . . . 75
5.7 Conclusion . . . 78

6 High-Order Quasi-Helmholtz Filters 81
6.1 Introduction . . . 81
6.2 Orthonormalization of the Basis Functions . . . 81
6.3 Analysis of the Spectrum of the Quasi-Helmholtz Decomposition . . . 83

6.3.1 Relationship with the Laplace-Beltrami Eigenvalue Problem 83
6.3.2 Convergence and Asymptotic Behavior of the Laplace-Beltrami Eigenvalues 84
6.4 Operator Filtering 85
6.4.1 Spectrum Filtering and Wavelet-Like Decomposition 85
6.4.2 Preconditioning the Electric Field Integral Equation 87
6.5 Implementation related details 89
6.6 Numerical results 89
6.7 Conclusion 94

7 Projector-Based Calderón Preconditioners for High-Order Discretizations 97
7.1 Introduction 97
7.2 Background and Notations 98
7.2.1 Extended Raviart-Thomas Space 100
7.3 The Novel Calderón Preconditioner for High-Order Discretization of the EFIE . 101
7.3.1 Proposed Formulation 101
7.3.2 Discretization of the Surface Rotation Operator 101
7.3.3 Suppression of the Static Nullspace Associated with the Quasi-Harmonic Subspace 103
7.4 Implementation Related Details 104
7.4.1 Enforcing the Numerical Stability at Low-Frequency 104
7.4.2 Efficiency of the Algorithm within Fast Iterative Frameworks 105
7.4.3 Replacing the Inverse Square-Roots by Sparse Transformation Matrices 107
7.5 Numerical Validations 110
7.5.1 Validation of the Simply-Connected Formulation 110
7.5.2 Validation of the Multiply-Connected Formulation 114
7.5.3 Validation for Less Regular Geometries 117
7.5.4 On the Use of Partial-Orthonormalization 119
7.6 Conclusion 122

8 Conclusion and Perspectives 123

A Physics Quantities 125

Bibliography 127

List of Tables

2.1 Location of the Lagrange interpolatory points in the reference triangle. 24
2.2 Location of the shifted Lagrange interpolatory points in the reference triangle. 35
4.1 Number of iterations achieved by AGMG for increasingly dense discretization with a residual error below 10^{-8}. 65
4.2 Number of iterations and run time of TFQMR to solve the EFIE with and without preconditioning at medium and low frequencies for the model of turbine with 9191 RWG functions. 68
7.1 Ratio between the dimensions of the auxiliary space (of order r) and of the primal space (of order p) for $C = 2E/3$ 106
7.2 Number of instances of each kind of matrix-vector product per iteration. 107
7.3 Comparision of the condition number betweem the Gram matrix associated with the GWP functions and its partially-orthonormalized counterpart, for different orders and mesh discretization of the unitary sphere. 119

List of Figures

1.1 Illustration of an incident field that impinges on a scattering object with a resulting scattered field. 6
2.1 Reference triangle and its geometrical elements. The size of the outgoing unit normals are shrunk and should equal the length of the edges. 22
2.2 Cross section of a closed PEC boundary of thickness $\delta > 0$ and associated thin layer obtained as the limit case where δ goes to zero. Dotted lines mean that the basis functions on that side are removed. 38
2.3 Cross section of a junction edge with different types of configuration as described in [146]: a) all surfaces are open, dotted lines mean that the basis functions on that side are removed, b) all surfaces are closed, c) one surface is closed and the other is open, d) more general case with several closed and open surfaces. In cases c) and d), the number of closed bodies connected to the edge e is $c_e = 2$. . 39
4.1 Notations of the vertices and cells used to define the RWG functions. 58
4.2 Two conventional ways to define an independent set of RWG functions on a junction-edge and their associated graph representation. It illustrates two extreme configuration scenarios: on the left, the one giving rise to the maximal degree of connectivity, with one cell connected to all the others; on the right, the one giving rise to the minimal degree of connectivity (but maximal graph distance), with each cell connected to at most two other cells through this edge. 59
4.3 Dual graph (in red) associated with a mesh (in gray) which contains open boundaries, a junction, and an aperture. This graph highlights the existence of two local Loops around the vertex belonging to the junction and one global Loop around the aperture. 60
4.4 Spectrum of the graph Laplacian $\boldsymbol{\Sigma}^\mathsf{T}\boldsymbol{\Sigma}$ for structures with multiple sheets, each discretized with two triangles. The two connectivity configurations of Fig. 4.2 are tested and yield different spectra for same structures. 63
4.5 Computational time required by AGMG to solve graph-Laplacian systems for increasing number of junctions. Configuration (1) and (2) are respectively corresponding to the maximal and minimal degree of connectivity that we get with a given RWG configuration on junctions. 65
4.6 Surface density current calculated along the surface of the turbine with the quasi-Helmholtz projectors at 1 Hz. 66
4.7 Condition number of the EFIE with and without preconditioning in function of the frequency for the model of turbine. 67
4.8 Bistatic RCS (dBsm) in the $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}})$ plane given the scattered far field produced by the turbine with an incident plane wave at a frequency of 1 Hz. 67
5.1 Condition number of the system matrices in function of the frequency for order zero to three. 76

5.2 Condition number of the system matrices in function of the inverse of the average cell diameter h; frequency of 1Hz. 76
5.3 Radar cross section (RCS) in dBsm at 100Hz. 77
5.4 Radar cross section (RCS) in dBsm at 3×10^8Hz. 77
5.5 Absolute value of the surface density current (in $\mathrm{dB_{A/m}}$) for the interlocked Möbius ring irradiated by a plane wave at 10Hz, with preconditioning (left) and without (right). 78
5.6 Absolute value of the surface density current (in $\mathrm{dB_{A/m}}$) for the airliner irradiated by a plane wave at 1Hz, with preconditioning (left) and without (right). 78
6.1 Singular value spectra of the continuous and discretized Laplacian operators along with their filtered counterparts for the sphere at order 2 with $h = 58$cm. . 90
6.2 Singular value spectra of $\mathbf{P}^{\widetilde{\Lambda}}\widetilde{\mathbf{T}}_s\mathbf{P}^{\widetilde{\Lambda}}$ and $\widetilde{\mathbf{T}}_h$ and of their filtered counterparts for the sphere at order 2 with $h = 58$cm. 90
6.3 Singular value spectra of the EFIE (without preconditioning), projector-based EFIE and filter-based EFIE for the sphere at order 2 with $h = 58$cm. Frequency is $f = 10$kHz. 90
6.4 Condition number of the different formulations matrices for the Sphere of radius 1m as a function of the number of unknowns (for decreasing h). Frequency is $f = 10$kHz. 91
6.5 Condition number of the different formulations matrices for the Torus of outer radius 50cm and inner radius 15cm as a function of the number of unknowns (for decreasing h). Frequency is $f = 10$kHz. 91
6.6 Condition number of the different formulations matrices for the NASA almond of dimensions 25.24cm $\times$ 9.75cm $\times$ 3.25cm as a function of the number of unknowns (for decreasing h). Frequency is $f = 10$kHz. 92
6.7 Residual error as a function of the number of unknowns for the iterative solver TFQMR for simulations of the unitary sphere discretized with $h = 10$cm. Frequency is $f = 10$ kHz. 92
6.8 Condition number of the different formulations matrices for the sphere of radius 1 m at different frequencies. Average mesh edge size is $h = 58$cm. 93
6.9 Condition number of the different formulations matrices for the torus of outer radius 50cm and inner radius 15cm at different frequencies. Average mesh edge size is $h = 8$cm. 93
6.10 Condition number of the different formulations matrices for the NASA almond of dimensions 25.24cm $\times$ 9.75cm $\times$ 3.25cm at different frequencies. Average mesh edge size is $h = 1.3$cm. 94
6.11 Condition number of the different formulations matrices for the Sphere of radius 1 m as a function of the number of unknowns (for decreasing h) without normalization. Frequency is $f = 10$kHz. 94
7.1 Moduli of the eigenvalues of the matrices involved in the preconditioned formulation (7.27) for simulations of the sphere at frequency $f = 1$MHz, with $p = 2$ and $h = 0.58$cm. 111
7.2 Eigenvalues of the matrices involved in the preconditioned formulation (7.27) for simulations of the sphere at frequency $f = 1$MHz, with $p = 2$ and $h = 0.58$cm. 112
7.3 Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the sphere ($f = 1$MHz). 113
7.4 Ratio between the maximal and minimal eigenvalue moduli as a function of the frequency for simulations of the sphere ($h = 58$cm). 113

7.5 Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the sphere (f = 1MHz, h = 0.58cm). . . . 114
7.6 Moduli of the eigenvalues of the matrices involved in the preconditioned formulation (7.43) for simulations of the torus of outer radius 0.5 m and inner radius 0.15 m at frequency f = 1Hz, with p = 2 and h = 8cm. . . . 115
7.7 Eigenvalues of the matrices involved in the preconditioned formulation (7.43) for simulations of the torus of outer radius 0.5 m and inner radius 0.15 m at frequency f = 1Hz, with p = 2 and h = 8cm. . . . 116
7.8 Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the torus (f = 1MHz). . . . 116
7.9 Ratio between the maximal and minimal eigenvalue moduli as a function of the frequency for simulations of the torus (h = 8cm). . . . 117
7.10 Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the torus (f = 1MHz and h = 8cm). . . . 117
7.11 Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the NASA almond (f = 1MHz). . . . 118
7.12 Ratio between the maximal and minimal eigenvalue moduli as a function of the frequency for simulations of the NASA almond (h = 1.3cm). . . . 118
7.13 Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the NASA almond (f = 1MHz and h = 1.3cm). . . . 118
7.14 Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the sphere, with partial orthonormalization (f = 1MHz). . . . 120
7.15 Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the torus, with partial orthonormalization (f = 1MHz). . . . 120
7.16 Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the NASA almond, with partial orthonormalization (f = 1MHz). . . . 120
7.17 Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the sphere, with partial orthonormalization (f = 1MHz, h = 58cm). . . . 121
7.18 Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the torus, with partial orthonormalization (f = 1MHz, h = 8cm). . . . 121
7.19 Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the NASA Almond, with partial orthonormalization (f = 1MHz, h = 1.3cm). . . . 121

Chapter 0

Introduction

Electromagnetic simulations play a significant role in a wide range of applications, from antenna design and radar systems to biomedical imaging and wireless communication. These applications often require accurate and efficient numerical methods that are capable of handling complex geometries, diverse materials, and a large range of frequencies. To address these challenges, computational electromagnetics (CEM) offers a variety of methods, including finite element method (FEM) [131, 100], finite difference time domain (FDTD) [145, 133], and boundary element method (BEM) [75, 106, 123]. Each technique has its strengths, and the choice of which to use strongly depends on the nature of the problem that we need to solve.

Among these methods, BEM stands out for several reasons. When modelling objects composed of several simple media (that are homogeneous, linear and isotropic), it only requires the discretization of surface boundaries, which significantly reduces the number of unknowns in comparison with volume methods such as FEM and FDTD. Additionally, BEM implicitly satisfies radiation conditions at infinity, which is a major advantage in electromagnetic scattering problems compared to the other techniques—that require meshing a relatively large region of the exterior domain and imposing approximating boundary conditions or perfectly matched layers at its truncating borders [15]. The efficiency of BEM can be further improved by resorting to high-order discretization [67, 115]. Very often, increasing the order of the boundary elements offers higher accuracy for a given number of unknowns, making them especially valuable to model complex and large-scale materials with high precision [108].

Despite these advantages, however, BEM faces several well-known challenges:

1. The method assembles very populated matrices, which lead to large computational time and storage costs [20].

2. It involves the evaluation of singular integrals, for which the accurate numerical approximation is critical and may be expensive to compute [52].

3. It suffers from ill-conditioning in certain scenarios, such as at low or high frequencies, in the presence of spurious resonances, with high-contrast media or for dense discretization [2]. These issues can limit the method's applicability, as the solution will suffer from numerical error and iterative solvers will not converge, or have very slow convergence [122, 70].

During the past years, several advances have been made to mitigate these drawbacks. Techniques such as the fast multipole method (FMM) [54, 126] and block low-rank compression of hierarchical matrices (H-mat) [20, 12, 97] have been developed to reduce memory and computational complexity of assembling and solving the BEM linear systems. Regarding the

computation of singular integrals, specialized integration rules have been developed to numerically approximate them. While they are currently quite reliable and effective, researches are still conducted nowadays to further improve these techniques [123, 52, 111].

Finally, preconditioning strategies have been introduced to address ill-conditioning and ensure the stability of BEM solvers. Among others, we can refer to the quasi-Helmholtz decompositions [138, 2], Calderón preconditioning [42, 7], multiresolution / hierarchical methods [4, 3], potential-based formulations [55, 38] and on-surface radiation conditions (OSRC) [56]. Unfortunately, few of these techniques are effective or even available with high-order discretizations. Moreover, handling complex scenarios such as the modelling of junctions between metallic and dielectric objects—which is a key situation in industrial applications—represents a major challenge for these methods [146]. These two limitations have been a significant barrier to the broader adoption of BEM in practical engineering problems. Specifically, they have prevented the use of efficient iterative solvers across a wide range of problems, often forcing reliance on direct solvers. While direct solvers can provide more robust solutions, their computational and memory costs grow prohibitively large for complex or large-scale problems, severely limiting BEM's ability to model such scenarios efficiently.

Within this context, this thesis will introduce new preconditioning techniques specifically designed for complex scenarios, including the modelling of metallic junctions and the use of high-order boundary elements. Through novel developments, especially in quasi-Helmholtz decompositions and Calderón preconditioners, this work aims at pushing the boundaries of what is achievable with BEM in both academic research and industry.

The following of this thesis is organized as follows. The three first chapters introduce the background and the notations. In Chapter 1, we derive the surface integral equations from the Maxwell's equations and we arrive at the variational formulations of these equations. In Chapter 2, we introduce the high-order boundary element method that consists in discretizing the variational formulations with mesh parametrizations and finite dimensional functional spaces defined from mappings of high-degree polynomial functions. In Chapter 3 we introduce the principles behind numerical methods and acceleration strategies to solve linear systems arising from BEM discretizations.

In Chapter 4, we study the application of the quasi-Helmholtz projectors to solve low-frequency problems when modelling metallic objects with junctions. In Chapter 5, we introduce a generalization of the quasi-Helmholtz projectors to precondition at low-frequency the linear systems arising from high-order discretization. The use of these high-order quasi-Hemholtz projectors is then extended in the two last chapters where we investigate the preconditioning of the electric field integral equation (EFIE) to cure its dense-refinement breakdown in the context of high-order discretization. In Chapter 6, we show that the filtering of the Laplacian operator appearing in the computation of the quasi-Helmholtz projectors enables the construction of a wavelet-like preconditioner which further stabilizes the EFIE. Finally, a novel Calderón preconditioning technique is presented in Chapter 7, where we show that the high-order quasi-Helmholtz projectors allow to build a dual space based on p-refinement, providing a stable and well-conditioned formulation. In Chapter 8 we conclude and discuss the next developments and possible extensions.

Chapter 1

From Maxwell's Equations to Surface Integral Equations

In this chapter, we introduce the fundamental equations of electromagnetics that we derive to obtain the surface integral equations on which we focus in this manuscript. We refer to Appendix A for a list of the physics quantities we describe hereafter.

1.1 Notations

Let us consider a smooth scalar function $f : \mathbb{R}^3 \times \mathbb{R}^+ \to \mathbb{C}$ that is function of the Cartesian coordinates $\boldsymbol{r} = (r_1, r_2, r_3)^\mathsf{T}$ and of the time variable $t \geq 0$. We denote the nabla operator by $\boldsymbol{\nabla} = (\partial_{r_1}, \partial_{r_2}, \partial_{r_3})^\mathsf{T}$, which allows us to define, in the Cartesian coordinates, the gradient and the Laplacian of f as $\boldsymbol{\nabla} f$ and $\Delta f = \boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} f)$, respectively. Let us consider now a smooth vector field $\boldsymbol{f} = (f_1, f_2, f_3)^\mathsf{T} : \mathbb{R}^3 \times \mathbb{R}^+ \to \mathbb{C}^3$, for which the divergence, rotational and Laplacian are defined as $\boldsymbol{\nabla} \cdot \boldsymbol{f}$, $\boldsymbol{\nabla} \times \boldsymbol{f}$ and $\boldsymbol{\Delta f}$ respectively, where $[\boldsymbol{\Delta f}]_i = \Delta f_i$. The partial derivatives with respect to the time variable are denoted by $\partial_t f$ and $\partial_t \boldsymbol{f} = (\partial_t f_1, \partial_t f_2, \partial_t f_3)^\mathsf{T}$. The real parts of complex-valued scalar and vector fields are denoted by $\mathrm{Re}(f)$ and $\mathrm{Re}(\boldsymbol{f})$ and their conjugate are $\overline{f}$ and $\overline{\boldsymbol{f}} = (\overline{f}_1, \overline{f}_2, \overline{f}_3)^\mathsf{T}$.

Let us consider a volume $\Omega \subseteq \mathbb{R}^3$. We denote by $L^2(\Omega)$ the space of square integrable functions in Ω. Let $f, g \in L^2(\Omega)$, and $\boldsymbol{f}, \boldsymbol{g} \in L^2(\Omega)^3$. We define the L^2 inner-product on Ω as

$$(f, g) \coloneqq \int_\Omega \overline{f(\boldsymbol{r})} g(\boldsymbol{r}) \, \mathrm{d}V(\boldsymbol{r}) \,, \qquad (\boldsymbol{f}, \boldsymbol{g}) \coloneqq \int_\Omega \overline{\boldsymbol{f}(\boldsymbol{r})} \cdot \boldsymbol{g}(\boldsymbol{r}) \, \mathrm{d}V(\boldsymbol{r}) \,. \tag{1.1}$$

We denote by δ_0 the distribution (called Dirac delta) that verifies

$$\int_\Omega \delta_0(|\boldsymbol{r}' - \boldsymbol{r}|) \boldsymbol{f}(\boldsymbol{r}') \, \mathrm{d}V(\boldsymbol{r}') \coloneqq \boldsymbol{f}(\boldsymbol{r}) \,. \tag{1.2}$$

Let Γ be a surface manifold contained in Ω. We denote by $L^2(\Gamma)$ the space of square integrable functions on Γ and $L^2_t(\Gamma)$ the space of square integrable tangent vector fields on Γ. Consider $f, g \in L^2(\Gamma)$ and $\boldsymbol{f}, \boldsymbol{g} \in L^2_t(\Gamma)$. We define the L^2 inner-product on Γ as

$$\langle f, g \rangle \coloneqq \int_\Gamma \overline{f(\boldsymbol{r})} g(\boldsymbol{r}) \, \mathrm{d}S(\boldsymbol{r}) \,, \qquad \langle \boldsymbol{f}, \boldsymbol{g} \rangle \coloneqq \int_\Gamma \overline{\boldsymbol{f}(\boldsymbol{r})} \cdot \boldsymbol{g}(\boldsymbol{r}) \, \mathrm{d}S(\boldsymbol{r}) \,. \tag{1.3}$$

We denote by δ_Γ the distribution (called surface Dirac delta) that verifies

$$\int_\Omega \boldsymbol{f}(\boldsymbol{r}) \delta_\Gamma(\boldsymbol{r}) \, \mathrm{d}V(\boldsymbol{r}) \coloneqq \int_\Gamma \boldsymbol{f}(\boldsymbol{r}) \, \mathrm{d}S(\boldsymbol{r}) \,. \tag{1.4}$$

We define the Gram matrices between two sets of scalar functions $\{f_m\}_{m=1}^M$, $\{g_n\}_{n=1}^N$ in $L^2(\Gamma)$ and two sets of tangent vector fields $\{\boldsymbol{f}_m\}_{m=1}^M$, $\{\boldsymbol{g}_n\}_{n=1}^N$ in $L_t^2(\Gamma)$, denoted by $\mathbf{G}_{f,g} \in \mathbb{R}^{M\times N}$ and $\mathbf{G}_{\boldsymbol{f},\boldsymbol{g}} \in \mathbb{R}^{M\times N}$, respectively, with components

$$\left[\mathbf{G}_{f,g}\right]_{mn} \coloneqq \langle f_m, g_n\rangle \,, \qquad \left[\mathbf{G}_{\boldsymbol{f},\boldsymbol{g}}\right]_{mn} \coloneqq \langle \boldsymbol{f}_m, \boldsymbol{g}_n\rangle \,. \tag{1.5}$$

1.2 Modelling of Electromagnetic Problems

1.2.1 Time-Domain Maxwell's Equations

In absence of electric charge and current excitations, the propagation of an electromagnetic field is given by the following homogeneous Maxwell's equations [40]

$$\begin{cases} \boldsymbol{\nabla}\times\boldsymbol{H} - \partial_t\boldsymbol{D} = \mathbf{0} & \text{(Maxwell-Ampère)}\,, \\ \boldsymbol{\nabla}\times\boldsymbol{E} + \partial_t\boldsymbol{B} = \mathbf{0} & \text{(Maxwell-Faraday)}\,, \\ \boldsymbol{\nabla}\cdot\boldsymbol{D} = 0 & \text{(Maxwell-Gauss)}\,, \\ \boldsymbol{\nabla}\cdot\boldsymbol{B} = 0 & \text{(Maxwell-Thomson)}\,. \end{cases} \tag{1.6}$$

In a simple medium (that is linear, isotropic, and homogeneous), the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{H}$ are linked to the electric induction $\boldsymbol{D}$ and magnetic induction $\boldsymbol{B}$ by the following relationships

$$\begin{aligned} \boldsymbol{D} &= \varepsilon\,\boldsymbol{E}\,, \\ \boldsymbol{B} &= \mu\,\boldsymbol{H}\,, \end{aligned} \tag{1.7}$$

where the scalar permittivity ε and the permeability μ are the characteristic constants of the medium.

1.2.2 Time-Harmonic Maxwell's Equations

We consider an electromagnetic field for which the time-dependency corresponds to a cosine wave at a given frequency f with associated angular frequency $\omega = 2\pi f$, and wavenumber $k = \omega\sqrt{\mu\varepsilon}$. The electric and magnetic fields can thus be expressed in the following form

$$\boldsymbol{u}(\boldsymbol{r},t) = \boldsymbol{u}_0(\boldsymbol{r})\cos(\omega t + \phi(\boldsymbol{r}))\,, \tag{1.8}$$

where $\boldsymbol{u}$ is a placeholder that stands for the vector fields $\boldsymbol{E}$, $\boldsymbol{H}$, $\boldsymbol{D}$ and $\boldsymbol{B}$.

Subsequently, we extend the electromagnetic field in the complex plane, by defining the phasors

$$\boldsymbol{u}(\boldsymbol{r}) = \boldsymbol{u}_0(\boldsymbol{r})e^{j\phi(\boldsymbol{r})}\,, \tag{1.9}$$

in such a way that

$$\boldsymbol{u}(\boldsymbol{r},t) = \mathrm{Re}\left(\boldsymbol{u}(\boldsymbol{r})e^{j\omega t}\right)\,. \tag{1.10}$$

We then have the additive rule

$$\boldsymbol{u}(\boldsymbol{r},t) + \boldsymbol{v}(\boldsymbol{r},t) = \mathrm{Re}\left(\left(\boldsymbol{u}(\boldsymbol{r}) + \boldsymbol{v}(\boldsymbol{r})\right)e^{j\omega t}\right)\,, \tag{1.11}$$

as well as the time-differentiation rule

$$\partial_t\boldsymbol{u}(\boldsymbol{r},t) = \mathrm{Re}\left(j\omega\,\boldsymbol{u}(\boldsymbol{r})e^{j\omega t}\right)\,. \tag{1.12}$$

From these two properties, we notice that the time-dependency here is simply expressed through the quantity $e^{j\omega t}$ that we can factorize in the Maxwell's equations when they are applied to the phasors. We thus obtain the time-harmonic Maxwell's equations by dividing the aforementioned equations by $e^{j\omega t} > 0$, which removes the time-dependency from these equations,

$$\begin{cases} \boldsymbol{\nabla} \times \boldsymbol{H} - j\omega \boldsymbol{D} = \boldsymbol{0} & \text{(Maxwell-Ampère)}\,, \\ \boldsymbol{\nabla} \times \boldsymbol{E} + j\omega \boldsymbol{B} = \boldsymbol{0} & \text{(Maxwell-Faraday)}\,, \\ \boldsymbol{\nabla} \cdot \boldsymbol{D} = 0 & \text{(Maxwell-Gauss)}\,, \\ \boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 & \text{(Maxwell-Thomson)}\,. \end{cases} \tag{1.13}$$

Once this system is solved, it is straightforward to retrieve the time-dependent vector fields of the original problem by computing for a fixed time t the real parts (1.10), for $\boldsymbol{E}(\boldsymbol{r}, t)$, $\boldsymbol{H}(\boldsymbol{r}, t)$, $\boldsymbol{D}(\boldsymbol{r}, t)$ and $\boldsymbol{B}(\boldsymbol{r}, t)$.

Alternatively, (1.8) can be thought as one of the modes of a Fourier series that we get by decomposing the time-dependent electromagnetic wave. This last statement shows that time-domain problems can be solved, or at least approximated, by combining the solutions of time-harmonic problems. In the following, we will focus on solving the time-harmonic equations only and we will not mention the time-domain equations anymore.

1.2.3 Jump Conditions

We consider that μ and ε are piecewise constants, which defines a set of simple media. We denote by Γ the interface between two of these media. For any vector field $\boldsymbol{u}$, its jump across the interface Γ is $\left[\boldsymbol{u}\right] \coloneqq (\boldsymbol{u}_1 - \boldsymbol{u}_2)_{|\Gamma}$, where $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ refer to the vector field $\boldsymbol{u}$ evaluated from outside and from inside Γ, respectively, according to the convention that has to be fixed for the direction of the unit normal vector to Γ, denoted $\boldsymbol{n}$. In absence of surface currents and surface charges, the jump conditions at every interface Γ between these media is [106]

$$\begin{aligned} \left[\boldsymbol{E} \times \boldsymbol{n}\right] = \boldsymbol{0}\,, \quad \left[\boldsymbol{D} \cdot \boldsymbol{n}\right] = 0\,, \\ \left[\boldsymbol{H} \times \boldsymbol{n}\right] = \boldsymbol{0}\,, \quad \left[\boldsymbol{B} \cdot \boldsymbol{n}\right] = 0\,. \end{aligned} \tag{1.14}$$

These boundary conditions express the fact that the electric and magnetic fields are tangent-continuous across the interface while the electric and magnetic inductions are normal-continuous.

Following the Maxwell-Ampère and Maxwell-Faraday equations (1.13), we also find

$$\left[\left(\frac{1}{j\omega\mu}\boldsymbol{\nabla} \times \boldsymbol{E}\right) \times \boldsymbol{n}\right] = \boldsymbol{0}\,, \tag{1.15a}$$

$$\left[\left(\frac{1}{j\omega\varepsilon}\boldsymbol{\nabla} \times \boldsymbol{H}\right) \times \boldsymbol{n}\right] = \boldsymbol{0}\,. \tag{1.15b}$$

Moreover, the jump condition (1.14) on the surface of a perfectly conducting (PEC) object can be further specified, accordingly to the fact that the electric field inside a PEC object is zero. The jump condition thus transforms into the PEC boundary condition, for $\boldsymbol{r} \in \Gamma$

$$\boldsymbol{E}(\boldsymbol{r}) \times \boldsymbol{n}(\boldsymbol{r}) = \boldsymbol{0}\,. \tag{1.16}$$

1.2.4 Radiation Conditions at Infinity

To ensure that the Maxwell's equations define a well-posed mathematical problem, additional propagation conditions need to be taken into account. In the following, we assume that the

electromagnetic field always satisfies the Sommerfeld's radiation condition, which ensures that electromagnetic waves radiate outward from sources and do not return from infinity [106]

$$\lim_{|\boldsymbol{r}|\to+\infty} |\boldsymbol{r}| \left(\partial_{|\boldsymbol{r}|}\boldsymbol{u}(\boldsymbol{r}) + jk\boldsymbol{u}(\boldsymbol{r})\right) = \boldsymbol{0}\,, \tag{1.17}$$

where $\boldsymbol{u}$ is, once again, a placeholder for the vector fields $\boldsymbol{E}$, $\boldsymbol{H}$, $\boldsymbol{D}$ and $\boldsymbol{B}$.

1.3 Derivation of the Vector Helmholtz Equation

1.3.1 Scattered Field Solving

In the following of the thesis, the electromagnetic field is decomposed in terms of an incident field—that is given in the definition of the electromagnetic problem we aim to solve—and a scattered field—that has to be determined to solve the electromagnetic problem. This terminology refers to the scattering of an incident wave that impinges on the surface of an object obstacle, but identical derivations can be done in a more general context, in terms of field and current excitations and radiated fields (for antenna problems for instance).

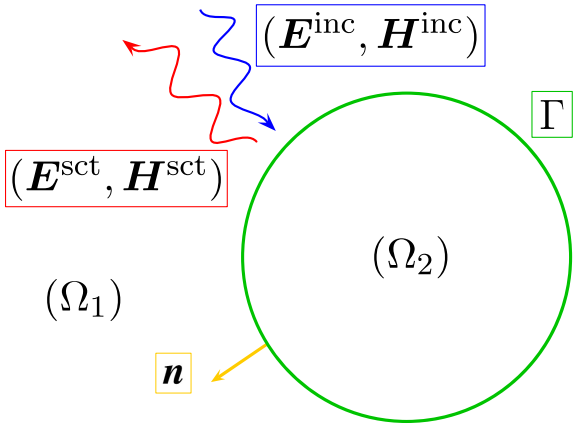


Figure 1.1: Illustration of an incident field that impinges on a scattering object with a resulting scattered field.

To start, we choose to focus on the model problem illustrated in Figure 1.1. Given an incident electromagnetic field $(\boldsymbol{E}^{\mathrm{inc}}, \boldsymbol{H}^{\mathrm{inc}})$ in a medium Ω_1—that is assumed to be independent of the obstacle—we aim to find the scattered electromagnetic field $(\boldsymbol{E}^{\mathrm{sct}}, \boldsymbol{H}^{\mathrm{sct}})$ that arises from the presence of an obstacle Ω_2. Subsequently, we express the total field $(\boldsymbol{E}, \boldsymbol{H}) = (\boldsymbol{E}^{\mathrm{inc}}, \boldsymbol{H}^{\mathrm{inc}}) + (\boldsymbol{E}^{\mathrm{sct}}, \boldsymbol{H}^{\mathrm{sct}})$ into the two domains as

$$(\boldsymbol{E}^{\mathrm{inc}}, \boldsymbol{H}^{\mathrm{inc}}) = \begin{cases} (\boldsymbol{E}_1^{\mathrm{inc}}, \boldsymbol{H}_1^{\mathrm{inc}}) & \text{in } \Omega_1, \\ (\boldsymbol{0}, \boldsymbol{0}) & \text{in } \Omega_2, \end{cases} \tag{1.18}$$

and

$$(\boldsymbol{E}^{\mathrm{sct}}, \boldsymbol{H}^{\mathrm{sct}}) = \begin{cases} (\boldsymbol{E}_1^{\mathrm{sct}}, \boldsymbol{H}_1^{\mathrm{sct}}) & \text{in } \Omega_1, \\ (\boldsymbol{E}_2^{\mathrm{sct}}, \boldsymbol{H}_2^{\mathrm{sct}}) & \text{in } \Omega_2, \end{cases} \tag{1.19}$$

where, by construction, $(\boldsymbol{E}_1^{\mathrm{sct}}, \boldsymbol{H}_1^{\mathrm{sct}})$ is only produced by the excitation source in Ω_1 and the jump conditions at the interface Γ, and $(\boldsymbol{E}_2^{\mathrm{sct}}, \boldsymbol{H}_2^{\mathrm{sct}})$ is only produced by the excitation source in Ω_2

(that is zero in our model problem) and the jump conditions at the interface Γ. We also impose that the magnetic permeability μ and electric permittivity ε are piecewise constant, i.e.,

$$\mu = \begin{cases} \mu_1 & \text{if } \boldsymbol{r} \in \Omega_1\,, \\ \mu_2 & \text{if } \boldsymbol{r} \in \Omega_2\,, \end{cases} \quad \text{and} \quad \varepsilon = \begin{cases} \varepsilon_1 & \text{if } \boldsymbol{r} \in \Omega_1\,, \\ \varepsilon_2 & \text{if } \boldsymbol{r} \in \Omega_2\,. \end{cases} \tag{1.20}$$

Moreover, the scattered field $(\boldsymbol{E}^{\text{sct}}, \boldsymbol{H}^{\text{sct}})$ has to verify the Sommerfeld's radiation condition (1.17) that describes its behavior at infinity.

1.3.2 Expressing the Maxwell's Equations in the Whole Space

We go back to the time-harmonic Maxwell's equations (1.13) that are first expressed separately in each domain

$$\begin{cases} \boldsymbol{\nabla} \times \boldsymbol{H}_1 - j\omega\varepsilon_1 \boldsymbol{E}_1 = \boldsymbol{0} \\ \boldsymbol{\nabla} \times \boldsymbol{E}_1 + j\omega\mu_1 \boldsymbol{H}_1 = \boldsymbol{0} \\ \varepsilon_1 \boldsymbol{\nabla} \cdot \boldsymbol{E}_1 = 0 \\ \mu_1 \boldsymbol{\nabla} \cdot \boldsymbol{H}_1 = 0 \end{cases} \quad \text{in } \Omega_1, \tag{1.21}$$

and

$$\begin{cases} \boldsymbol{\nabla} \times \boldsymbol{H}_2 - j\omega\varepsilon_2 \boldsymbol{E}_2 = \boldsymbol{0} \\ \boldsymbol{\nabla} \times \boldsymbol{E}_2 + j\omega\mu_2 \boldsymbol{H}_2 = \boldsymbol{0} \\ \varepsilon_2 \boldsymbol{\nabla} \cdot \boldsymbol{E}_2 = 0 \\ \mu_2 \boldsymbol{\nabla} \cdot \boldsymbol{H}_2 = 0 \end{cases} \quad \text{in } \Omega_2, \tag{1.22}$$

with $(\boldsymbol{E}_1, \boldsymbol{H}_1) = (\boldsymbol{E}_1^{\text{sct}}, \boldsymbol{H}_1^{\text{sct}}) + (\boldsymbol{E}_1^{\text{inc}}, \boldsymbol{H}_1^{\text{inc}})$ in Ω_1, and $(\boldsymbol{E}_2, \boldsymbol{H}_2) = (\boldsymbol{E}_2^{\text{sct}}, \boldsymbol{H}_2^{\text{sct}})$ in Ω_2.

To be able to recast these partial differential equations into surface integral equations, we first need to extend their domains of application to the whole three-dimensional space $\mathbb{R}^3$. After extending the incident field $(\boldsymbol{E}_1^{\text{inc}}, \boldsymbol{H}_1^{\text{inc}})$ into the whole space $\mathbb{R}^3$—that would exist in absence of any obstacle—we extend the solution in Ω_1 by the field in $\mathbb{R}^3$

$$(\tilde{\boldsymbol{E}}_1, \tilde{\boldsymbol{H}}_1) = \begin{cases} (\boldsymbol{E}_1^{\text{sct}}, \boldsymbol{H}_1^{\text{sct}}) & \text{in } \Omega_1\,, \\ (-\boldsymbol{E}_1^{\text{inc}}, -\boldsymbol{H}_1^{\text{inc}}) & \text{in } \Omega_2\,, \end{cases} \tag{1.23}$$

and the solution in Ω_2 by the field in $\mathbb{R}^3$

$$(\tilde{\boldsymbol{E}}_2, \tilde{\boldsymbol{H}}_2) = \begin{cases} (\boldsymbol{0}, \boldsymbol{0}) & \text{in } \Omega_1\,, \\ (\boldsymbol{E}_2^{\text{sct}}, \boldsymbol{H}_2^{\text{sct}}) & \text{in } \Omega_2\,. \end{cases} \tag{1.24}$$

These vector fields are constructed so that the total field in $\mathbb{R}^3$ is

$$(\boldsymbol{E}, \boldsymbol{H}) = (\tilde{\boldsymbol{E}}_1, \tilde{\boldsymbol{H}}_1) + (\tilde{\boldsymbol{E}}_2, \tilde{\boldsymbol{H}}_2) + (\boldsymbol{E}_1^{\text{inc}}, \boldsymbol{H}_1^{\text{inc}}) = \begin{cases} (\boldsymbol{E}_1, \boldsymbol{H}_1) & \text{in } \Omega_1\,, \\ (\boldsymbol{E}_2, \boldsymbol{H}_2) & \text{in } \Omega_2\,, \end{cases} \tag{1.25}$$

where, to insist, $(\boldsymbol{E}_1^{\text{inc}}, \boldsymbol{H}_1^{\text{inc}})$ is continuous and is defined in the whole space.

This modification of the original problem relies on the superposition principle, which shows that we can introduce an arbitrary incident field in a given domain, providing the corresponding scattered field is adjusted to cancel it inside that domain. This approach is motivated by the need for a continuous incident field in $\mathbb{R}^3$, ensuring that the discontinuities of the total field across interfaces are solely represented by the scattered fields (1.23) and (1.24) in the equivalent problem.

1.3.3 Jump Formulae

Let us consider a piecewise smooth function u in $\mathbb{R}^3$ that has a discontinuity across a smooth and oriented surface Γ—which separates two media Ω_1 and Ω_2. We define

$$u := \begin{cases} u_1 & \text{in } \Omega_1\,, \\ u_2 & \text{in } \Omega_2\,, \end{cases} \tag{1.26}$$

and we denote by $\boldsymbol{n} = (n_1, n_2, n_3)^\intercal$ the outgoing unit normal to Γ—that goes from Ω_2 to Ω_1 in our model problem.

Proposition 1.1 (Jump formula for scalar functions)**.** The jump formula gives the partial derivatives of u, in the sense of distributions, in terms of the partial derivatives of u_1 in Ω_1 and u_2 in Ω_2 and of the jump $[u] := (u_1 - u_2)_{|\Gamma}$ across Γ

$$\partial_{r_i} u = \{\partial_{r_i} u\} + [u]\, n_i\, \delta_\Gamma\,, \tag{1.27}$$

where

$$\{\partial_{r_i} u\} = \begin{cases} \partial_{r_i} u_1 & \text{in } \Omega_1\,, \\ \partial_{r_i} u_2 & \text{in } \Omega_2\,. \end{cases} \tag{1.28}$$

Proof. For any smooth vector field $\boldsymbol{\varphi}$, the integration by parts gives us, for $d \in \{1,2\}$,

$$\begin{aligned} \int_{\Omega_d} \boldsymbol{\nabla}\cdot\boldsymbol{\varphi}\, u \,\mathrm{d}V &= -\int_{\Omega_d} \boldsymbol{\varphi}\cdot\boldsymbol{\nabla} u \,\mathrm{d}V \pm \int_\Gamma \boldsymbol{\varphi}\cdot(-\boldsymbol{n})\, u \,\mathrm{d}S \\ &= -\int_{\Omega_d} \boldsymbol{\varphi}\cdot(\boldsymbol{\nabla} u \pm u\,\boldsymbol{n}\,\delta_\Gamma)\,\mathrm{d}V\,, \end{aligned} \tag{1.29}$$

where the sign in front of the surface integral is positive if the integration is done in Ω_1 and negative if it is done in Ω_2. This proves a vector field identity that is the jump formula for the gradient, for which each component gives (1.27). □

Proposition 1.2 (Jump formulae for the rotational and the divergence)**.** Consider a piecewise smooth vector field

$$\boldsymbol{u} := \begin{cases} \boldsymbol{u}_1 & \text{in } \Omega_1\,, \\ \boldsymbol{u}_2 & \text{in } \Omega_2\,, \end{cases} \tag{1.30}$$

that has a discontinuity across Γ. From the scalar case (1.27), we deduce the jump formula for its rotational

$$\boldsymbol{\nabla}\times\boldsymbol{u} = \{\boldsymbol{\nabla}\times\boldsymbol{u}\} + \left[\boldsymbol{n}\times\boldsymbol{u}\right]\delta_\Gamma\,, \tag{1.31}$$

where

$$\{\boldsymbol{\nabla}\times\boldsymbol{u}\} = \begin{cases} \boldsymbol{\nabla}\times\boldsymbol{u}_1 & \text{in } \Omega_1\,, \\ \boldsymbol{\nabla}\times\boldsymbol{u}_2 & \text{in } \Omega_2\,, \end{cases} \tag{1.32}$$

and $\left[\boldsymbol{n}\times\boldsymbol{u}\right] := (\boldsymbol{n}\times\boldsymbol{u}_1 - \boldsymbol{n}\times\boldsymbol{u}_2)_{|\Gamma}$ along Γ, while the jump formula for its divergence is

$$\boldsymbol{\nabla}\cdot\boldsymbol{u} = \{\boldsymbol{\nabla}\cdot\boldsymbol{u}\} + \left[\boldsymbol{u}\cdot\boldsymbol{n}\right]\delta_\Gamma\,, \tag{1.33}$$

where

$$\{\boldsymbol{\nabla}\cdot\boldsymbol{u}\} = \begin{cases} \boldsymbol{\nabla}\cdot\boldsymbol{u}_1 & \text{in } \Omega_1\,, \\ \boldsymbol{\nabla}\cdot\boldsymbol{u}_2 & \text{in } \Omega_2\,, \end{cases} \tag{1.34}$$

and $\left[\boldsymbol{u}\cdot\boldsymbol{n}\right] := (\boldsymbol{u}_1\cdot\boldsymbol{n} - \boldsymbol{u}_2\cdot\boldsymbol{n})_\Gamma$.

From the extension of the total field (1.25) in $\mathbb{R}^3$, from the boundary condition (1.14), and because—by construction—the extension of the incident field in $\mathbb{R}^3$ is continuous across Γ, we have the following jump relations

$$\begin{aligned} \left[\boldsymbol{n} \times \widetilde{\boldsymbol{H}}_1\right] + \left[\boldsymbol{n} \times \widetilde{\boldsymbol{H}}_2\right] &= \boldsymbol{0}\,, \\ \left[\widetilde{\boldsymbol{E}}_1 \times \boldsymbol{n}\right] + \left[\widetilde{\boldsymbol{E}}_2 \times \boldsymbol{n}\right] &= \boldsymbol{0}\,, \end{aligned} \tag{1.35}$$

and

$$\begin{aligned} \left[\varepsilon_1 \widetilde{\boldsymbol{E}}_1 \cdot \boldsymbol{n}\right] + \left[\varepsilon_2 \widetilde{\boldsymbol{E}}_2 \cdot \boldsymbol{n}\right] &= 0\,, \\ \left[\mu_1 \widetilde{\boldsymbol{H}}_1 \cdot \boldsymbol{n}\right] + \left[\mu_2 \widetilde{\boldsymbol{H}}_2 \cdot \boldsymbol{n}\right] &= 0\,. \end{aligned} \tag{1.36}$$

This shows that we can define, at the interface Γ, the electric and magnetic surface current densities, that verify

$$\begin{aligned} \boldsymbol{J} &= \left[\boldsymbol{n} \times \widetilde{\boldsymbol{H}}_1\right] = -\left[\boldsymbol{n} \times \widetilde{\boldsymbol{H}}_2\right]\,, \\ \boldsymbol{M} &= \left[\widetilde{\boldsymbol{E}}_1 \times \boldsymbol{n}\right] = -\left[\widetilde{\boldsymbol{E}}_2 \times \boldsymbol{n}\right]\,, \end{aligned} \tag{1.37}$$

as well as the electric and magnetic surface charge densities, that verify

$$\begin{aligned} \rho_e &= \left[\varepsilon_1 \widetilde{\boldsymbol{E}}_1 \cdot \boldsymbol{n}\right] = -\left[\varepsilon_2 \widetilde{\boldsymbol{E}}_2 \cdot \boldsymbol{n}\right]\,, \\ \rho_m &= \left[\mu_1 \widetilde{\boldsymbol{H}}_1 \cdot \boldsymbol{n}\right] = -\left[\mu_2 \widetilde{\boldsymbol{H}}_2 \cdot \boldsymbol{n}\right]\,. \end{aligned} \tag{1.38}$$

Moreover, the incident field $(\boldsymbol{E}_1^{\text{inc}}, \boldsymbol{H}_1^{\text{inc}})$ is a solution of the homogeneous Maxwell's equations (1.13) as it has no jump across Γ. By application of the jump formulae (1.31) and (1.33), we subsequently extend the systems (1.21) and (1.22) in $\mathbb{R}^3$, in the sense of distributions, as

$$\begin{cases} \boldsymbol{\nabla} \times \widetilde{\boldsymbol{H}}_1 - j\omega\varepsilon_1 \widetilde{\boldsymbol{E}}_1 = \boldsymbol{J}\,\delta_\Gamma\,, \\ \boldsymbol{\nabla} \times \widetilde{\boldsymbol{E}}_1 + j\omega\mu_1 \widetilde{\boldsymbol{H}}_1 = -\boldsymbol{M}\,\delta_\Gamma\,, \\ \varepsilon_1 \boldsymbol{\nabla} \cdot \widetilde{\boldsymbol{E}}_1 = \rho_e\,\delta_\Gamma\,, \\ \mu_1 \boldsymbol{\nabla} \cdot \widetilde{\boldsymbol{H}}_1 = \rho_h\,\delta_\Gamma\,, \end{cases} \tag{1.39}$$

and

$$\begin{cases} \boldsymbol{\nabla} \times \widetilde{\boldsymbol{H}}_2 - j\omega\varepsilon_2 \widetilde{\boldsymbol{E}}_2 = -\boldsymbol{J}\,\delta_\Gamma\,, \\ \boldsymbol{\nabla} \times \widetilde{\boldsymbol{E}}_2 + j\omega\mu_2 \widetilde{\boldsymbol{H}}_2 = \boldsymbol{M}\,\delta_\Gamma\,, \\ \varepsilon_2 \boldsymbol{\nabla} \cdot \widetilde{\boldsymbol{E}}_2 = -\rho_e\,\delta_\Gamma\,, \\ \mu_2 \boldsymbol{\nabla} \cdot \widetilde{\boldsymbol{H}}_2 = -\rho_m\,\delta_\Gamma\,. \end{cases} \tag{1.40}$$

1.3.4 Vector Helmholtz Equation

We thus have two systems of partial differential equations, with non-zero righ-hand sides, in the form

$$\begin{cases} \boldsymbol{\nabla} \times \widetilde{\boldsymbol{H}} - j\omega\varepsilon\, \widetilde{\boldsymbol{E}} = \boldsymbol{J}\delta_\Gamma\,, \\ \boldsymbol{\nabla} \times \widetilde{\boldsymbol{E}} + j\omega\mu\, \widetilde{\boldsymbol{H}} = -\boldsymbol{M}\delta_\Gamma\,, \\ \varepsilon \boldsymbol{\nabla} \cdot \widetilde{\boldsymbol{E}} = \rho_e\,\delta_\Gamma\,, \\ \mu \boldsymbol{\nabla} \cdot \widetilde{\boldsymbol{H}} = \rho_h\,\delta_\Gamma\,, \end{cases} \tag{1.41}$$

where we omitted to write the subscripts for the two domains. Simply notice that the signs convention here corresponds to that of Ω_1 and that we simply have to take the right-hand side with opposite signs for Ω_2. To solve this system, we express its solution as the sum of the two solutions $(\tilde{\boldsymbol{E}}_e, \tilde{\boldsymbol{H}}_e)$ and $(\tilde{\boldsymbol{E}}_h, \tilde{\boldsymbol{H}}_h)$ found when $\rho_h = 0$ and $\rho_e = 0$, respectively. This technique follows the steps derived in [40].

Case where the magnetic surface charge density is zero.

For twice differentiable vector and scalar functions $\boldsymbol{A}$ and ψ (without discontinuity), we have the following two properties

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \boldsymbol{A}) = 0\,, \tag{1.42}$$

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla}\psi) = \boldsymbol{0}\,. \tag{1.43}$$

These identities, compared with the second and fourth equations of the system (1.41) (with $\rho_h = 0$) show that $\tilde{\boldsymbol{E}}_e$ and $\tilde{\boldsymbol{H}}_e$ can be written as

$$\tilde{\boldsymbol{E}}_e = -\boldsymbol{\nabla}\tilde{\psi} - j\omega\tilde{\boldsymbol{A}}\,, \tag{1.44}$$

$$\tilde{\boldsymbol{H}}_e = \frac{1}{\mu}\boldsymbol{\nabla} \times \tilde{\boldsymbol{A}}\,, \tag{1.45}$$

where $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{\psi}}$ are continuous but their derivatives have a jump across Γ, so that,

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \tilde{\boldsymbol{A}}) = 0\,, \tag{1.46}$$

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla}\tilde{\psi}) = \boldsymbol{M}\delta_\Gamma\,. \tag{1.47}$$

The vector field $\tilde{\boldsymbol{A}}$ and the scalar $\tilde{\psi}$ are called the magnetic vector potential and the electric scalar potential, respectively.

Moreover, we have some degrees of freedom in the choice of the scalar and vector potentials so that we can arbitrarily set the Lorenz gauge

$$\boldsymbol{\nabla} \cdot \tilde{\boldsymbol{A}} = -j\omega\varepsilon\mu\,\tilde{\psi}\,. \tag{1.48}$$

By inserting (1.44), (1.45) and (1.48) into the first equation in (1.41) and by using the property

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \tilde{\boldsymbol{A}}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \tilde{\boldsymbol{A}}) - \boldsymbol{\Delta}\tilde{\boldsymbol{A}}\,, \tag{1.49}$$

we find the following vector Helmholtz equation for the magnetic vector potential

$$-(\boldsymbol{\Delta} + k^2)\tilde{\boldsymbol{A}} = \mu\boldsymbol{J}\delta_\Gamma\,, \tag{1.50}$$

while using the property

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla}\psi) = \Delta\tilde{\psi}\,, \tag{1.51}$$

together with the third equation in (1.41) and the Lorenz gauge (1.48) gives us the scalar Helmholtz equation for the electric scalar potential

$$-(\Delta + k^2)\tilde{\psi} = \frac{\rho_e}{\varepsilon}\,\delta_\Gamma\,. \tag{1.52}$$

Case where the electric surface charge density is zero.

Equally, we use the vector identities (1.42) applied to the first and third equations of the system (1.41) (with $\rho_e = 0$) to show that $\tilde{\boldsymbol{E}}_h$ and $\tilde{\boldsymbol{H}}_h$ can be written as

$$\tilde{\boldsymbol{E}}_h = -\frac{1}{\varepsilon}\boldsymbol{\nabla} \times \tilde{\boldsymbol{F}} \tag{1.53}$$

$$\tilde{\boldsymbol{H}}_h = -\boldsymbol{\nabla}\tilde{\varphi} - j\omega\tilde{\boldsymbol{F}}\,, \tag{1.54}$$

where $\tilde{\boldsymbol{F}}$ and $\tilde{\boldsymbol{\varphi}}$ are continuous but their derivatives have a jump across Γ, so that,

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \tilde{\boldsymbol{F}}) = 0\,, \tag{1.55}$$

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla}\tilde{\varphi}) = -\boldsymbol{J}\delta_\Gamma\,. \tag{1.56}$$

The vector field $\tilde{\boldsymbol{F}}$ and the scalar $\tilde{\varphi}$ are called the electric vector potential and the magnetic scalar potential, respectively. Further, as in the previous case, we set the Lorenz gauge

$$\boldsymbol{\nabla} \cdot \tilde{\boldsymbol{F}} = -j\omega\varepsilon\mu\,\tilde{\varphi}\,. \tag{1.57}$$

By inserting (1.53), (1.54) and (1.57) into the second equation in (1.41) and by using the property (1.49), we find the following vector Helmholtz equation for the electric vector potential

$$-(\boldsymbol{\Delta} + k^2)\tilde{\boldsymbol{F}} = \varepsilon\boldsymbol{M}\delta_\Gamma\,, \tag{1.58}$$

while using the property (1.51) together with the fourth equation in (1.41) and the Lorenz gauge (1.57) gives us the scalar Helmholtz equation for the magnetic scalar potential

$$-(\Delta + k^2)\tilde{\varphi} = \frac{\rho_h}{\mu}\,\delta_\Gamma\,. \tag{1.59}$$

Case where both magnetic and electric surface charge densities are non-zero.

The general case where neither $\rho_e = 0$ nor $\rho_h = 0$ is found as the summation of the solutions (1.44)–(1.45) and (1.53)–(1.54) and by using the Lorenz gauges (1.48), (1.57)

$$\tilde{\boldsymbol{E}} = \tilde{\boldsymbol{E}}_e + \tilde{\boldsymbol{E}}_h = -j\omega\tilde{\boldsymbol{A}} + \frac{1}{j\omega\mu\varepsilon}\boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \tilde{\boldsymbol{A}}) - \frac{1}{\varepsilon}\boldsymbol{\nabla} \times \tilde{\boldsymbol{F}}\,, \tag{1.60}$$

$$\tilde{\boldsymbol{H}} = \tilde{\boldsymbol{H}}_e + \tilde{\boldsymbol{H}}_h = \frac{1}{\mu}\boldsymbol{\nabla} \times \tilde{\boldsymbol{A}} - j\omega\tilde{\boldsymbol{F}} + \frac{1}{j\omega\mu\varepsilon}\boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \tilde{\boldsymbol{F}})\,. \tag{1.61}$$

Solving the vector Helmholtz equations (1.50), (1.58) for the magnetic and electric vector potentials $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{F}}$ will finally allow us to express the electromagnetic field $(\tilde{\boldsymbol{E}}, \tilde{\boldsymbol{H}})$ in terms of surface potentials applied to the electric and magnetic surface current densities $\boldsymbol{J}$ and $\boldsymbol{M}$. The determination of these surface current densities, that will become the functional unknowns of surface integral equations, will allow us to find the solution of the Maxwell's equations.

1.3.5 Functional Spaces

Now that we have set the surface current density unknowns $\boldsymbol{J}$ and $\boldsymbol{M}$, we may be interested in defining the correct spaces they belong to.

For physics reasons, we look for a finite energy electromagnetic field, which translates into mathematics as $\widetilde{\boldsymbol{E}} \in L^2(\mathbb{R}^3)^3$ and $\widetilde{\boldsymbol{H}} \in L^2(\mathbb{R}^3)^3$. From the Maxwell-Ampère and Maxwell-Faraday equations (1.13), this also implies that $\boldsymbol{\nabla} \times \widetilde{\boldsymbol{E}} \in L^2(\mathbb{R}^3)^3$ and $\boldsymbol{\nabla} \times \widetilde{\boldsymbol{H}} \in L^2(\mathbb{R}^3)^3$. We thus define

$$H(\boldsymbol{\nabla}\times, \mathbb{R}^3) \coloneqq \left\{ \boldsymbol{u} \in L^2(\mathbb{R}^3)^3 \ : \ \boldsymbol{\nabla} \times \boldsymbol{u} \in L^2(\mathbb{R}^3)^3 \right\} \tag{1.62}$$

as the functional space to which $\widetilde{\boldsymbol{E}}$ and $\widetilde{\boldsymbol{H}}$ belong, that verifies the inclusions $H^1(\mathbb{R}^3)^3 \subset H(\boldsymbol{\nabla}\times, \mathbb{R}^3) \subset L^2(\mathbb{R}^3)^3$, where $H^1(\mathbb{R}^3)$ is the space of functions in $L^2(\mathbb{R}^3)$ for which the gradient is in $L^2(\mathbb{R}^3)^3$.

For a scalar function $u \in H^1(\mathbb{R}^3)$, the trace theorems [106] on the surface boundary Γ state that $u_{|\Gamma} \in H^{1/2}(\Gamma)$ and $\partial_{\boldsymbol{n}} u_{|\Gamma} \equiv (\boldsymbol{\nabla} u \cdot \boldsymbol{n})_{|\Gamma} \in H^{-1/2}(\Gamma)$ with

$$H^{1/2}(\Gamma) \coloneqq \left\{ u \in L^2(\Gamma) \ : \ \int_\Gamma \int_\Gamma \frac{|u(\boldsymbol{r}) - u(\boldsymbol{r}')|^2}{|\boldsymbol{r}' - \boldsymbol{r}|^3} \, \mathrm{d}S(\boldsymbol{r}) \mathrm{d}S(\boldsymbol{r}') < +\infty \right\} , \tag{1.63}$$

and where $H^{-1/2}(\Gamma)$ is the functional space that is dual to $H^{1/2}(\Gamma)$ [29]. These spaces verify the inclusions $H^{1/2}(\Gamma) \subset L^2(\Gamma) \subset H^{-1/2}(\Gamma)$.

Next, we define the surface gradient $\mathbf{grad}_\Gamma$ as the surface tangent component of the "standard" gradient $\boldsymbol{\nabla}$. For a smooth function φ defined in the neighborhood of Γ, we thus have, for $r \in \Gamma$,

$$\mathbf{grad}_\Gamma \varphi(\boldsymbol{r}) \coloneqq \boldsymbol{n}(\boldsymbol{r}) \times (\boldsymbol{\nabla}\varphi(\boldsymbol{r}) \times \boldsymbol{n}(\boldsymbol{r})) \, . \tag{1.64}$$

Conversely, if we study a function φ that is only defined on the surface Γ, we can think of a smooth prolongation $\widetilde{\varphi}$ defined in the neighborhood of Γ such that $\mathbf{grad}_\Gamma \varphi \equiv \mathbf{grad}_\Gamma \widetilde{\varphi}$. By duality, we define the surface divergence $\mathrm{div}_\Gamma \boldsymbol{u}$ of a sufficiently smooth surface tangent vector field $\boldsymbol{u}$ so that for every smooth function φ

$$\int_\Gamma \varphi \, \mathrm{div}_\Gamma \boldsymbol{u} \, \mathrm{d}S(\boldsymbol{r}) = - \int_\Gamma \mathbf{grad}_\Gamma \varphi \cdot \boldsymbol{u} \, \mathrm{d}S(\boldsymbol{r}) \, . \tag{1.65}$$

From the trace theorem [106], the surface current densities $\boldsymbol{J}$ and $\boldsymbol{M}$ belong to the functional space

$$H^{-1/2}(\mathrm{div}_\Gamma, \Gamma) \coloneqq \left\{ \boldsymbol{u} \in H_t^{-1/2}(\Gamma) \ : \ \mathrm{div}_\Gamma \boldsymbol{u} \in H^{-1/2}(\Gamma) \right\} , \tag{1.66}$$

where $H_t^{-1/2}$ is the space dual to

$$H_t^{1/2}(\Gamma) \coloneqq \left\{ \boldsymbol{u} \in L_t^2(\Gamma) \ : \ \int_\Gamma \int_\Gamma \frac{|\boldsymbol{u}(\boldsymbol{r}) - \boldsymbol{u}(\boldsymbol{r}')|^2}{|\boldsymbol{r}' - \boldsymbol{r}|^3} \, \mathrm{d}S(\boldsymbol{r}) \mathrm{d}S(\boldsymbol{r}') < +\infty \right\} . \tag{1.67}$$

Furthermore, we define the surface vector and scalar curl operators for sufficiently smooth scalar function φ and vector field $\boldsymbol{u}$ as

$$\mathbf{curl}_\Gamma \varphi = \mathbf{grad}_\Gamma \varphi \times \boldsymbol{n} \, , \tag{1.68}$$

$$\mathrm{curl}_\Gamma \boldsymbol{u} = \mathrm{div}_\Gamma (\boldsymbol{u} \times \boldsymbol{n}) \, , \tag{1.69}$$

respectively, with the duality stemming from (1.65)

$$\int_\Gamma \varphi \, \mathrm{curl}_\Gamma \boldsymbol{u} \, \mathrm{d}S(\boldsymbol{r}) = \int_\Gamma \mathbf{curl}_\Gamma \varphi \cdot \boldsymbol{u} \, \mathrm{d}S(\boldsymbol{r}) \, . \tag{1.70}$$

Finally, it can be shown that [29]

$$\begin{aligned} H^{-1/2}(\mathrm{curl}_\Gamma, \Gamma) &:= \left\{ \boldsymbol{v} \in H_t^{-1/2}(\Gamma) \ : \ \mathrm{curl}_\Gamma \boldsymbol{v} \in H^{-1/2}(\Gamma) \right\} \\ &= \left\{ \boldsymbol{v} = \boldsymbol{n} \times \boldsymbol{u} \ : \ \boldsymbol{u} \in H^{-1/2}(\mathrm{div}_\Gamma, \Gamma) \right\} \, . \end{aligned} \tag{1.71}$$

1.4 Stratton-Chu Representation Formula

1.4.1 Green's Kernel for the Helmholtz Equation

The Green's kernel is a mathematical tool that allows for solving linear partial differential equations with non-zero right-hand side that are expressed in the sense of distributions as

$$\mathfrak{D}_{\boldsymbol{r}} \boldsymbol{u}(\boldsymbol{r}) = \boldsymbol{f}(\boldsymbol{r}) \, , \tag{1.72}$$

where $\mathfrak{D}_{\boldsymbol{r}}$ is a linear differential operator applied to functions in $\mathbb{R}^3$ that incorporates boundary conditions in its definition. It is possible to express the solution $\boldsymbol{u}$ from the solution of the fundamental equation that is given, in the sense of distributions, as

$$\mathfrak{D}_{\boldsymbol{r}} g(|\boldsymbol{r}|) = \delta_0(|\boldsymbol{r}|) \, , \tag{1.73}$$

where g is called the Green's kernel. The solution of (1.72) is indeed given, up to the solutions of the homogeneous equation $\mathfrak{D}_{\boldsymbol{r}} g(|\boldsymbol{r}|) = 0$, by

$$\boldsymbol{u}(\boldsymbol{r}) = \int_{\mathbb{R}^3} g(|\boldsymbol{r}' - \boldsymbol{r}|) \boldsymbol{f}(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}' \, , \tag{1.74}$$

as we check that

$$\begin{aligned} \mathfrak{D}_{\boldsymbol{r}} \boldsymbol{u}(\boldsymbol{r}) &= \int_{\mathbb{R}^3} \mathfrak{D}_{\boldsymbol{r}} g(|\boldsymbol{r}' - \boldsymbol{r}|) \boldsymbol{f}(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}' \\ &= \int_{\mathbb{R}^3} \delta_0(|\boldsymbol{r}' - \boldsymbol{r}|) \boldsymbol{f}(\boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}' = \boldsymbol{f}(\boldsymbol{r}) \, . \end{aligned} \tag{1.75}$$

Proposition 1.3 (Green's kernel of the Helmholtz equation [106])**.** Except for some wavenumbers k_*, for which the squares are the eigenvalues of the Laplacian operator $-\Delta$, the unique fundamental solution of the Helmholtz equation, which solves the fundamental equation in the sense of distributions

$$-\left(\Delta + k^2\right) g = \delta_0 \, , \tag{1.76}$$

and that verifies the Sommerfeld's condition (1.17) is

$$g(|\boldsymbol{r}|) = \frac{e^{-jk|\boldsymbol{r}|}}{4\pi|\boldsymbol{r}|} \, . \tag{1.77}$$

This fundamental solution is called the Green's kernel of the Helmholtz equation, that will also refer to the function of two-variables

$$G(\boldsymbol{r}, \boldsymbol{r}') = g(|\boldsymbol{r}' - \boldsymbol{r}|) = \frac{e^{-jk|\boldsymbol{r}' - \boldsymbol{r}|}}{4\pi|\boldsymbol{r}' - \boldsymbol{r}|} \, . \tag{1.78}$$

Proof. In spherical reference, we notice that the right-hand side $\delta_0(|\boldsymbol{r}|)$ is independent of the polar and azimuthal angles. Hence, the equation is invariant by rotation and becomes

$$-\left(\frac{1}{r^2}\partial_r\left(r^2\partial_r g(r)\right)+k^2 g(r)\right)=\delta_0(r)\,, \tag{1.79}$$

where $r\equiv|\boldsymbol{r}|$ is the radial distance of the spherical coordinates.

One can check that the solution of the associated homogeneous equation (with zero right-hand side) is in the form

$$g(r)=A\frac{e^{jkr}}{r}+B\frac{e^{-jkr}}{r}\,, \tag{1.80}$$

where A and B are two constants to be determined. Moreover, in spherical coordinates, the Sommerfeld's condition becomes

$$\lim_{r\to+\infty} r\left(\partial_r g(r)+jkg(r)\right)=0\,, \tag{1.81}$$

which implies $A=0$.

Notice that it is relevant to consider the solution of the homogeneous equation as the right-hand side is zero whenever r is not zero. Taking into account $\delta_0(r)$ as the right-hand side will allow us to determine B. To do so, we first define a ball $B(\epsilon)$ of radius ϵ centered at the origin $r=0$, and we denote by $S(\epsilon)$ its surface boundary, that is the sphere of radius ϵ. One can easily show that

$$\lim_{\epsilon\to0}\int_{B(\epsilon)} g(|\boldsymbol{r}|)\,\mathrm{d}V(\boldsymbol{r})=\lim_{\epsilon\to0}\int_0^{2\pi}\int_0^{\pi}\int_0^{\epsilon} g(r)\,r^2 sin\theta\,\mathrm{d}r\mathrm{d}\theta\mathrm{d}\phi=0\,, \tag{1.82}$$

as $g(r)r^2$ is bounded in the limit of $r\to0$ (and it goes to zero, in fact). Moreover, it follows from the divergence theorem that

$$\begin{aligned}\int_{B(\epsilon)}\Delta g(|\boldsymbol{r}|)\,\mathrm{d}V(\boldsymbol{r}) &= \int_{S(\epsilon)}\boldsymbol{\nabla}g(|\boldsymbol{r}|)\cdot\boldsymbol{n}_\epsilon(\boldsymbol{r})\,\mathrm{d}S(\boldsymbol{r})\\ &= \int_0^{2\pi}\int_0^{\pi}\partial_r g(\epsilon)\,\epsilon^2\sin\theta\,\mathrm{d}\theta\mathrm{d}\phi\\ &= -4\pi B\,(1+jk\epsilon)\,e^{-jk\epsilon}\,,\end{aligned} \tag{1.83}$$

where $\boldsymbol{n}_\epsilon(\boldsymbol{r})=\boldsymbol{r}/|\boldsymbol{r}|$, for $\boldsymbol{r}\in S(\epsilon)$, is the outgoing unit normal to the sphere $S(\epsilon)$. From (1.82) and (1.83), and by definition of the fundamental equation (1.76), we find

$$\begin{aligned}\lim_{\epsilon\to0}\int_{B(\epsilon)}-\left(\Delta+k^2\right)g(|\boldsymbol{r}|)\,\mathrm{d}V(\boldsymbol{r}) &= 4\pi B\\ &= \lim_{\epsilon\to0}\int_{B(\epsilon)}\delta_0(|\boldsymbol{r}|)\,\mathrm{d}V(\boldsymbol{r})=1\,,\end{aligned} \tag{1.84}$$

which proves $B=1/(4\pi)$. □

1.4.2 Expressing the Electromagnetic Field in Terms of Surface Potentials

By application of Proposition 1.3, the magnetic and electric vector potentials, solutions of the vector Helmholtz equations (1.50) and (1.58) are given by, for $\boldsymbol{r} \notin \Gamma$,

$$\tilde{\boldsymbol{A}}(\boldsymbol{r}) = \mu \int_\Gamma G(\boldsymbol{r},\boldsymbol{r}')\boldsymbol{J}(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,, \tag{1.85}$$

$$\tilde{\boldsymbol{F}}(\boldsymbol{r}) = \varepsilon \int_\Gamma G(\boldsymbol{r},\boldsymbol{r}')\boldsymbol{M}(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,. \tag{1.86}$$

By inserting these solutions into (1.60) and (1.61) we find an expression of the electric and magnetic fields $\tilde{\boldsymbol{E}}$ and $\tilde{\boldsymbol{H}}$ in terms of surface potentials applied to the electric and magnetic surface current densities $\boldsymbol{J}$ and $\boldsymbol{M}$. Before giving an explicit expression of the solution of the original problem (1.21)–(1.22) we give the following result that will allow to slightly rearrange the fomulation.

Proposition 1.4 (Divergence of the vector potentials)**.** Consider the divergence of the surface vector potential applied to a surface current density $\boldsymbol{u} \in H^{-1/2}(\mathrm{div}_\Gamma,\Gamma)$. The following property holds

$$\boldsymbol{\nabla}\cdot\int_\Gamma G(\boldsymbol{r},\boldsymbol{r}')\boldsymbol{u}(r')\,\mathrm{d}S(\boldsymbol{r}') = \int_\Gamma G(\boldsymbol{r},\boldsymbol{r}')\mathrm{div}_\Gamma\boldsymbol{u}(r')\,\mathrm{d}S(\boldsymbol{r}')\,. \tag{1.87}$$

Proof. We first write

$$\boldsymbol{\nabla}\cdot\int_\Gamma G(\boldsymbol{r},\boldsymbol{r}')\boldsymbol{u}(r')\,\mathrm{d}S(\boldsymbol{r}') = \int_\Gamma \boldsymbol{\nabla}G(\boldsymbol{r},\boldsymbol{r}')\cdot\boldsymbol{u}(r')\,\mathrm{d}S(\boldsymbol{r}'). \tag{1.88}$$

Then, we use the fact that $\boldsymbol{\nabla}G(\boldsymbol{r},\boldsymbol{r}') = -\boldsymbol{\nabla}'G(\boldsymbol{r},\boldsymbol{r}')$ and it follows that $\boldsymbol{\nabla}'G(\boldsymbol{r},\boldsymbol{r}')\cdot\boldsymbol{u}(\boldsymbol{r}') = \mathrm{grad}'_\Gamma G(\boldsymbol{r},\boldsymbol{r}')\cdot\boldsymbol{u}(\boldsymbol{r}')$, because $\boldsymbol{u}$ is tangent to Γ. We finally proceed to an integration by parts according to the definition of the surface divergence (1.65). Remark that the prime on div_Γ is omitted for simplicity of notations, as there is no ambiguity with $\boldsymbol{u}$ only depending on $\boldsymbol{r}'$. □

From equations (1.60)–(1.61) and by using the solutions of the vector potentials (1.85)–(1.86) together with the result of Proposition (1.4), we find the following expression for the solutions of (1.21) and (1.22), that is the so-called Stratton-Chu representation formula [40]

$$\begin{cases} \tilde{\boldsymbol{E}}_1 &= -j\omega\mu_1\mathcal{S}_{s,1}\boldsymbol{J} + \frac{1}{j\omega\varepsilon_1}\mathcal{S}_{h,1}\boldsymbol{J} - \mathcal{D}_1\boldsymbol{M}\,,\\ \tilde{\boldsymbol{H}}_1 &= -j\omega\varepsilon_1\mathcal{S}_{s,1}\boldsymbol{M} + \frac{1}{j\omega\mu_1}\mathcal{S}_{h,1}\boldsymbol{M} + \mathcal{D}_1\boldsymbol{J}\,,\\ \tilde{\boldsymbol{E}}_2 &= j\omega\mu_2\mathcal{S}_{s,2}\boldsymbol{J} - \frac{1}{j\omega\varepsilon_2}\mathcal{S}_{h,2}\boldsymbol{J} + \mathcal{D}_2\boldsymbol{M}\,,\\ \tilde{\boldsymbol{H}}_2 &= j\omega\varepsilon_2\mathcal{S}_{s,2}\boldsymbol{M} - \frac{1}{j\omega\mu_2}\mathcal{S}_{h,2}\boldsymbol{M} - \mathcal{D}_2\boldsymbol{J}\,, \end{cases} \tag{1.89}$$

with the potentials, for $\boldsymbol{u} \in H^{-1/2}(\mathrm{div}_\Gamma,\Gamma)$, $d \in \{1,2\}$ and $\boldsymbol{r} \notin \Gamma$,

$$\left(\mathcal{S}_{s,d}\boldsymbol{u}\right)(\boldsymbol{r}) \coloneqq \int_\Gamma G_d(\boldsymbol{r},\boldsymbol{r}')\boldsymbol{u}(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,, \tag{1.90}$$

$$\left(\mathcal{S}_{h,d}\boldsymbol{u}\right)(\boldsymbol{r}) \coloneqq \boldsymbol{\nabla}\int_\Gamma G_d(\boldsymbol{r},\boldsymbol{r}')\mathrm{div}_\Gamma\boldsymbol{u}(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,, \tag{1.91}$$

$$\left(\mathcal{D}_d\boldsymbol{u}\right)(\boldsymbol{r}) \coloneqq \boldsymbol{\nabla}\times\int_\Gamma G_d(\boldsymbol{r},\boldsymbol{r}')\boldsymbol{u}(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,, \tag{1.92}$$

where we have indexed the Green's kernel by the indices of the media Ω_1 and Ω_2 as it depends on the wavenumbers $k_1 := \omega\sqrt{\varepsilon_1\mu_1}$ and $k_2 := \omega\sqrt{\varepsilon_2\mu_2}$.

Using the definitions (1.23) and (1.24), we find the following integral representations of the electromagnetic fields outside Γ

$$\begin{cases} \boldsymbol{E}_1 = \boldsymbol{E}_1^{\text{inc}} - j\omega\mu_1\mathcal{S}_{s,1}\boldsymbol{J} + \frac{1}{j\omega\varepsilon_1}\mathcal{S}_{h,1}\boldsymbol{J} - \mathcal{D}_1\boldsymbol{M} \\ \boldsymbol{H}_1 = \boldsymbol{H}_1^{\text{inc}} - j\omega\varepsilon_1\mathcal{S}_{s,1}\boldsymbol{M} + \frac{1}{j\omega\mu_1}\mathcal{S}_{h,1}\boldsymbol{M} + \mathcal{D}_1\boldsymbol{J} \end{cases} \quad \text{in } \Omega_1 \setminus \Gamma\,, \tag{1.93}$$

and inside Γ

$$\begin{cases} \boldsymbol{E}_2 = j\omega\mu_2\mathcal{S}_{s,2}\boldsymbol{J} - \frac{1}{j\omega\varepsilon_2}\mathcal{S}_{h,2}\boldsymbol{J} + \mathcal{D}_2\boldsymbol{M} \\ \boldsymbol{H}_2 = j\omega\varepsilon_2\mathcal{S}_{s,2}\boldsymbol{M} - \frac{1}{j\omega\mu_2}\mathcal{S}_{h,2}\boldsymbol{M} - \mathcal{D}_2\boldsymbol{J} \end{cases} \quad \text{in } \Omega_2 \setminus \Gamma\,. \tag{1.94}$$

1.4.3 Tangent Trace of the Surface Potentials

In order to derive the surface integral equations—that will leverage the boundary conditions (1.14)—we need to express the tangent trace on Γ of the electromagnetic field, from inside and outside Γ. From (1.93) and (1.94), we can express them as functions of the tangent trace of the surface potentials (1.90)–(1.92).

Definition 1.1 (Principal value [106])**.** Let us consider an integrable function $f(\boldsymbol{r},\boldsymbol{r}')$ on Γ and the associated surface integral operator

$$\mathcal{L}(\boldsymbol{r}) = \int_\Gamma f(\boldsymbol{r},\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,. \tag{1.95}$$

The Cauchy principal value of $\mathcal{L}$ is defined as the following limit

$$\text{p.v.}\,\mathcal{L}(\boldsymbol{r}) = \lim_{\epsilon\to 0}\int_{\Gamma\setminus B(\boldsymbol{r},\epsilon)} f(\boldsymbol{r},\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,, \tag{1.96}$$

where $B(\boldsymbol{r},\epsilon)$ is the ball centered on $\boldsymbol{r}$ and of radius ϵ.

Proposition 1.5 (Tangent traces of the potentials, [106])**.** Let us denote by Γ the interface between two domains Ω_1 and Ω_2. The tangent traces of the potentials $\mathcal{S}_{s,d}$ and $\mathcal{S}_{h,d}$, with $d \in \{1,2\}$, are continuous across Γ and for every $\boldsymbol{r} \in \Gamma$ where Γ admits a tangent plane, we find

$$\begin{aligned} \left(\mathcal{T}_{s,d}\boldsymbol{u}\right)(\boldsymbol{r}) &:= \boldsymbol{n}(\boldsymbol{r}) \times \left(\lim_{\widetilde{\boldsymbol{r}}\in\Omega_d\to\boldsymbol{r}} \left(\mathcal{S}_{s,d}\boldsymbol{u}\right)(\widetilde{\boldsymbol{r}})\right) \\ &= \boldsymbol{n}(\boldsymbol{r}) \times \int_\Gamma G_d(\boldsymbol{r},\boldsymbol{r}')\boldsymbol{u}(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,, \end{aligned} \tag{1.97}$$

$$\begin{aligned} \left(\mathcal{T}_{h,d}\boldsymbol{u}\right)(\boldsymbol{r}) &:= \boldsymbol{n}(\boldsymbol{r}) \times \left(\lim_{\widetilde{\boldsymbol{r}}\in\Omega_d\to\boldsymbol{r}} \left(\mathcal{S}_{h,d}\boldsymbol{u}\right)(\widetilde{\boldsymbol{r}})\right) \\ &= \boldsymbol{n}(\boldsymbol{r}) \times \mathbf{grad}_\Gamma \int_\Gamma G_d(\boldsymbol{r},\boldsymbol{r}')\mathrm{div}_\Gamma\boldsymbol{u}(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,. \end{aligned} \tag{1.98}$$

Differently, the tangent trace of the potentials $\mathcal{D}_d$ (for $d \in \{1,2\}$) is discontinuous across Γ and for every $\boldsymbol{r} \in \Gamma$ where Γ admits a tangent plane, we find

$$\begin{aligned} \boldsymbol{n}(\boldsymbol{r}) \times \left(\lim_{\widetilde{\boldsymbol{r}}\in\Omega_1\to\boldsymbol{r}} \left(\mathcal{D}_1\boldsymbol{u}\right)(\widetilde{\boldsymbol{r}})\right) &= \left(\mathcal{K}_1\boldsymbol{u}\right)(\boldsymbol{r}) - \frac{\boldsymbol{u}(\boldsymbol{r})}{2}\,, \\ \boldsymbol{n}(\boldsymbol{r}) \times \left(\lim_{\widetilde{\boldsymbol{r}}\in\Omega_2\to\boldsymbol{r}} \left(\mathcal{D}_2\boldsymbol{u}\right)(\widetilde{\boldsymbol{r}})\right) &= \left(\mathcal{K}_2\boldsymbol{u}\right)(\boldsymbol{r}) + \frac{\boldsymbol{u}(\boldsymbol{r})}{2}\,, \end{aligned}$$

where for $d \in \{1,2\}$

$$(\mathcal{K}_d \boldsymbol{u})(\boldsymbol{r}) \coloneqq \boldsymbol{n}(\boldsymbol{r}) \times \text{p.v.} \int_\Gamma \boldsymbol{\nabla} G_d(\boldsymbol{r}, \boldsymbol{r}') \times \boldsymbol{u}(\boldsymbol{r}') \, \mathrm{d}S(\boldsymbol{r}') \,, \tag{1.100}$$

and where the gradient of the Green's kernel (1.78) is found as, for $d \in \{1,2\}$

$$\boldsymbol{\nabla} G_d(\boldsymbol{r}, \boldsymbol{r}') = \left(\frac{jk_d|\boldsymbol{r}' - \boldsymbol{r}| + 1}{4\pi|\boldsymbol{r}' - \boldsymbol{r}|^2} \right) \frac{\boldsymbol{r}' - \boldsymbol{r}}{|\boldsymbol{r}' - \boldsymbol{r}|} e^{-jk_d|\boldsymbol{r}' - \boldsymbol{r}|} \,. \tag{1.101}$$

Proof. At a point $\boldsymbol{r} \in \Gamma$ where Γ admits a tangent plane, the surface integral of a function $f_d(\boldsymbol{r}, \boldsymbol{r}')$ can be split in the form

$$\begin{aligned} \boldsymbol{n}(\boldsymbol{r}) \times \int_\Gamma f_d(\boldsymbol{r}, \boldsymbol{r}') \, \mathrm{d}S(\boldsymbol{r}') = \boldsymbol{n}(\boldsymbol{r}) \times \text{p.v.} \int_\Gamma f_d(\boldsymbol{r}, \boldsymbol{r}') \, \mathrm{d}S(\boldsymbol{r}') \\ \pm \boldsymbol{n}_d(\boldsymbol{r}) \times \lim_{\epsilon \to 0} \int_{hS_d(\boldsymbol{r},\epsilon)} f_d(\boldsymbol{r}, \boldsymbol{r}') \, \mathrm{d}S(\boldsymbol{r}') \,, \end{aligned} \tag{1.102}$$

where $hS_d(\boldsymbol{r}, \epsilon)$ is the half-sphere of radius ϵ centered on $\boldsymbol{r} \in \Gamma$ that is located in Ω_d and where the sign in front of the second integral is positive if $\boldsymbol{r}' \in \Omega_1$ and negative if $\boldsymbol{r}' \in \Omega_2$ so that $\boldsymbol{n}_d = \pm \boldsymbol{n}$ is the unit normal going from inside to outside Ω_d. The integral on $hS_d(\boldsymbol{r}, \epsilon)$ can be expressed in spherical coordinates as

$$\int_{h\mathcal{S}(\boldsymbol{r},\epsilon)} f_d(\boldsymbol{r}, \boldsymbol{r}') \, \mathrm{d}S(\boldsymbol{r}') = \int_0^{2\pi} \int_0^{\pi/2} f_d(\boldsymbol{r}, \boldsymbol{r}') \, \epsilon^2 \sin\theta \, \mathrm{d}\theta \mathrm{d}\phi \,, \tag{1.103}$$

where

$$\boldsymbol{r}' = \boldsymbol{r} + \epsilon \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix} . \tag{1.104}$$

For the integral operator $\mathcal{S}_{s,d}$, the limit of this integral is zero, while it tends towards a vector orthogonal to the surface for $\mathcal{S}_{h,d}$, so that there is no discontinuity in the tangent direction for both operators. For $\mathcal{D}_d$, however, we can write from (1.101) and $|\boldsymbol{r}' - \boldsymbol{r}| = \epsilon$

$$\begin{aligned} &\boldsymbol{\nabla} \times \int_{hS_d(\boldsymbol{r},\epsilon)} G_d(\boldsymbol{r}, \boldsymbol{r}') \boldsymbol{u}(\boldsymbol{r}') \, \mathrm{d}S(\boldsymbol{r}') \\ &= \int_0^{2\pi} \int_0^{\pi/2} \boldsymbol{\nabla} G_d(\boldsymbol{r}, \boldsymbol{r}') \times \boldsymbol{u}(\boldsymbol{r}') \, \epsilon^2 \sin\theta \, \mathrm{d}\theta \mathrm{d}\phi \\ &= \int_0^{2\pi} \int_0^{\pi/2} \left(\frac{jk_d\epsilon + 1}{4\pi} \right) \boldsymbol{n}_\epsilon(\boldsymbol{r}) \, e^{-jk_d\epsilon} \times \boldsymbol{u}(\boldsymbol{r}') \sin\theta \, \mathrm{d}\theta \mathrm{d}\phi \,, \end{aligned} \tag{1.105}$$

where the outgoing unit normal to $hS_d(\boldsymbol{r}, \epsilon)$ is

$$\boldsymbol{n}_\epsilon(\boldsymbol{r}) \coloneqq \frac{\boldsymbol{r}' - \boldsymbol{r}}{|\boldsymbol{r}' - \boldsymbol{r}|} \underset{\epsilon \to 0}{\to} \boldsymbol{n}_d(\boldsymbol{r}) \,. \tag{1.106}$$

Taking the limit of (1.105) when ϵ goes to zero, we find

$$\boldsymbol{n}_d(\boldsymbol{r}) \times \lim_{\epsilon \to 0} \boldsymbol{\nabla} \times \int_{h\mathcal{S}(\boldsymbol{r},\epsilon)} G_d(\boldsymbol{r}, \boldsymbol{r}') \boldsymbol{u}(\boldsymbol{r}') \, \mathrm{d}S(\boldsymbol{r}') = -\frac{\boldsymbol{u}(\boldsymbol{r})}{2} \,, \tag{1.107}$$

where we used $\boldsymbol{n}_d \times (\boldsymbol{n}_d \times \boldsymbol{u}) = -\boldsymbol{u}$ as $\boldsymbol{u}$ is a tangent vector.

□

We also define the surface integral operator

$$\mathcal{S}_d \coloneqq -jk_d\mathcal{S}_{s,d} + (jk_d)^{-1}\mathcal{S}_{h,d}\,. \tag{1.108}$$

From Proposition 1.5, the tangent trace of $\mathcal{S}_d$ is continuous across Γ where Γ admits a tangent plane, and we have for $d \in \{1,2\}$

$$\begin{aligned}(\mathcal{T}_d\boldsymbol{u})(\boldsymbol{r}) &\coloneqq \boldsymbol{n}(\boldsymbol{r}) \times \left(\lim_{\widetilde{\boldsymbol{r}}\in\Omega_d\to\boldsymbol{r}} (\mathcal{S}_d\boldsymbol{u})(\widetilde{\boldsymbol{r}})\right)\\ &= -jk_d\left(\mathcal{T}_{s,d}\boldsymbol{u}\right)(\boldsymbol{r}) + (jk_d)^{-1}\left(\mathcal{T}_{h,d}\boldsymbol{u}\right)(\boldsymbol{r})\,.\end{aligned} \tag{1.109}$$

1.5 Surface Integral Equations and Variational Problems for Dielectric and Metallic Boundaries

1.5.1 Case of a Dielectric Interface

The Integral equations relating the surface current densities $\boldsymbol{J}$ and $\boldsymbol{M}$ are obtained by using the Stratton-Chu representation formula (1.89) in conjunction with the boundary conditions (1.14). Subsequently, we obtain the system of integral equations that is widely-known as the Poggio-Miller-Chang-Harrington-Wu-Tsai (PMCHWT) formulation [146]

$$\begin{cases} \left(\eta_1\mathcal{T}_1 + \eta_2\mathcal{T}_2\right)\boldsymbol{J} - \left(\mathcal{K}_1 + \mathcal{K}_2\right)\boldsymbol{M} &= -\boldsymbol{n}\times\boldsymbol{E}^{\text{inc}}\,,\\ -\left(\mathcal{K}_1 + \mathcal{K}_2\right)\boldsymbol{J} - \left(\dfrac{1}{\eta_1}\mathcal{T}_1 + \dfrac{1}{\eta_2}\mathcal{T}_2\right)\boldsymbol{M} &= \boldsymbol{n}\times\boldsymbol{H}^{\text{inc}}\,,\end{cases} \tag{1.110}$$

where $\eta_d = \sqrt{\mu_d/\varepsilon_d}$, for $d \in \{1,2\}$, are the charecteristic impedances of the media Ω_1 and Ω_2.

The variational problem associated with the PMCHWT formulation is

Find $\boldsymbol{J}, \boldsymbol{M} \in H^{-1/2}(\text{div}_\Gamma, \Gamma)$ such that for every $\boldsymbol{J}', \boldsymbol{M}' \in H^{-1/2}(\text{div}_\Gamma, \Gamma)$:

$$\begin{cases} \left\langle \boldsymbol{n}\times\boldsymbol{J}', \left(\eta_1\mathcal{T}_1 + \eta_2\mathcal{T}_2\right)\boldsymbol{J}\right\rangle - \left\langle \boldsymbol{n}\times\boldsymbol{J}', \left(\mathcal{K}_1 + \mathcal{K}_2\right)\boldsymbol{M}\right\rangle &= -\left\langle \boldsymbol{J}', \boldsymbol{E}^{\text{inc}}\right\rangle\\ -\left\langle \boldsymbol{n}\times\boldsymbol{M}', \left(\mathcal{K}_1 + \mathcal{K}_2\right)\boldsymbol{J}\right\rangle - \left\langle \boldsymbol{n}\times\boldsymbol{M}', \left(\dfrac{1}{\eta_1}\mathcal{T}_1 + \dfrac{1}{\eta_2}\mathcal{T}_2\right)\boldsymbol{M}\right\rangle &= \left\langle \boldsymbol{M}', \boldsymbol{H}^{\text{inc}}\right\rangle\end{cases} \tag{1.111}$$

1.5.2 Case of a PEC Boundary

When Γ is a perfectly electrically conducting boundary (that is used to model the surface of metallic objects), the PEC boundary condition (1.16) implies that $\boldsymbol{M} = \boldsymbol{0}$. From this same boundary condition, and in conjunction with the Stratton-Chu representation formula (1.89), we get two surface integral equations known as the electric field integral equation (EFIE) [146]

$$\mathcal{T}\boldsymbol{J} = -\frac{1}{\eta}\boldsymbol{n}\times\boldsymbol{E}^{\text{inc}}\,, \tag{1.112}$$

and the magnetic field integral equation (MFIE) [146]

$$\left(\tfrac{1}{2}\mathcal{I} - \mathcal{K}\right)\boldsymbol{J} = \boldsymbol{n} \times \boldsymbol{H}^{\text{inc}}\,, \tag{1.113}$$

where $\mathcal{I}$ is the identity operator. Notice that for both equations we removed the indices relating the media Ω_1 and Ω_2, as there is no ambiguity with the electric field being zero in Ω_2 (the characteristic constants without indices are thus those of Ω_1).

As discussed in Proposition 1.3, the Green's kernel is not unique for some frequencies related to the eigenvalues of the Laplacian operator. It results that the EFIE and MFIE suffer from numerical instability due to the presence of spurious resonances at some frequencies [2]. Instead, the weighted sum of these two equations, known as the combined field integral equation (CFIE), is free from spurious resonances, for $\alpha \in]0,1[$, [146]

$$\begin{aligned}\alpha\mathcal{T}\boldsymbol{J} + (1-\alpha)\boldsymbol{n} \times \left(\tfrac{1}{2}\mathcal{I} - \mathcal{K}\right)\boldsymbol{J} \\ = -\frac{\alpha}{\eta}\boldsymbol{n} \times \boldsymbol{E}^{\text{inc}} + (1-\alpha)\boldsymbol{n} \times \boldsymbol{n} \times \boldsymbol{H}^{\text{inc}}\,.\end{aligned} \tag{1.114}$$

Subsequently, we provide the variational problems for the EFIE and MFIE.

Electric Field Integral Equation

Find $\boldsymbol{J} \in H^{-1/2}(\text{div}_\Gamma, \Gamma)$ such that for every $\boldsymbol{J}' \in H^{-1/2}(\text{div}_\Gamma, \Gamma)$:

$$\left\langle \boldsymbol{n} \times \boldsymbol{J}', \mathcal{T}\boldsymbol{J} \right\rangle = -\frac{1}{\eta}\left\langle \boldsymbol{n} \times \boldsymbol{J}', \boldsymbol{n} \times \boldsymbol{E}^{\text{inc}} \right\rangle \tag{1.115}$$

Magnetic Field Integral Equation

Find $\boldsymbol{J} \in H^{-1/2}(\text{div}_\Gamma, \Gamma)$ such that for every $\boldsymbol{J}' \in H^{-1/2}(\text{div}_\Gamma, \Gamma)$:

$$\left\langle \boldsymbol{n} \times \boldsymbol{J}', \left(\tfrac{1}{2}\mathcal{I} - \mathcal{K}\right)\boldsymbol{J} \right\rangle = \left\langle \boldsymbol{n} \times \boldsymbol{J}', \boldsymbol{n} \times \boldsymbol{H}^{\text{inc}} \right\rangle \tag{1.116}$$

1.6 Calderón Identities

From the definition of the surface current densities (1.37) and from the Stratton-Chu representation formula (1.89), we can write the following identity, that stands for any $\boldsymbol{J}, \boldsymbol{M} \in H^{-1/2}(\text{div}_\Gamma, \Gamma)$, for $d \in \{1,2\}$,

$$\begin{pmatrix} \frac{1}{2}\mathcal{I} \mp \mathcal{K}_d & \mp\mathcal{T}_d \\ \pm\mathcal{T}_d & \frac{1}{2}\mathcal{I} \mp \mathcal{K}_d \end{pmatrix} \begin{pmatrix} \boldsymbol{M} \\ \eta_d \boldsymbol{J} \end{pmatrix} = \begin{pmatrix} \boldsymbol{M} \\ \eta_d \boldsymbol{J} \end{pmatrix}, \tag{1.117}$$

where in "±" and "∓" it is understood that the "upper" sign stands for $d = 1$ and the "lower" sign for $d = 2$. This shows that the operators

$$C_d := \begin{pmatrix} \frac{1}{2}\mathcal{I} \mp \mathcal{K}_d & \mp\mathcal{T}_d \\ \pm\mathcal{T}_d & \frac{1}{2}\mathcal{I} \mp \mathcal{K}_d \end{pmatrix} \tag{1.118}$$

are projectors, that are known as the Calderón projectors [82].

It follows that $(\mathcal{I} - C_d) \circ C_d = 0$, which can be written as

$$\begin{aligned} &\begin{pmatrix} -\frac{1}{2}\mathcal{I} \mp \mathcal{K}_d & \mp\mathcal{T}_d \\ \pm\mathcal{T}_d & -\frac{1}{2}\mathcal{I} \mp \mathcal{K}_d \end{pmatrix} \circ \begin{pmatrix} \frac{1}{2}\mathcal{I} \mp \mathcal{K}_d & \mp\mathcal{T}_d \\ \pm\mathcal{T}_d & \frac{1}{2}\mathcal{I} \mp \mathcal{K}_d \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{K}_d \circ \mathcal{K}_d - \frac{1}{4}\mathcal{I} - \mathcal{T}_d \circ \mathcal{T}_d & \mathcal{T}_d \circ \mathcal{K}_d + \mathcal{K}_d \circ \mathcal{T}_d \\ -\mathcal{T}_d \circ \mathcal{K}_d - \mathcal{K}_d \circ \mathcal{T}_d & \mathcal{K}_d \circ \mathcal{K}_d - \frac{1}{4}\mathcal{I} - \mathcal{T}_d \circ \mathcal{T}_d \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned} \tag{1.119}$$

which in turns implies the two Calderón identities

$$\mathcal{T}_d \circ \mathcal{T}_d = -\tfrac{1}{4}\mathcal{I} + \mathcal{K}_d \circ \mathcal{K}_d\,, \tag{1.120}$$

and

$$\mathcal{T}_d \circ \mathcal{K}_d = -\mathcal{K}_d \circ \mathcal{T}_d\,. \tag{1.121}$$

For smooth enough surface boundaries Γ, the operator $\mathcal{K}_d$ is compact so that $\mathcal{T}_d \circ \mathcal{T}_d$ is a scaling of the identity operator plus a compact perturbation. Its spectrum is thus bounded with eigenvalues clustering at $-1/4$ [42].

Chapter 2

High-Order Boundary Element Method

In this chapter, we describe the boundary element method that allows to discretize the variational problems of the integral equations that we previously derived. Overall, the method needs the following three ingredients:

- An approximation of the surface boundary Γ by a set of adjacent geometrical elements that is referred to as a mesh discretization.
- An approximation of the Sobolev functional spaces by finite dimensional subspaces associated with the elements of the mesh, that are referred to as finite element spaces, or boundary element spaces in the context of surface boundary equations.
- Numerical methods to approximate the surface integrals appearing in integral operators and inner-products of the variational problems, that are referred to as quadrature rules.

2.1 Mesh Discretization of Surface Boundaries

2.1.1 Reference Triangle

In this thesis, we consider the case where the surface Γ is approximated by a mesh of curved triangles [67, 115]. We start by defining a reference triangle $\widehat{K}$ that is parameterized with local coordinates

$$\widehat{\boldsymbol{r}} = (u, v)^{\mathsf{T}}, \tag{2.1}$$

such that

$$\widehat{K} := \{0 \leq u \leq 1;\ 0 \leq v \leq 1 - u\}\ . \tag{2.2}$$

Here and in the following of the document, the symbols "with hat" will refer to local variables while the symbols "without hat" will refer to global variables.

We consider the three vertices of the reference triangle

$$\widehat{\boldsymbol{p}}_1 = (1,0)^{\mathsf{T}}, \quad \widehat{\boldsymbol{p}}_2 = (0,1)^{\mathsf{T}}, \quad \widehat{\boldsymbol{p}}_3 = (0,0)^{\mathsf{T}}, \tag{2.3}$$

and we denote by $\widehat{e}_b$ the edge that is opposite to $\widehat{\boldsymbol{p}}_b$, i.e., $\widehat{e}_1$ is the edge between $\widehat{\boldsymbol{p}}_2$ and $\widehat{\boldsymbol{p}}_3$, $\widehat{e}_2$ is the edge between $\widehat{\boldsymbol{p}}_3$ and $\widehat{\boldsymbol{p}}_1$, while $\widehat{e}_3$ is the edge between $\widehat{\boldsymbol{p}}_1$ and $\widehat{\boldsymbol{p}}_2$. Finally, we denote by $\widehat{\boldsymbol{\nu}}$ the unit outgoing normal to the edges of $\widehat{K}$, and more specifically, $\widehat{\boldsymbol{\nu}}_b$ is the unit outgoing normal to edge $\widehat{e}_b$.

These different elements are illustrated in Figure 2.1.

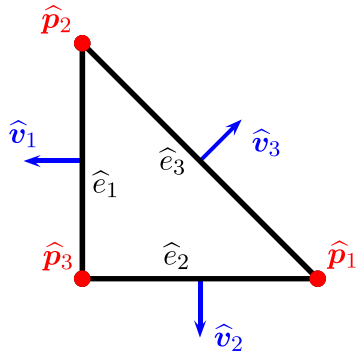


Figure 2.1: Reference triangle and its geometrical elements. The size of the outgoing unit normals are shrunk and should equal the length of the edges.

2.1.2 Local-to-Global Parametrization

For each cell K of the mesh, we build a local-to-global parametrization that maps the local coordinates $\widehat{\boldsymbol{r}} \in \widehat{K}$ to the global points $\boldsymbol{r} \in K$

$$\begin{aligned} \widehat{F}_K : \widehat{K} &\longrightarrow K \\ \widehat{\boldsymbol{r}} &\longmapsto \boldsymbol{r} = \widehat{F}_K(\widehat{\boldsymbol{r}}) \end{aligned} , \tag{2.4}$$

and that ensures the continuity of the mesh along the borders of the triangles. Such parametrization can be obtained with the Lagrange interpolation of control points that are located on the surface Γ. The lowest order $q = 1$ of such interpolation provides flat triangles, where the interpolatory points are located on the vertices of the reference triangle. For quadratic triangles ($q = 2$), the interpolatory points are the vertices and the edge-centroids of the reference triangle. Notice that for general geometries, it is not possible to perfectly fit the mesh to the surface Γ with Lagrange interpolation, and only the control points are ensured to be located on Γ. A more general setting is provided by the non-uniform rationale B-splines (NURBS) that offer additional degrees of freedom in choosing the control points and imposing regularity on the borders between mesh elements [137].

2.1.3 Unit Normal and Jacobian Matrices

Given this parametrization, we denote by

$$\mathbf{J}_{\widehat{F}_K} = \begin{pmatrix} \partial_u \widehat{F}_K & \partial_v \widehat{F}_K \end{pmatrix}^\mathsf{T} \tag{2.5}$$

the Jacobian of $\widehat{F}_K$ (a 2×3 matrix) and by

$$\mathbf{J}_{\widehat{F}_K^{-1}} = \begin{pmatrix} \mathbf{grad}_\Gamma u & \mathbf{grad}_\Gamma v \end{pmatrix} \tag{2.6}$$

the Jacobian of the inverse mapping $\widehat{F}_K^{-1}$ (a 3×2 matrix). Moreover, we have [115]

$$\mathbf{J}_{\widehat{F}_K} \mathbf{J}_{\widehat{F}_K^{-1}} = \mathbf{I} , \tag{2.7}$$

where $\mathbf{I}$ is the two-dimensional identity matrix.

The area of the elementary surface is

$$\widehat{\mathcal{J}}_K = \sqrt{\det\left(\mathbf{J}_{\widehat{F}_K}\mathbf{J}_{\widehat{F}_K}^{\mathsf{T}}\right)} = |\partial_u \widehat{F}_K \times \partial_v \widehat{F}_K|\,, \tag{2.8}$$

and the outgoing unit normal to K is

$$\widehat{\boldsymbol{n}}_K = \frac{\partial_u \widehat{F}_K \times \partial_v \widehat{F}_K}{\widehat{\mathcal{J}}_K}\,, \tag{2.9}$$

that is expressed in global coordinates as $\boldsymbol{n}(\boldsymbol{r}) = \widehat{\boldsymbol{n}}_K \circ \widehat{F}_K^{-1}(\boldsymbol{r})$, for $\boldsymbol{r} \in K$, which is therefore the outgoing unit normal to the mesh approximation of Γ.

Furthermore, the property (2.7) implies that

$$\begin{aligned} \mathbf{grad}_\Gamma u &= \frac{1}{\widehat{\mathcal{J}}_K}\partial_v \widehat{F}_K \times \widehat{\boldsymbol{n}}_K\,, \\ \mathbf{grad}_\Gamma v &= \frac{1}{\widehat{\mathcal{J}}_K}\widehat{\boldsymbol{n}}_K \times \partial_u \widehat{F}_K\,, \end{aligned} \tag{2.10}$$

and it follows

$$\widehat{\boldsymbol{n}}_K = \widehat{\mathcal{J}}_K \mathbf{grad}_\Gamma u \times \mathbf{grad}_\Gamma v\,. \tag{2.11}$$

We can now prove the following relationship between the Jacobian of $\widehat{F}_K$ and that of $\widehat{F}_K^{-1}$. Let us define an arbitrary vector $\widehat{\boldsymbol{x}} = (x_1, x_2)^{\mathsf{T}} \in \widehat{K}$ and the ninety-degree rotation matrix

$$\widehat{\mathbf{R}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{2.12}$$

which is the counterpart of the $\boldsymbol{u} \mapsto \boldsymbol{n} \times \boldsymbol{u}$ operator in the reference triangle. We thus have from (2.10)

$$\begin{aligned} \frac{1}{\widehat{\mathcal{J}}_K}\mathbf{J}_{\widehat{F}_K}^{\mathsf{T}}\widehat{\mathbf{R}}\widehat{\boldsymbol{x}} &= \frac{1}{\widehat{\mathcal{J}}_K}\mathbf{J}_{\widehat{F}_K}^{\mathsf{T}}\begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} = \frac{1}{\widehat{\mathcal{J}}_K}\left(x_1\partial_v \widehat{F}_K - x_2\partial_u \widehat{F}_K\right) \\ &= x_1\widehat{\boldsymbol{n}}_K \times \mathbf{grad}_\Gamma u + x_2\widehat{\boldsymbol{n}}_K \times \mathbf{grad}_\Gamma v \\ &= \widehat{\boldsymbol{n}}_K \times \left(\mathbf{J}_{F_K^{-1}}\widehat{\boldsymbol{x}}\right). \end{aligned} \tag{2.13}$$

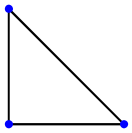
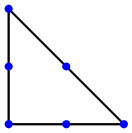
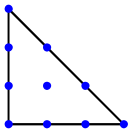
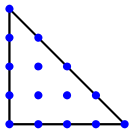
Remark 2.1. If the geometry is planar, we can define a mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$, so that $\mathbf{J}_{\widehat{F}_K}$ and $\mathbf{J}_{\widehat{F}_K^{-1}}$ are two-dimensional square matrices, and $\widehat{\mathcal{J}}_K = \det \mathbf{J}_{\widehat{F}_K}$. As we are in two dimensions, $\widehat{\boldsymbol{n}}_K\times$ is replaced by $\widehat{\mathbf{R}}$, so that we get

$$\widehat{\mathbf{R}}\mathbf{J}_{\widehat{F}_K^{-1}} = \frac{1}{\widehat{\mathcal{J}}_K}\mathbf{J}_{\widehat{F}_K}^{\mathsf{T}}\widehat{\mathbf{R}}\,. \tag{2.14}$$

By using $\widehat{\mathbf{R}}^{-1} = \widehat{\mathbf{R}}^{\mathsf{T}}$ and the inverse formula for two-dimensional matrices, we finally obtain

$$\mathbf{J}_{\widehat{F}_K^{-1}} = \frac{1}{\widehat{\mathcal{J}}_K}\widehat{\mathbf{R}}^{\mathsf{T}}\mathbf{J}_{\widehat{F}_K}^{\mathsf{T}}\widehat{\mathbf{R}} = \mathbf{J}_{\widehat{F}_K}^{-1} \quad \text{(only in 2D!)}\,. \tag{2.15}$$

Table 2.1: Location of the Lagrange interpolatory points in the reference triangle.

	$q=1$	$q=2$	$q=3$	$q=4$
$\widehat{\boldsymbol{r}}^{(q)}_{m,n}$				

2.1.4 Example with the Lagrange interpolation

In this section, we present the parametrization of the mesh triangles by uniform Lagrange interpolation—with interpolatory points on the triangles that are uniformly distributed, as shown in Table 2.1.

For a given order of interpolation $q \geq 1$, we first define the Silvester-Lagrange polynomials of a single variable $u \in [0,1]$ as

$$\mathscr{L}^{(q)}_n(u) := \begin{cases} 1 & n = 0\,, \\ \dfrac{1}{n!}\displaystyle\prod_{m=0}^{n-1}(qu-m) & n = 1,\dots,q\,. \end{cases} \tag{2.16}$$

These functions are polynomials of degree n. For $1 \leq n \leq q$, they are unity at $\frac{n}{q}$ and zero at $u = 0, \frac{1}{q}, \dots, \frac{n-1}{q}$, while it is constant equal to one for $n = 0$. On the reference triangle of local coordinates $\widehat{\boldsymbol{r}} = (u, v)$, we define the Lagrange polynomials of two variables as

$$\mathscr{L}^{(q)}_{\ell,m,n}(\widehat{\boldsymbol{r}}) = \mathscr{L}^{(q)}_{\ell}(u)\mathscr{L}^{(q)}_{m}(v)\mathscr{L}^{(q)}_{n}(1-u-v)\,, \tag{2.17}$$

with $0 \leq \ell, m, n \leq q$ and $\ell + m + n = q$. These functions interpolate at the points $\widehat{\boldsymbol{r}}^{(q)}_{s,t} = (\frac{s}{q}, \frac{t}{q})$, with $0 \leq s + t \leq q$ (see Table 2.1), i.e.,

$$\mathscr{L}^{(q)}_{\ell,m,n}(\widehat{\boldsymbol{r}}^{(q)}_{s,t}) = \delta_{\ell,s}\delta_{m,t}\,, \tag{2.18}$$

where $\delta_{m,n}$ is the Kronecker symbol.

The parametrization of the triangles K in the mesh approximation of Γ is thus given by

$$\widehat{F}_K(\widehat{\boldsymbol{r}}) = \sum_{\ell=0}^{q}\sum_{n=0}^{q-\ell} \boldsymbol{p}^K_{\ell,m,q-\ell-m}\mathscr{L}^{(q)}_{\ell,m,q-\ell-m}(\widehat{\boldsymbol{r}})\,, \tag{2.19}$$

where the control points $\boldsymbol{p}^K_{\ell,m,n}$ are built such that $\widehat{F}_K(\widehat{\boldsymbol{r}}_{s,t}) = \boldsymbol{p}^K_{s,t,q-s-t} \in \Gamma$. These points can be obtained from various algorithms that are called meshers. In this thesis, the meshes used for the numerical experiments are made with the software Gmsh [60] which provides Lagrange interpolatory meshes up to order 10 and features to design geometries.

Next, we compute the partial derivatives of the mesh parametrization

$$\begin{aligned}\partial_u\widehat{F}_K(\widehat{\boldsymbol{r}}) = \sum_{\ell=0}^{q}\sum_{m=0}^{q-\ell} \boldsymbol{p}^K_{\ell,m,n}\Big(&\dot{\mathscr{L}}^{(q)}_{\ell}(u)\mathscr{L}^{(q)}_{m}(v)\mathscr{L}^{(q)}_{n}(1-u-v) \\ &- \mathscr{L}^{(q)}_{\ell}(u)\mathscr{L}^{(q)}_{m}(v)\dot{\mathscr{L}}^{(q)}_{n}(1-u-v)\Big)\,, \qquad (n = q-\ell-m)\,,\end{aligned} \tag{2.20}$$

$$\partial_v \widehat{F}_K(\widehat{\boldsymbol{r}}) = \sum_{\ell=0}^{q} \sum_{m=0}^{q-\ell} \boldsymbol{p}^K_{\ell,m,n} \Big(\mathscr{L}^{(q)}_\ell(u) \dot{\mathscr{L}}^{(q)}_m(v) \mathscr{L}^{(q)}_n(1-u-v) - \mathscr{L}^{(q)}_\ell(u) \mathscr{L}^{(q)}_m(v) \dot{\mathscr{L}}^{(q)}_n(1-u-v) \Big) , \qquad (n = q-\ell-m) , \tag{2.21}$$

where the first derivatives of the Silvester-Lagrange functions are found as

$$\dot{\mathscr{L}}^{(q)}_n(u) = \begin{cases} 0 & n = 0 , \\ \dfrac{q}{n!} \displaystyle\sum_{\ell=0}^{n-1} \prod_{\substack{m=0 \\ m\neq\ell}}^{n-1} (qu - m) & n > 0 . \end{cases} \tag{2.22}$$

The computation of the outgoing unit normal, area of the elementary surface, Jacobians of the mapping and inverse mapping is now a simple application of the above definitions.

2.1.5 Integration and Differentiation on Surface Boundaries

With this mapping, integration on Γ can be approximated by integration on the mesh, for any integrable function f in Γ as

$$\int_\Gamma f(\boldsymbol{r}) \, \mathrm{d}S(\boldsymbol{r}) \approx \sum_K \int_{\widehat{K}} f \circ \widehat{F}_K(\widehat{\boldsymbol{r}}) \, \widehat{\mathcal{J}}_K(\widehat{\boldsymbol{r}}) \, \mathrm{d}\widehat{\boldsymbol{r}} , \tag{2.23}$$

where for any integrable function g in $\widehat{K}$

$$\int_{\widehat{K}} g(\widehat{\boldsymbol{r}}) \, \mathrm{d}\widehat{\boldsymbol{r}} = \int_{u=0}^{1} \int_{v=0}^{1-u} g(u,v) \, \mathrm{d}v \mathrm{d}u . \tag{2.24}$$

In the following of the thesis, we will not repeat that the integration on the mesh is an approximation of the integration on Γ, and the symbol "Γ" will now refer to its mesh approximation, by abuse of notation.

Remark 2.2. It is very common to make use of the fact that the integrals are systematically computed via global-to-local variable transformations, which introduce the composition of functions in the form

$$f \circ \widehat{F}_K(\widehat{\boldsymbol{r}}) \coloneqq f(\widehat{F}_K(\widehat{\boldsymbol{r}})) . \tag{2.25}$$

As it will be shown hereafter, the boundary element functions in global coordinates will always be defined as mappings of local functions that depend on the local coordinates $\widehat{\boldsymbol{r}} = (u, v)$, which will introduce a composition with the inverse parametrization $\widehat{F}^{-1}_K(\boldsymbol{r})$. Very often in practice, this inverse parametrization is not analytically available, and the explicit computation of such mapping may resort to numerically solving a minimization problem giving the local coordinates $\widehat{\boldsymbol{r}}$ that solve $|\widehat{F}_K(\widehat{\boldsymbol{r}}) - \boldsymbol{r}| = 0$. While such computation is needed for some methods approximating singular integrals [52], it is not to directly evaluate the integrand in (2.23) as we will always have mappings of functions that are composed with $\widehat{F}^{-1}_K \circ \widehat{F}_K(\widehat{\boldsymbol{r}}) = \widehat{\boldsymbol{r}}$ (see for instance the next example (2.26) and replace f by λ in (2.23); this also works for every vector and scalar functions introduced thereafter). This also shows that the term $F_K(\widehat{\boldsymbol{r}})$ is not explicitly computed neither and that integration is directly done in local coordinates.

Now, let us consider a piecewise smooth function in the form , for $\boldsymbol{r} \in K$,

$$\lambda(\boldsymbol{r}) = \widehat{\lambda}_K \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,, \tag{2.26}$$

which is continuous across the edges of the mesh, i.e., $\widehat{\lambda}_K(\widehat{\boldsymbol{r}}) = \widehat{\lambda}_L(\widehat{\boldsymbol{r}}')$ for any $\widehat{\boldsymbol{r}} \in K$ and $\widehat{\boldsymbol{r}}' \in L$ such that $\widehat{F}_K(\widehat{\boldsymbol{r}}) = \widehat{F}_L(\widehat{\boldsymbol{r}}')$, where K and L are two adjacent triangles.

Proposition 2.1 (Surface gradient and vector curl of continuous functions)**.** The surface gradient and surface vector curl of λ are well defined and are given, for $\boldsymbol{r} \in K$, by

$$\mathbf{grad}_\Gamma \lambda(\boldsymbol{r}) = \left(\mathbf{J}_{\widehat{F}_K^{-1}} \widehat{\boldsymbol{\nabla}} \widehat{\lambda}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,, \tag{2.27}$$

$$\mathbf{curl}_\Gamma \lambda(\boldsymbol{r}) = \left(\frac{1}{\mathcal{J}_K} \mathbf{J}_{F_K}^\mathsf{T} \widehat{\mathbf{R}}^\mathsf{T} \widehat{\boldsymbol{\nabla}} \widehat{\lambda}_K \right) \circ F_K^{-1}(\boldsymbol{r}) \,. \tag{2.28}$$

Proof. Consider a continuous function $\tilde{\lambda}$ defined for $\boldsymbol{r} = (x, y, z)^\mathsf{T} \in \mathbb{R}^3$ and denote by λ its trace on Γ such that $\lambda(\boldsymbol{r}) = \tilde{\lambda}(\boldsymbol{r})$ for $r \in \Gamma$. Using the chain rule and the definition of the surface gradient (1.64), we have, for $\boldsymbol{r} \in K$,

$$\begin{aligned} \mathbf{grad}_\Gamma \lambda(\boldsymbol{r}) &= \boldsymbol{n}(\boldsymbol{r}) \times \boldsymbol{\nabla} \tilde{\lambda}(\boldsymbol{r}) \times \boldsymbol{n}(\boldsymbol{r}) \\ &= \boldsymbol{n}(\boldsymbol{r}) \times \left(\partial_u \widehat{\lambda}_K(\widehat{\boldsymbol{r}})\, \boldsymbol{\nabla} u + \partial_v \widehat{\lambda}_K(\widehat{\boldsymbol{r}}) \boldsymbol{\nabla} v \right) \times \boldsymbol{n}(\boldsymbol{r}) \\ &= \partial_u \widehat{\lambda}_K(\widehat{\boldsymbol{r}})\, \mathbf{grad}_\Gamma u + \partial_v \widehat{\lambda}_K(\widehat{\boldsymbol{r}})\, \mathbf{grad}_\Gamma v \\ &= \left(\mathbf{J}_{\widehat{F}_K^{-1}} \widehat{\boldsymbol{\nabla}} \widehat{\lambda}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,. \end{aligned} \tag{2.29}$$

The identity for the curl comes from (2.13) which shows, for $\boldsymbol{r} \in K$,

$$\begin{aligned} \mathbf{curl}_\Gamma \lambda(\boldsymbol{r}) = \mathbf{grad}_\Gamma \lambda(\boldsymbol{r}) \times \boldsymbol{n}(\boldsymbol{r}) &= - \left(\widehat{\boldsymbol{n}}_K(\boldsymbol{r}) \times \mathbf{J}_{\widehat{F}_K^{-1}} \widehat{\boldsymbol{\nabla}} \widehat{\lambda}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \\ &= \left(\frac{1}{\widehat{\mathcal{J}}_K} \mathbf{J}_{\widehat{F}_K}^\mathsf{T} \widehat{\mathbf{R}}^\mathsf{T} \widehat{\boldsymbol{\nabla}} \widehat{\lambda}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,. \end{aligned} \tag{2.30}$$

□

We now consider a piecewise smooth vector field in the form, for $\boldsymbol{r} \in K$,

$$\boldsymbol{\psi}(\boldsymbol{r}) = \left(\frac{1}{\widehat{\mathcal{J}}_K} \mathbf{J}_{\widehat{F}_K} \widehat{\boldsymbol{\psi}}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,, \tag{2.31}$$

which is normal-continuous across the edges of the mesh, i.e., $\widehat{\boldsymbol{\psi}}_K(\widehat{\boldsymbol{r}}) \cdot \widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{r}}) = -\widehat{\boldsymbol{\psi}}_L(\widehat{\boldsymbol{r}}') \cdot \widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{r}}')$ for any $\widehat{\boldsymbol{r}} \in K$ and $\widehat{\boldsymbol{r}}' \in L$ such that $\widehat{F}_K(\widehat{\boldsymbol{r}}) = \widehat{F}_L(\widehat{\boldsymbol{r}}')$, where K and L are two adjacent triangles.

Proposition 2.2 (Surface divergence of div-conforming functions)**.** The surface divergence of $\boldsymbol{\psi}$ is well defined and is given, for $\boldsymbol{r} \in K$, by

$$\mathrm{div}_\Gamma \boldsymbol{\psi}(\boldsymbol{r}) = \left(\frac{1}{\widehat{\mathcal{J}}_K} \widehat{\boldsymbol{\nabla}} \cdot \widehat{\boldsymbol{\psi}}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,. \tag{2.32}$$

Such a vector $\boldsymbol{\psi}$ is said to be div-conforming.

Proof. Consider a continuous function in the form $\lambda(\boldsymbol{r}) = \lambda_K \circ F_K^{-1}(\boldsymbol{r})$, for $\boldsymbol{r} \in K$. By definition of the surface divergence (1.65) and from (2.7) and (2.27)

$$\begin{aligned}\langle \boldsymbol{\psi}, \mathbf{grad}_\Gamma \lambda \rangle &= \sum_K \int_{\widehat{K}} \left(\mathbf{J}^{\mathsf{T}}_{\widehat{F}_K} \widehat{\boldsymbol{\psi}} \right)(\widehat{\boldsymbol{r}}) \cdot \left(\mathbf{J}_{\widehat{F}_K^{-1}} \widehat{\boldsymbol{\nabla}} \widehat{\lambda} \right)(\widehat{\boldsymbol{r}}) \, \mathrm{d}\widehat{\boldsymbol{r}} = \sum_K \int_{\widehat{K}} \widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{r}}) \cdot \widehat{\boldsymbol{\nabla}} \, \widehat{\lambda}(\widehat{\boldsymbol{r}}) \, \mathrm{d}\widehat{\boldsymbol{r}} \\ &= -\sum_K \int_{\widehat{K}} \widehat{\boldsymbol{\nabla}} \cdot \widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{r}}) \widehat{\lambda}(\widehat{\boldsymbol{r}}) \, \mathrm{d}\widehat{\boldsymbol{r}} + \sum_K \int_{\partial \widehat{K}} \widehat{\lambda}(\widehat{\boldsymbol{r}}) \, \widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{r}}) \cdot \widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{r}}) \, \mathrm{d}l(\widehat{\boldsymbol{r}}) \\ &= -\sum_K \int_{\widehat{K}} \widehat{\boldsymbol{\nabla}} \cdot \widehat{\boldsymbol{\psi}}(\widehat{\boldsymbol{r}}) \, \widehat{\lambda}(\widehat{\boldsymbol{r}}) \, \mathrm{d}\widehat{\boldsymbol{r}} = -\langle \mathrm{div}_\Gamma \boldsymbol{\psi}, \lambda \rangle \,,\end{aligned} \tag{2.33}$$

where $\mathrm{div}_\Gamma \boldsymbol{\psi}$ is defined as in (2.32). The cancellation of the line integrals along the boundary of $\widehat{K}$ is a consequence of the normal-continuity of $\boldsymbol{\psi}$. □

Finally, we consider a vector field in the form

$$\boldsymbol{\varphi}(\boldsymbol{r}) = \boldsymbol{n}(\boldsymbol{r}) \times \boldsymbol{\psi}(\boldsymbol{r}) = \left(\mathbf{J}_{\widehat{F}_K^{-1}} \widehat{\boldsymbol{\varphi}}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,, \tag{2.34}$$

with $\widehat{\boldsymbol{\varphi}}_K = \widehat{\mathbf{R}} \widehat{\boldsymbol{\psi}}_K$, that is a consequence of (2.13). As it is the ninety-degree rotation of a normal-continuous vector field, $\boldsymbol{\varphi}$ is tangent-continuous across the edges of the mesh, i.e., $\widehat{\mathbf{R}} \widehat{\boldsymbol{\varphi}}_K(\widehat{\boldsymbol{r}}) \cdot \widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{r}}) = -\widehat{\mathbf{R}} \widehat{\boldsymbol{\varphi}}_L(\widehat{\boldsymbol{r}}') \cdot \widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{r}}')$ for any $\widehat{\boldsymbol{r}} \in K$ and $\widehat{\boldsymbol{r}}' \in L$ such that $\widehat{F}_K(\widehat{\boldsymbol{r}}) = \widehat{F}_L(\widehat{\boldsymbol{r}}')$, where K and L are two adjacent triangles.

Proposition 2.3 (Surface scalar curl of curl-conforming functions)**.** The surface scalar curl of $\boldsymbol{\varphi}$ is well defined and is given, for $\boldsymbol{r} \in K$, by

$$\mathrm{curl}_\Gamma \boldsymbol{\varphi}(\boldsymbol{r}) = \left(\frac{1}{\widehat{\mathcal{J}}_K} \widehat{\boldsymbol{\nabla}} \cdot \widehat{\mathbf{R}}^{\mathsf{T}} \widehat{\boldsymbol{\varphi}}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,. \tag{2.35}$$

Such a vector $\boldsymbol{\varphi}$ is said to be curl-conforming.

Proof. By definition of the surface scalar curl (1.69)

$$\mathrm{curl}_\Gamma \boldsymbol{\varphi} = \mathrm{div}_\Gamma(\boldsymbol{\varphi} \times \boldsymbol{n}) = \mathrm{div}_\Gamma \boldsymbol{\psi} \,, \tag{2.36}$$

and it follows from (2.32) and $\widehat{\mathbf{R}}^{-1} = \widehat{\mathbf{R}}^{\mathsf{T}}$ that, for $\boldsymbol{r} \in K$,

$$\begin{aligned}\mathrm{curl}_\Gamma \boldsymbol{\varphi}(\boldsymbol{r}) &= \left(\frac{1}{\widehat{\mathcal{J}}_K} \widehat{\boldsymbol{\nabla}} \cdot \widehat{\boldsymbol{\psi}}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \\ &= \left(\frac{1}{\widehat{\mathcal{J}}_K} \widehat{\boldsymbol{\nabla}} \cdot \widehat{\mathbf{R}}^{\mathsf{T}} \widehat{\boldsymbol{\varphi}}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,.\end{aligned} \tag{2.37}$$

□

Proposition 2.4 (Nullspaces of the surface divergence and scalar curl)**.** Let us consider a continuous function λ, we have the following properties

$$\mathrm{div}_\Gamma \mathbf{curl}_\Gamma \lambda = 0 \,, \tag{2.38}$$

$$\mathrm{curl}_\Gamma \mathbf{grad}_\Gamma \lambda = 0 \,. \tag{2.39}$$

Proof. We prove the first relation by using (2.28) and (2.32), for $\boldsymbol{r} \in K$,

$$\begin{aligned} \mathrm{div}_\Gamma \mathbf{curl}_\Gamma \lambda(\boldsymbol{r}) &= \mathrm{div}_\Gamma \left(\frac{1}{\widehat{\mathcal{J}}_K} \mathbf{J}_{\widehat{F}_K}^\intercal \widehat{\mathbf{R}}^\intercal \widehat{\boldsymbol{\nabla}} \widehat{\lambda}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \\ &= \left(\frac{1}{\widehat{\mathcal{J}}_K} \widehat{\boldsymbol{\nabla}} \cdot \widehat{\mathbf{R}}^\intercal \widehat{\boldsymbol{\nabla}} \widehat{\lambda}_K \right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) = 0 \,, \end{aligned} \tag{2.40}$$

as a consequence of $\widehat{\boldsymbol{\nabla}} \cdot \widehat{\mathbf{R}}^\intercal \widehat{\boldsymbol{\nabla}} = \partial_v \partial_u - \partial_u \partial_v = 0$.

The second relation thus follows by definition of the surface vector and scalar curls (1.68)–(1.69)

$$\mathrm{curl}_\Gamma \mathbf{grad}_\Gamma \lambda = \mathrm{div}_\Gamma \left(\mathbf{grad}_\Gamma \lambda \times \boldsymbol{n} \right) = \mathrm{div}_\Gamma \mathbf{curl}_\Gamma \lambda = 0 \,. \tag{2.41}$$

□

2.2 Boundary element spaces

In this section, we denote by E, V, and C the numbers of edges, vertices, and cells of the mesh, respectively.

2.2.1 Definitions

The definition of the following boundary element spaces are motivated by the previous mapping properties of continuous, div-conforming and curl-conforming vector fields. In order to discretize the integral equations, a usual choice is to select subspaces of $H^{1/2}(\Gamma)$, $H^{-1/2}(\Gamma)$, $H^{-1/2}(\mathrm{div}_\Gamma, \Gamma)$ and $H^{-1/2}(\mathrm{curl}_\Gamma, \Gamma)$ that are obtained as local-to-global mappings of polynomial spaces in the local coordinates.

Let us first denote by

$$\widehat{\mathbb{P}}_p := \mathrm{span} \left\{ u^i v^j, \quad 0 \leq i + j \leq p \right\} \tag{2.42}$$

the space of polynomials of degree up to p on the reference triangle, which is of dimension

$$\dim \widehat{\mathbb{P}}_p = \frac{(p+1)(p+2)}{2} \,. \tag{2.43}$$

The continuous Galerkin space of order p, denoted as $\mathbb{CG}_p \subset H^{1/2}(\Gamma)$, is the space of functions (for $\boldsymbol{r} \in K$)

$$\lambda^{(p)}(\boldsymbol{r}) = \widehat{\lambda}_K \circ \widehat{F}_K^{-1}(\boldsymbol{r}) = \widehat{\lambda}_K(\widehat{\boldsymbol{r}}) \,, \tag{2.44}$$

where

$$\widehat{\lambda}_K(\widehat{\boldsymbol{r}}) \in \widehat{\mathbb{P}}_{p+1} \,, \tag{2.45}$$

with continuity across the edges of the mesh. This space is of dimension

$$L_p = V + pE + \frac{p(p-1)}{2} C \,. \tag{2.46}$$

The bubble functions $\beta^{(p)} \in \mathbb{B}_p \subset \mathbb{CG}_{p+1}$ are the continuous Galerkin functions of order $p+1$ that vanish along the edges of the cells. This space is of dimension

$$B_p = \frac{p(p+1)}{2} C \,. \tag{2.47}$$

The discontinuous Galerkin space of order p, denoted as $\mathbb{DG}_p \subset H^{-1/2}(\Gamma)$, is the space of functions (for $\boldsymbol{r} \in K$)

$$\sigma^{(p)}(\boldsymbol{r}) = \left(\frac{\widehat{\sigma}_K}{\widehat{\mathcal{J}}_K}\right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,, \tag{2.48}$$

where

$$\widehat{\sigma}_K(\widehat{\boldsymbol{r}}) \in \widehat{\mathbb{P}}_p \,, \tag{2.49}$$

without any imposed continuity across the edges of the mesh. This space is of dimension

$$S_p = \frac{(p+1)(p+2)}{2} C \,. \tag{2.50}$$

The Raviart-Thomas space of order p, denoted as $\mathbb{RT}_p \subset H^{-1/2}(\mathrm{div}_\Gamma, \Gamma)$, is the space of vector fields (for $\boldsymbol{r} \in K$)

$$\boldsymbol{\psi}^{(p)}(\boldsymbol{r}) = \left(\frac{\mathbf{J}^{\mathsf{T}}_{\widehat{F}_K} \widehat{\boldsymbol{\psi}}_K}{\widehat{\mathcal{J}}_K}\right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) \,, \tag{2.51}$$

where

$$\widehat{\boldsymbol{\psi}}_K(\widehat{\boldsymbol{r}}) \in (\widehat{\mathbb{P}}_p)^2 + \begin{pmatrix} u \\ v \end{pmatrix} \widehat{\mathbb{P}}_p \,, \tag{2.52}$$

with normal-continuity across the edges of the mesh, This space is of dimension

$$N_p = (p+1)E + p(p+1)C \,. \tag{2.53}$$

The rotated Raviart-Thomas functions $\boldsymbol{\varphi}^{(p)} \in \boldsymbol{n} \times \mathbb{RT}_p \subset H^{-1/2}(\mathrm{curl}_\Gamma, \Gamma)$ are in the form (for $\boldsymbol{r} \in K$)

$$\boldsymbol{\varphi}^{(p)}(\boldsymbol{r}) = \left(\mathbf{J}_{\widehat{F}_K^{-1}} \widehat{\boldsymbol{\varphi}}_K\right) \circ F_K^{-1}(\boldsymbol{r}) \,. \tag{2.54}$$

If $\boldsymbol{\varphi}^{(p)} = \boldsymbol{n} \times \boldsymbol{\psi}^{(p)}$, with $\boldsymbol{\psi}^{(p)} \in \mathbb{RT}_p$, we have

$$\widehat{\boldsymbol{\varphi}}_K = \widehat{\mathbf{R}} \widehat{\boldsymbol{\psi}}_K \in (\widehat{\mathbb{P}}_p)^2 + \begin{pmatrix} -v \\ u \end{pmatrix} \widehat{\mathbb{P}}_p \,, \tag{2.55}$$

with tangent-continuity across the edges of the mesh. This space is also of dimension N_p.

Proposition 2.5 (Relationship Between the Boundary Element Spaces)**.** These boundary element spaces are related to each other through the surface differential operators

$$\mathbf{curl}_\Gamma \mathbb{CG}_p \subset \mathbb{RT}_p \,, \tag{2.56}$$
$$\mathbf{grad}_\Gamma \mathbb{CG}_p \subset \boldsymbol{n} \times \mathbb{RT}_p \,, \tag{2.57}$$
$$\mathrm{div}_\Gamma \mathbb{RT}_p \subset \mathbb{DG}_p \,, \tag{2.58}$$
$$\mathrm{curl}_\Gamma(\boldsymbol{n} \times \mathbb{RT}_p) \subset \mathbb{DG}_p \,. \tag{2.59}$$

Moreover,

$$\mathrm{div}_\Gamma \mathbf{curl}_\Gamma \mathbb{CG}_p = 0 \,, \tag{2.60}$$
$$\mathrm{curl}_\Gamma \mathbf{grad}_\Gamma \mathbb{CG}_p = 0 \,. \tag{2.61}$$

Proof. These are direct applications of propositions 2.1, 2.2, 2.3 and 2.4, together with the fact that $\widehat{\boldsymbol{\nabla}} \widehat{\mathbb{P}}_{p+1} \subseteq \widehat{\mathbb{P}}_p^2$ and $\widehat{\boldsymbol{\nabla}} \cdot \left(\widehat{\mathbb{P}}_p^2 + \begin{pmatrix} u \\ v \end{pmatrix} \widehat{\mathbb{P}}_p\right) = \widehat{\mathbb{P}}_p$. □

2.2.2 Quasi-Helmholtz Decompositions of the Raviart-Thomas and Rotated Raviart-Thomas Spaces

It is known that any smooth vector field $\boldsymbol{f}$ can be expressed in terms of its Helmholtz decomposition [29]

$$\boldsymbol{f} = \mathbf{curl}_\Gamma \phi + \mathbf{grad}_\Gamma \gamma + \boldsymbol{h} \tag{2.62}$$

where ϕ and γ are twice differentiable scalar functions. This states that $\boldsymbol{f}$ can be decomposed as a sum of a solenoidal (divergence-free) component $\mathbf{curl}_\Gamma \phi$, an irrotational (curl-free) component $\mathbf{grad}_\Gamma \gamma$, and an harmonic (both divergence- and curl-free) component $\boldsymbol{h}$, that are orthogonal to each other

$$\langle \mathbf{grad}_\Gamma \gamma, \mathbf{curl}_\Gamma \phi \rangle = \langle \mathbf{grad}_\Gamma \gamma, \boldsymbol{h} \rangle = \langle \boldsymbol{h}, \mathbf{curl}_\Gamma \phi \rangle = 0\,. \tag{2.63}$$

This decomposition does not hold for div- and curl-conforming vector fields, but a similar result can still be obtained, which is the so-called quasi-Helmholtz decomposition.

For div-conforming vector fields $\boldsymbol{\psi} \in \mathbb{RT}_p$, the scalar surface curl is not well-defined, and it turns out that their irrotational and harmonic components cannot be defined neither. Nevertheless, there exists a quasi-irrotational vector field $\boldsymbol{u}_\perp \in \mathbb{RT}_p$ and a quasi-harmonic vector field $\boldsymbol{h}_\perp \in \mathbb{RT}_p$ such that $\boldsymbol{\psi} \in \mathbb{RT}_p$ can be decomposed in terms of its quasi-Helmholtz decomposition [2]

$$\boldsymbol{\psi} = \mathbf{curl}_\Gamma \lambda + \boldsymbol{u}_\perp + \boldsymbol{h}_\perp\,, \tag{2.64}$$

where $\mathbf{curl}_\Gamma \lambda \in \mathbb{RT}_p$, for some $\lambda \in \mathbb{CG}_p$, is the (actual) solenoidal component.

This decomposition verifies the following properties closely related to those of the continuous Helmholtz decomposition:

(i) The three components are orthogonal to each other, i.e., $\langle \boldsymbol{u}_\perp, \mathbf{curl}_\Gamma \lambda \rangle = \langle \boldsymbol{u}_\perp, \boldsymbol{h}_\perp \rangle = \langle \boldsymbol{h}_\perp, \mathbf{curl}_\Gamma \lambda \rangle = 0$.

(ii) The vector fields $\boldsymbol{h}_\perp$ and $\mathbf{curl}_\Gamma \lambda$ are divergence-free.

(iii) The vector field $\boldsymbol{u}_\perp$ is a weak surface gradient, in the sense that it is orthogonal to divergence-free functions.

(iv) There exists $\sigma \in \mathbb{DG}_p$ such that $\mathrm{div}_\Gamma \boldsymbol{\psi} = \mathrm{div}_\Gamma \boldsymbol{u}_\perp = \sigma$.

For curl-conforming vectors $\boldsymbol{\varphi} \in \boldsymbol{n} \times \mathbb{RT}_p$, the quasi-Helmholtz decomposition becomes

$$\boldsymbol{\varphi} \coloneqq \boldsymbol{n} \times \boldsymbol{\psi} = \mathbf{grad}_\Gamma \lambda + \boldsymbol{n} \times \boldsymbol{u}_\perp + \boldsymbol{n} \times \boldsymbol{h}_\perp\,, \tag{2.65}$$

which shows that it can be decomposed in terms of an irrotational component $\mathbf{grad}_\Gamma \lambda$, a quasi-solenoidal component $\boldsymbol{n} \times \boldsymbol{u}_\perp \in \boldsymbol{n} \times \mathbb{RT}_p$ and a quasi-harmonic component $\boldsymbol{n} \times \boldsymbol{h}_\perp \in \boldsymbol{n} \times \mathbb{RT}_p$.

Equally:

(i) The orthogonality between the three components is preserved.

(ii) The vector fields $\boldsymbol{n} \times \boldsymbol{h}_\perp$ and $\mathbf{grad}_\Gamma \lambda$ are curl-free.

(iii) The vector field $\boldsymbol{n} \times \boldsymbol{u}_\perp$ is a weak vector surface curl, in the sense that it is orthogonal to curl-free functions.

(iv) There exists $\sigma \in \mathbb{DG}_p$ such that $\mathrm{curl}_\Gamma \boldsymbol{\varphi} = \mathrm{curl}_\Gamma(\boldsymbol{n} \times \boldsymbol{u}_\perp) = \mathrm{div}_\Gamma(\boldsymbol{u}_\perp) = \sigma$.

Proposition 2.6 (Completeness of the quasi-Helmholtz decomposition)**.** Let us consider a closed surface Γ discretized by a mesh of V vertices, E edges and C triangular cells, that is possibly multiply connected with genus H. We have

$$N_p = (S_p - 1) + (L_p - 1) + 2H\,, \tag{2.66}$$

where for div-conforming / curl-conforming functions, $S_p - 1$ is the number of independent quasi-irrotational / quasi-solenoidal functions, $L_p - 1$ is the number of independent solenoidal / irrotational functions and $2H$ is the number of quasi-Harmonic functions. The case of open boundaries, junctions and multiple bodies will be covered in chapter 4 for order 0. Similar relationship is obtained at higher order.

Proof of Equation (2.66). For closed geometry, the Euler-Poincaré formula states that

$$E = (C - 1) + (V - 1) + 2H\,, \tag{2.67}$$

which gives (2.66) at order $p = 0$. We then compute from (2.50) and (2.46)

$$L_p + S_p = V + pE + (p^2 + p + 1)C\,. \tag{2.68}$$

Hence,

$$\begin{aligned}(S_p - 1) + (L_p - 1) + 2H &= (C - 1) + (V - 1) + 2H + pE + p(p+1)C \\ &= E + pE + p(p+1)C = N_p\,.\end{aligned} \tag{2.69}$$

□

2.2.3 Interpolatory Basis Functions

In this section we describe a particular choice of basis functions, that are interpolatory basis functions—i.e., they are built from the definition of a set of interpolatory points on which they are zero, except on one of these points [67]. Other choices of basis functions are possible, such as the hierarchical basis functions, for which the basis functions are built hierarchically order by order, where a set of basis functions is defined at each order to complete the lower order space so as to span the higher order space [115]. Differently, the interpolatory basis functions of a given order are all different from those of the previous orders [115].

Discontinuous Galerkin basis functions

The discontinuous Galerkin basis functions are the more trivial to implement as no continuity needs to be imposed. Such basis functions can be built from the previous definition of the Silvester-Lagrange interpolatory functions (2.16) of order p [69]

$$\widehat{\sigma}_{\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) = \mathscr{L}_{\boldsymbol{i}[1]}^{(p)}(u)\mathscr{L}_{\boldsymbol{i}[2]}^{(p)}(v)\mathscr{L}_{\boldsymbol{i}[3]}^{(p)}(1-u-v)\,, \tag{2.70}$$

where $\boldsymbol{i}$, is a three-dimensional array of indices such that $\boldsymbol{i}[1] + \boldsymbol{i}[2] + \boldsymbol{i}[3] = p$. On each cell K, we can define $(p+1)(p+2)/2$ functions in global coordinates such that

$$\sigma_n^{(p)}(\boldsymbol{r}) = \begin{cases} \left(\dfrac{1}{\widehat{\mathcal{J}}_K}\widehat{\sigma}_{\boldsymbol{i}_n}^{(p)}\right) \circ \widehat{F}_K^{-1}(\boldsymbol{r})\,, & \text{if } \boldsymbol{r} \in K\,, \\ 0\,, & \text{otherwise}\,, \end{cases} \tag{2.71}$$

where $\boldsymbol{i}_n$ is the array of indices associated with the basis function $\sigma_n^{(p)}$ in local coordinates. For $p \geq 1$, the functions $\widehat{\sigma}_{\boldsymbol{i}_n}^{(p)}$ interpolate in the reference triangle at the points $\widehat{\boldsymbol{r}}_{\boldsymbol{\ell}}^{(p)} = \left(\frac{\boldsymbol{\ell}[1]}{p}, \frac{\boldsymbol{\ell}[2]}{p}\right)^{\mathsf{T}}$, with $\boldsymbol{\ell}[1] + \boldsymbol{\ell}[2] \leq p$, i.e.,

$$\widehat{\sigma}_{\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}_{\boldsymbol{\ell}}^{(p)}) = \delta_{\boldsymbol{i}[1],\boldsymbol{\ell}[1]}\delta_{\boldsymbol{i}[2],\boldsymbol{\ell}[2]} \,. \tag{2.72}$$

For $p = 0$, these functions are constant equal to one on a given triangle and are zero otherwise. For that reason they are often called patch functions. We set $\widehat{\boldsymbol{r}}_{\boldsymbol{\ell}}^{(0)} \equiv \widehat{\boldsymbol{r}}^{(0)} = \left(\frac{1}{3}, \frac{1}{3}\right)$ to be the barycenter of the reference triangle.

Continuous Galerkin basis functions

We define the continuous Galerkin basis functions in local coordinates by using the Silvester-Lagrange basis functions of order $p + 1$ [69]

$$\widehat{\lambda}_{\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) = \mathscr{L}_{\boldsymbol{i}[1]}^{(p+1)}(u)\mathscr{L}_{\boldsymbol{i}[2]}^{(p+1)}(v)\mathscr{L}_{\boldsymbol{i}[3]}^{(p+1)}(1-u-v)\,, \tag{2.73}$$

with $\boldsymbol{i}[1] + \boldsymbol{i}[2] + \boldsymbol{i}[3] = p + 1$. The implementation of continuous Galerkin basis functions is a bit more complicated as it requires the finding for common interpolatory points at the boundary of the triangles, wherever the basis function is not zero. Thus, there exists three types of continuous basis functions:

- **Vertex-based functions:** if $\boldsymbol{i} = (p, 0, 0)$ or $\boldsymbol{i} = (0, p, 0)$ or $\boldsymbol{i} = (0, 0, p)$, the function is unity at one of the vertices of the reference triangle and equals zero along the edge opposite to that vertex. We need to connect this function to the functions associated with the same vertex in the other triangles so as to impose continuity.

- **Edge-based functions:** if only a single index in $\boldsymbol{i}$ equals zero, the function is unity on one point along one of the edges of the reference triangle, while it is zero along the other edges. We need to connect this function to the function that equals one at the same point on the other side of that edge so as to impose continuity.

- **Cell-based functions:** otherwise, the function is zero along all the edges of the reference triangle, and the interpolatory point where it is not zero is inside of the cell. As a consequence, we do not need to associate this function with any other function. These functions are bubble functions, and these are the basis functions we need to span the subspace $\mathbb{B}_p \subset \mathbb{CG}_{p+1}$.

For each vertex of the mesh, we can define a single vertex-based function, for which the domain consists of all the triangles adjacent to the vertex. For each edge of the mesh, we can define $p+1$ edge-based functions, for which the domain consists of the two triangles adjacent to the edge. For each triangle of the mesh, we can define $p(p+1)/2$ cell-based (bubble) functions, for which the domain consists of that single triangle. Let us denote by $K_1, \ldots, K_M$ the triangles composing the domain of a continuous function (where $M = 1$ for cell-based functions, $M = 2$ for edge-based functions and $M \geq 3$ for vertex-based functions). The continuous functions in global coordinates are thus found as

$$\lambda_n^{(p)}(\boldsymbol{r}) = \begin{cases} \widehat{\lambda}_{\boldsymbol{i}_n^{K_1}}^{(p)} \circ \widehat{F}_{K_1}^{-1}(\boldsymbol{r}) & \text{if } \boldsymbol{r} \in K_1, \\ \vdots & \vdots \\ \widehat{\lambda}_{\boldsymbol{i}_n^{K_M}}^{(p)} \circ \widehat{F}_{K_M}^{-1}(\boldsymbol{r}) & \text{if } \boldsymbol{r} \in K_M, \\ 0 & \text{otherwise}, \end{cases} \tag{2.74}$$

where $\boldsymbol{i}_n^{K_\ell}$, for $\ell = 1, \ldots, M$, is the array of local indices for the interpolatory point associated with $\lambda_n^{(p)}(\boldsymbol{r})$ in triangle K_ℓ. Notice that there is no reason these indices are the same for each triangle $K_1, \ldots K_M$, and that they are found, by construction, to match with the same interpolatory point in the global coordinates of each triangle.

At order zero we only have vertex-based basis functions. They are the linear piecewise functions that equal one at a given vertex and go to zero on the opposite edges to the vertex. For that reason, they are often called pyramid functions.

Raviart-Thomas basis functions

Lowest order ($p = 0$):

At order zero, we define in the reference triangle the three edge-based basis functions [115]

$$\widehat{\boldsymbol{\psi}}_1^{(0)}(\widehat{\boldsymbol{r}}) = (u-1, v)^\mathsf{T}, \tag{2.75a}$$
$$\widehat{\boldsymbol{\psi}}_2^{(0)}(\widehat{\boldsymbol{r}}) = (u, v-1)^\mathsf{T}, \tag{2.75b}$$
$$\widehat{\boldsymbol{\psi}}_3^{(0)}(\widehat{\boldsymbol{r}}) = (u, v)^\mathsf{T}. \tag{2.75c}$$

which are linear vector fields going from the vertex $\widehat{\boldsymbol{p}}_b$ to the opposite edge $\widehat{e}_b$, with $b = 1,2,3$, respectively (see the notations in Figure 2.1). In global coordinates, the associated basis functions are defined on both sides of a given edge that separates two triangles K and L. They are obtained by identifying the local basis function associated with that edge, and they are multiplied by opposite signs so that the basis function goes outside K and goes inside L across their common edge, in order to ensure normal-continuity. We can write

$$\boldsymbol{\psi}_n^{(0)}(\boldsymbol{r}) = \begin{cases} \left(\dfrac{1}{\widehat{\mathcal{J}}_K}\mathbf{J}_{\widehat{F}_K}^\mathsf{T}\widehat{\boldsymbol{\psi}}_{b_n^K}^{(0)}\right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) & \text{if } \boldsymbol{r} \in K, \\ -\left(\dfrac{1}{\widehat{\mathcal{J}}_L}\mathbf{J}_{\widehat{F}_L}^\mathsf{T}\widehat{\boldsymbol{\psi}}_{b_n^L}^{(0)}\right) \circ \widehat{F}_L^{-1}(\boldsymbol{r}) & \text{if } \boldsymbol{r} \in L, \\ 0 & \text{otherwise,} \end{cases} \tag{2.76}$$

where b_n^K and b_n^L are the edge indices ($b = 1,2,3$) in the reference triangle that are associated with the basis function $\boldsymbol{\psi}_n^{(0)}(\boldsymbol{r})$ for triangle K and L, respectively. These basis functions are known as the Rao-Wilton-Glisson (RWG) basis functions [118]. They are div-conforming vector fields, as their normal component is continuous across edge $\widehat{e}_b$ and zero along the others.

Subsequently, we compute their divergences in local coordinates

$$\boldsymbol{\nabla} \cdot \widehat{\boldsymbol{\psi}}_i^{(0)}(\widehat{\boldsymbol{r}}) = 2, \qquad (i = 1,2,3), \tag{2.77}$$

and in global coordinates

$$\mathrm{div}_\Gamma \boldsymbol{\psi}_n^{(0)}(\boldsymbol{r}) = \begin{cases} \left(2/\widehat{\mathcal{J}}_K\right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) & \text{if } \boldsymbol{r} \in K, \\ -\left(2/\widehat{\mathcal{J}}_K\right) \circ \widehat{F}_K^{-1}(\boldsymbol{r}) & \text{if } \boldsymbol{r} \in L, \\ 0 & \text{otherwise.} \end{cases} \tag{2.78}$$

Higher orders ($p \geq 1$):

The high-order div-conforming basis functions, are obtained by multiplying the RWG functions with high-order scalar basis functions [67]. We first define the shifted Silvester-Lagrange polynomials of order p (in one dimension), that are obtained from the definition of the Silvester-Lagrange polynomials (2.16), for $p \geq 2$ [69]

$$\mathscr{R}_n^{(p)}(u) = \mathscr{L}_{n-1}^{(p)}\left(u - \frac{1}{p}\right), \qquad 1 \leq n \leq p+1, \tag{2.79}$$

These functions are polynomials of degree $n-1$. For $2 \leq n \leq p+1$. they are unity at $\frac{n-1}{p}$ and zero at $u = \frac{1}{p}, \ldots, \frac{n-2}{p}$, while it is constant equal to one for $n = 1$. Additionally, their derivatives are computed as

$$\dot{\mathscr{R}}_n^{(p)}(u) = \dot{\mathscr{L}}_{n-1}^{(p)}\left(u - \frac{1}{p}\right), \tag{2.80}$$

where $\dot{\mathscr{L}}_{n-1}^{(p)}$ is defined in 2.22.

We now define the shifted Silvester-Lagrange polynomials in the reference triangle, for $p \geq 0$ [69]

$$\mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) = \mathscr{L}_{\boldsymbol{i}[b]}^{(p+2)}(\boldsymbol{\xi}[b])\,\mathscr{R}_{\boldsymbol{i}[b-1]}^{(p+2)}(\boldsymbol{\xi}[b-1])\,\mathscr{R}_{\boldsymbol{i}[b+1]}^{(p+2)}(\boldsymbol{\xi}[b+1]), \tag{2.81}$$

where we have set $\boldsymbol{\xi} = (\widehat{\boldsymbol{r}}_1, \widehat{\boldsymbol{r}}_2, 1 - \widehat{\boldsymbol{r}}_1 - \widehat{\boldsymbol{r}}_2)$ and where $\boldsymbol{i}$ is a three-dimensional array of indices such that $\boldsymbol{i}[1] + \boldsymbol{i}[2] + \boldsymbol{i}[3] = p+2$, $0 \leq \boldsymbol{i}[b] \leq p$ and $1 \leq \boldsymbol{i}[b+1], \boldsymbol{i}[b-1] \leq p+1$. Notice that with these notations, the entries of $\boldsymbol{\xi}$ and $\boldsymbol{i}$ have to be considered with a modulo of three, i.e., $\boldsymbol{\xi}[b] = \boldsymbol{\xi}[b+3] = \boldsymbol{\xi}[b-3]$ and $\boldsymbol{i}[b] = \boldsymbol{i}(b+3) = \boldsymbol{i}(b-3)$, for $b = 1,2,3$.

These functions interpolate at the shifted interpolatory points on the reference triangle which correspond to the uniform Lagrange interpolatory points in the reference triangle, less the vertices of the triangle and less the points on the opposite edges to $\widehat{e}_b$, that we denote as $\widehat{\boldsymbol{r}}_{b,m,n}^{(q)}$. (see Table 2.2 for an illustration of the distribution of these points).

Subsequently, we compute their partial derivatives with respect to $\boldsymbol{\xi}$ as

$$\begin{aligned}
\partial_{\boldsymbol{\xi}[b]}\mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) &= \dot{\mathscr{L}}_{\boldsymbol{i}[b]}^{(p+2)}(\boldsymbol{\xi}[b])\,\mathscr{R}_{\boldsymbol{i}[b+1]}^{(p+2)}(\boldsymbol{\xi}[b+1])\,\mathscr{R}_{\boldsymbol{i}[b-1]}^{(p+2)}(\boldsymbol{\xi}[b-1]),\\
\partial_{\boldsymbol{\xi}[b+1]}\mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) &= \mathscr{L}_{\boldsymbol{i}[b]}^{(p+2)}(\boldsymbol{\xi}[b])\,\dot{\mathscr{R}}_{\boldsymbol{i}[b+1]}^{(p+2)}(\boldsymbol{\xi}[b+1])\,\mathscr{R}_{\boldsymbol{i}[b-1]}^{(p+2)}(\boldsymbol{\xi}[b-1]),\\
\partial_{\boldsymbol{\xi}[b-1]}\mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) &= \mathscr{L}_{\boldsymbol{i}[b]}^{(p+2)}(\boldsymbol{\xi}[b])\,\mathscr{R}_{\boldsymbol{i}[b+1]}^{(p+2)}(\boldsymbol{\xi}[b+1])\,\dot{\mathscr{R}}_{\boldsymbol{i}[b-1]}^{(p+2)}(\boldsymbol{\xi}[b-1]).
\end{aligned} \tag{2.82}$$

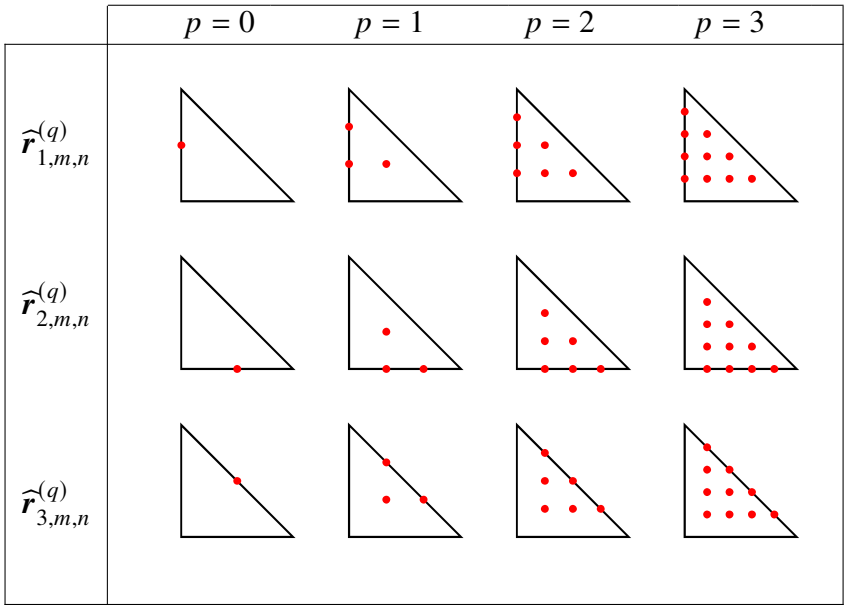
Finally, we define the high-order Raviart-Thomas basis functions in the local coordinates as

$$\widehat{\boldsymbol{\psi}}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) = \mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}})\widehat{\boldsymbol{\psi}}_b^{(0)}(\widehat{\boldsymbol{r}}), \tag{2.83}$$

where, again, $\boldsymbol{i}[1] + \boldsymbol{i}[2] + \boldsymbol{i}[3] = p+2$, $0 \leq \boldsymbol{i}[b] \leq p$ and $1 \leq \boldsymbol{i}[b+1], \boldsymbol{i}[b-1] \leq p+1$. Additionally, their surface divergences can be found as

$$\begin{aligned}
\widehat{\boldsymbol{\nabla}} \cdot \widehat{\boldsymbol{\psi}}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) &= \widehat{\boldsymbol{\nabla}}\mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) \cdot \widehat{\boldsymbol{\psi}}_b^{(0)}(\widehat{\boldsymbol{r}}) + \mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}})\widehat{\boldsymbol{\nabla}} \cdot \widehat{\boldsymbol{\psi}}_b^{(0)}(\widehat{\boldsymbol{r}})\\
&= \boldsymbol{\xi}[b+1]\left(\partial_{\boldsymbol{\xi}[b+1]}\mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) - \partial_{\boldsymbol{\xi}[b]}\mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}})\right)\\
&\quad + \boldsymbol{\xi}[b-1]\left(\partial_{\boldsymbol{\xi}[b-1]}\mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}) - \partial_{\boldsymbol{\xi}[b]}\mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}})\right) + 2\mathscr{R}_{b,\boldsymbol{i}}^{(p)}(\widehat{\boldsymbol{r}}).
\end{aligned} \tag{2.84}$$

Table 2.2: Location of the shifted Lagrange interpolatory points in the reference triangle.



These basis functions are known as the Graglia-Wilton-Peterson (GWP) functions and they span the Raviart-Thomas space of order p [67]. They form a natural extension of the RWG functions as the definition of RWG functions is retrieved for $p = 0$ (in that case $\mathscr{R}^{(0)}_{b,\boldsymbol{i}}(\widehat{\boldsymbol{r}}) = 1$ for every $b \in \{1,2,3\}$, and we notice $\boldsymbol{i}[b] = 0$, $\boldsymbol{i}[b+1] = 1$ and $\boldsymbol{i}[b-1] = 1$ even though it is useless at order zero).

In global coordinates, there are two types of GWP basis functions:

- **Edge-based functions:**

For all edge e that is shared between two triangles K and L, we can define $p+1$ basis functions

$$\boldsymbol{\psi}^{(p)}_n(\boldsymbol{r}) = \begin{cases} \left(\dfrac{1}{\widehat{\mathcal{J}}_K}\mathscr{R}^{(p)}_{b^K_n,\boldsymbol{i}^K_n}\mathbf{J}^{\mathsf{T}}_{\widehat{F}_K}\widehat{\boldsymbol{\psi}}^{(0)}_{b^K_n}\right)\circ\widehat{F}^{-1}_K(\boldsymbol{r}) & \text{if } \boldsymbol{r}\in K, \\ -\left(\dfrac{1}{\widehat{\mathcal{J}}_L}\mathscr{R}^{(p)}_{b^L_n,\boldsymbol{i}^L_n}\mathbf{J}^{\mathsf{T}}_{\widehat{F}_L}\widehat{\boldsymbol{\psi}}^{(0)}_{b^L_n}\right)\circ\widehat{F}^{-1}_L(\boldsymbol{r}) & \text{if } \boldsymbol{r}\in L, \\ 0 & \text{otherwise,} \end{cases} \tag{2.85}$$

with surface divergence

$$\operatorname{div}_\Gamma\boldsymbol{\psi}^{(p)}_n(\boldsymbol{r}) = \begin{cases} \left(\dfrac{1}{\widehat{\mathcal{J}}_K}\widehat{\boldsymbol{\nabla}}\cdot\widehat{\boldsymbol{\psi}}^{(p)}_{b^K_n,\boldsymbol{i}^K_n}(\widehat{\boldsymbol{r}})\right)\circ\widehat{F}^{-1}_K(\boldsymbol{r}) & \text{if } \boldsymbol{r}\in K, \\ -\left(\dfrac{1}{\widehat{\mathcal{J}}_L}\widehat{\boldsymbol{\nabla}}\cdot\widehat{\boldsymbol{\psi}}^{(p)}_{b^K_n,\boldsymbol{i}^L_n}(\widehat{\boldsymbol{r}})\right)\circ\widehat{F}^{-1}_L(\boldsymbol{r}) & \text{if } \boldsymbol{r}\in L, \\ 0 & \text{otherwise,} \end{cases} \tag{2.86}$$

where b^K_n and b^L_n are the local indices of edge e in triangles K and L and where $\boldsymbol{i}^K_n$ and $\boldsymbol{i}^L_n$ are the arrays of local indices associated with the same interpolatory point defining the basis function

$\boldsymbol{\psi}_n^{(p)}$, in K and L, such that

$$\begin{aligned} \boldsymbol{i}_n^K[b_n^K] = \boldsymbol{i}_n^L[b_n^L] = 0\,, \\ 1 \leq \boldsymbol{i}_n^K[b_n^K+1], \boldsymbol{i}_n^L[b_n^K+1] \leq p+1\,, \\ \boldsymbol{i}_n^K[b_n^K-1] = p+2-\boldsymbol{i}_n^K[b_n^K+1]\,, \\ \boldsymbol{i}_n^L[b_n^L-1] = p+2-\boldsymbol{i}_n^L[b_n^L+1]\,. \end{aligned} \tag{2.87}$$

These functions are div-conforming as they are normal-continuous across edge e and tangent to the other edges of K and L.

- **Cell-based functions:**

For every cell K, we can define $3p(p+1)/2$ additional basis functions

$$\boldsymbol{\psi}_n^{(p)}(\boldsymbol{r}) = \begin{cases} \left(\dfrac{1}{\widehat{\mathcal{J}}_K}\mathscr{R}^{(p)}_{b_n^K,\boldsymbol{i}_n^K}\mathbf{J}^{\mathsf{T}}_{\widehat{F}_K}\widehat{\boldsymbol{\psi}}^{(0)}_{b_n^K}\right)\circ\widehat{F}_K^{-1}(\boldsymbol{r}) & \text{if } \boldsymbol{r}\in K, \\ 0 & \text{otherwise,} \end{cases} \tag{2.88}$$

with surface divergence

$$\mathrm{div}_\Gamma\boldsymbol{\psi}_n^{(p)}(\boldsymbol{r}) = \begin{cases} \left(\dfrac{1}{\widehat{\mathcal{J}}_K}\widehat{\boldsymbol{\nabla}}\cdot\widehat{\boldsymbol{\psi}}^{(p)}_{b_n^K,\boldsymbol{i}_n^K}(\widehat{\boldsymbol{r}})\right)\circ\widehat{F}_K^{-1}(\boldsymbol{r}) & \text{if } \boldsymbol{r}\in K, \\ 0 & \text{otherwise,} \end{cases} \tag{2.89}$$

where $b_n^K = 1,2,3$ is the local index of the edges of triangle K and $\boldsymbol{i}_n^K$ is the array of local indices which is associated with an interpolatory point defining the basis function $\boldsymbol{\psi}_n^{(p)}$, such that

$$\begin{aligned} 1 \leq \boldsymbol{i}_n^K[1] \leq p\,, \\ 1 \leq \boldsymbol{i}_n^K[2] \leq p+1-\boldsymbol{i}_n^K[1]\,, \\ \boldsymbol{i}_n^K[3] = p+2-\boldsymbol{i}_n^K[1]-\boldsymbol{i}_n^K[2]\,. \end{aligned} \tag{2.90}$$

These functions are div-conforming as they equal zero on one edge of K and are tangent along the two others. With this definition, we obtain three different basis functions for each interpolatory point located inside the cell (for each possible value of b_n^K). Because they are two-dimensional vectors, only two basis functions are independent so that one cell-based basis function needs to be removed for each interpolatory point located inside the cell [67]. The choice of the basis function to remove can be done arbitrarily, and for instance, we can systematically drop the case $b_n^K = 3$. In this way, we only build $p(p+1)$ instead of $3p(p+1)/2$ basis functions for each cell of the mesh.

2.3 Discretization of the Variational Problem

2.3.1 Standard Discretization of the EFIE

We focus here on the discretization of the electric field integral equation in its variational form (1.115).

For a given mesh approximation of Γ, and for a given order p, the Sobolev space $H^{-1/2}(\mathrm{div}_\Gamma, \Gamma)$ is restricted to the Raviart-Thomas space $\mathbb{RT}_p$. This technique is known as the Galerkin method, where the same finite/boundary elements are used to approximate the testing and source functions. More specifically, we approximate the surface current density (the source function) as

$$\boldsymbol{J} \approx \sum_{n=1}^{N_p} [\boldsymbol{j}_p]_n \boldsymbol{\psi}_n^{(p)} , \tag{2.91}$$

and the EFIE is tested with rotated basis functions $\{\boldsymbol{n} \times \boldsymbol{\psi}^{(p)}\}$ (the testing functions). This defines N_p equations, for $m = 1, \ldots, N_p$,

$$\left\langle \boldsymbol{n} \times \boldsymbol{\psi}_m^{(p)}, \mathcal{T} \sum_{n=1}^{N_p} [\boldsymbol{j}_p]_n \boldsymbol{\psi}_n^{(p)} \right\rangle = -\frac{1}{\eta} \left\langle \boldsymbol{n} \times \boldsymbol{\psi}_m^{(p)}, \boldsymbol{n} \times \boldsymbol{E}^{\mathrm{inc}} \right\rangle . \tag{2.92}$$

By linearity of the EFIE operator $\mathcal{T}$, we can finally express this system of equations in terms of a square matrix system of size N_p

$$\mathbf{T}_p \boldsymbol{j}_p = \boldsymbol{e}_p . \tag{2.93}$$

The system matrix is

$$\mathbf{T}_p \coloneqq -jk\mathbf{T}_{s|p} + (jk)^{-1}\mathbf{T}_{h|p} , \tag{2.94}$$

with the discretization of the integral operators

$$\begin{aligned} [\mathbf{T}_{s|p}]_{mn} &\coloneqq \langle \boldsymbol{n} \times \boldsymbol{\psi}_m^{(p)}, \mathcal{T}_s \boldsymbol{\psi}_n^{(p)} \rangle \\ &= \int_\Gamma \int_\Gamma G(\boldsymbol{r}, \boldsymbol{r}') \boldsymbol{\psi}_m^{(p)}(\boldsymbol{r}) \cdot \boldsymbol{\psi}_n^{(p)}(\boldsymbol{r}') \, \mathrm{d}S(\boldsymbol{r}') \mathrm{d}S(\boldsymbol{r}) , \end{aligned} \tag{2.95}$$

$$\begin{aligned} [\mathbf{T}_{h|p}]_{mn} &\coloneqq \langle \boldsymbol{n} \times \boldsymbol{\psi}_m^{(p)}, \mathcal{T}_h \boldsymbol{\psi}_n^{(p)} \rangle \\ &= -\int_\Gamma \int_\Gamma G(\boldsymbol{r}, \boldsymbol{r}') \mathrm{div}_\Gamma \boldsymbol{\psi}_m^{(p)}(\boldsymbol{r}) \mathrm{div}_\Gamma \boldsymbol{\psi}_n^{(p)}(\boldsymbol{r}') \, \mathrm{d}S(\boldsymbol{r}') \mathrm{d}S(\boldsymbol{r}) , \end{aligned} \tag{2.96}$$

where an integration by parts (1.65) has been used in (2.96), allowing to symmetrize the expression and to remove the hypersingularity of the gradient of the Green's kernel. The discretization of the right-hand side is

$$\begin{aligned} [\boldsymbol{e}_p]_m &\coloneqq -\frac{1}{\eta} \langle \boldsymbol{n} \times \boldsymbol{\psi}_m^{(p)}, \boldsymbol{n} \times \boldsymbol{E}^{\mathrm{inc}} \rangle \\ &= -\frac{1}{\eta} \int_\Gamma \boldsymbol{\psi}_m^{(p)}(\boldsymbol{r}) \cdot \boldsymbol{E}^{\mathrm{inc}}(\boldsymbol{r}) \, \mathrm{d}S(\boldsymbol{r}) . \end{aligned} \tag{2.97}$$

In the following of the thesis, when there will be no ambiguity about the order, we will omit the index p in the subscripts of the system matrix, unknown and right-hand side vectors. The distinction will only be needed in Chapter 7 where a refinement in order (p-refinement) will be needed to generate an auxiliary space. We will also refer to the column elements as expansion / source functions, and the row elements as testing functions and we will use the same terminology for the integration points $\boldsymbol{r}'$ and $\boldsymbol{r}$, respectively.

2.3.2 The case of Thin PEC Layers

As the operators appearing in the EFIE are continuous across the boundary, the same equation can be used to model thin PEC layers, that form open surface manifolds. The same derivation, however, cannot be done with the MFIE and CFIE that can be used only to model closed PEC boundaries [146].

Thin PEC layers model the limit case where the thickness of the PEC boundary is negligible with respect to the other dimensions of the object and with respect to the wavelength. In that limit, the two parallel surfaces merge together, and a thin layer is formed [146] (see Figure 2.2 below). We refer to [27] and references therein for mathematical analysis of the EFIE on open structures.

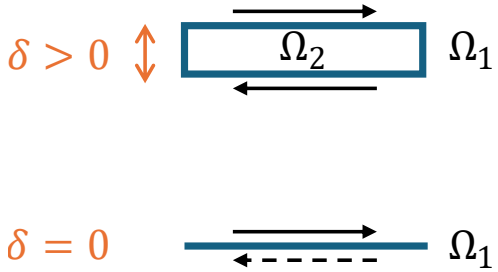


Figure 2.2: Cross section of a closed PEC boundary of thickness $\delta > 0$ and associated thin layer obtained as the limit case where δ goes to zero. Dotted lines mean that the basis functions on that side are removed.

From the discretization point of view, no basis function is associated with the edges along the border of the open surface, as the basis functions need to be normal-continuous across the edges. This can also be seen from the fact that the basis functions that were living on the corner of the boundary cancel out when the two triangles forming their domains are merged.

In the interior of the surface, each basis function is duplicated on the same interpolatory point with, for instance, opposite signs, corresponding to the currents on the "upper" and "lower" sides of the boundary. To build a linearly independent system, these basis functions need to be computed only once, and the coefficient unknown associated with the remaining basis function becomes the combination of one less the other coefficient associated with the two original basis functions. In other words, if we have two basis functions $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ with associated coefficients α_1 and α_2, respectively, one possible choice is to replace them by $\boldsymbol{\psi} = \boldsymbol{\psi}_1 = -\boldsymbol{\psi}_2$ with associated coefficient $\alpha = \alpha_1 - \alpha_2$ as we get $\alpha\boldsymbol{\psi} = \alpha_1\boldsymbol{\psi}_1 + \alpha_2\boldsymbol{\psi}_2$. With this choice, the representation of the unknowns in terms of surface current densities on both sides of the PEC layer becomes a representation in terms of jumps of the surface current densities across the PEC thin layer. Notice that removing α_1 and α_2 from the unknowns of the linear system by replacing them with α will prevent their recovery. Nevertheless, as the solution of the EFIE approximation is the sum of the basis functions times their associated coefficients (2.91), finding the coefficient α is all we need.

2.3.3 The Case of PEC Junctions

With thin PEC Layers, we may also model structures with junctions, that correspond to the limit case where the thickness of several connected PEC objects goes to zero along the same edge. We thus obtain edges in the mesh that are shared by more than two triangles.

We first consider the case where all the objects connected to the junction are thin PEC layers. For an edge e shared by n_e triangles, we automatically have $(p+1)n_e$ basis functions along that edge. For the same reason than explained in the case of a simple open boundary (without junctions), one basis function per interpolatory point needs to be removed in order to provide a linearly independent system [89]. It can be seen, indeed, that summing the basis functions associated with the same interpolatory point located on an edge will give zero, so that one of these basis functions is minus the sum of all the others. For that interpolatory point located on an edge, we can then build $(n_e - 1)$ basis functions, with associated coefficients for which the sum is the same than that of the n_e original unknowns. With these notations, the number of basis functions—and the dimension of the Raviart-Thomas space—is

$$N_p = (p+1)\sum_{e=1}^{E}(n_e - 1) + p(p+1)C\,. \tag{2.98}$$

Notice that $n_e = 2$ in the case where edge e does not form a junction and is inside the open surface, and $n_e = 1$ when it is on the border of the surface, which gives the standard number of unknowns (2.53) if there is no junction at all.

Finally, another type of junction configuration can also be modeled, where not all the PEC objects connected to the edge are open boundaries. In that case, no basis function connecting the two sides of any closed object is defined, as there was no current there before merging the edges of the mesh (i.e., before considering the thickness of the PEC to be zero at that location). In that configuration, there is no linear dependency between the basis functions, so that none of them is removed [146]. We finally denote by c_e the number of closed bodies connected to the edge e. With these notations

$$N_p = (p+1)\sum_{e=1}^{E}(n_e - \max(1, c_e)) + p(p+1)C\,, \tag{2.99}$$

where $\max(1, c_e) = 1$ if there is no closed boundary connected to the edge e.

To recap, different possible cases of PEC junctions are illustrated in Figure 2.3.

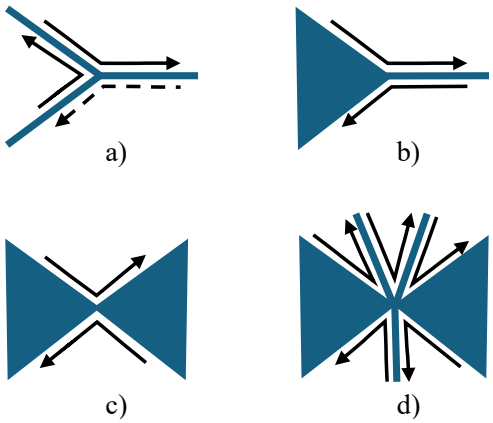


Figure 2.3: Cross section of a junction edge with different types of configuration as described in [146]: a) all surfaces are open, dotted lines mean that the basis functions on that side are removed, b) all surfaces are closed, c) one surface is closed and the other is open, d) more general case with several closed and open surfaces. In cases c) and d), the number of closed bodies connected to the edge e is $c_e = 2$.

2.4 Numerical Integration

The computation of the boundary element matrix discretization of the integral equations requires the use of numerical quadrature methods. Notice that the entries of the system matrices arising from the boundary element method are evaluations of double integrals, which imply that most of the elements of the matrix are non-zero. This is different from the more standard finite element method, for which the Galerkin discretization of a differential operator introduces a single integral only—which implies the computation of a (very) sparse matrix.

Moreover, the integrals that we need to evaluate to assemble the system matrices are singular. In particular, standard quadrature rules would imply the evaluation of singular functions of order of magnitude $O(|\boldsymbol{r}' - \boldsymbol{r}|^{-1})$ in (2.95)–(2.96) and $O(|\boldsymbol{r}' - \boldsymbol{r}|^{-2})$ (1.92), where $\boldsymbol{r}'$ and $\boldsymbol{r}$ are the source and testing points, respectively. This is not a desirable feature to apply standard Gaussian quadrature rules, as it would result in the evaluation of functions that are divided by zero, or by values close to zero. Therefore, we need to use specialized methods to approximate the integral of singular functions.

Let us define an integrable function of two variables $f(\boldsymbol{r}, \boldsymbol{r}')$ with singularity at most $O(|\boldsymbol{r}' - \boldsymbol{r}|^{-2})$—which encompasses the different integral operators considered in that thesis, including the cases where the functions to integrate behave as $O(|\boldsymbol{r}' - \boldsymbol{r}|^{-1})$ and $O(1)$. The double integrals on the surface boundary Γ can thus be approximated as

$$\begin{aligned}\int_\Gamma \int_\Gamma f(\boldsymbol{r}, \boldsymbol{r}')\, \mathrm{d}S(\boldsymbol{r}')\mathrm{d}S(\boldsymbol{r}) &\approx \sum_K \sum_L \int_K \int_L f(\boldsymbol{r}, \boldsymbol{r}')\, \mathrm{d}S(\boldsymbol{r}')\mathrm{d}S(\boldsymbol{r})\,, \\ &= \sum_K \sum_L \int_{\widehat{K}} \int_{\widehat{K}} \widehat{f}_{L,M}(\widehat{\boldsymbol{r}}, \widehat{\boldsymbol{r}}')\, \mathrm{d}\widehat{\boldsymbol{r}}'\, \mathrm{d}\widehat{\boldsymbol{r}}\,,\end{aligned} \tag{2.100}$$

where

$$\widehat{f}_{L,M}(\widehat{\boldsymbol{r}}, \widehat{\boldsymbol{r}}') = f(\widehat{F}_K(\widehat{\boldsymbol{r}}), \widehat{F}_L(\widehat{\boldsymbol{r}}'))\widehat{\mathcal{J}}_K(\widehat{\boldsymbol{r}})\widehat{\mathcal{J}}_L(\widehat{\boldsymbol{r}}')\,. \tag{2.101}$$

Notice that when the elements K and L are touching, the inner integral (on triangle K) is considered as a principal value.

The approximation of these integrals is then separated into three cases:

1) When the two triangles L and K are far enough from each other, standard Gaussian quadratures can be used. We define Gaussian quadrature points $\widehat{\boldsymbol{r}}_i = (u_i, v_i)$ in the reference triangle, associated with weights w_i such that

$$\int_{\widehat{K}} \int_{\widehat{K}} \widehat{f}_{L,M}(\widehat{\boldsymbol{r}}, \widehat{\boldsymbol{r}}')\, \mathrm{d}\widehat{\boldsymbol{r}}'\, \mathrm{d}\widehat{\boldsymbol{r}} \approx \sum_i w_i \sum_j w_j\, \widehat{f}_{L,M}(\widehat{\boldsymbol{r}}_i, \widehat{\boldsymbol{r}}_j)\,. \tag{2.102}$$

In general, these quadrature rules are able to exactly integrate polynomials of a given order, and the approximation error still reduces with the number of points for non-polynomial functions. Notice that the more $f(\boldsymbol{r}, \boldsymbol{r}')$ is having variations—because of increasing order of the basis functions and because of increasing wavenumber in the Green's kernel—the more quadrature points we need. For instance, the number of Gaussian points has to be increased quadratically with respect to the order of basis functions [115].

2) When the geometrical elements are touching each other, different methods can be used, such as the so-called singularity subtraction and singularity cancellation methods. The singularity subtraction [74] consists in removing the singular part of the function $f(\boldsymbol{r}, \boldsymbol{r}')$ for which we know the analytical solution, and to use Gaussian quadrature methods to evaluate the remaining part. The singularity cancellation [102] consists in applying a coordinate transform to the

integration variables, with an associated Jacobian determinant that cancels out the singularity in $f(\boldsymbol{r}, \boldsymbol{r}')$. In [123], Sauter and Schwab propose three singularity cancellation techniques for the cases of common triangle, common edge and common vertex. For each of these specific cases, the technique proceeds to a mapping of the pair of reference triangles $\widehat{K} \times \widehat{K}$ into a combination of several square domains, for which the associated Jacobian determinant cancels the $O(|\boldsymbol{r}' - \boldsymbol{r}|^{-2})$ singularity. Gaussian quadrature rules for the square are then used to approximate these integrals.

3) When the triangles K and L are too close to apply the Gaussian quadrature, but are not touching, different techniques need to be applied. These integrals are often referred to as near singular integrals. A technique may consist in splitting $f(\boldsymbol{r}', \boldsymbol{r})$ into two contributions [101]. The first contribution is $f(\boldsymbol{r}'_\perp, \boldsymbol{r})$, where $\boldsymbol{r}'_\perp$ is the orthogonal projection of $\boldsymbol{r}' \in L$ onto the triangle K, for which the associated integral can be computed via singularity cancellation or singularity extraction. The second contribution is the remainder $f(\boldsymbol{r}', \boldsymbol{r}) - f(\boldsymbol{r}'_\perp, \boldsymbol{r})$, for which the associated integral can be evaluated with standard Gaussian quadrature rules.

2.5 Discussion on the Discretization of the Other Variational Problems

In this section, we recap the differences between the discretization of the EFIE variational formulation and the discretization of the other variational formulations (for the MFIE, CFIE and PMCHWT). These formulations will not be further explored in the following of the thesis and this is done only briefly to introduce the possible extensions of our work and the difficulties that may be encountered while doing so.

2.5.1 Discretization of the MFIE and CFIE

The discretization of the MFIE introduces the need for the computation of a so-called mixed Gram matrix, that evaluates the inner products between $H^{-1/2}(\mathrm{div}_\Gamma, \Gamma)$ and $H^{-1/2}(\mathrm{curl}_\Gamma, \Gamma)$ boundary elements. While this inner product is L^2-stable with the continuous spaces, it produces an ill-conditioned Gram matrix when using the same boundary element spaces (up to the rotation) to discretize both $H^{-1/2}(\mathrm{div}_\Gamma, \Gamma)$ and $H^{-1/2}(\mathrm{curl}_\Gamma, \Gamma)$ [42, 28].

At order zero, when using RWG functions (2.76) to expand the current (2.91), this problem has been solved by introducing the Buffa-Christiansen (BC) functions [7], which are found as a linear combination of RWG functions built on the barycentric refinement of the mesh. These functions are said to be dual to the RWG functions in the sense that the dimension of its solenoidal and quasi-irrotational subspaces are exchanged with respect to the dimension of the solenoidal and quasi-irrotational subspaces of the Raviart-Thomas space of order 0, respectively. This property is the key to obtain a well-conditioned Gram matrix, that evaluates the inner products between RWG and rotated BC functions [7, 46]. At higher order, the construction of a dual space sharing similar properties has been introduced in [135]. Another construction will be proposed in Chapter 7 and will be applied to the discretization of the Calderón identity in the context of operator preconditioning of the EFIE.

2.5.2 Discretization of PMCHWT

The discretization of the PMCHWT variational formulation can be done by using the Raviart-Thomas elements for both expansion and testing spaces, and both magnetic and electric unknowns.

The modelling of more than two regions is obtained by superposing all the problems that we obtain when Ω_1 represents one of the regions, where the excitation fields and currents are non-zero, and Ω_2 contains all the other regions, where the excitation fields and currents are set to zero (see Section 1). One major challenge in that case, however, is related to the presence of several dielectric interfaces touching at the same edge, that we refer to as dielectric junctions.

Differently from the case of PEC junctions, the basis functions at a dielectric junction need to share the same unknowns when they are associated with the same interpolatory point, which is a consequence of the boundary conditions (1.14). From the implementation point of view, this can be done by assembling the matrix discretization separately for each domain and by summing at the same row and same column the contributions of the basis functions associated with an identical interpolatory point, located at the interface between several domains.

The situation complexifies further when the junction is made of several dielectric interfaces, thin PEC layers and closed PEC boundaries. The rules to set the coefficient unknowns can still be obtained as a consequence of the boundary conditions (1.14), and can be found, for instance, in [146]. Notably, when the edge is shared with PEC boundaries, the magnetic surface current density $\boldsymbol{M}$ is zero and no magnetic basis functions are set for the interpolatory points along the junction, in every domain. This means that the number of basis functions to discretize the electric and magnetic surface current densities may not be the same when dealing with composite structures. Along PEC boundaries, and at junctions with PEC boundaries, the PMCHWT formulation is replaced by the EFIE, MFIE or CFIE. It is advised in general to chose the EFIE for thin PEC layers and the CFIE for closed PEC objects [146] in order to avoid spurious resonances. In any case, this results in an overall system discretization that relies on hybrid formulations.

Very few preconditioning techniques are compatible with the modelling of dielectric junctions. An alternative proposed in [43, 47] is to get rid of the junctions by considering a fictitious space of zero thickness between the different objects, so that the interfaces are duplicated and the junctions are artificially removed. This way of modelling the interfaces is referred to as the multi-trace formulation, in opposition to the single-trace formulation that refers to the technique of enforcing continuity by merging the unknowns, as presented above. However, the multi-trace formulation results in two major drawbacks: (i) every unknown is duplicated on each interface, including the case where there is no junction, which implies a system that is at least twice bigger than with the single-trace formulation; and (ii) this fills the entire matrix of non-zero coefficients, which is not the case with the single-trace formulation as interactions associated with non-touching regions are zero in that case.

Chapter 3

Numerical solvers

3.1 Spectral Decompositions, Pseudoinversion and Orthogonal Projections

In this section, we collect the different notions of linear algebra that allow us to define the eigenvalue and singular value spectra of a matrix as well as their connections to inversion, pseudoinversion, orthogonal projections and stability of a linear system. These results can be found in the books [14, 96, 130].

Definition 3.1 (Eigenvalues)**.** Let $\mathbf{A}$ be a matrix of size $N \times N$. We say that $\upsilon \in \mathbb{C}$ is an eigenvalue of $\mathbf{A}$ if their exists an eigenvector $\boldsymbol{x}$ such that $\mathbf{A}\boldsymbol{x} = \upsilon\boldsymbol{x}$.

Theorem 3.1 (Cayley-Hamilton)**.** The matrix $\mathbf{A}$ possesses N eigenvalues (possibly equal to the same value multiple times, or equal to zero), that are the roots of the characteristic polynomial $\det(\upsilon\mathbf{I} - \mathbf{A})$. In the following we denote by $\{\upsilon_i\}_{i=1}^N$ the set of eigenvalues of $\mathbf{A}$, such that $|\upsilon_1| \geq \ldots \geq |\upsilon_N|$.

Theorem 3.2 (Courant-Fischer-Weyl min-max principle)**.** Let $\mathbf{A}$ be an Hermitian matrix (i.e., $\mathbf{A} = \mathbf{A}^*$) of size N and let us consider the Rayleigh-Ritz quotient (for $\boldsymbol{x} \neq 0$)

$$R_{\mathbf{A}}(\boldsymbol{x}) = \frac{\boldsymbol{x}^*\mathbf{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}\,. \tag{3.1}$$

Then,

$$\upsilon_i = \max_{\substack{\mathcal{M}\subset\mathbb{R}^N \\ \dim(\mathcal{M})=i}} \min_{\boldsymbol{x}\in\mathcal{M}} R_{\mathbf{A}}(\boldsymbol{x}) = \min_{\substack{\mathcal{M}\subset\mathbb{R}^N \\ \dim(\mathcal{M})=N-i+1}} \max_{\boldsymbol{x}\in\mathcal{M}} R_{\mathbf{A}}(\boldsymbol{x})\,, \tag{3.2}$$

with, in particular,

$$\upsilon_1 = \max_{\boldsymbol{x}\in\mathbb{R}^N} R_{\mathbf{A}}(\boldsymbol{x})\,, \qquad \upsilon_N = \min_{\boldsymbol{x}\in\mathbb{R}^N} R_{\mathbf{A}}(\boldsymbol{x})\,. \tag{3.3}$$

Theorem 3.3 (Diagonalization of an Hermitian Matrix)**.** Let $\mathbf{A}$ be an Hermitian matrix of size N. There exists a unitary matrix $\mathbf{U}$ such that $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$, where $\mathbf{D}$ is a diagonal matrix containing the eigenvalues of $\mathbf{A}$ (i.e., $[\mathbf{D}]_{ii} = \upsilon_i$). Notice that the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{A}$, which form an orthonormal basis of $\mathbb{R}^N$.

Definition 3.2 (Singular Value Decomposition (SVD))**.** Let $\mathbf{A}$ be any matrix of size $M \times N$. There exist a unitary matrix $\mathbf{U}$ of size $M \times M$ ($\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}$), a unitary matrix $\mathbf{V}$ of size

$N \times N$ ($\mathbf{V}\mathbf{V}^* = \mathbf{V}^*\mathbf{V} = \mathbf{I}$) and a diagonal matrix $\mathbf{S}$ of size $M \times N$ with non-negative diagonal entries, such that $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^*$. We denote by $\{s_i\}_{i=1}^{\min(M,N)}$ the diagonal entries of $\mathbf{S}$, which are, by definition, the singular values of $\mathbf{A}$, with $s_1 \geq \ldots \geq s_{\min(M,N)} \geq 0$, while the columns of $\mathbf{U}$ and $\mathbf{V}$ are there associated singular vectors. Notice that if $\mathbf{A}$ is Hermitian, then $\mathbf{V} = \mathbf{U}$ and its singular value decomposition matches with its eigenvalue decomposition.

Proposition 3.1 (Characterization of the SVD)**.** Let us consider the singular value decomposition $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^*$. Then, $\mathbf{A}\mathbf{A}^* = \mathbf{U}\mathbf{S}\mathbf{S}^\mathsf{T}\mathbf{U}^*$ and $\mathbf{A}^*\mathbf{A} = \mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{S}\mathbf{V}^*$ are two Hermitian semidefinite positive matrices, with eigenvalues $\{s_i^2\}_{i=1}^{\min(M,N)}$ (completed with zeros to match the dimension of $\mathbf{V}$ and $\mathbf{U}$). Conversely, the non-zero singular values of $\mathbf{A}$ are the square roots of the non-zero eigenvalues of $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$.

Theorem 3.4 (Courant-Fischer-Weyl min-max principle for singular values)**.** Let $\mathbf{A}$ be any matrix of size $M \times N$ with associated normal Rayleigh-Ritz quotient (for $\boldsymbol{x} \neq 0$)

$$R_{\mathbf{A}^*\mathbf{A}}(\boldsymbol{x}) = \frac{\boldsymbol{x}^*\mathbf{A}^*\mathbf{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}} = \frac{|\mathbf{A}\boldsymbol{x}|^2}{|\boldsymbol{x}|^2}\,. \tag{3.4}$$

Then, for $i = 1, \ldots, \min(M, N)$,

$$s_i = \max_{\substack{\mathcal{M}\subset\mathbb{R}^N \\ \dim(\mathcal{M})=i}} \min_{\boldsymbol{x}\in\mathcal{M}} \frac{|\mathbf{A}\boldsymbol{x}|}{|\boldsymbol{x}|} = \min_{\substack{\mathcal{M}\subset\mathbb{R}^N \\ \dim(\mathcal{M})=\min(M,N)-i+1}} \max_{\boldsymbol{x}\in\mathcal{M}} \frac{|\mathbf{A}\boldsymbol{x}|}{|\boldsymbol{x}|}\,, \tag{3.5}$$

with, in particular,

$$s_1 = \max_{\boldsymbol{x}\in\mathbb{R}^N} \frac{|\mathbf{A}\boldsymbol{x}|}{|\boldsymbol{x}|}\,, \qquad s_N = \min_{\boldsymbol{x}\in\mathbb{R}^N} \frac{|\mathbf{A}\boldsymbol{x}|}{|\boldsymbol{x}|}\,. \tag{3.6}$$

Definition 3.3 (Rank of a Matrix)**.** Let us consider the matrix $\mathbf{A}$ of size $M \times N$ with singular values $\{s_i\}_{i=1}^{\min(M,N)}$. There exists $r \leq \min(M, N)$ such that $s_i = 0$ for all $i > r$. This value corresponds to the rank of $\mathbf{A}$, which is the number of linearly independent columns and rows of $\mathbf{A}$. The matrix is said to be full rank / invertible when $r = M = N$, full row rank when $r = M < N$, full column rank when $r = N < M$, and rank-deficient otherwise.

Theorem 3.5 (Spectral Norm and Spectral Radius)**.** For any matrix $\mathbf{A}$, its spectral norm is

$$\|\mathbf{A}\| = s_1\,, \tag{3.7}$$

where s_1 is its maximal singular value. For a square matrix, its spectral radius is defined as

$$\rho(\mathbf{A}) = |\upsilon_1|\,, \tag{3.8}$$

where υ_1 is its eigenvalue of maximal modulus. Furthermore, it can be shown that

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\|\,. \tag{3.9}$$

Definition 3.4 (Reduced SVD)**.** Let us consider a matrix $\mathbf{A}$ of rank r with singular value decomposition $\mathbf{U}\mathbf{S}\mathbf{V}^*$. We can set $\mathbf{U} = (\widetilde{\mathbf{U}} \;\; \widetilde{\mathbf{U}}_\perp)$ and $\mathbf{V} = (\widetilde{\mathbf{V}} \;\; \widetilde{\mathbf{V}}_\perp)$ such that

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} \widetilde{\mathbf{U}} & \widetilde{\mathbf{U}}_\perp \end{pmatrix} \begin{pmatrix} \widetilde{\mathbf{S}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \widetilde{\mathbf{V}}^* \\ \widetilde{\mathbf{V}}_\perp^* \end{pmatrix} \\ &= \widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*\,, \end{aligned} \tag{3.10}$$

where $\widetilde{\mathbf{S}}$ is a square diagonal matrix of size $r \times r$ containing the positive singular values of $\mathbf{A}$, and where $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ are rectangular matrices of size $M \times r$ and $N \times r$ respectively. This decomposition is referred to as the reduced singular value decomposition. Notice that $\mathbf{U}$ and $\mathbf{V}$ being unitary matrices, we have $\widetilde{\mathbf{U}}^*\widetilde{\mathbf{U}} = \widetilde{\mathbf{V}}^*\widetilde{\mathbf{V}} = \mathbf{I}$, $\widetilde{\mathbf{U}}^*\widetilde{\mathbf{U}}_\perp = \widetilde{\mathbf{V}}^*\widetilde{\mathbf{V}}_\perp = \mathbf{0}$, $\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^* = \mathbf{I} - \widetilde{\mathbf{U}}_\perp\widetilde{\mathbf{U}}_\perp^*$ and $\widetilde{\mathbf{V}}\widetilde{\mathbf{V}}^* = \mathbf{I} - \widetilde{\mathbf{V}}_\perp\widetilde{\mathbf{V}}_\perp^*$.

Definition 3.5. The range (or the column space) of a matrix $\mathbf{A}$ refers to the vector space spanned by its columns, i.e.,

$$\text{range}\,\mathbf{A} \coloneqq \left\{\boldsymbol{x} \in \mathbb{C}^M \,:\, \exists \boldsymbol{y} \in \mathbb{C}^N \text{ such that } \boldsymbol{x} = \mathbf{A}\boldsymbol{y}\right\} . \tag{3.11}$$

If we denote by $\widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*$ the reduced SVD of $\mathbf{A}$, then range $\mathbf{A}$ = range $\widetilde{\mathbf{U}}$.

Definition 3.6 (Moore-Penrose Pseudoinverse [14])**.** Let us consider a matrix $\mathbf{A}$ with a reduced singular value decomposition $\widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*$. The Moore-Penrose pseudoinverse of $\mathbf{A}$ is defined as

$$\mathbf{A}^+ = \widetilde{\mathbf{V}}\widetilde{\mathbf{S}}^{-1}\widetilde{\mathbf{U}}^* . \tag{3.12}$$

Notice that if $\mathbf{A}$ is a (square) invertible matrix, this definition matches with the definition of the inverse of $\mathbf{A}$, as in that case $\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^* = \mathbf{I}$ and $\widetilde{\mathbf{V}}\widetilde{\mathbf{V}}^* = \mathbf{I}$. The Moore-Penrose pseudoinverse allows to solve ill-posed systems, i.e., when $\mathbf{A}$ is rectangular or square and not invertible, but the system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ can be solved because $\boldsymbol{b}$ belongs to the range of $\mathbf{A}$. More generally, the vector $\boldsymbol{x} = \mathbf{A}^+\boldsymbol{b}$ minimizes $|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}|$ (in the Euclidean norm).

Proposition 3.2 (Properties of the Moore-Penrose Pseudoinverse [14])**.** We can check that

1. $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$.
2. $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$.
3. $\mathbf{A}\mathbf{A}^+$ and $\mathbf{A}^+\mathbf{A}$ are Hermitian matrices.
4. If $\mathbf{A}$ has linearly independent columns (full column rank) and $\mathbf{B}$ has linearly independent rows (full row rank), then $(\mathbf{A}\mathbf{B})^+ = \mathbf{B}^+\mathbf{A}^+$.
5. If $\mathbf{A}$ is full column rank, then $\mathbf{A}^+\mathbf{A} = \mathbf{I}$.
6. If $\mathbf{A}$ is full row rank, then $\mathbf{A}\mathbf{A}^+ = \mathbf{I}$.

Proposition 3.3. Let us consider a matrix $\mathbf{A}$ with a singular value decomposition $\mathbf{U}\mathbf{S}\mathbf{V}$ and a reduced singular value decomposition $\widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*$. We also define the orthogonal complement to $\widetilde{\mathbf{U}}$ as $\widetilde{\mathbf{U}}_\perp$ such that $\mathbf{U} = \begin{pmatrix}\widetilde{\mathbf{U}} & \widetilde{\mathbf{U}}_\perp\end{pmatrix}$ and $\widetilde{\mathbf{U}}^*\widetilde{\mathbf{U}}_\perp = \mathbf{0}$. We thus have

$$\begin{aligned} \mathbf{A}\left(\mathbf{A}^*\mathbf{A}\right)^+\mathbf{A}^* &= \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^* , \\ \mathbf{I} - \mathbf{A}\left(\mathbf{A}^*\mathbf{A}\right)^+\mathbf{A}^* &= \widetilde{\mathbf{U}}_\perp\widetilde{\mathbf{U}}_\perp^* . \end{aligned} \tag{3.13}$$

Proof. We check from the definition of the pseudoinverse and from $\widetilde{\mathbf{U}}^*\widetilde{\mathbf{U}} = \widetilde{\mathbf{V}}^*\widetilde{\mathbf{V}} = \mathbf{I}$ that

$$\begin{aligned} \mathbf{A}\left(\mathbf{A}^*\mathbf{A}\right)^+\mathbf{A}^* &= \widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*\left(\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}\widetilde{\mathbf{U}}^*\widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*\right)^+\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}\widetilde{\mathbf{U}}^* = \widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*\left(\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}^2\widetilde{\mathbf{V}}^*\right)^+\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}\widetilde{\mathbf{U}}^* \\ &= \widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*\left(\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}^{-2}\widetilde{\mathbf{V}}^*\right)\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}\widetilde{\mathbf{U}}^* = \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^* . \end{aligned} \tag{3.14}$$

Further, from the definition of $\mathbf{U}$, we have

$$\mathbf{U}\mathbf{U}^* = \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^* + \widetilde{\mathbf{U}}_\perp\widetilde{\mathbf{U}}_\perp^* = \mathbf{I} , \tag{3.15}$$

such that $\mathbf{I} - \mathbf{A}\left(\mathbf{A}^*\mathbf{A}\right)^+\mathbf{A}^* = \widetilde{\mathbf{U}}_\perp\widetilde{\mathbf{U}}_\perp^*$. □

Definition 3.7. The orthogonal projector onto the range of $\mathbf{A}$ is given by

$$\mathbf{P}^A = \mathbf{A}\,(\mathbf{A}^*\mathbf{A})^+\,\mathbf{A}^*\,, \tag{3.16}$$

while the orthogonal projector onto the space orthogonal to the range of $\mathbf{A}$ is $\mathbf{I} - \mathbf{P}^A$.

Proof. Let us define two matrices $\mathbf{A} \in \mathbb{C}^{N\times M}$ and $\mathbf{B} \in \mathbb{C}^{N\times(N-M)}$ that are orthogonal to each other (i.e., $\mathbf{A}^*\mathbf{B} = \mathbf{0}$). Then, for any vector $\boldsymbol{x} \in \mathbb{C}^N$, the rank theorem states that there exist $\boldsymbol{y}_A \in \mathbb{C}^M$ and $\boldsymbol{y}_B \in \mathbb{C}^{N-M}$ such that the orthogonal decomposition $\boldsymbol{x} = \mathbf{A}\boldsymbol{y}_A + \mathbf{B}\boldsymbol{y}_B$ is unique. We denote by $\widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}$ the reduced SVD of $\mathbf{A}$. We thus have from Proposition 3.3

$$\begin{aligned}\mathbf{P}^A\boldsymbol{x} &= \left(\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^*\right)\widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*\boldsymbol{y}_A + \left(\mathbf{A}\,(\mathbf{A}^*\mathbf{A})^+\,\mathbf{A}^*\right)\mathbf{B}\boldsymbol{y}_B \\ &= \widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*\boldsymbol{y}_A = \mathbf{A}\boldsymbol{y}_A\,,\end{aligned} \tag{3.17}$$

and

$$(\mathbf{I} - \mathbf{P}^A)\boldsymbol{x} = \mathbf{B}\boldsymbol{y}_B = \mathbf{P}^B\boldsymbol{x}\,. \tag{3.18}$$

□

Theorem 3.6. [Truncated SVD] Let us consider a matrix $\mathbf{A}$ of rank r, with reduced singular value decomposition $\widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^*$. The best approximation of $\mathbf{A}$ by a matrix of rank $r' < r$ is given by the truncated SVD

$$\mathbf{A}_{r'} = \widetilde{\mathbf{U}}_{r'}\widetilde{\mathbf{S}}_{r'}\widetilde{\mathbf{V}}^*_{r'} \tag{3.19}$$

where $\widetilde{\mathbf{U}}_{r'}$ and $\widetilde{\mathbf{V}}_{r'}$ are the matrices containing the r' first columns of $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{S}}_{r'}$ is the diagonal matrix containing the upper left $r' \times r'$ block of $\widetilde{\mathbf{S}}$. Moreover, the truncating error is

$$\|\mathbf{A} - \mathbf{A}_{r'}\| = s_{r'+1}\,, \tag{3.20}$$

where $s_{r'+1}$ is the $(r'+1)$-th highest singular value of $\mathbf{A}$.

Definition 3.8 (Condition number)**.** Let us consider a (square) invertible matrix $\mathbf{A}$ of size N. Its condition number is defined as

$$\mathrm{cond}(\mathbf{A}) \coloneqq \|\mathbf{A}\|\,\|\mathbf{A}^{-1}\| = \frac{s_1}{s_N}\,, \tag{3.21}$$

where s_1 ans s_N are the maximal and minimal singular values of $\mathbf{A}$, respectively. For a (possibly rectangular) matrix of rank r, we define the pseudo condition number as

$$\mathrm{cond}^+(\mathbf{A}) \coloneqq \|\mathbf{A}\|\,\|\mathbf{A}^+\| = \frac{s_1}{s_r}\,, \tag{3.22}$$

where s_r is the smallest non-zero singular value of $\mathbf{A}$. A matrix with a high condition number (or pseudo condition number) is said to be ill-conditioned.

Theorem 3.7 (Error bound on the solution of invertible linear systems)**.** Let us consider the system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ where $\mathbf{A}$ is a (square) invertible matrix and $\boldsymbol{b}$ is a non-zero vector. If we consider a perturbation $\delta\boldsymbol{b}$ on $\boldsymbol{b}$, the perturbation $\delta\boldsymbol{x}$ on the solution $\boldsymbol{x}$, that satisfies $\mathbf{A}(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b}$, is bounded in relative norm by the condition number as

$$\frac{|\delta\boldsymbol{x}|}{|\boldsymbol{x}|} \leq \mathrm{cond}(\mathbf{A})\frac{|\delta\boldsymbol{b}|}{|\boldsymbol{b}|}\,. \tag{3.23}$$

If we consider a perturbation $\delta\mathbf{A}$ on $\mathbf{A}$, the perturbation $\delta\boldsymbol{x}$ on the solution $\boldsymbol{x}$, that satisfies $(\mathbf{A} + \delta\mathbf{A})(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b}$, is bounded in relative norm by the condition number as

$$\frac{|\delta\boldsymbol{x}|}{|\boldsymbol{x} + \delta\boldsymbol{x}|} \leq \mathrm{cond}(\mathbf{A})\frac{\|\delta\mathbf{A}\|}{\|\mathbf{A}\|}\,. \tag{3.24}$$

3.2 Matrix Factorizations and Direct Solvers

Let us consider the linear system

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b}\,. \tag{3.25}$$

Direct solvers refer to the computation of the solution of (3.25) with algorithms that are done in a finite and predetermined number of steps. According to Theorem 3.7, the accuracy of the numerical solution depends on the working machine precision in addition to the condition number of $\mathbf{A}$. To solve the system, these algorithms usually rely on the pre-computation of a factorization of $\mathbf{A}$ [65], with matrices that are straightforward to invert. Examples of such matrices are

- Diagonal matrices, that imply multiplying the right-hand side by the inverse of the diagonal entries.
- Orthogonal / unitary matrices, that imply multiplying the right hand side by their transposes / conjugate transposes.
- Upper and lower triangular matrices, that imply solving the system with backward and forward substitutions.
- Permutation matrices, that imply reordering the elements of the solution.

The singular value decomposition (SVD), for instance, has these properties. If the system matrix is written as $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^*$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{S}$ is a diagonal matrix of non-zero entries, the solution of the system is formally given by $\boldsymbol{x} = \left(\mathbf{V}\left(\mathbf{S}^{-1}\left(\mathbf{U}^*\boldsymbol{b}\right)\right)\right)$, where two matrix-vector products with $\mathbf{U}^*$ and $\mathbf{V}$ and division by the diagonal elements of $\mathbf{S}$ are carried out.

The QR-decomposition is based on Gram-Schmidt orthogonalization and provides a factorization of the system matrix in the form $\mathbf{A} = \mathbf{Q}\mathbf{R}$, where $\mathbf{Q}$ is a unitary matrix and $\mathbf{R}$ is an upper triangular matrix. The solution of the system is formally given by $\boldsymbol{x} = \left(\mathbf{R}^{-1}\left(\mathbf{Q}^*\boldsymbol{b}\right)\right)$, where a matrix-vector product with the conjugate transpose of $\mathbf{Q}$ and a backward substitution with the elements of $\mathbf{R}$ are carried out.

The LU-factorization gives a factorization of $\mathbf{A}$ in terms of a multiplication between a lower triangular matrix $\mathbf{L}$ and an upper triangle matrix $\mathbf{U}$. This decomposition is not always possible and the most general case resorts to a permutation matrix $\mathbf{P}$, such that $\mathbf{A} = \mathbf{P}\mathbf{L}\mathbf{U}$. The solution of the linear system can be formally written as $\boldsymbol{x} = \left(\mathbf{U}^{-1}\left(\mathbf{L}^{-1}\left(\mathbf{P}^{\mathsf{T}}\boldsymbol{b}\right)\right)\right)$, which implies permuting the elements of $\mathbf{b}$ (with the map given by $\mathbf{P}$), and applying forward and backward substitutions with the matrices $\mathbf{L}$ and $\mathbf{U}$, respectively.

These direct solvers share the same computational complexity, which is cubic with respect to the size of the system [65]. More precisely, the factorization of $\mathbf{A}$ requires a cubic amount of operations while the solving steps require a quadratic number of operations. Nevertheless, it should be noted that the LU-factorization is faster to compute than the SVD and QR-decomposition, and it is usually preferred when solving well-posed linear systems (with a square and invertible matrix $\mathbf{A}$). The SVD and QR-decomposition are however useful to solve rectangular systems, or ill-posed systems associated with non-invertible square matrices, as these factorizations allow to compute $\mathbf{A}^+\boldsymbol{x}$ [14].

Despite their high complexity, direct solvers are widely used in computational electromagnetics for two main reasons:

- They are quite well adapted to deal with a large number of right-hand side vectors (i.e., for several excitation field and current sources in the electromagnetic problem) as the factorization can be done once for all.
- They are more robust to ill-conditioning than iterative solvers as the termination of the algorithm is ensured.

3.3 Iterative Solvers

Differently from direct solvers, iterative solvers are algorithms that compute an approximation of the solution step-by-step with a sequence of vector iterates $\{\boldsymbol{x}_i\}$ that converge to the exact solution of the linear system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ under some circumstances [122]. The number of iterations is not known in advance and it is not even ensured that the convergence will finally occur, except for certain solvers applied to some classes of matrices. For instance, the conjugate gradient (CG), which belongs to the family of Krylov solvers, converges in at most N iterations (where N is the size of the system) when solving Hermitian positive definite systems in exact arithmetic [77]. However, this is not even true with finite machine precision as small errors in the computation of the iterates, especially for ill-conditioned matrices, will hinder the computation of the actual converging sequence, and at least slow down that convergence, if not jeopardizing it [107]. Most of these solvers rely on stopping criteria, for instance when the relative residual norm

$$\frac{|\mathbf{A}\boldsymbol{x}_i - \boldsymbol{b}|}{|\boldsymbol{b}|} \tag{3.26}$$

or the relative difference between two iterates

$$\frac{|\boldsymbol{x}_i - \boldsymbol{x}_{i-1}|}{|\boldsymbol{x}_{i-1}|} \tag{3.27}$$

are lower than a given threshold $\mathrm{tol} > 0$. In this section, we briefly describe the principles of some of these iterative solvers.

3.3.1 Stationary methods

Stationary methods generate a sequence of iterates in the form $\boldsymbol{x}_{i+1} = F(\boldsymbol{x}_i)$ which, in case of convergence, tend towards the stationary point $\boldsymbol{x} = F(\boldsymbol{x})$ that matches with the solution of the linear system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ [122].

A first example is the Richardson method, where given a constant $\alpha > 0$,

$$\boldsymbol{x}_{i+1} = \mathrm{Rich}(\boldsymbol{x}_i, \boldsymbol{A}, \boldsymbol{b}) \coloneqq \boldsymbol{x}_i + \alpha(\boldsymbol{b} - \mathbf{A}\boldsymbol{x}_i)\,. \tag{3.28}$$

The method converges if $\|\mathbf{I} - \alpha\mathbf{A}\| < 1$ and for Hermitian positive definite matrices this is ensured if $\alpha = 2/(\upsilon_1 + \upsilon_N)$ where $\upsilon_1, \upsilon_N > 0$ are the maximal and minimal eigenvalues of $\mathbf{A}$, respectively.

Other examples of stationary methods are based on the decomposition of $\mathbf{A}$ in terms of its lower, diagonal and upper parts, that is $\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{U}$ where $\mathbf{D}$ is diagonal, $\mathbf{L}$ is lower triangular (without diagonal), and $\mathbf{U}$ is upper triangular (without diagonal). For instance, the Jacobi method is given by the sequence of iterates

$$\boldsymbol{x}_{i+1} = \mathrm{Jac}(\boldsymbol{x}_i, \boldsymbol{A}, \boldsymbol{b}) \coloneqq \mathbf{D}^{-1}(\boldsymbol{b} - (\mathbf{L} + \mathbf{U})\boldsymbol{x}_i) \tag{3.29}$$

that converges if and only if $\rho(\mathbf{D}^{-1}(\mathbf{L}+\mathbf{U})) < 1$ and the Gauss-Seidel method is given by the sequence of iterates

$$\boldsymbol{x}_{i+1} = \mathrm{GS}(\boldsymbol{x}_i, \boldsymbol{A}, \boldsymbol{b}) \coloneqq (\mathbf{L}+\mathbf{D})^{-1}(\boldsymbol{b}-\mathbf{U})\boldsymbol{x}_i) \tag{3.30}$$

that converges if and only if $\rho((\mathbf{L}+\mathbf{D})^{-1}\mathbf{U}) < 1$.

The convergence of these methods is clearly not ensured for every matrix, and it is often very slow when it occurs. The stationary methods can however be used in multigrid methods, which provide more efficient iterative solvers.

3.3.2 Multigrid methods

Multigrid methods are a widely used class of algorithms that are particularly well-suited to iteratively solve matrix systems arising from the discretization of elliptic partial differential equations [123, 104, 109, 19, 120, 134].

Given a matrix $\mathbf{A}^{(0)}$ of size N associated with the system $\mathbf{A}^{(0)}\boldsymbol{x}^{(0)} = \boldsymbol{b}^{(0)}$, the objective is to look for restriction and prolongation matrices $\mathbf{R}_{1,0}$ and $\mathbf{P}_{0,1}$ such that $\mathbf{A}^{(1)} = \mathbf{R}_{1,0}\mathbf{A}^{(0)}\mathbf{P}_{0,1}$ is a matrix of smaller size that shares some spectral properties with $\mathbf{A}^{(0)}$. In particular, if we denote by

$$\boldsymbol{r}_i^{(0)} = \boldsymbol{b}^{(0)} - \mathbf{A}^{(0)}\boldsymbol{x}_i^{(0)} \tag{3.31}$$

the residual associated with the first system for a given iterate $\boldsymbol{x}_i^{(0)}$, and we set

$$\boldsymbol{r}_i^{(1)} = \mathbf{R}_{1,0}\boldsymbol{r}_i^{(0)}\,, \tag{3.32}$$

then the solution of the system

$$\mathbf{A}^{(1)}\boldsymbol{e}_i^{(1)} = \boldsymbol{r}_i^{(1)} \tag{3.33}$$

allows to define the correction

$$\boldsymbol{x}_{i+1}^{(0)} = \boldsymbol{x}_i^{(0)} + \mathbf{P}_{0,1}\boldsymbol{e}_i^{(1)} \tag{3.34}$$

that should have a smaller residual norm than $\boldsymbol{x}_i$, i.e., $|\boldsymbol{r}_{i+1}^{(0)}| \leq |\boldsymbol{r}_i^{(0)}|$.

The construction of the restriction and prolongation matrices depends on the type of system we are solving. For instance, when $\mathbf{A}^{(0)}$ is obtained as the Galerkin discretization of a differential equation, these matrices can be built as the projection of the basis functions onto the functions defined on another grid level. In a more general context, these matrices can be built by inspecting the coefficients in the matrix $\mathbf{A}^{(0)}$ in order to reproduce the effects of a grid coarsening. Such techniques can be applied to a wider variety of matrices and are referred to as algebraic multigrids.

Further, this procedure can be reproduced at multiple levels. We denote by L the number of levels and by $\mathbf{R}_{\ell+1,\ell}$ and $\mathbf{P}_{\ell,\ell+1}$ the restriction and prolongation matrices for each level $\ell = 0, \ldots, L-1$ such that $\mathbf{A}^{(\ell+1)} = \mathbf{R}_{\ell+1,\ell}\mathbf{A}^{(\ell)}\mathbf{P}_{\ell,\ell+1}$ are matrices of decreasing size for which the solution of the associated linear systems allows to reduce the residual norm at the lower levels. Moreover, the coarsening matrices are made in a way that the size of the system at level $\ell+1$ is reduced by an almost constant factor with respect to the size of the system at level ℓ. If $L = O(\log N)$, the size of $\mathbf{A}_L$ is close to that factor, and its associated system can be solved with a direct solver at little cost.

For solving the system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, multigrid methods then consist in repeating a recursive procedure (for instance the V-cycle described in Algorithm 2) until reaching the chosen tolerance threshold (parameter tol) or the maximum number of iterations (parameter maxiter); as shown in Algorithm 1.

At each level of the V-cycle, a stationary method referred to as smoother is employed to approximate the solution at that given level. We can choose for instance the Richardson, Jacobi or Gauss-Seidel methods to define the smoothing function (with the functions Rich, Jac and GS in (3.28), (3.29) and (3.30)). Only few iterations of the smoother are done at the beginning (parameter ν_1) and at the end (parameter ν_2) of the level. In between of these two smoothing steps, restriction (3.32), correction (3.33) and prolongation (3.34) are computed in order to improve the convergence obtained with the smoother. The direct inversion of the coarsened system in the correction step is only done at the highest level $\ell = L$ and the V-cycle is called recursively at the other levels in order to get an approximate solution of the system.

For some class of matrices, such as the Galerkin discretization of the Laplacian equation [120] or the Laplacian of graphs [19], this method is very effective and can be carried out in $\log N$ iterations, where N is the size of the system, such that the overall complexity is $N \log N$ (as $\mathbf{A}$ is sparse).

Algorithm 1 Multigrid($\mathbf{A}, \boldsymbol{b}, \boldsymbol{x}_0$, maxiter, tol, L)

Initialization of level 0:
$\mathbf{A}^{(0)} \leftarrow \mathbf{A}$
$\boldsymbol{x}_0^{(0)} \leftarrow \boldsymbol{x}_0$
Coarsening:
Compute $\mathbf{R}_{\ell+1,\ell}$, $\mathbf{P}_{\ell,\ell+1}$ and $\mathbf{A}^{(\ell+1)} = \mathbf{R}_{\ell+1,\ell}\mathbf{A}^{(\ell)}\mathbf{P}_{\ell,\ell+1}$, for $\ell = 0, \ldots, L-1$.
for $i = 0, \ldots,$ maxiter $- 1$ **do**
  $\boldsymbol{x}_{i+1}^{(0)} =$ V-cycle$(\boldsymbol{x}_i^{(0)}, \boldsymbol{b}, 0, L)$
  if $|\boldsymbol{b} - \mathbf{A}^{(0)}\boldsymbol{x}_{i+1}^{(0)}|/|\boldsymbol{b}| <$ tol **then**
    return $\boldsymbol{x}_{i+1}^{(0)}$
  end if
end for
return *No convergence*

Algorithm 2 V-cycle($\boldsymbol{x}^{(\ell)}, \boldsymbol{r}^{(\ell)}, \ell, L$)

Pre-smoothing: $\boldsymbol{x}^{(\ell)} \leftarrow$ smoother$(\boldsymbol{x}^{(\ell)}, \mathbf{A}^{(\ell)}, \boldsymbol{r}^{(\ell)})$ [ν_1 times]
Restriction: $\boldsymbol{r}^{(\ell+1)} = \mathbf{R}_{\ell+1,\ell}\left(\boldsymbol{r}^{(\ell)} - \mathbf{A}^{(\ell)}\boldsymbol{x}^{(\ell)}\right)$
Correction:
if $q < L$ **then**
  $\boldsymbol{e}^{(\ell+1)} \leftarrow$ V-cycle $\left(\boldsymbol{r}^{(\ell+1)}, \ell+1, L\right)$
else
  $\boldsymbol{e}^{(\ell+1)} \leftarrow$ solve $(\mathbf{A}_L, \boldsymbol{r}_L)$
end if
Prolongation: $\boldsymbol{x}^{(\ell)} \leftarrow \boldsymbol{x}^{(\ell)} + \mathbf{P}_{\ell,\ell+1}\boldsymbol{e}^{(\ell+1)}$
Post-smoothing: $\boldsymbol{x}^{(\ell)} \leftarrow$ smoother$(\boldsymbol{x}^{(\ell)}, \mathbf{A}^{(\ell)}, \boldsymbol{r}^{(\ell)})$ [ν_2 times]
return $\boldsymbol{x}^{(\ell)}$

3.3.3 Krylov methods

Another family of iterative solvers is based on the use of the Krylov subspaces [73]

$$\mathbb{K}_i(\mathbf{A}, \boldsymbol{r}) = \text{span}\left(\boldsymbol{r}, \mathbf{A}\boldsymbol{r}, \mathbf{A}^2\boldsymbol{r}, \dots, \mathbf{A}^i\boldsymbol{r}\right) , \tag{3.35}$$

as they generate a sequence of iterates in the form

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i , \tag{3.36}$$

where $\boldsymbol{p}_i \in \mathbb{K}_i(\mathbf{A}, \boldsymbol{r}_0)$ is the search direction, $\alpha_i \in \mathbb{R}$ is the step length at iteration i, and where the residuals are $\boldsymbol{r}_i = \boldsymbol{b} - \mathbf{A}\boldsymbol{x}_i$ (with $\boldsymbol{r}_0$ defining the Krylov space that is considered here).

The distinction between the different Krylov methods resides in the definition of the search direction and step length. For instance, in the Hermitian positive definite case (i.e., when $\boldsymbol{x}^*\mathbf{A}\boldsymbol{x}$ defines a norm) the method of the conjugate gradient (CG) [77] generates a sequence of search directions $\boldsymbol{p}_{i+1} = \boldsymbol{r}_{i+1} + \beta_i \boldsymbol{p}_i$, where β_i is computed to form $\mathbf{A}$-conjugate vectors

$$\boldsymbol{p}_i^* \mathbf{A} \boldsymbol{p}_j = 0 , \qquad i \neq j , \tag{3.37}$$

and computes the step length α_i to minimize the $\mathbf{A}$-norm of the residual

$$\alpha_i \coloneqq \underset{\alpha}{\text{argmin}}\, \boldsymbol{r}_{i+1}^* \mathbf{A} \boldsymbol{r}_{i+1} . \tag{3.38}$$

In particular, these two properties together with $\mathbf{A}$ being Hermitian positive definite imply the orthogonality between the residuals

$$\boldsymbol{r}_i^* \boldsymbol{r}_j = 0 , \qquad i \neq j , \tag{3.39}$$

and the method requires at most N iterations to find the exact solution assuming exact arithmetic, where N is the size of the system. This still holds for Hermitian indefinite matrices, providing $\boldsymbol{p}_i^* \mathbf{A} \boldsymbol{p}_i \neq 0$ for every iteration i until convergence. Also the following convergence estimate can be proven [107]

$$\sqrt{(\boldsymbol{x} - \boldsymbol{x}_i)^\mathsf{T} \mathbf{A} (\boldsymbol{x} - \boldsymbol{x}_i)} \leq 2 \left(\frac{\sqrt{\text{cond}(\mathbf{A})} - 1}{\sqrt{\text{cond}(\mathbf{A})} + 1} \right)^i \sqrt{(\boldsymbol{x} - \boldsymbol{x}_0)^\mathsf{T} \mathbf{A} (\boldsymbol{x} - \boldsymbol{x}_0)} , \tag{3.40}$$

where $\boldsymbol{x}$ is the exact solution of the system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$. This convergence estimate indicates that the smaller is the condition number of $\mathbf{A}$, the faster is the convergence in the norm induced by matrix $\mathbf{A}$. Moreover, it is known that ill-conditioning of $\mathbf{A}$ introduces numerical errors in the computation of the iterates, which implies loss of orthogonality between the residuals, compromising the convergence of the method [107].

In the more general case, when $\mathbf{A}$ is not Hermitian, the conjugate gradient method is not suitable to solve the underlying system anymore, but other Krylov methods with similar properties can be employed instead.

The generalized minimal residual (GMRES) method [121] minimizes the residual norm $|\boldsymbol{r}_i|$ with a sequence of orthonormal search directions $\boldsymbol{p}_i^* \boldsymbol{p}_j = \delta_{i,j}$. In this way, the method converges in exact arithmetic in at most N iterations. One major drawback, however, is that it requires to store all the previous search directions to build an orthonormal basis with the updated search direction, leading to the storage of an increasingly large number of vectors. To tackle this issue,

the orthogonalization process may be restarted after a given number of iterations. While that strategy may be effective for some systems, this does not ensure the convergence of the algorithm in a general context however.

The biconjugate gradient (BiCG) method [122] generates two sequences of iterates to solve $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ and $\mathbf{A}^*\widetilde{\boldsymbol{x}} = \boldsymbol{b}$, respectively, with bi-conjugate search directions

$$\boldsymbol{p}_i^* \mathbf{A} \widetilde{\boldsymbol{p}}_j = 0\,, \quad i \neq j\,, \tag{3.41}$$

and bi-orthogonal residuals

$$\boldsymbol{r}_i^* \widetilde{\boldsymbol{r}}_j = 0\,, \quad i \neq j\,, \tag{3.42}$$

while the residual norm is not minimized. This method suffers from severe numerical instabilities, and it is usually stabilized by adding one iteration of GMRES after each iteration, which is referred to as biconjugate gradient stabilized (BiCGSTAB) method. The quasi-minimal residual (QMR) method [58] generates the same type of bi-conjugate and bi-orthogonal sequences than BiCG, but it additionally solves a least square problem in order to build residuals with a minimal norm. Differently from GMRES, the minimization of the residuals is carried out within a norm induced by a tridiagonal matrix that simplifies the minimization problem and that avoids the need to store all the previous search directions. However, this modified norm is different at each iteration, so that the full-orthogonalization between the residuals is not ensured. The transpose-free quasi-minimal residual (TFQMR) method [57] is a variant of QMR that avoids computing products with the conjugate transpose of $\mathbf{A}$.

Unfortunately, few convergence results have been established for these methods, and they are not as sharp as for the conjugate gradient (3.40). Nevertheless, it has been proven for some classes of matrices, that a small condition number, or, in some other cases, bounded eigenvalues clustered away from zero, imply rapid convergence of Krylov methods [57, 70, 99].

When the matrix $\mathbf{A}$ is ill-conditioned, the preconditioning of the system is needed to enhance the convergence of Krylov solvers [113, 99]. Left and right preconditioners $\mathbf{C}_L$ and $\mathbf{C}_R$ are found—one potentially being the identity matrix—that allow to define a preconditioned matrix $\mathbf{C}_R\mathbf{A}\mathbf{C}_L$ for which the spectrum is bounded, and for which the solution of associated linear systems with Krylov solvers should require less iterations. We then solve the preconditioned system

$$\mathbf{C}_L\mathbf{A}\mathbf{C}_R\boldsymbol{y} = \mathbf{C}_L\boldsymbol{b}\,, \tag{3.43}$$

with $\boldsymbol{x} = \mathbf{C}_R\boldsymbol{y}$, that is equivalent to the original system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ whenever $\mathbf{C}_L$ and $\mathbf{C}_R$ are full rank matrices. Notice that during the iterative process, we only carry out matrix-vector products, first with $\mathbf{C}_R$, then with $\mathbf{A}$, and finally with $\mathbf{C}_L$, so that the two matrix-matrix products in $\mathbf{C}_L\mathbf{A}\mathbf{C}_R$ are never explicitly computed. Ideally, the cost to assemble these matrices and to iteratively solve the underlying preconditioned system should lead to a faster computation than the direct inversion, or than using iterative solvers without preconditioning (when convergence still occurs). Moreover, it is not rare that $\mathbf{C}_L$ and $\mathbf{C}_R$ are the inverse of some operators, as a good preconditioner would consist in approximating the inverse of $\mathbf{A}$ by $\mathbf{C}_L\mathbf{C}_R$. Hence, the matrix-vector products associated with these two preconditioning matrices are sometimes computed via numerical solvers as well. For instance, backward/forward substitutions with incomplete (sparse) LU-factorizations or multiplication with sparse approximation of the inverse (SAI) can be used as preconditioners [113, 33]. For Laplacian operators, few iterations of multigrid methods can be used as a preconditioner for the conjugate gradient, instead of being used as a stand-alone iterative solver [103]. In the following chapters, we will present preconditioners that are more specific to the integral equations.

3.4 Fast Matrix-Vector Products

In this section, we introduce some principles related to the implementation of fast matrix-vector products, which are essential for accelerating the operations performed at each step of iterative solvers—as these methods rely on repeating matrix-vector multiplications with the system matrix. The combination of fast matrix-vector products with preconditioning techniques—that aim at reducing and controlling the number of iterations—is the backbone of fast and scalable solvers. It must be noted that any efficient solver should additionally rely on effective parallelization of the matrix-vector products, and the adaptation of these algorithms to a parallel framework must be additionally taken into account. This parallelization can be further improved via domain decomposition techniques, that consist in partitioning the simulation into smaller problems associated with subdomains of the object—that can be solved independently, in parallel—with additional coupling problems to be solved in order to impose continuity at their interfaces [110, 151, 80].

In the context of surface integral equations, most of the fast matrix-vector products algorithms are based on the decomposition of the matrix system into near-field and far-field interaction blocks; which are blocks of matrices that contain interactions between groups of functions that are near or far from each other, respectively [20]. No quantitative details is given here about that definition, and we refer to [20, 12] for more details.

With a proper ordering of the basis functions, the associated blocks of near-field interactions are mostly located in the diagonal of the system matrix. Their evaluations are done in a standard way, with specialized quadrature rules for singular and near-singular integrals, as described in Section 2.4. These blocks can be stored separately from the far-interaction blocks. If the object of the simulation is large enough with respect to the wavelength, only a small part of the matrix entries correspond to near-field interactions, so that the matrix containing these associated blocks is sparse (with $\mathcal{O}(N)$ non-zero entries, where N is the size of the system).

For far-field interactions, efficient computation can be achieved by exploiting the separability of the Green's kernel for two sufficiently distant points $\boldsymbol{r}$ and $\boldsymbol{r}'$ [20, 54]:

$$G(\boldsymbol{r},\boldsymbol{r}') \approx \sum_{i=1}^{r_k} f_i(\boldsymbol{r}) g_i(\boldsymbol{r}')\,, \tag{3.44}$$

where r_k is the approximating (truncated) rank of the Green's kernel. This separation allows to compute integrals in a factorized form, if the supports of two basis functions ψ and φ are far enough from each other:

$$\int_{\Gamma\times\Gamma} G(\boldsymbol{r},\boldsymbol{r}')\psi(\boldsymbol{r})\varphi(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\mathrm{d}S(\boldsymbol{r}) \approx \sum_{i=1}^{r_k} \int_{\Gamma} f_i(\boldsymbol{r})\psi(\boldsymbol{r})\,\mathrm{d}S(\boldsymbol{r}) \int_{\Gamma} g_i(\boldsymbol{r}')\varphi(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,. \tag{3.45}$$

Based on this principle, the fast multipole method (FMM) [54] is a powerful algorithm widely used for accelerating matrix-vector products. The core idea of the FMM is to approximate interactions between two distant clusters of basis functions using multipole expansions of the Green's kernel. This process aggregates the interactions at the source cluster, translates them to the target cluster, and disaggregates them for evaluation at the finer level, thereby replacing dense "basis-to-basis" computations with computationally cheaper "group-to-group" interactions [125]. This structured approach dramatically reduces the number of direct evaluations, making the computation much more efficient.

The FMM's efficiency can be further enhanced using a hierarchical scheme, commonly implemented with recursive octree partitioning of the domain. This extension, known as the multilevel fast multipole algorithm (MLFMA)[126], accelerates aggregation and disaggregation via fast interpolation and anterpolation techniques. For example, the Fast Fourier Transform (FFT) can be leveraged to efficiently evaluate these operations [132]. MLFMA achieves $O(N \log N)$ complexity for both the storage of matrix elements and the computation of matrix-vector products, making it highly suitable for large-scale problems.

An alternative yet complementary approach leverages algebraic matrix compression. This method approximates the far-interaction blocks $\mathbf{A}$ of the system matrix in low-rank form: $\mathbf{A} \approx \mathbf{U}\mathbf{V}^*$, where $\mathbf{U}$ and $\mathbf{V}$ are matrices with a much smaller number of columns than $\mathbf{A}$ [20]. The truncated SVD is the optimal method for constructing such approximations, as it minimizes the approximation error for a given rank in the spectral norm (see Proposition 3.6). However, computing the SVD is computationally expensive and requires full knowledge of all matrix entries.

To circumvent this bottleneck, the adaptive cross approximation (ACA) [12] can be employed. The ACA builds a low-rank approximation iteratively by selecting pivotal rows and columns of the matrix block, which are computed on the fly, and that stops before assembling the whole matrix. This approach avoids evaluating the full matrix and provides a practical trade-off between computational cost and accuracy. The approximation process uses a stopping criterion to control the accuracy by estimating the residual error without explicitly constructing the full matrix block. Similar to the FMM, hierarchical clustering via recursive octree partitioning can be applied to the ACA. The resulting multilevel ACA (MLACA) achieves $O(N \log N)$ complexity for both storage and matrix-vector product computations [150].

Finally, these two algorithms can be combined together to provide a more robust algorithm. More precisely, the FMM is more adapted to high-frequency while the low-rank block compressions are more effective for low-frequency and more distant interactions.

3.5 Discussion and Objectives of the Next Chapters

In summary, iterative solvers present several advantages over direct solvers. While direct solvers provide solutions in a finite number of operations, they have a computational complexity that grows prohibitively large as the size of the system increases, with $O(N^3)$ complexity for the computational cost and $O(N^2)$ for the storage cost. In contrast, iterative solvers combining fast matrix-vector products and preconditioning techniques, offer a promising alternative with the potential for quasi-linear ($O(N \log N)$) complexity, for both storage and computational costs.

The effectiveness of iterative solvers, however, is challenged by the need for efficient and robust preconditioners within a generalized framework, in particular in the context of arbitrary frequency, geometry, materials, and discretization density.

More particularly, the EFIE faces limitations in low-frequency and dense-refinement scenarios [2], which can lead to convergence breakdowns of iterative solvers and loss of accuracy of the solution. The PMCHWT formulation inherits these breakdowns too [45]. While spurious resonances do not significantly affect the scattered field produced by the EFIE solution [41, 9], they also contribute to an unbounded spectrum in the neighborhood of certain resonant frequencies [2]. The CFIE offers a potential remedy to this issue, but it also suffers from low-frequency and dense-refinement breakdowns. Also, the MFIE and Calderón approaches for the EFIE, CFIE and PMCHWT formulations encounter challenges due to the presence of topological cycles in the geometry, causing some eigenvalues associated with these cycles to

approach zero [2]. There also exist high-contrast [66] and high-conductivity [64] breakdowns of the PMCHWT formulation.

In the next chapters we will introduce preconditioning formulations allowing to cure the low-frequency and dense-refinement breakdowns of the EFIE. In particular, we will focus on the development of new techniques to precondition problems involving junctions and linear systems arising from high-order discretizations, for which only few strategies have been investigated in the literature as of now. These constitute important steps in developing the same preconditioning framework for the other aforementioned formulations, as most of their stabilization related techniques rely on the schemes developed for the EFIE [2].

Chapter 4

Quasi-Helmholtz Projectors for Open Boundaries with Junctions

Preliminary results of this work were presented at the 2022 IEEE International Symposium on Antennas and Propagation [24]. A revised version of this chapter has been published as a communication letter in IEEE Transactions on Antennas and Propagation under creative commons license [23].

4.1 Introduction

Time-harmonic electromagnetic scattering and radiation problems for perfectly electrically conducting (PEC) objects are often solved using the electric field integral equation (EFIE) when spurious resonances cannot arise. Unfortunately, as the modelling frequency decreases, this formulation is subject to increasingly severe ill-conditioning and yields increasingly less accurate results because of loss of significance in floating point operations [2]. A classical approach to fix the low-frequency breakdown is to leverage quasi-Helmholtz decompositions [2] to treat independently the solenoidal and non-solenoidal contributions of the elements of the discrete system. This can be achieved, for instance, using Loop-Star decompositions [2, 138, 87, 8]. However, this approach worsens the conditioning of the system for dense meshes and requires the identification of the global cycles in the structure, which is often a burdensome operation [2]. This burden can significantly worsen in the presence of open boundaries with junctions since they can give rise to a significant number of global cycles associated with apertures [142, 87].

The quasi-Helmholtz projectors [8] exploits orthogonal projections over the Loop and Star subspaces to decompose and rescale the problematic contributions of the system matrices, solution, and right-hand side. This technique does not require an explicit identification of the global cycles and does not worsen the dense discretization conditioning of the preconditioned equations but its generalization to the presence of junctions had not been investigated yet.

The contribution exposed in the chapter is twofold: (i) the usage of quasi-Helmholtz projector strategies for structure containing junctions is investigated, addressing the questions of completeness and orthogonality of the underlying decompositions and (ii) the computational cost of the approach is analyzed both by studying the effect of the presence of junctions on the inversion of the graph Laplacian that appears in the definition of the projectors and by verifying the overall numerical performance of the scheme. In particular, we study numerically the performance of the aggregation-based algebraic multigrid (AGMG) [104, 109], which is one

possible implementation of multigrid methods, well-known for providing an efficient preconditioner for graph Laplacian systems [19, 84]. Our observations show that the effectiveness of the quasi-Helmholtz projectors approach is not jeopardized in the presence of junctions and that it can be successfully used with realistic structures of industrial relevance.

4.2 RWG connectivity and dual graph at low-order

In this Section, we investigate the relationship between the lowest order Raviart-Thomas basis functions (the RWG functions) and the graph connectivity of the element of the mesh. First, we notice that the RWG basis functions (2.76) can be rewritten as

$$\boldsymbol{\psi}_n(\boldsymbol{r}) = \begin{cases} \pm\dfrac{\boldsymbol{r}-\boldsymbol{r}_n^{\pm}}{\mathcal{J}^{\pm}} & \text{if } \boldsymbol{r} \in c_n^{\pm} \\ 0 & \text{otherwise} \end{cases}, \tag{4.1}$$

where the notation for the geometrical elements describing a pair of adjacent triangular cells $c_n^{\pm}$ is given in Fig. 4.1 and $\mathcal{J}^{\pm}$ denotes their respective elementary surface areas. With these notations, the RWG functions can be seen as the link between two adjacent cells of the mesh. Hence, we can compare these basis functions to the edges of an associated graph, for which the nodes are the cells of the mesh.

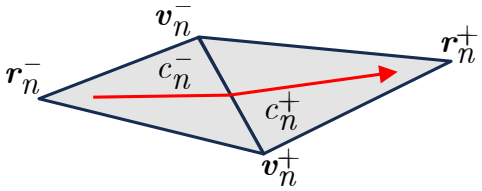


Figure 4.1: Notations of the vertices and cells used to define the RWG functions.

The EFIE can be used for both open and closed manifolds that can contain junctions, i.e. structures for which the associated mesh contains edges shared by more than two cells (see for instance Fig. 4.2, Fig. 4.3 and [34]). A popular approach to handle structures with junctions is to associate to a mesh edge e which connects together n_e cells, $n_e - 1$ independent RWG functions [34, 146]. These functions are defined by considering $n_e - 1$ independent couples of cells sharing the edge [146]. Different choices of configurations that define RWG functions spanning the same space are possible, but they lead to different connectivity properties between cells. Here and in the rest of this contribution we consider that two adjacent cells c_m, c_n are connected if there exists an RWG function $\boldsymbol{\psi}_i$ such that $c_i^+ = c_m$ and $c_i^- = c_n$ or $c_i^+ = c_n$ and $c_i^- = c_m$. In particular, the degrees of connectivity—the number of cells connected to a given cell via RWG functions—and the graph distance—the smallest number of RWG functions needed to link a given pair of cells—depend on this choice. We provide in Fig. 4.2 a visual representation of two of the possible configurations along with their consequences on the connectivity of the underlying graph. The nodes of this graph, that we will refer to as the dual graph of the discretization, correspond to the cells c_m of the mesh and its edges link pairs of nodes (c_m, c_n) if there exists an RWG function $\boldsymbol{\psi}_i$ such that $c_i^+ = c_m$ and $c_i^- = c_n$. With this formalism, the total number of basis functions associated with a mesh containing E edges is $N = \sum_{e=1}^{E}(n_e - 1)$.

We recap, here, that after replacing (2.91) into (1.112) and by testing the equation with rotated RWG functions $\{\boldsymbol{n} \times \boldsymbol{\psi}_m\}_{m=1}^{N}$, we obtain the $N \times N$ linear system

$$\mathbf{T}\boldsymbol{j} = \boldsymbol{e}\,, \tag{4.2}$$

where $\mathbf{T} = -jk\,\mathbf{T}_s + \frac{1}{jk}\,\mathbf{T}_h$ with $[\mathbf{T}_s]_{mn} = \langle \boldsymbol{n} \times \boldsymbol{\psi}_m, \mathcal{T}_s \boldsymbol{\psi}_n \rangle$, $[\mathbf{T}_h]_{mn} = \langle \boldsymbol{n} \times \boldsymbol{\psi}_m, \mathcal{T}_h \boldsymbol{\psi}_n \rangle$, and $[\boldsymbol{e}]_m = -\langle \boldsymbol{n} \times \boldsymbol{\psi}_m, \boldsymbol{n} \times \frac{\boldsymbol{E}^{\text{inc}}}{\eta} \rangle$, and where $\langle \boldsymbol{f}, \boldsymbol{g} \rangle := \int_\Gamma \boldsymbol{f}(\boldsymbol{r}) \cdot \boldsymbol{g}(\boldsymbol{r})\, \mathrm{d}S(\boldsymbol{r})$.

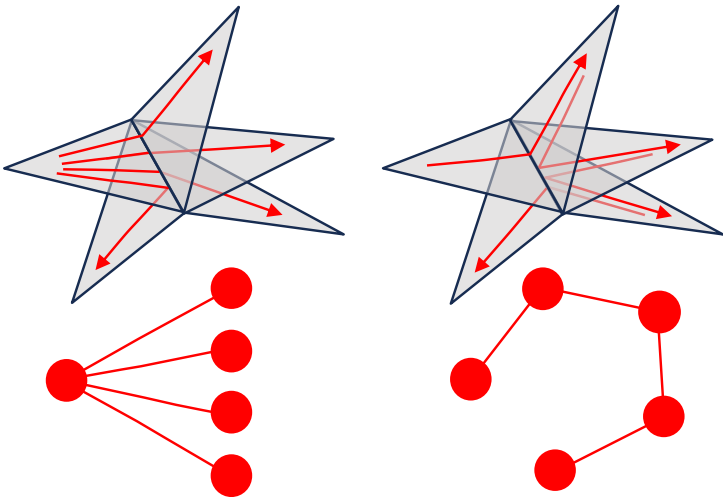


Figure 4.2: Two conventional ways to define an independent set of RWG functions on a junction-edge and their associated graph representation. It illustrates two extreme configuration scenarios: on the left, the one giving rise to the maximal degree of connectivity, with one cell connected to all the others; on the right, the one giving rise to the minimal degree of connectivity (but maximal graph distance), with each cell connected to at most two other cells through this edge.

4.3 Loop and Star Decomposition with Junctions

The EFIE (2.93) is known to suffer from severe ill-conditioning and dramatic losses of accuracy at low-frequency (see [2] and references therein). These problems can be tackled starting with a Loop-Star decomposition of the solution vector, i.e. by decomposing the solution vector $\boldsymbol{j}$ in (2.91) as

$$\boldsymbol{j} = \boldsymbol{\Sigma}\boldsymbol{s} + \boldsymbol{\Lambda}\boldsymbol{l} + \mathbf{H}\boldsymbol{h}\,, \tag{4.3}$$

where $\boldsymbol{\Sigma}$, $\boldsymbol{\Lambda}$, and $\mathbf{H}$ are the Star-to-RWG, Loop-to-RWG, and quasi-Harmonic-functions-to-RWG transformation matrices, respectively, and $\boldsymbol{s}$, $\boldsymbol{l}$ and $\boldsymbol{h}$ are the vectors containing the associated expansion coefficients. In the presence of junctions, the matrix $\boldsymbol{\Sigma}$ is defined as in the junction-less case [2] as

$$[\boldsymbol{\Sigma}]_{mn} = \begin{cases} \pm 1 & \text{if } c_m^{\pm} \text{ is cell } n \\ 0 & \text{otherwise} \end{cases}\,, \tag{4.4}$$

which is also referred to as the charge matrix, in the sense that it corresponds to the discretization of the divergence operator [2] and its rank equals the number of independent charges [87]. With this definition, $\boldsymbol{\Sigma}$ contains the information on the connectivity between cells. This connectivity depends on the choice of the RWG configuration along the junctions (see the discussion above and Fig. 4.2) and impacts the spectral properties of $\boldsymbol{\Sigma}$. These properties and their influence on the performance of the proposed method are further investigated in Section 4.5.

Regarding the Loops, we propose the following definition based on the dual graph faces (see

Fig. 4.3), for compactness

$$[\mathbf{\Lambda}]_{mn} = \begin{cases} \pm 1 & \text{if } \boldsymbol{v}_m^{\pm} \text{ belongs to the dual-face } n \\ 0 & \text{otherwise} \end{cases} . \tag{4.5}$$

The construction of the quasi-Harmonic-functions-to-RWG transformation matrix $\mathbf{H}$ requires the identification of the global cycles associated with the genus of the geometry (see [2] and references therein).

The use of the Loop-Star basis is known to be problematic under certain circumstances because, on the one hand, it worsens the conditioning of the original EFIE [2] and, on the other hand, it requires the detection of topological Loops in $\mathbf{H}$ that can jeopardize the overall linear complexity of the scheme. Moreover, even the presence of open structure cycles (that in the formalism of this chapter are associated with the matrix $\mathbf{\Lambda}$) can be problematic and it can require an expensive cycle detection procedure. The reader should also note that the recovery of even standard Loops (around the vertices of the junction edges) can be problematic when several junctions are present.

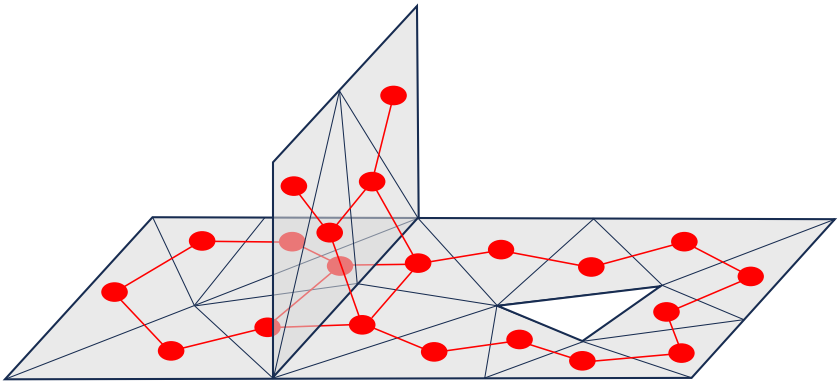


Figure 4.3: Dual graph (in red) associated with a mesh (in gray) which contains open boundaries, a junction, and an aperture. This graph highlights the existence of two local Loops around the vertex belonging to the junction and one global Loop around the aperture.

4.4 Quasi-Helmholtz Projectors

4.4.1 Definition

The above drawbacks can be tackled by leveraging on the quasi-Helmholtz projectors. Similarly to what is done in the junction-less case, we define the projector to the non-solenoidal subspace as

$$\mathbf{P}^{\Sigma} = \mathbf{\Sigma}\left(\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}\right)^{+}\mathbf{\Sigma}^{\mathsf{T}} , \tag{4.6}$$

where "+" denotes the Moore-Penrose pseudoinverse (3.12), and the projector to the solenoidal subspace as

$$\mathbf{P}^{\Lambda H} = \mathbf{I} - \mathbf{P}^{\Sigma} . \tag{4.7}$$

Finally, we use the quasi-Helmholtz decomposition provided by the projectors to compensate the frequency scaling of the solenoidal and non-solenoidal parts of the EFIE system when the frequency is decreasing [2, 87]. To this end, we define the preconditioning matrix

$$\mathbf{P} = \frac{1}{\sqrt{k}}\mathbf{P}^{\Lambda H} + j\sqrt{k}\mathbf{P}^{\Sigma} \tag{4.8}$$

to form the preconditioned EFIE

$$\mathbf{P}\left(-jk\,\mathbf{T}_s + \frac{1}{jk}\,\mathbf{T}_h\right)\mathbf{P}\boldsymbol{y} = \mathbf{P}\boldsymbol{e}\,, \tag{4.9}$$

that exhibits a bounded condition number at low-frequency [2], and where

$$\mathbf{P}\boldsymbol{y} = \boldsymbol{j}\,. \tag{4.10}$$

Additionally, to avoid loss of significant digits at low-frequency, (4.9) needs to be expanded in order to impose $\mathbf{P}^{\Lambda H}\mathbf{T}_h = \mathbf{T}_h\mathbf{P}^{\Lambda H} = \mathbf{0}$ and the Loop contributions of right-hand side and post-processing field solutions need to be computed using specific techniques [2]. More precisely, we write

$$\mathbf{PTP} = -jC\,\mathbf{P}^{\Lambda H}\mathbf{T}_s\mathbf{P}^{\Lambda H} - j/C\,\mathbf{T}_h + k\,\mathbf{P}^{\Lambda H}\mathbf{T}_s\mathbf{P}^{\Sigma} + k\,\mathbf{P}^{\Sigma}\mathbf{T}_s\mathbf{P}^{\Lambda H} + jk^2/C\,\mathbf{P}^{\Sigma}\mathbf{T}_s\mathbf{P}^{\Sigma}. \tag{4.11}$$

and the right-hand side is expressed as

$$\mathbf{P}\boldsymbol{e} = j\sqrt{k/C}\,\mathbf{P}^{\Sigma}\boldsymbol{e} + \sqrt{C/k}\,\mathbf{P}^{\Lambda H}\boldsymbol{e}_{\text{sub}}, \tag{4.12}$$

where $\boldsymbol{e}_{\text{sub}}$ is the right-hand side subtracted from its divergence-free (static) component [2]. Finally, the post-processing computation of the electric field requires a similar treatment, by using separately the solenoidal and non-solenoidal contributions of the solution

$$\boldsymbol{j}_{\text{sol}} = \sqrt{C/k}\,\mathbf{P}^{\Lambda H}\boldsymbol{y} \quad \text{and} \quad \boldsymbol{j}_{\text{nsol}} = j\sqrt{k/C}\,\mathbf{P}^{\Sigma}\boldsymbol{y} \tag{4.13}$$

and subtracting the static part of the Green's kernel when integrating with the solenoidal contribution [2].

4.4.2 Study of orthogonality and completeness

To ensure the proper applicability of the quasi-Helmholtz projectors to arbitrary problems with junctions, orthogonality between the Loop and Star bases and the completeness of the proposed quasi-Helmholtz decomposition must be verified.

The orthogonality between the Loop / quasi-Harmonic-functions and the Star transformation matrices follows from the fact that, across a triangle, a cycle has to enter through one edge and to exit through another one, whereas a Star associated with a given cell exits across all the branches made by the RWG functions defined on that cell, even in the presence of junctions. It results in a null dot product between the columns of $\mathbf{\Lambda}$ / $\mathbf{H}$ and $\mathbf{\Sigma}$, which finally gives $\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Lambda} = \mathbf{0}$ and $\mathbf{\Sigma}^{\mathsf{T}}\mathbf{H} = \mathbf{0}$.

Next, we show the completeness of the quasi-Helmholtz decomposition, i.e., that the Loops, quasi-Harmonic functions and Stars span the entire RWG space. We establish our result for connected geometries. The generalization to the non-connected case is straightforward, by individually applying the following result to each connected subdomain. We first describe the elements of the dual graph: S is the number of nodes that correspond to the cells of the mesh, F is the number of faces that correspond to the local Loops and apertures, including the exterior boundary of open objects, and H is the number of handles of the structure, present in both the mesh and its dual graph, which corresponds to $2H$ quasi-harmonic functions; N denotes the number of RWG functions and so the number of edges of the dual graph. We can then relate these quantities using the Euler-Poincaré formula [51] applied to the dual graph

$$N = (S-1) + (F-1) + 2H. \tag{4.14}$$

First, the number of independent cycles (Loops and quasi-Harmonic functions) is known to be $(F-1)+2H$ [71] where the linear dependency between the non-Harmonic cycles is taken into account. Next, the rank of $\mathbf{\Sigma}$ is $S-1$ since it has a nullspace associated with the constant vector. This is related to the fact that a div-conforming current has a neutral charge along the whole structure [87], as a consequence of the divergence theorem and by definition of the div-conforming functions (see Section 2.2.3)

$$\int_{\Gamma} \mathrm{div}_{\Gamma}\psi_i(\boldsymbol{r})\,\mathrm{d}S(\boldsymbol{r}) = -\sum_{K}\int_{\partial K} \mathrm{div}_{\Gamma}\psi_i(\boldsymbol{r})\cdot\boldsymbol{\nu}_K(\boldsymbol{r})\,\mathrm{d}l(\boldsymbol{r}) = 0\,, \tag{4.15}$$

where $\boldsymbol{\nu}_K$ is the outgoing unit normal to the border of triangle K. Hence, the rank of the subspace orthogonal to the range of $\mathbf{\Sigma}$ is $N-(S-1) = F-1+2H$, which corresponds to the total number of independent cycles. Because $\mathbf{\Lambda}$ and $\mathbf{H}$ are both orthogonal to $\mathbf{\Sigma}$, it results the completeness between the Loop and Star bases and that $\mathbf{P}^{\Lambda H}$ is the projector onto the solenoidal subspace (local Loops and global cycles).

4.5 Spectral Analysis of Graph Laplacians with Junctions

The computation of the Star projector (4.6) requires solving systems arising from the pseudoinversion of $\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}$. Thus, the spectral properties of this matrix directly impact the computational efficiency of this procedure. In particular, the definition (4.4) of $\mathbf{\Sigma}$ is the one of the adjacency matrix of the dual graph described above, which in turn means that $\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}$ is the graph Laplacian associated with this graph [49]. The spectral properties of this class of matrices have been thoroughly studied [95, 98, 49].

First, the graph Laplacian $\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}$ has a nullspace associated with the constant functions restricted to each connected subdomain of Γ. Because the pseudoinverse of $\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}$ in (4.6) will only be applied to vectors belonging to the range of $\mathbf{\Sigma}^{\mathsf{T}}$, we will focus on the spectral properties of the graph Laplacian restricted to this subspace. In particular, we will consider the pseudo condition number (3.22) to that subspace that we define as the ratio between the largest singular value $\upsilon_{\max}$ and the smallest non-zero singular value $\upsilon_{\min}$ of $\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}$, i.e.

$$\mathrm{cond}^{+}\left(\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}\right) = \frac{\upsilon_{\max}}{\upsilon_{\min}}. \tag{4.16}$$

In practice, the convergence of several iterative methods will be faster if this pseudo condition number is close to one, provided that the initial guess belongs to the range of $\mathbf{\Sigma}^{\mathsf{T}}$ (see for instance the convergence results on the conjugate gradient method in solving singular symmetric systems [86]). The maximal singular value can be bounded as [95]

$$\delta \leq \upsilon_{\max} \leq 2\delta\,, \tag{4.17}$$

where δ denotes the maximal degree of connectivity of the graph associated with $\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}$. Regarding the minimal non-zero singular value, we have the upper bound [49]

$$\upsilon_{\min} \leq 1 - \frac{\sqrt{\delta-1}}{\delta}\left(1-\frac{2}{D}\right)+\frac{2}{D}\,, \tag{4.18}$$

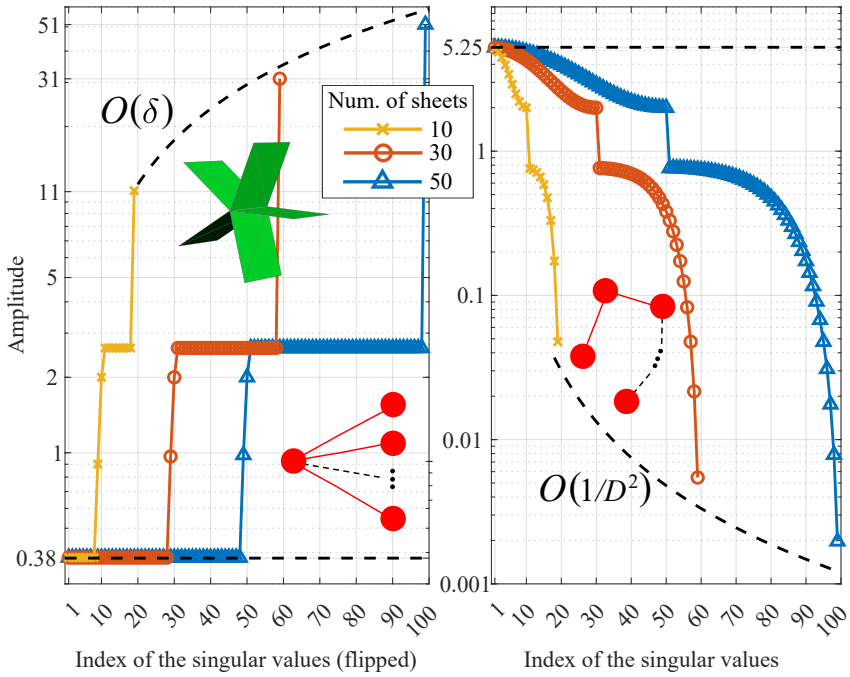


Figure 4.4: Spectrum of the graph Laplacian $\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}$ for structures with multiple sheets, each discretized with two triangles. The two connectivity configurations of Fig. 4.2 are tested and yield different spectra for same structures.

where D denotes the diameter of the graph associated with $\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}$. Let us first consider the case of δ growing. Combining (4.17) and (4.18) we get

$$\mathrm{cond}^+\left(\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}\right) \geq \frac{\delta}{1-\frac{\sqrt{\delta-1}}{\delta}\left(1-\frac{2}{D}\right)+\frac{2}{D}}. \tag{4.19}$$

This inequality shows that the pseudo condition number increases asymptotically with the maximal degree of connectivity δ. When δ is bounded and D grows, however, (4.19) does not provide a useful bound, but the following [98]

$$2\left\lceil\sqrt{\mathrm{cond}^+\left(\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}\right)}\sqrt{\frac{\alpha^2-1}{4\alpha}}+1\right\rceil\left\lceil\log_\alpha\frac{S}{2}\right\rceil \geq D \tag{4.20}$$

can be used instead, where α is any arbitrary real number greater than 1, and S is the dimension of $\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}$. This shows that the pseudo condition number increases asymptotically with respect to the graph diameter D, provided that $\log_\alpha \frac{S}{2}$ increases slower than D. This requirement is met whenever δ is bounded, since in that case D increases as S^d, with $d = 1$ for elongated (quasi-lineic) surfaces, $d = 1/2$ for surfaces of balanced length and width and $d = 1/3$ for three dimensional lattice (quasi-volumic) structures. Equations (4.19) and (4.20) only imply an increase of the pseudo condition number in the asymptotic limit. However, these bounds are sharp enough that this increase has been observed numerically for small degrees of connectivity and small graph diameters, as shown in Fig. 4.4.

A direct consequence of (4.20), is that $\mathrm{cond}^+\left(\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}\right)$ increases asymptotically when refining the mesh, since D will be increasing. The second consequence from the above analysis is that the pseudo condition number $\mathrm{cond}^+\left(\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}\right)$ also increases asymptotically when inserting additional

sheets to the structure. To further clarify this, let us consider the two extreme configurations of RWG functions along the junctions, illustrated in Fig. 4.2, both implying different asymptotic behavior for the degrees of connectivity and graph diameters. In the first case (as in Fig. 4.2, left), the graph diameter is not impacted by the presence of junctions but the connectivity increases with the maximal number of sheets connected together, which results in an increase of the pseudo condition number. In the second case (as in Fig. 4.2, right), the connectivity is bounded but the graph diameter increases with the maximal number of sheets connected together, which also results in an increase of the pseudo condition number.

Because of this conditioning behavior, the computation of the quasi-Helmholtz projectors requires an efficient method to precondition and solve graph Laplacian systems. In Section 4.6, we will investigate numerically the effects of using multigrid approaches [19, 84, 104, 109] for this purpose.

4.6 Numerical Results

This section will first show the performance of algebraic multigrids in performing the graph Laplacian pseudoinverses required by quasi-Helmholtz projectors, then the overall preconditioning approach will be validated on simply and non-simply connected structures containing junctions.

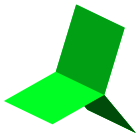
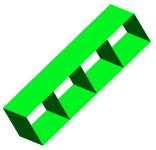
4.6.1 Pseudoinversion of the Graph Laplacian

First, we investigate the use of the aggregation-based algebraic multigrid (AGMG), described in [109, 104] and its robustness in preconditioning the graph Laplacian $\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}$ in the presence of junctions. We study the number of iterations as a function of the number of unknowns, when refining mesh of structures having a bounded number of junctions and with the RWG configuration on junctions that maximizes the connectivity (Fig. 4.2, left). Table 4.1 shows that the number of iterations of AGMG increases logarithmically with the number of unknowns for various structures while the pseudo condition number increases linearly with the number of unknowns. This results in a quasi-linear number of operations since each iteration requires sparse matrix vector products.

Next, we address the question of the AGMG performance when the number of connected sheets increases. Fig. 4.5 shows the time required by the setup and solving phases of AGMG as a function of the number of sheets, when the graph Laplacian is built using the two extremal RWG configurations (minimal and maximal connectivity). We observe that the setup time increases almost quadratically when the number of sheets and the connectivity both increase, while it remains linear when the number of sheets increases but the connectivity is bounded. In both cases, the solving time is quasi-linear with respect to the number of unknowns. Note that for a smaller number of junctions, we however observe better performance with the configuration having the maximal connectivity. The computation time for the first experiment has not been reported above because the scenario is similar to the second configuration in Fig. 4.5, and numerical observations have shown a similar behavior.

For this particular version of algebraic multigrid, our numerical observations thus point out that the connectivity plays an important role in the performance of the graph Laplacian inversion using AGMG and that the proper RWG configuration on junctions can be chosen to get a robust algorithm.

Table 4.1: Number of iterations achieved by AGMG for increasingly dense discretization with a residual error below 10^{-8}.

Structure	Unk.	Cond.	Iter.	Structure	Unk.	Cond.	Iter.
	7.2E3 6.9E4 7.0E5	1.1E4 1.1E5 1.0E6	21 22 22		2.9E3 2.8E5 2.7E6	2.5E3 2.3E5 2.3E6	18 24 22
	1.2E3 1.2E5 1.2E6	1.5E3 1.3E5 1.3E6	18 22 23		3.2E3 3.0E5 3.0E6	8.4E3 7.5E5 7.5E6	19 22 23
	9.6E3 2.4E5 2.7E6	5.1E3 1.2E5 1.2E6	19 20 24		4.0E3 4.0E5 1.0E6	7.5E3 7.1E5 1.8E6	19 22 24

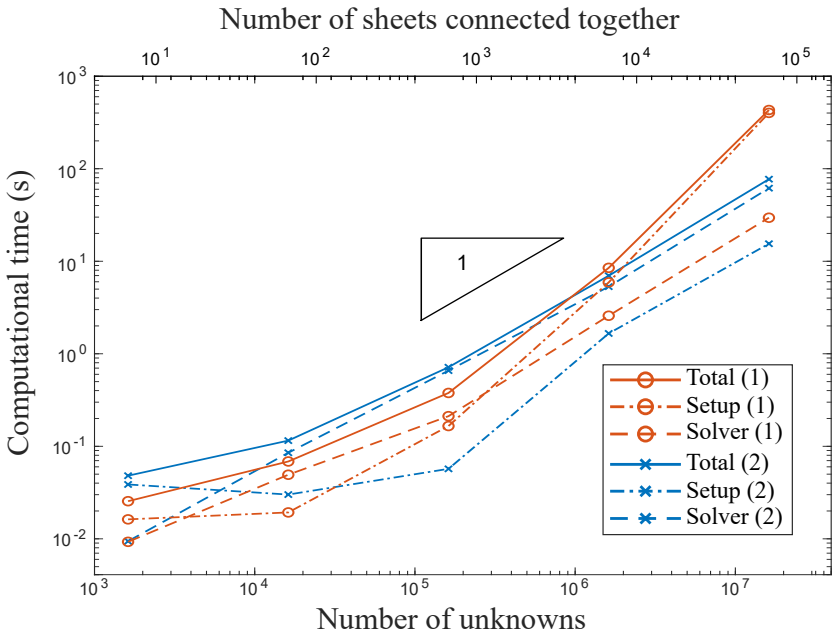


Figure 4.5: Computational time required by AGMG to solve graph-Laplacian systems for increasing number of junctions. Configuration (1) and (2) are respectively corresponding to the maximal and minimal degree of connectivity that we get with a given RWG configuration on junctions.

4.6.2 Efficiency of the Quasi-Helmholtz Preconditioning for Structures with Junctions

We now numerically test the efficiency of the proposed method with application to the computation of the radar cross section (RCS) of a jet engine illuminated by an incident plane wave $\boldsymbol{E}^{\text{inc}}(\boldsymbol{r}) = z/|z|e^{-jk\boldsymbol{r}\cdot\boldsymbol{x}/|\boldsymbol{x}|}$. The shell of the engine is a toroidal geometry that gives two quasi-Harmonic functions. On the front, the blades of the turbine are approximated by metallic sheets and are connected to the rotor by junctions. This part of the structure is separated from the shell and generates an independent set of Loops and Stars (with respect to the Euler-Poincaré formula). In particular, $\boldsymbol{\Sigma}$ has a nullspace of dimension 2 generated by the constant functions restricted to each separated subdomain. On the back, a fence made of metallic sheets connects the shell to the nozzle forming junctions at both extremities of the sheets as well as apertures between them. The nozzle itself is an open surface with an aperture at its end. The object is discretized using 9191 RWG functions that can be split into 3011 and 6180 Loops and Stars respectively. The average number of iterations of AGMG required to solve a single graph Laplacian system with a desired accuracy of 10^{-6} is 19, which is repeated at every multiplication involving the projectors in the iterative inversion of the preconditioned EFIE. The surface current density calculated at a frequency of 1 Hz using the EFIE preconditioned with the quasi-Helmholtz projectors is given in Fig. 4.6.

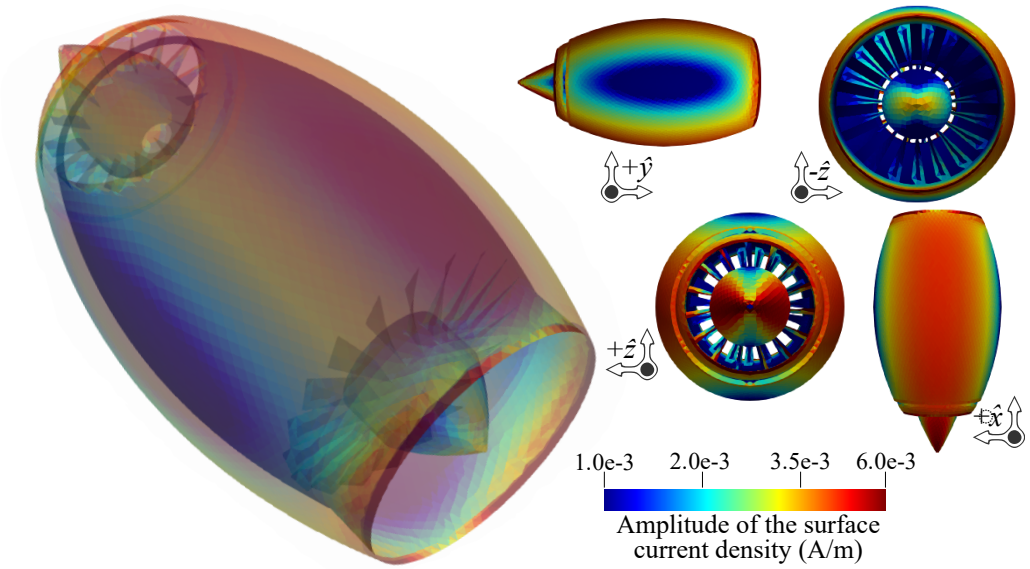


Figure 4.6: Surface density current calculated along the surface of the turbine with the quasi-Helmholtz projectors at 1 Hz.

Our work (4.9) is compared with the standard EFIE (without preconditioning) as well as the Loop-Star decomposed EFIE [2]. In Fig. 4.7, we observe that the two preconditioning techniques give rise to systems with a bounded condition number when the frequency decreases, while the condition number dramatically increases for the standard formulation until it saturates due to numerical effects. In addition, the condition numbers obtained with the quasi-Helmholtz projectors are significantly lower than the Loop-Star decomposed EFIE. Solving with the solver TFQMR at 1 Hz for a desired accuracy of 10^{-6}, the iterative procedure requires less iterations when the system is preconditioned with the quasi-Helmholtz projectors than when preconditioned using a Loop-Star decomposition, since only 419 iteration are required for the former

instead of 4416 for the latter. Under these conditions, the EFIE does not converge without preconditioning. We also note that the two preconditioning methods give comparable results at low-frequency, unlike the standard formulation (solved via direct factorization), as is shown in Fig. 4.8.

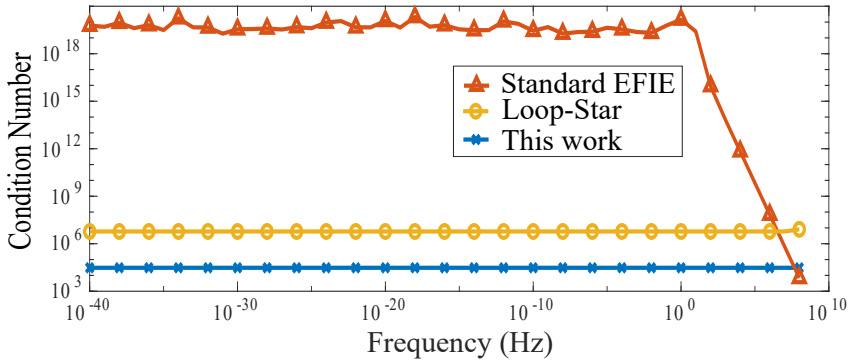


Figure 4.7: Condition number of the EFIE with and without preconditioning in function of the frequency for the model of turbine.

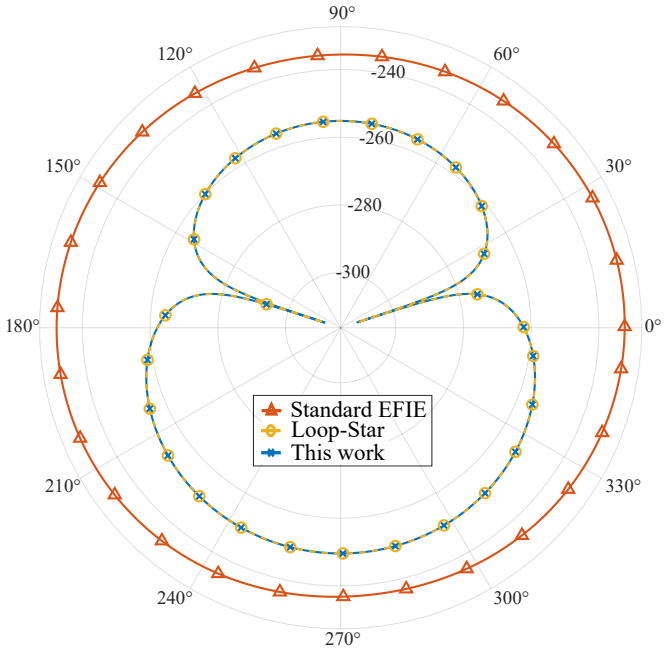


Figure 4.8: Bistatic RCS (dBsm) in the $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{z}})$ plane given the scattered far field produced by the turbine with an incident plane wave at a frequency of 1 Hz.

Finally, in Table 4.2 we report the number of iterations and computation time required to solve the EFIE with and without preconditioning, showing better performance for the proposed scheme at a frequency of 5 MHz.

Table 4.2: Number of iterations and run time of TFQMR to solve the EFIE with and without preconditioning at medium and low frequencies for the model of turbine with 9191 RWG functions.

	Standard EFIE		**Loop-Star**		**This work**	
Freq.	**Iter.**	**Time** (s)	**Iter.**	**Time** (s)	**Iter.**	**Time** (s)
5 MHz	1632	176.7	4064	690.8	393	137.2
1 Hz	*No convergence*		4416	779.0	419	142.3

4.7 Conclusion

We theoretically and numerically showed how the quasi-Helmholtz projectors method can be used to model low frequency problems containing junctions with the EFIE. Explanations concerning the generalization of the Loops and Stars for such structures were given and we showed that multigrid solvers can be used to efficiently compute the projectors, even in the presence of a large number of junctions and densely discretized open structures.

This work also opens new perspectives in preconditioning problems arising from the modelling of relatively complex and industrially relevant structures, and shows that such remedies are possible to build in that context.

Chapter 5

High-Order Quasi-Helmholtz Projectors

Preliminary results of this work were presented at the XXXVth URSI General Assembly and Scientific Symposium 2023 [21]. A revised version of this chapter has been published as a journal paper in IEEE Transactions on Antennas and Propagation under creative commons license [22].

5.1 Introduction

As introduced in the previous chapter, the stabilization of the EFIE using quasi-Helmholtz decompositions [138, 149, 87, 53] is well-known and has also been investigated in the high-order case [141, 136, 63]. It consists in a change of basis that allows to reorganize the system into solenoidal and non-solenoidal contributions and to rescale them appropriately to cure the problematic behavior of the EFIE at low-frequency. However, the construction of the solenoidal basis functions can be burdensome because it requires the identification of the global cycles; a task that might be challenging when modelling complex geometries (see Chapter 4).

More recently, the method of the quasi-Helmholtz projectors has been developed for the stabilization of the EFIE discretized with RWG functions [2]. The method generates orthogonal projectors over the solenoidal and non-solenoidal subspaces from the computation of the Star matrix, without having to explicitly identify the cycles. They are subsequently used to separate and to rescale the different contributions of the EFIE, without further degrading the condition number when increasing the mesh-density [2].

This work proposes for the first time the high-order counterpart of the quasi-Helmholtz projectors for the stabilization of the EFIE when using high-order basis functions. The contribution will first present a general definition of the Star basis functions in high-order that encompasses several choices in terms of basis elements and testing in the charge space. The contribution will then show that the consequential definition of the quasi-Helmholtz projectors is not dependent on any of these specific choices. Finally the new projectors will be used to regularize the electric field integral equation discretized with high-order elements.

5.2 Definition

In this section we will define high-order quasi-Helmholtz projectors to address the above described low-frequency limitations of the EFIE in the high-order case. From the zeroth-order case [2] we learn that primal projectors can be obtained from a properly chosen *Star matrix*.

The problem is that a unique definition of a Star matrix in high-order can be challenging [141]. In this contribution, we will propose one approach to define a Star matrix that, indeed, does not lead to a unique definition. We will equally show in this chapter, however, that the corresponding Star (non-solenoidal) projector will be invariant, regardless of the non-uniqueness of the Star definition we will adopt.

Given the coefficients $\boldsymbol{j}$ of a function expressed as a linear combination of GWP functions, one could think of building a high-order Loop-Star decomposition in the form

$$\boldsymbol{j} = \boldsymbol{\Sigma}\boldsymbol{s} + \boldsymbol{\Lambda}\boldsymbol{l} + \mathbf{H}\boldsymbol{h}. \tag{5.1}$$

Similarly to the zeroth-order case, $\boldsymbol{\Lambda}$ and $\mathbf{H}$ express the high-order local-Loops-to-GWP and global-cycles-to-GWP change of bases, respectively [2]. As pertains to $\boldsymbol{\Sigma}$, we will refer to this matrix as the high-order *Star* matrix, with an abuse of terminology steming from the zeroth-order case. It is a matrix orthogonal to $\boldsymbol{\Lambda}$ and $\mathbf{H}$ that complements the space of coefficients so that (5.1) stands for every $\boldsymbol{j} \in \mathbb{C}^{N_p}$, where N_p is the dimension of the Raviart-Thomas space, as defined in Chapter 2.

The columns of the matrix $\boldsymbol{\Lambda}$ can be found as the expansion coefficients of the set of solenoidal functions $\{\mathbf{curl}_\Gamma \lambda_i^{(p)}\}_{i=1}^{L_p}$ expressed in the GWP basis, where $\{\lambda_i^{(p)}\}_{i=1}^{L_p}$ is a set of basis functions for the continuous space $\mathbb{CG}_p$, for instance the interpolatory functions defined in Section 2.2.3. As the GWP functions are interpolatory, it is possible to find these coefficients by evaluating the solenoidal functions at the interpolatory points associated with the GWP functions. In a more general case, this matrix can be found as the solution of the system

$$\mathbf{G}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}} \boldsymbol{\Lambda} = \mathbf{G}_{\boldsymbol{\psi}^{(p)},\mathbf{curl}_\Gamma \lambda^{(p)}}\,, \tag{5.2}$$

that can be inverted by block as the solenoidal functions $\{\mathbf{curl}_\Gamma \lambda_i^{(p)}\}_{i=1}^{L_p}$ are generated by a local combination of GWP functions (the GWP functions that share the same support than the solenoidal functions). Of course, we can also directly compute

$$\boldsymbol{\Lambda} = \mathbf{G}^{-1}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}} \mathbf{G}_{\boldsymbol{\psi}^{(p)},\mathbf{curl}_\Gamma \lambda^{(p)}}\,, \tag{5.3}$$

if we do not care about computational performance and are looking for an explicit expression of that matrix. The columns of $\mathbf{H}$ are less straightforward to define. One possible choice is to express the global cycles in terms of an RWG combination, as this corresponds to the identification of cycles in the dual graph (see Section 4.3) and to express the RWG functions in the GWP basis of order p (via interpolation or by solving a system with Gram matrices as in (5.2)).

Nevertheless, in the following, we will not need an explicit definition of these two matrices. We will only be using the fact that both $\boldsymbol{\Lambda}$ and $\mathbf{H}$ contain coefficient representations of solenoidal functions, that are orthogonal to the coefficients of Star functions, given by the columns of $\boldsymbol{\Sigma}$.

Consider now the injective linear function $\mathcal{L}$

$$\mathcal{L}: \begin{array}{ccc} \mathbb{DG}_p & \longrightarrow & \mathbb{R}^{S_p} \\ q & \longmapsto & \mathcal{L}q \end{array} \tag{5.4}$$

with $\mathcal{L}_m q := [\mathcal{L}q]_m$. It should be noted that injectivity and linearity imply

$$\mathcal{L}q = 0 \iff q(\boldsymbol{r}) = 0 \quad \forall \boldsymbol{r} \in \Gamma. \tag{5.5}$$

Now we can propose the following general definition for $\boldsymbol{\Sigma}$

$$\left[\boldsymbol{\Sigma}\right]_{m,n} = \mathcal{L}_n \mathrm{div}_\Gamma \boldsymbol{\psi}_m^{(p)}. \tag{5.6}$$

By the construction of $\boldsymbol{\Sigma}$ and because $\boldsymbol{\Lambda}$ and $\mathbf{H}$ describe solenoidal functions we get

$$\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Lambda} = \mathbf{0} \tag{5.7}$$

and

$$\boldsymbol{\Sigma}^{\mathsf{T}}\mathbf{H} = \mathbf{0}. \tag{5.8}$$

The above conditions result in that all coefficients of solenoidal functions are in the null-space of $\boldsymbol{\Sigma}^{\mathsf{T}}$. Also the converse statement, all elements of the null space of $\boldsymbol{\Sigma}^{\mathsf{T}}$ are coefficients of solenoidal functions, is true following from (5.5) and (5.6). Since the solenoidal subspace of the GWP functions has a dimension $N_p - S_p + N_{\text{bodies}}$ [141], the null-space of $\boldsymbol{\Sigma}^{\mathsf{T}}$ has dimension $N_p - S_p + N_{\text{bodies}}$, where N_{bodies} is the number of disconnected bodies in Γ. As a consequence, the dimension of the range of $\boldsymbol{\Sigma}^{\mathsf{T}}$, which is also the dimension of the range of $\boldsymbol{\Sigma}$, equals to $N_p - (N_p - S_p + N_{\text{bodies}}) = S_p - N_{\text{bodies}}$. This fact proves that $\left[\boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \mathbf{H}\right]$, a rectangular matrix, has a rank equal to the number of basis functions N_p in (2.53). This shows that our definition of the high-order Stars produces a valid complement of the solenoidal subspaces as the columns of $\left[\boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \mathbf{H}\right]$ are coefficient representations of functions that generate the entire GWP space.

We are now ready to define the quasi-Helmoltz projectors. In particular, mimicking the zeroth-order case, define

$$\mathbf{P}^{\Sigma} = \boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}\right)^{+}\boldsymbol{\Sigma}^{\mathsf{T}} \tag{5.9}$$

and

$$\mathbf{P}^{\Lambda H} = \mathbf{I} - \mathbf{P}^{\Sigma}, \tag{5.10}$$

in which "+" is used to denote the Moore-Penrose pseudoinverse [14]. It should be noted that, because of the orthogonality of the projectors, i.e.

$$\mathbf{P}^{\Sigma}\mathbf{P}^{\Lambda H} = \mathbf{P}^{\Sigma}(\mathbf{I} - \mathbf{P}^{\Sigma}) = \mathbf{P}^{\Lambda H}\mathbf{P}^{\Sigma} = \mathbf{P}^{\Sigma} - \mathbf{P}^{\Sigma} = \mathbf{0}\,, \tag{5.11}$$

and because of completeness of the decomposition, $\mathbf{P}^{\Lambda H}$ is the projector onto the solenoidal (local and global Loops) subspace.

5.3 Uniqueness of the Projectors

From the generality of our definition of $\boldsymbol{\Sigma}$ in (5.6) one could think of obtaining a different set of projectors for each specific choice $\boldsymbol{\Sigma}$. We will now prove instead that the high-order quasi-Helmholtz projectors, with our definition, are invariant under any particular choice of the high-order Star matrix. In particular

Proposition 5.1. Consider two different operators $\mathcal{L}$ and $\widehat{\mathcal{L}}$ both satisfying definition and properties in (5.4) with associated matrices $\boldsymbol{\Sigma}$ and $\widehat{\boldsymbol{\Sigma}}$, respectively defined via (5.6), then there exists an invertible matrix $\mathbf{M}$ so that $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\mathbf{M}$.

Proof. Consider an arbitrary basis $\left\{b_\ell^{(p)}\right\}_{\ell=1}^{S_p}$ of $\mathbb{DG}_p$. We can then express the divergence of each $\boldsymbol{\psi}_m^{(p)}(\boldsymbol{r})$ as

$$\operatorname{div}_\Gamma \boldsymbol{\psi}_m^{(p)}(\boldsymbol{r}) = \sum_{\ell=1}^{S_p} a_{m\ell} b_\ell^{(p)}(\boldsymbol{r})\,, \tag{5.12}$$

with a proper set of coefficients $\{a_{m\ell}\}_{\ell=1}^{S_p}$ for each $\boldsymbol{\psi}_m^{(p)}(\boldsymbol{r})$. Since both $\boldsymbol{\Sigma}$ and $\widehat{\boldsymbol{\Sigma}}$ are defined on the same GWP set $\boldsymbol{\psi}_m^{(p)}(\boldsymbol{r})$ and assuming the same ordering of the basis functions in both cases, from (5.12) we have

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{L} \text{ and } \widehat{\boldsymbol{\Sigma}} = \mathbf{A}\widehat{\mathbf{L}}\,, \tag{5.13}$$

with $\left[\mathbf{A}\right]_{m\ell} = a_{m\ell}$, $\left[\mathbf{L}\right]_{\ell n} = \mathcal{L}_n b_\ell^{(p)}$, and $\left[\widehat{\mathbf{L}}\right]_{\ell n} = \widehat{\mathcal{L}}_n b_\ell^{(p)}$.

Using the definition, linearity, and injectivity of $\mathcal{L}$, together with the completeness of the basis $\left\{b_\ell^{(p)}\right\}_{\ell=1}^{S_p}$, we obtain

$$\begin{aligned}
\mathbf{L}^{\mathsf{T}}\boldsymbol{x} = \mathbf{0} &\Rightarrow \sum_{\ell=1}^{S_p} \left(\mathcal{L}_m b_\ell^{(p)}\right) x_\ell = 0\,, \qquad \forall m = 1,\ldots,S_p \\
&\Rightarrow \mathcal{L}_m \left(\sum_{\ell=1}^{S_p} x_\ell b_\ell^{(p)} \right) = 0\,, \qquad \forall m = 1,\ldots,S_p \\
&\Rightarrow \sum_{\ell=1}^{S_p} x_\ell b_\ell^{(p)} = 0 \Rightarrow \boldsymbol{x} = \mathbf{0}\,,
\end{aligned} \tag{5.14}$$

which, with linearity, establishes the invertibility of $\mathbf{L}$. The same strategy shows the invertibility of $\widehat{\mathbf{L}}$. We now have

$$\mathbf{A} = \mathbf{A} \Rightarrow \mathbf{A}\widehat{\mathbf{L}}\widehat{\mathbf{L}}^{-1} = \mathbf{A}\mathbf{L}\mathbf{L}^{-1} \Rightarrow \widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{L}}^{-1} = \boldsymbol{\Sigma}\mathbf{L}^{-1}\,, \tag{5.15}$$

from which we obtain the looked for relationship

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\mathbf{M}\,, \tag{5.16}$$

with $\mathbf{M} = \mathbf{L}^{-1}\widehat{\mathbf{L}}$, an invertible matrix. □

The well-posedness of the definition of the projectors in (5.9) and (5.10) now follows. In particular

Corollary 5.1. The quasi-Helmholtz projectors are invariant on any specific choice of a high-order Star matrix satisfying (5.6).

Proof. Given two matrices as above, we have from the previous proposition that $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}\mathbf{M}$ with $\mathbf{M}$ invertible. First note that, letting

$$\boldsymbol{\Sigma} = \widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^{\mathsf{T}} \tag{5.17}$$

be the reduced singular value decomposition (SVD) of $\boldsymbol{\Sigma}$ (see Definition 3.4), then

$$\mathbf{M}^{\mathsf{T}}\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{M} = \left(\mathbf{M}^{\mathsf{T}}\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}\right)\left(\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^{\mathsf{T}}\mathbf{M}\right)\,, \tag{5.18}$$

where the left matrix in parenthesis is a full column rank matrix and the right matrix in parenthesis is a full row rank matrix. By using the properties of the Moore-Penrose pseudoinverse (see Proposition 3.2) we get

$$\left(\mathbf{M}^{\mathsf{T}}\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{M}\right)^{+} = \left(\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^{\mathsf{T}}\mathbf{M}\right)^{+}\left(\mathbf{M}^{\mathsf{T}}\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}\right)^{+}\,, \tag{5.19}$$

as well as

$$\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^{\mathsf{T}}\mathbf{M}\left(\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^{\mathsf{T}}\mathbf{M}\right)^{+} = \left(\mathbf{M}^{\mathsf{T}}\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}\right)^{+}\mathbf{M}^{\mathsf{T}}\widetilde{\mathbf{V}}\widetilde{\mathbf{S}} = \mathbf{I}\,, \tag{5.20}$$

from which it follows that

$$\begin{aligned}\mathbf{P}^{\widehat{\Sigma}} &= \widehat{\mathbf{\Sigma}}\left(\widehat{\mathbf{\Sigma}}^{\mathsf{T}}\widehat{\mathbf{\Sigma}}\right)^{+}\widehat{\mathbf{\Sigma}}^{\mathsf{T}} \\ &= \mathbf{\Sigma}\mathbf{M}\left(\mathbf{M}^{\mathsf{T}}\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}\mathbf{M}\right)^{+}\mathbf{M}^{\mathsf{T}}\mathbf{\Sigma}^{\mathsf{T}} \\ &= \widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^{\mathsf{T}}\mathbf{M}\left(\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^{\mathsf{T}}\mathbf{M}\right)^{+}\left(\mathbf{M}^{\mathsf{T}}\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}\right)^{+}\mathbf{M}^{\mathsf{T}}\widetilde{\mathbf{V}}\widetilde{\mathbf{S}}\widetilde{\mathbf{U}}^{\mathsf{T}} \\ &= \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^{\mathsf{T}} = \mathbf{\Sigma}\left(\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}\right)^{+}\mathbf{\Sigma}^{\mathsf{T}} = \mathbf{P}^{\Sigma}\,.\end{aligned} \tag{5.21}$$

The uniqueness of $\mathbf{P}^{\Sigma}$ combined with (5.10) directly implies that of $\mathbf{P}^{\Lambda H}$. □

To fix ideas and obtain an algorithm for the projectors, we propose to choose the following explicit definition of the Star matrix $\mathbf{\Sigma}$

$$\left[\mathbf{\Sigma}\right]_{mn} = \int_{\Gamma} \mathrm{div}_{\Gamma}\boldsymbol{\psi}_m^{(p)}(\boldsymbol{r})\sigma_n^{(p)}(\boldsymbol{r})\,\mathrm{d}S(\boldsymbol{r})\,, \tag{5.22}$$

where $\{\sigma_n^{(p)}\}_{n=1}^{S_p}$ are the interpolatory basis functions of $\mathbb{DG}_p$ defined in Section 2.2.3. We will see in the next chapters that this definition allows to connect the Star matrix with a mixed discretization of the Laplacian operator. This matrix can also be seen as the testing with GWP functions of a weak gradient applied to the discontinuous functions, which shows that the quasi-Helmholtz decomposition relates a decomposition of Raviart-Thomas functions in terms of divergence free functions and weak gradient of discontinuous functions (see (2.64)).

Another possible choice of Star matrix is

$$\left[\widehat{\mathbf{\Sigma}}\right]_{mn} = \begin{cases} \widehat{\boldsymbol{\nabla}}\cdot\widehat{\psi}^{(p)}_{b_m^K,\boldsymbol{i}_m^K}(\widehat{\boldsymbol{r}}^{(p)}_{\boldsymbol{i}_n})\,, & \text{if } \sigma_n \text{ is not zero on triangle } K \textit{ [cell- and edge-based GWP]} \\ -\widehat{\boldsymbol{\nabla}}\cdot\widehat{\psi}^{(p)}_{b_m^L,\boldsymbol{i}_m^L}(\widehat{\boldsymbol{r}}^{(p)}_{\boldsymbol{i}_n})\,, & \text{if } \sigma_n \text{ is not zero on triangle } L \textit{ [edge-based GWP only]} \\ 0\,, & \text{otherwise} \end{cases} \tag{5.23}$$

where we used the notations of Section 2.2.3. This second definition gives the expansion coefficients of the GWP functions in the basis of discontinuous Garlerkin interpolatory functions, i.e.,

$$\mathrm{div}_{\Gamma}\psi_m^{(p)}(\boldsymbol{r}) = \sum_{n=1}^{S_p}\left[\widehat{\mathbf{\Sigma}}\right]_{mn}\sigma_n^{(p)}(\boldsymbol{r})\,. \tag{5.24}$$

Moreover, this generalizes the definition we gave at order zero (4.4) (up to a scaling factor), as we have

$$\left[\widehat{\mathbf{\Sigma}}\right]_{mn} = \begin{cases} 2\,, & \text{if } K \text{ is triangle } n\,, \\ -2\,, & \text{if } L \text{ is triangle } n\,, \\ 0\,, & \text{otherwise}\,, \end{cases} \tag{5.25}$$

for $p = 0$.

Finally, we check that these two definitions verify Proposition 5.1. From (5.24), it is direct to see that

$$\mathbf{\Sigma} = \widehat{\mathbf{\Sigma}}\mathbf{G}_{\sigma^{(p)},\sigma^{(p)}}\,, \tag{5.26}$$

and we have

$$\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma}\mathbf{M}\,, \tag{5.27}$$

where $\mathbf{M} = \mathbf{G}^{-1}_{\sigma^{(p)},\sigma^{(p)}}$ is an invertible matrix.

5.4 Stabilization of the EFIE

Now, from (5.11), the high-order quasi-Helmholtz projectors satisfy

$$\mathbf{P}^{\Lambda H}\mathbf{T}_h = \mathbf{T}_h\mathbf{P}^{\Lambda H} = \mathbf{0} \quad \text{and} \quad \mathbf{T}_h\mathbf{P}^{\Sigma} = \mathbf{P}^{\Sigma}\mathbf{T}_h = \mathbf{T}_h\,, \tag{5.28}$$

which generalizes the analogous property valid in the zeroth-order case [2]. As a matter of fact, from (5.24), the matrix $\mathbf{T}_h$ can be expressed in the following form

$$\mathbf{T}_h = \widehat{\mathbf{\Sigma}}\mathbf{W}\widehat{\mathbf{\Sigma}}^{\mathsf{T}}\,, \tag{5.29}$$

where

$$\left[\mathbf{W}\right]_{mn} = \int_{\Gamma}\int_{\Gamma} G(\boldsymbol{r},\boldsymbol{r}')\sigma_m^{(p)}(\boldsymbol{r})\sigma_n^{(p)}(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r}')\,\mathrm{d}S(\boldsymbol{r})\,, \tag{5.30}$$

is a full-rank matrix that discretizes the single layer potential with $\mathbb{DG}_p$ functions. This shows that $\mathbf{T}_h$ has the same range than $\widehat{\mathbf{\Sigma}}$, which is also the range of $\mathbf{\Sigma}$ and $\mathbf{P}^{\Sigma}$.

This allows to proceed with the same formal strategy developed for zeroth-order projectors that here will work for the high-order EFIE. Define the preconditioning matrix

$$\mathbf{P} = j\sqrt{k/C}\,\mathbf{P}^{\Sigma} + \sqrt{C/k}\,\mathbf{P}^{\Lambda H}, \tag{5.31}$$

with the scaling factor

$$C = \sqrt{\frac{\|\mathbf{T}_h\|}{\|\mathbf{P}^{\Lambda H}\mathbf{T}_s\mathbf{P}^{\Lambda H}\|}}. \tag{5.32}$$

The preconditioned system reads

$$\mathbf{PTP}\boldsymbol{y} = \mathbf{P}\boldsymbol{e}, \tag{5.33}$$

with

$$\boldsymbol{j} = \mathbf{P}\boldsymbol{y}. \tag{5.34}$$

The proof of low-frequency well-conditioning and stability of (5.33) is formally identical to the one for the zeroth-order case [2] and we omit it here for the sake of conciseness.

5.5 Implementation Details

In this section, we deal with implementation related details that could be useful to the reader while implementing all new techniques described here.

We assume that all matrix-vector products are done in a quasi-linear number of operations, using multipole or compression-based approaches [112, 119] for the integral operators and a sparse algorithm for the Star matrices. Note that, to match the computational complexity of fast matrix-vector product algorithms, the naive computation of the norms in (5.32) should be avoided and iterative approaches, such as power iterations [129], should be used instead. Moreover, the numerical (pseudo-)inversion of $\mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma}$ in (5.9) has to be done iteratively whenever a product with the projectors is involved. As in the zeroth-order case, this matrix is similar to the one we get by discretizing the Laplacian equation with finite elements of order p. We can thus rely on algorithms tailored to solve such systems from other fields of applications [76]. In particular, multigrid algorithms have proven to be quite effective. Because there exists several variants of multigrids, one could use the p-multigrid versions [134, 76, 83] (tailored for high-order finite

elements) but there are also robust implementations of the standard algebraic multigrid [103] generalized to different scenarios that are often easier to use in a black-box fashion.

Finally, to maximize the effect of our regularization to all frequencies, the computation of the projected EFIE has to be done carefully. For numerical purposes, the properties (5.28) have to be enforced explicitly in the computation of the preconditioned matrix [2], which reads

$$\begin{aligned}\mathbf{PTP} = &-jC\,\mathbf{P}^{\Lambda H}\mathbf{T}_s\mathbf{P}^{\Lambda H} - j/C\,\mathbf{T}_h + k\,\mathbf{P}^{\Lambda H}\mathbf{T}_s\mathbf{P}^{\Sigma} \\ &+ k\,\mathbf{P}^{\Sigma}\mathbf{T}_s\mathbf{P}^{\Lambda H} + jk^2/C\,\mathbf{P}^{\Sigma}\mathbf{T}_s\mathbf{P}^{\Sigma}.\end{aligned} \tag{5.35}$$

The loss of significant digits in the computation of the right-hand side also has to be avoided, as is well-known when dealing with very low-frequency numerical strategies [2]. In particular, the solenoidal and non-solenoidal contributions of the right-hand side must be computed separately, and the static part of the excitation source is subtracted when tested with the solenoidal functions [149, 2]. More explicitly, we write

$$\mathbf{P}\boldsymbol{e} = j\sqrt{k/C}\,\mathbf{P}^{\Sigma}\boldsymbol{e} + \sqrt{C/k}\,\mathbf{P}^{\Lambda H}\boldsymbol{e}_{\text{sub}}, \tag{5.36}$$

where $\boldsymbol{e}_{\text{sub}}$ is the subtracted right-hand side [2]. Finally, the post-processing computation of the electric field requires a similar treatment, by using separately the solenoidal and non-solenoidal contributions of the solution

$$\boldsymbol{j}_{\text{sol}} = \sqrt{C/k}\,\mathbf{P}^{\Lambda H}\boldsymbol{y} \quad \text{and} \quad \boldsymbol{j}_{\text{nsol}} = j\sqrt{k/C}\,\mathbf{P}^{\Sigma}\boldsymbol{y} \tag{5.37}$$

and subtracting the static part of the Green's kernel when integrating with the solenoidal contribution [2].

5.6 Numerical Results

In this section, we give numerical results that validate the use of the quasi-Helmholtz projectors by comparing it against the standard EFIE and against a standard quasi-Helmholtz decomposition. The quasi-Helmholtz decomposition we adopted for comparison is a generalization of the Loop/Star technique (L/S-EFIE) from the zeroth-order: the matrix $\boldsymbol{\Sigma}$ is used for generating the Stars and the Loops are computed as described in [141]. The meshes are generated from the software Gmsh that provides quadratic (curvilinear) triangles [60]. The singular integrals are computed using the singularity cancellation scheme described in [123] while near-singular and far interactions are computed with Gaussian quadratures. Moreover, because we are interested in studying the condition number of the underlying systems, the projectors need to be explicitly computed, which prevents the use of a fast strategy as described in Section 5.5. To this extent, we rely on a direct pseudo-inversion of $\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}$ based on the computation of its reduced SVD.

Our first validation is done over the unitary sphere for different frequencies. Figure 5.1 shows the condition number of the system matrix for each different method as a function of frequency and for different orders. We observe that the condition number of the standard EFIE increases dramatically, as expected, while it remains constant for the standard Loop/Star technique and for the approach proposed here based on high-order quasi-Helmholtz projectors. For the latter case, however, the condition number is lower. This effect can be better understood from the test in Figure 5.2 where the condition number is shown in function of the inverse mesh size. It is clear that, differently from standard Loop/Star, the new projectors do not worsen the spectral behavior of the original EFIE generalizing what happens in the zeroth-order case [2]. The radar cross

sections (RCS) at 100Hz and at 3×10^8Hz are obtained in Figures 5.3 and 5.4, respectively. This shows that the low frequency-breakdown, absent at higher frequencies, is instead corrupting the results at low-frequency when no low-frequency treatment is employed. At the same time, these results validate the stability of our scheme in a wide frequency range.

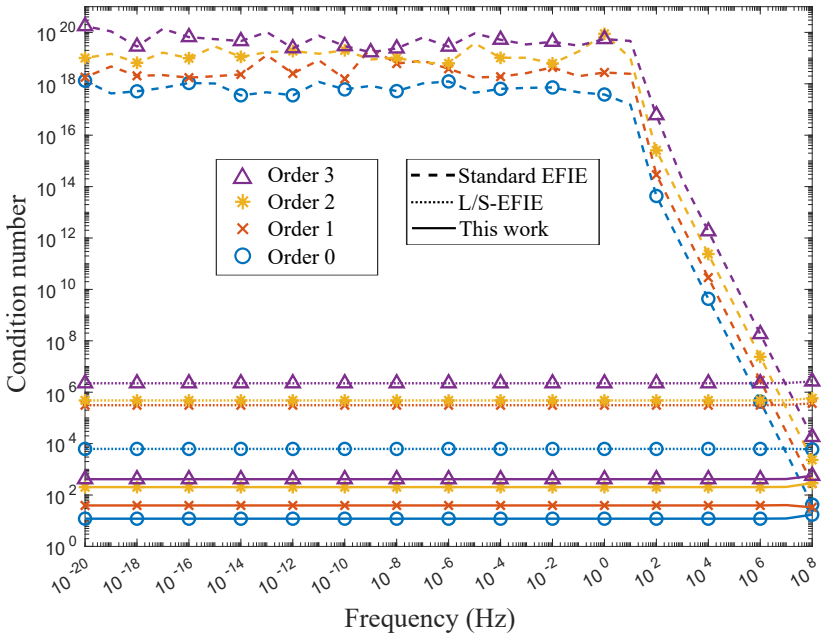


Figure 5.1: Condition number of the system matrices in function of the frequency for order zero to three.

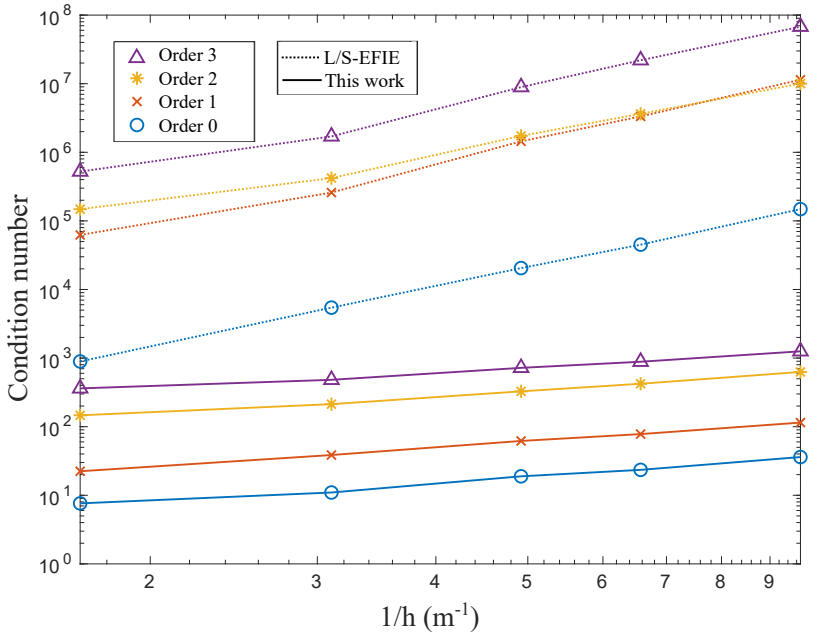


Figure 5.2: Condition number of the system matrices in function of the inverse of the average cell diameter h; frequency of 1Hz.

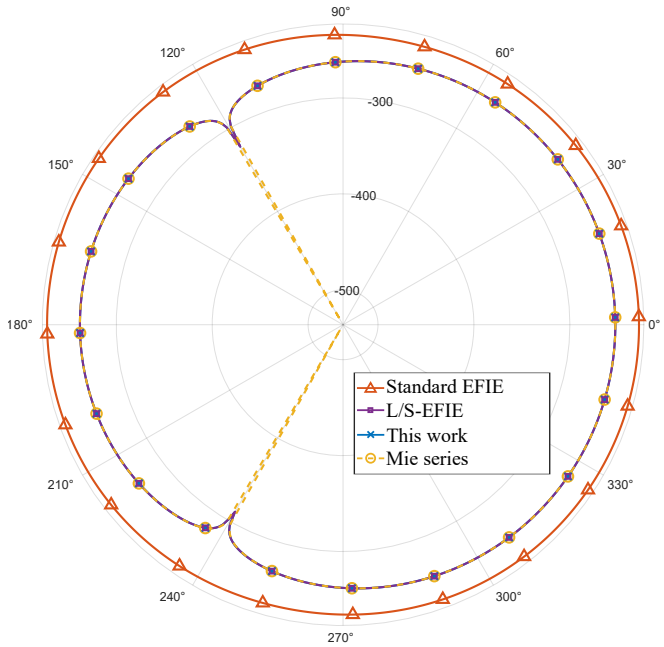


Figure 5.3: Radar cross section (RCS) in dBsm at 100Hz.

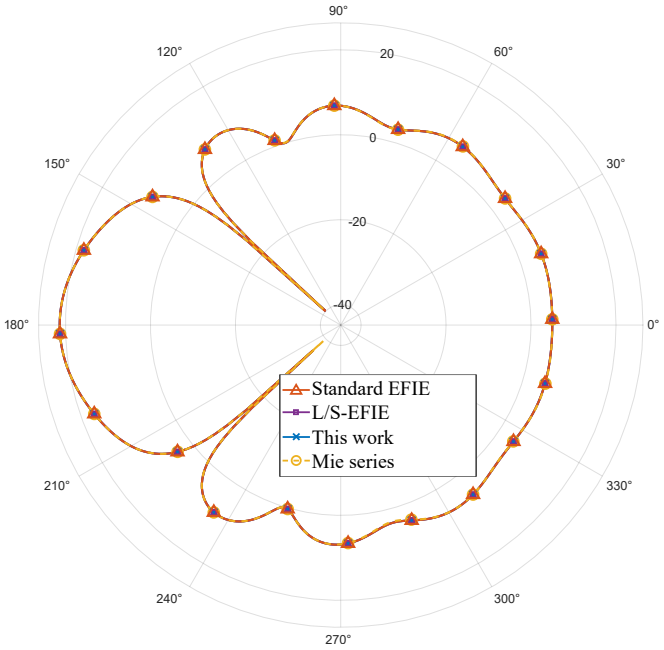


Figure 5.4: Radar cross section (RCS) in dBsm at 3×10^8Hz.

To test the performance of our scheme on a non-simply connected geometry we used the "two interlocked Möbius ring" structure, that consists of two separate objects with interlaced holes and handles. The rings are of radius 1m and are discretized with 3780 cells and 5670 edges. In total, there are 72 global Loops associated with this topology. Figure 5.5 shows the surface density current obtained by solving the EFIE with and without preconditioning at 10Hz using basis functions of order two (39690 unknowns). We get, without the need of detecting the global Loops, the same results with the quasi-Helmholtz projectors and the Loop/Star-EFIE for which however the global Loops must be explicitly detected. It should also be noted that the accuracy is completely lost without preconditioning.

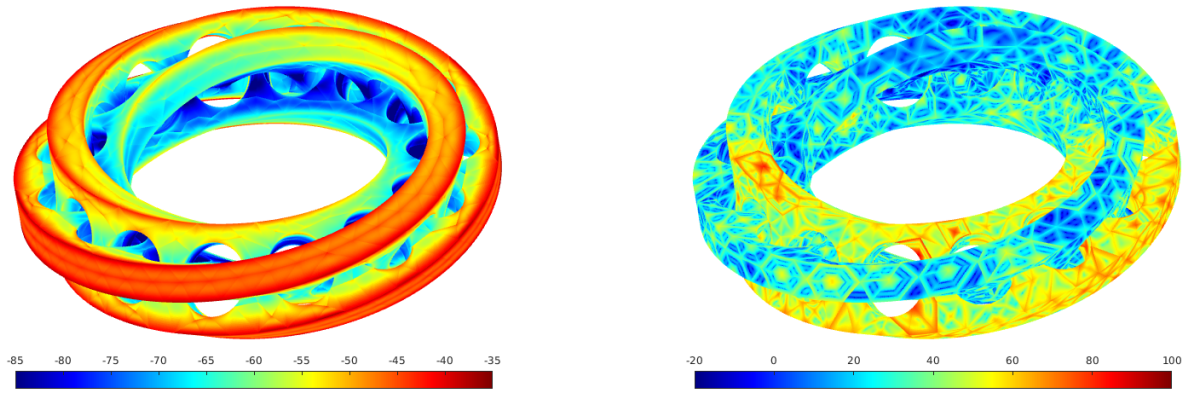


Figure 5.5: Absolute value of the surface density current (in $\mathrm{dB_{A/m}}$) for the interlocked Möbius ring irradiated by a plane wave at 10Hz, with preconditioning (left) and without (right).

Our final validation test scenario is on the model of an airliner with 2D apertures corresponding to the windows and the shell of the jet engine. The mesh contains 3942 cells and 5750 internal edges, and we use basis functions of order two (42900 unknowns). In addition to show the relevance of our method to solve industrial scenarios, this example completes the numerical study with problems containing global Loops around apertures, which are of different kind than the harmonic Loops of the previous example (see the definition of $\mathbf{\Lambda}$ in Chapter 4). Figure 5.6 shows the surface density current obtained by solving the EFIE with and without preconditioning at 1Hz. As for the previous case, the quasi-Helmholtz projectors give an identical solution as the one obtained with Loop-Star decompositions (which however requires Loop detection), while the solution without preconditioning is completely jeopardized.

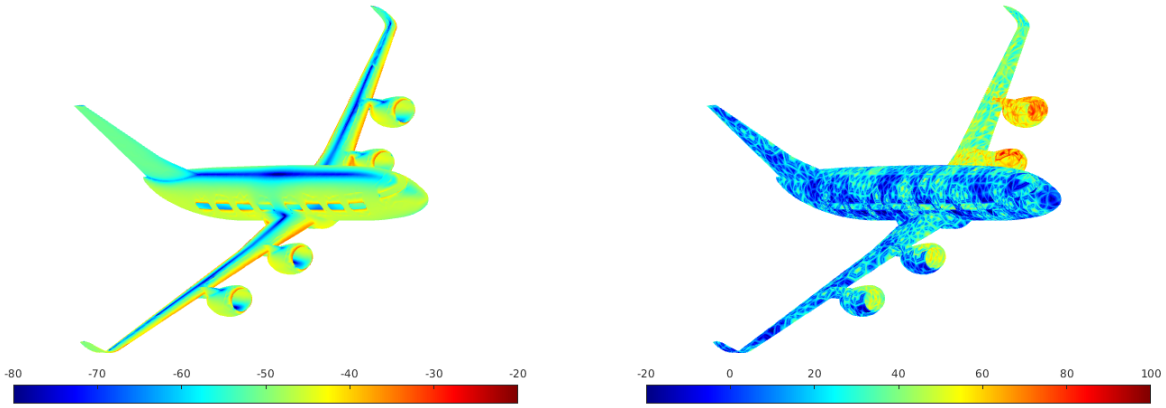


Figure 5.6: Absolute value of the surface density current (in $\mathrm{dB_{A/m}}$) for the airliner irradiated by a plane wave at 1Hz, with preconditioning (left) and without (right).

5.7 Conclusion

In this chapter, we have extended the use of the quasi-Helmholtz projectors for stabilizing the EFIE when discretized with high-order basis functions. The new scheme is based on a generalized definition of Star matrix that is wide enough to encompass numerous relevant scenarios, including the use of GWP basis elements. The contribution has shown that this

results into unique quasi-Helmholtz projectors irrespectively of the specific choice of Star matrix. Numerical results have shown the effectiveness of the new approach.

In addition to enable the low-frequency preconditioning of the EFIE, high-order quasi-Helmholtz projectors will allow us to provide efficient preconditioners to cure the dense-refinement breakdown of the EFIE (with respect to the mesh density, and with respect to the basis functions order), as it will be presented in the two next chapters.

Chapter 6

High-Order Quasi-Helmholtz Filters

Preliminary results of this work were presented at the 2024 IEEE International Symposium on Antennas and Propagation [26].

6.1 Introduction

In addition to suffer from the low-frequency breakdown, the EFIE suffers from the so-called h-refinement breakdown, that occurs when the average edge size h of the mesh elements decreases for a fixed frequency. This ill-conditioning is related to the fact that the EFIE operator acts differently on the solenoidal and non-solenoidal functions. More specifically, the singular values associated with solenoidal functions decrease as $O(h)$, while the eigenvalues associated with non-solenoidal functions increase as $O(1/h)$ [2]. Other ill-conditioning effects due to the discretization are related to non uniform scalings between the basis functions. Notably, the condition number deteriorates for meshes that are ill shaped (presence of sharp elements) and non-uniform (high contrast between the size of the elements), and the condition number increases almost exponentially with the basis functions order, if the basis functions are not orthogonal, as it is the case with the interpolatory functions [2].

Several strategies have been proposed to address the low-frequency breakdown, the h-refinement breakdown, or both (the reader may refer to [2] for a complete review). Recently, quasi-Helmholtz filters have been proposed as an alternative approach to treating both low-frequency and h-refinement breakdowns of the EFIE [94]. These are obtained by filtering the graph Laplacian matrices appearing in the quasi-Helmholtz projectors definitions to further flatten the spectrum of the preconditioned system. This approach can be seen as a direct extension of the low-frequency stabilization leveraging quasi-Helmholtz projectors, where filtering operations allow additional curing of the h-refinement breakdown. Following the generalization of the quasi-Helmholtz projectors to the high-order case, this work addresses the treatment of the low-frequency and the h-refinement breakdown by using the quasi-Helmholtz filters in a strategy compatible with arbitrary basis function order.

6.2 Orthonormalization of the Basis Functions

The linear systems arising from the BEM may suffer from ill-conditioning because of the lack of orthogonality between the basis functions [94]. The problem further worsens for non-uniform meshes and for increasing orders. This can be tackled by resorting to the transformation matrices

that orthonormalize the basis functions with respect to the inner product (1.3), which are given by the inverse square roots of the Gram matrices associated with these basis functions (1.5): we compute $\mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}}$ for the continuous scalar functions, $\mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}$ for the div-conforming vector functions and $\mathbf{G}^{-1/2}_{\sigma^{(p)},\sigma^{(p)}}$ for the discontinuous scalar functions.

Subsequently, the normalized EFIE is

$$\widetilde{\mathbf{T}}\widetilde{\boldsymbol{j}} = \widetilde{\boldsymbol{e}}\,, \tag{6.1}$$

where

$$\widetilde{\mathbf{T}} = -jk\widetilde{\mathbf{T}}_s + \frac{1}{jk}\widetilde{\mathbf{T}}_h\,, \tag{6.2}$$

$$\widetilde{\mathbf{T}}_s = \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\mathbf{T}_s\mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\,, \tag{6.3}$$

$$\widetilde{\mathbf{T}}_h = \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\mathbf{T}_h\mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\,, \tag{6.4}$$

$$\boldsymbol{j} = \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\widetilde{\boldsymbol{j}}\,, \tag{6.5}$$

$$\widetilde{\boldsymbol{e}} = \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\boldsymbol{e}\,, \tag{6.6}$$

while the normalized GWP-to-Star and GWP-to-Loop transformation matrices are

$$\widetilde{\boldsymbol{\Sigma}} = \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\boldsymbol{\Sigma}\mathbf{G}^{-1/2}_{\sigma^{(p)},\sigma^{(p)}} = \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\mathbf{G}_{\mathrm{div}_\Gamma\boldsymbol{\psi}^{(p)},\sigma^{(p)}}\mathbf{G}^{-1/2}_{\sigma^{(p)},\sigma^{(p)}}, \tag{6.7}$$

$$\widetilde{\boldsymbol{\Lambda}} = \mathbf{G}^{1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\boldsymbol{\Lambda}\mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}} = \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\mathbf{G}_{\boldsymbol{\psi}^{(p)},\mathbf{curl}_\Gamma\lambda^{(p)}}\mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}}\,, \tag{6.8}$$

with the orthogonal property analog to (5.7)

$$\widetilde{\boldsymbol{\Sigma}}^{\mathsf{T}}\widetilde{\boldsymbol{\Lambda}} = \mathbf{0}\,. \tag{6.9}$$

Consequently, the quasi-Helmholtz decomposition of the expansion coefficients expressed in the orthonormalized basis is

$$\mathbb{C}^N = \widetilde{\Sigma} \oplus \widetilde{\Lambda} \oplus \widetilde{H}\,, \tag{6.10}$$

where

$$\widetilde{\Sigma} = \mathrm{range}(\widetilde{\boldsymbol{\Sigma}}), \tag{6.11}$$

$$\widetilde{\Lambda} = \mathrm{range}(\widetilde{\boldsymbol{\Lambda}}), \tag{6.12}$$

and

$$\widetilde{H} = \mathrm{null}(\widetilde{\boldsymbol{\Sigma}}) \cap \mathrm{null}(\widetilde{\boldsymbol{\Lambda}}). \tag{6.13}$$

Finally, the normalized quasi-Helmholtz projectors are

$$\mathbf{P}^{\widetilde{\Sigma}} = \widetilde{\boldsymbol{\Sigma}}\left(\widetilde{\boldsymbol{\Sigma}}^{\mathsf{T}}\widetilde{\boldsymbol{\Sigma}}\right)^{+}\widetilde{\boldsymbol{\Sigma}}\,, \tag{6.14}$$

$$\mathbf{P}^{\widetilde{\Lambda}} = \widetilde{\boldsymbol{\Lambda}}\left(\widetilde{\boldsymbol{\Lambda}}^{\mathsf{T}}\widetilde{\boldsymbol{\Lambda}}\right)^{+}\widetilde{\boldsymbol{\Lambda}}\,, \tag{6.15}$$

$$\mathbf{P}^{\widetilde{H}} = \mathbf{I} - \mathbf{P}^{\widetilde{\Sigma}} - \mathbf{P}^{\widetilde{\Lambda}}\,, \tag{6.16}$$

with the analog properties derived from (5.28)

$$\mathbf{P}^{\widetilde{\Lambda}}\widetilde{\mathbf{T}}_h = \widetilde{\mathbf{T}}_h\mathbf{P}^{\widetilde{\Lambda}} = \mathbf{P}^{\widetilde{H}}\widetilde{\mathbf{T}}_h = \widetilde{\mathbf{T}}_h\mathbf{P}^{\widetilde{H}} = \mathbf{0}\,, \tag{6.17}$$

$$\mathbf{P}^{\widetilde{\Sigma}}\widetilde{\mathbf{T}}_h = \widetilde{\mathbf{T}}_h\mathbf{P}^{\widetilde{\Sigma}} = \widetilde{\mathbf{T}}_h\,, \tag{6.18}$$

and the normalized projector-based EFIE becomes

$$\widetilde{\mathbf{P}}\widetilde{\mathbf{T}}\widetilde{\mathbf{P}}\widetilde{\boldsymbol{y}} = \widetilde{\mathbf{P}}\widetilde{\boldsymbol{e}}\,, \tag{6.19}$$

with

$$\widetilde{\mathbf{P}} = j\sqrt{k/\widetilde{C}}\mathbf{P}^{\widetilde{\Sigma}} + \sqrt{\widetilde{C}/k}\mathbf{P}^{\widetilde{\Lambda}\widetilde{H}}\,, \tag{6.20a}$$

$$\widetilde{\boldsymbol{j}} = \widetilde{\mathbf{P}}\widetilde{\boldsymbol{y}}\,, \tag{6.20b}$$

$$\mathbf{P}^{\widetilde{\Lambda}\widetilde{H}} = \mathbf{P}^{\widetilde{\Lambda}} + \mathbf{P}^{\widetilde{H}} = \mathbf{I} - \mathbf{P}^{\widetilde{\Sigma}}\,, \tag{6.20c}$$

$$\widetilde{C} = \sqrt{\frac{\|\widetilde{\mathbf{T}}_h\|}{\|\mathbf{P}^{\widetilde{\Lambda}\widetilde{H}}\widetilde{\mathbf{T}}_s\mathbf{P}^{\widetilde{\Lambda}\widetilde{H}}\|}}\,. \tag{6.20d}$$

In addition to improve the conditioning of the EFIE and that of its projector-based counterpart, the spectrum analysis of the orthonormalized Loop and Star bases will lead to an effective strategy to precondition the underlying systems for dense mesh discretization. This issue is addressed in the following sections.

6.3 Analysis of the Spectrum of the Quasi-Helmholtz Decomposition

In this section, we investigate the spectrum properties of the operators that we need to pseudo-invert in order to compute the normalized projectors (6.15) and (6.14). We show that they are Galerkin discretization matrices of the Laplace-Beltrami eigenvalue problem in standard [131] and mixed [92, 18] finite element methods (FEM), respectively. Consequently, we leverage results arising from FEM together with operator spectral theory to analyze the spectrum of the orthonormalized quasi-Helmholtz decomposition.

6.3.1 Relationship with the Laplace-Beltrami Eigenvalue Problem

The Laplace-Beltrami eigenvalue problem reads [106]

$$-\mathrm{div}_\Gamma\mathbf{grad}_\Gamma f_n = \upsilon_n f_n\,, \tag{6.21}$$

where $\{\upsilon_n\}_{n\in\mathbb{N}}$ are the Laplace-Beltrami eigenvalues associated with the eigenfunctions $\{f_n\}_{n\in\mathbb{N}}$. In standard FEM, this problem is discretized as [131, 17]

$$\mathbf{G}_{\mathbf{grad}_\Gamma\lambda^{(p)},\mathbf{grad}_\Gamma\lambda^{(p)}}\mathbf{U} = \mathbf{G}_{\lambda^{(p)},\lambda^{(p)}}\mathbf{U}\mathbf{D}^{\widetilde{\Lambda}} \tag{6.22}$$

where $\mathbf{U}$ contains the eigenvectors (i.e., the expansion coefficients of approximations of the Laplace-Beltrami eigenfunctions) and $\mathbf{D}^{\widetilde{\Lambda}}$ is a diagonal matrix that contains their associated eigenvalues $\upsilon_n^{\widetilde{\Lambda}}$ (i.e., approximations of the Laplace-Beltrami eigenvalues). This generalized eigenvalue equation can be recast into the standard eigenvalue equation

$$\begin{gathered}\left(\mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}}\mathbf{G}_{\mathbf{grad}_\Gamma\lambda^{(p)},\mathbf{grad}_\Gamma\lambda^{(p)}}\mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}}\right)\mathbf{V} = \mathbf{V}\mathbf{D}^{\widetilde{\Lambda}} \\ \text{with}\quad \mathbf{U} = \mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}}\mathbf{V}\,.\end{gathered} \tag{6.23}$$

It follows that

$$\begin{aligned}\widetilde{\mathbf{\Lambda}}^{\mathsf{T}}\widetilde{\mathbf{\Lambda}} &= \mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}}\mathbf{\Lambda}^{\mathsf{T}}\mathbf{G}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\mathbf{\Lambda}\mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}} \\ &= \mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}}\mathbf{G}_{\mathbf{curl}_\Gamma\lambda^{(p)},\mathbf{curl}_\Gamma\lambda^{(p)}}\mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}} \\ &= \mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}}\mathbf{G}_{\mathbf{grad}_\Gamma\lambda^{(p)},\mathbf{grad}_\Gamma\lambda^{(p)}}\mathbf{G}^{-1/2}_{\lambda^{(p)},\lambda^{(p)}}\end{aligned} \tag{6.24}$$

is a Galerkin discretization matrix of the Laplace-Beltrami eigenvalue problem.

Alternatively, the Laplace-Beltrami eigenvalue problem can be discretized via mixed FEM [92, 18]. More precisely, by introducing the auxiliary vector $\boldsymbol{u}_n = \mathbf{grad}_\Gamma f_n$, the Laplace eigenvalue problem can be rewritten as

$$\begin{cases}\boldsymbol{u}_n - \mathbf{grad}_\Gamma f_n &= \mathbf{0} \\ \mathrm{div}_\Gamma \boldsymbol{u}_n &= -\upsilon_n f_n\end{cases}. \tag{6.25}$$

The unknown $\boldsymbol{u}_n$ is approximated by a linear combination of Raviart-Thomas functions, while f_n is approximated by a linear combination of discontinuous Galerkin functions (therefore, the gradient in the first row has to be considered in the distributional sense). After testing the first row by Raviart-Thomas elements and the second row by discontinuous Galerkin functions, and by applying integration by parts, we find the saddle point system

$$\begin{pmatrix}\mathbf{G}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}} & \mathbf{\Sigma} \\ \mathbf{\Sigma}^{\mathsf{T}} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{A} \\ \mathbf{B}\end{pmatrix} = \begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{G}_{\sigma^{(p)},\sigma^{(p)}}\end{pmatrix}\begin{pmatrix}\mathbf{A} \\ \mathbf{B}\end{pmatrix}\mathbf{D}^{\widetilde{\Sigma}}, \tag{6.26}$$

where $\mathbf{A}$ and $\mathbf{B}$ contain the expansion coefficients of the approximated eigenfunctions and $\mathbf{D}^{\widetilde{\Sigma}}$ contains their associated eigenvalues $\upsilon_n^{\widetilde{\Sigma}}$. Rearranging the first row, we find

$$\mathbf{A} = -\mathbf{G}^{-1}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\mathbf{\Sigma}\mathbf{B}, \tag{6.27}$$

while the second row is

$$\mathbf{\Sigma}^{\mathsf{T}}\mathbf{A} = -\mathbf{G}_{\sigma^{(p)},\sigma^{(p)}}\mathbf{B}\mathbf{D}^{\widetilde{\Sigma}}, \tag{6.28}$$

which finally gives

$$\left(\mathbf{\Sigma}^{\mathsf{T}}\mathbf{G}^{-1}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\mathbf{\Sigma}\right)\mathbf{B} = \mathbf{G}_{\sigma^{(p)},\sigma^{(p)}}\mathbf{B}\mathbf{D}^{\widetilde{\Sigma}}. \tag{6.29}$$

This generalized eigenvalue equation can be recast into the standard eigenvalue equation

$$\begin{gathered}\left(\mathbf{G}^{-1/2}_{\sigma^{(p)},\sigma^{(p)}}\mathbf{\Sigma}^{\mathsf{T}}\mathbf{G}^{-1}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\mathbf{\Sigma}\mathbf{G}^{-1/2}_{\sigma^{(p)},\sigma^{(p)}}\right)\mathbf{C} = \mathbf{C}\mathbf{D}^{\widetilde{\Sigma}} \\ \text{with} \quad \mathbf{B} = \mathbf{G}^{-1/2}_{\sigma^{(p)},\sigma^{(p)}}\mathbf{C}.\end{gathered} \tag{6.30}$$

It follows that

$$\widetilde{\mathbf{\Sigma}}^{\mathsf{T}}\widetilde{\mathbf{\Sigma}} = \mathbf{G}^{-1/2}_{\sigma^{(p)},\sigma^{(p)}}\mathbf{\Sigma}^{\mathsf{T}}\mathbf{G}^{-1}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\mathbf{\Sigma}\mathbf{G}^{-1/2}_{\sigma^{(p)},\sigma^{(p)}} \tag{6.31}$$

is a Galerkin discretization matrix of the Laplace-Beltrami eigenvalue problem.

6.3.2 Convergence and Asymptotic Behavior of the Laplace-Beltrami Eigenvalues

The two discretizations of the Laplace-Beltrami eigenvalue problems (6.22) and (6.26) have been intensively studied in the past decades [17, 18, 92, 131, 148]. In particular, the convergence

of the eigenvalues of the discretized systems (6.24) and (6.31) to the eigenvalues of the Laplace-Beltrami operator (6.21) are well known, and proved to be (for h small enough)

$$0 < \upsilon_n^{\widetilde{\Lambda}} - \upsilon_n \leq C_1 h^{2(p+1)}\,, \tag{6.32}$$

$$0 < \upsilon_n^{\widetilde{\Sigma}} - \upsilon_n \leq C_2 h^{2(p+1)}\,, \tag{6.33}$$

where C_1 and C_2 are two coefficients independent of h but dependent on other parameters, such as the order p, the shape of the mesh, and the eigenvalue that we want to approximate [148]. In particular, it has been shown [131] that $C_1 \leq C_3 \upsilon_n^{p+2}$ where C_3 is a coefficient independent of p and υ_n. These bounds show that convergence is ensured for $h \to 0$. Additionally, the convergence rate with respect to h is improved when p increases. Yet, the convergence pertaining only to p (for h fixed) remains ambiguous [148].

Finally, we rely on the spectral theory on the Laplace-Beltrami operator to analyze the asymptotic behavior of the discrete eigenvalues. Assuming that the eigenvalues are given in increasing amplitude, the Weyl's law is a classical result that establishes [10]

$$\upsilon_n \sim \frac{4\pi n}{\text{area}\,(\Gamma)}\,. \tag{6.34}$$

This shows, together with (6.32) and (6.33), that the discrete eigenvalues $\upsilon_n^{\widetilde{\Lambda}}$ and $\upsilon_n^{\widetilde{\Sigma}}$ will grow as their index n when refining in h.

6.4 Operator Filtering

6.4.1 Spectrum Filtering and Wavelet-Like Decomposition

Consider a generic matrix $\mathbf{X} \in \mathbb{R}^{N\times M}$ of rank r (used as a placeholder for either $\boldsymbol{\Sigma}$ or $\boldsymbol{\Lambda}$ as well as for their respective normalized counterparts $\widetilde{\boldsymbol{\Sigma}}$ or $\widetilde{\boldsymbol{\Lambda}}$) with associated column space $X = \text{range}(\mathbf{X})$. Denote its *reduced* singular value decomposition (SVD) $\mathbf{X} = \mathbf{U}_\mathbf{X}\mathbf{S}_\mathbf{X}\mathbf{V}_\mathbf{X}^\mathsf{T}$ (i.e., the SVD that does not explicitly compute the nullspace of $\mathbf{X}$), where $\mathbf{U}_\mathbf{X} \in \mathbb{R}^{N\times r}$ and $\mathbf{V}_\mathbf{X} \in \mathbb{R}^{M\times r}$ are unitary matrices with columns being the singular vectors of $\mathbf{X}$ and where $\mathbf{S}_\mathbf{X} \in \mathbb{R}^{r\times r}$ is a diagonal matrix containing the non-negative singular values $\{s_i\}_{i=1}^r$ of $\mathbf{X}$ in an increasing amplitude (see Definition 3.4). The reduced SVD of the symmetric positive semi-definite matrix $\mathbf{X}^\mathsf{T}\mathbf{X}$ (which matches with its eigenvalue decomposition) is thus given by

$$\mathbf{X}^\mathsf{T}\mathbf{X} = \mathbf{V}_\mathbf{X}\mathbf{S}_\mathbf{X}^2\mathbf{V}_\mathbf{X}^\mathsf{T}\,. \tag{6.35}$$

To derive our filtering strategy, we first define the truncated diagonal matrix $\mathbf{S}_{\mathbf{X},n}$

$$\left[\mathbf{S}_{\mathbf{X},n}\right]_{ii} = \begin{cases} s_i & \text{if } i \leq n \\ 0 & \text{otherwise} \end{cases}\,, \tag{6.36}$$

which allows for the following definition of the truncated SVD of $\mathbf{X}$ [94]

$$\mathbf{X}_n = \mathbf{U}_\mathbf{X}\mathbf{S}_{\mathbf{X},n}\mathbf{V}_\mathbf{X}^\mathsf{T}\,, \tag{6.37}$$

where we used $\mathbf{U}_\mathbf{X}^\mathsf{T}\mathbf{U}_\mathbf{X} = \mathbf{I}$. Accordingly, the truncated SVD of $\mathbf{X}^\mathsf{T}\mathbf{X}$ is

$$\left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)_n = \mathbf{V}_\mathbf{X}\mathbf{S}_{\mathbf{X},n}^2\mathbf{V}_\mathbf{X}^\mathsf{T} = \mathbf{X}_n^\mathsf{T}\mathbf{X}_n\,, \tag{6.38}$$

and the truncated version of the projector on X, i.e., the projector on $X_n = \text{range}(\mathbf{X}_n)$ is

$$\begin{aligned}\mathbf{P}_n^X &= \mathbf{X}_n\left(\mathbf{X}_n^\mathsf{T}\mathbf{X}_n\right)^+\mathbf{X}_n^\mathsf{T} = \mathbf{U}_\mathbf{X}\mathbf{S}_{\mathbf{X},n}\mathbf{V}_\mathbf{X}^\mathsf{T}\left(\mathbf{V}_\mathbf{X}\mathbf{S}_{\mathbf{X},n}^2\mathbf{V}_\mathbf{X}^\mathsf{T}\right)^+\mathbf{V}_\mathbf{X}\mathbf{S}_{\mathbf{X},n}\mathbf{U}_\mathbf{X}^\mathsf{T} \\ &= \mathbf{U}_\mathbf{X}\mathbf{S}_{\mathbf{X},n}\mathbf{V}_\mathbf{X}^\mathsf{T}\left(\mathbf{V}_\mathbf{X}\mathbf{S}_{\mathbf{X},n}^{-2}\mathbf{V}_\mathbf{X}^\mathsf{T}\right)\mathbf{V}_\mathbf{X}\mathbf{S}_{\mathbf{X},n}\mathbf{U}_\mathbf{X}^\mathsf{T} = \mathbf{U}_\mathbf{X}\mathbf{I}_n\mathbf{U}_\mathbf{X}^\mathsf{T} \\ &= \mathbf{U}_\mathbf{X}\mathbf{S}_\mathbf{X}\mathbf{V}_\mathbf{X}^\mathsf{T}\left(\mathbf{V}_\mathbf{X}\mathbf{S}_\mathbf{X}^2\mathbf{V}_\mathbf{X}^\mathsf{T}\right)_n^+\mathbf{V}_\mathbf{X}\mathbf{S}_{\mathbf{X},n}\mathbf{U}_\mathbf{X}^\mathsf{T} = \mathbf{X}\left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)_n^+\mathbf{X}^\mathsf{T},\end{aligned} \tag{6.39}$$

where we used $\mathbf{V}_\mathbf{X}^\mathsf{T}\mathbf{V}_\mathbf{X} = \mathbf{I}$, $\mathbf{S}_{\mathbf{X},n}^{-1}\mathbf{S}_\mathbf{X} = \mathbf{S}_{\mathbf{X},n}^{-1}\mathbf{S}_{\mathbf{X},n} = \mathbf{I}_n$ and $\mathbf{S}_\mathbf{X}\mathbf{S}_{\mathbf{X},n}^{-1} = \mathbf{S}_{\mathbf{X},n}\mathbf{S}_{\mathbf{X},n}^{-1} = \mathbf{I}_n$, with the truncated inverse of $\mathbf{S}_\mathbf{X}$ defined as

$$\left[\mathbf{S}_{\mathbf{X},n}^{-1}\right]_{ii} = \begin{cases} 1/s_i & \text{if } i \leq n \\ 0 & \text{otherwise} \end{cases}, \tag{6.40}$$

and the truncated identity defined as

$$\left[\mathbf{I}_n\right]_{ii} = \begin{cases} 1 & \text{if } i \leq n \\ 0 & \text{otherwise} \end{cases}. \tag{6.41}$$

Then, for $m > n$

$$\begin{aligned}\mathbf{P}_m^X\mathbf{P}_n^X &= \mathbf{U}_\mathbf{X}\mathbf{I}_m\mathbf{U}_\mathbf{X}{}^\mathsf{T}\mathbf{U}_\mathbf{X}\mathbf{I}_n\mathbf{U}_\mathbf{X}{}^\mathsf{T} = \mathbf{U}_\mathbf{X}\mathbf{I}_n\mathbf{U}_\mathbf{X}{}^\mathsf{T} \\ &= \mathbf{P}_n^X\mathbf{P}_m^X = \mathbf{P}_n^X.\end{aligned} \tag{6.42}$$

This implies the following property (for $m > n > p > q$)

$$\begin{aligned}\left(\mathbf{P}_m^X - \mathbf{P}_n^X\right)\left(\mathbf{P}_p^X - \mathbf{P}_q^X\right) &= \mathbf{P}_m^X\mathbf{P}_p^X - \mathbf{P}_m^X\mathbf{P}_q^X - \mathbf{P}_n^X\mathbf{P}_p^X + \mathbf{P}_n^X\mathbf{P}_q^X \\ &= \mathbf{P}_p^X - \mathbf{P}_q^X - \mathbf{P}_p^X + \mathbf{P}_q^X = \mathbf{0},\end{aligned} \tag{6.43}$$

as well as (for $m > n$)

$$\begin{aligned}\left(\mathbf{P}_m^X - \mathbf{P}_n^X\right)\left(\mathbf{P}_m^X - \mathbf{P}_n^X\right) &= \mathbf{P}_m^X\mathbf{P}_m^X + \mathbf{P}_n^X\mathbf{P}_n^X - \mathbf{P}_n^X\mathbf{P}_m^X - \mathbf{P}_m^X\mathbf{P}_n^X \\ &= \mathbf{P}_m^X + \mathbf{P}_n^X - 2\mathbf{P}_n^X = \mathbf{P}_m^X - \mathbf{P}_n^X.\end{aligned} \tag{6.44}$$

For a logarithm base α, we define the wavelet-like operators on $\{X_\ell\}_{\ell=1}^{N_{X,\alpha}+1}$, with $N_{X,\alpha} = \lfloor \log_\alpha(r) \rfloor$, as

$$\mathbf{W}_1^X = \mathbf{P}_{\alpha-1}^X, \tag{6.45a}$$

$$\mathbf{W}_\ell^X = \mathbf{P}_{\alpha^\ell-1}^X - \mathbf{P}_{\alpha^{\ell-1}-1}^X, \quad \ell = 2, \ldots, N_{X,\alpha}, \tag{6.45b}$$

$$\mathbf{W}_{N_{X,\alpha}+1}^X = \mathbf{P}^X - \mathbf{P}_{\alpha^{N_{X,\alpha}}-1}^X. \tag{6.45c}$$

These operators are projectors, i.e.,

$$\mathbf{W}_\ell^X\mathbf{W}_\ell^X = \mathbf{W}_\ell^X \quad \forall \ell = 1, \ldots, N_{X,\alpha}+1 \tag{6.46}$$

which is a consequence of (6.44). They are also orthogonal to each other, i.e.,

$$\mathbf{W}_n^X\mathbf{W}_m^X = \mathbf{0}, \quad n \neq m, \tag{6.47}$$

which is a consequence of (6.43). Furthermore, they form by construction a partition of the projector on X, i.e.,

$$\sum_{\ell=1}^{N_{X,\alpha}+1} \mathbf{W}_\ell^X = \mathbf{P}^X \,. \tag{6.48}$$

This defines a wavelet-like partition of the subspace X

$$X = \bigoplus_{\ell=1}^{N_{X,\alpha}+1} W_\ell^X \,, \tag{6.49}$$

with $W_\ell^X = \text{range}\left(\mathbf{W}_\ell^X\right)$.

From (5.7) and (6.9), it can be shown that for any integer m and n

$$\boldsymbol{\Sigma}_n^\mathsf{T} \boldsymbol{\Lambda}_m = \widetilde{\boldsymbol{\Sigma}}_n^\mathsf{T} \widetilde{\boldsymbol{\Lambda}}_m = \mathbf{0} \,, \tag{6.50}$$

which also implies

$$\mathbf{P}_n^\Sigma \mathbf{P}_m^\Lambda = \mathbf{P}_n^{\widetilde{\Sigma}} \mathbf{P}_n^{\widetilde{\Lambda}} = \mathbf{0} \,. \tag{6.51}$$

We additionally get for any integer n

$$\mathbf{P}_n^\Sigma \mathbf{P}^H = \mathbf{P}_n^\Lambda \mathbf{P}^H = \mathbf{P}_n^{\widetilde{\Sigma}} \mathbf{P}^{\widetilde{H}} = \mathbf{P}_n^{\widetilde{\Lambda}} \mathbf{P}^{\widetilde{H}} = \mathbf{0} \,. \tag{6.52}$$

According to these orthogonality properties, the quasi-Helmholtz decomposition (5.1) and its normalized counterpart (6.10) can be further decomposed into wavelet-like quasi-Helmholtz decompositions

$$\mathbb{C}^N = \bigoplus_{\ell=1}^{N_{\Sigma,\alpha}+1} W_\ell^\Sigma \bigoplus_{\ell=1}^{N_{\Lambda,\alpha}+1} W_\ell^\Lambda \bigoplus H \,, \tag{6.53}$$

and

$$\mathbb{C}^N = \bigoplus_{\ell=1}^{N_{\widetilde{\Sigma},\alpha}+1} W_\ell^{\widetilde{\Sigma}} \bigoplus_{\ell=1}^{N_{\widetilde{\Lambda},\alpha}+1} W_\ell^{\widetilde{\Lambda}} \bigoplus \widetilde{H} \,, \tag{6.54}$$

respectively.

6.4.2 Preconditioning the Electric Field Integral Equation

The wavelet-like decompositions enable us to define the quasi-Helmholtz filters [94]

$$\mathbf{Q}^\Lambda = \sum_{\ell=1}^{N_{\Lambda,\alpha}+1} \beta_\ell^\Lambda \mathbf{W}_\ell^\Lambda \,, \tag{6.55}$$

$$\mathbf{Q}^\Sigma = \sum_{\ell=1}^{N_{\Sigma,\alpha}+1} \beta_\ell^\Sigma \mathbf{W}_\ell^\Sigma \,, \tag{6.56}$$

with

$$\beta_\ell^\Lambda = \|\mathbf{W}_\ell^\Lambda \mathbf{T}_s \mathbf{W}_\ell^\Lambda\|^{-1/2} \,, \quad \ell = 1, \ldots, N_{\Lambda,\alpha} + 1 \,, \tag{6.57a}$$

$$\beta_\ell^\Sigma = \|\mathbf{W}_\ell^\Sigma \mathbf{T}_h \mathbf{W}_\ell^\Sigma\|^{-1/2} \,, \quad \ell = 1, \ldots, N_{\Sigma,\alpha} + 1 \,. \tag{6.57b}$$

as well as the normalized quasi-Helmholtz filters

$$\mathbf{Q}^{\widetilde{\Lambda}} = \sum_{\ell=1}^{N_{\widetilde{\Lambda},\alpha}+1} \beta_\ell^{\widetilde{\Lambda}} \mathbf{W}_\ell^{\widetilde{\Lambda}}\,, \tag{6.58}$$

$$\mathbf{Q}^{\widetilde{\Sigma}} = \sum_{\ell=1}^{N_{\widetilde{\Sigma},\alpha}+1} \beta_\ell^{\widetilde{\Sigma}} \mathbf{W}_\ell^{\widetilde{\Sigma}}\,, \tag{6.59}$$

with

$$\beta_\ell^{\widetilde{\Lambda}} = \|\mathbf{W}_\ell^{\widetilde{\Lambda}}\widetilde{\mathbf{T}}_s\mathbf{W}_\ell^{\widetilde{\Lambda}}\|^{-1/2}\,,\quad \ell = 1,\ldots,N_{\widetilde{\Lambda},\alpha}+1\,, \tag{6.60a}$$

$$\beta_\ell^{\widetilde{\Sigma}} = \|\mathbf{W}_\ell^{\widetilde{\Sigma}}\widetilde{\mathbf{T}}_h\mathbf{W}_\ell^{\widetilde{\Sigma}}\|^{-1/2}\,,\quad \ell = 1,\ldots,N_{\widetilde{\Sigma},\alpha}+1\,. \tag{6.60b}$$

The preconditioning matrices are finally given by

$$\mathbf{Q} = \gamma^{\Sigma}\,\mathbf{Q}^{\Sigma} + \gamma^{\Lambda}\,\mathbf{Q}^{\Lambda} + \gamma^{H}\mathbf{P}^{H}\,, \tag{6.61}$$

with

$$\gamma^{\Sigma} = jk^{1/2}\|\mathbf{Q}^{\Sigma}\mathbf{T}_h\mathbf{Q}^{\Sigma}\|^{-1/2} \tag{6.62a}$$

$$\gamma^{\Lambda} = k^{-1/2}\|\mathbf{Q}^{\Lambda}\mathbf{T}_s\mathbf{Q}^{\Lambda}\|^{-1/2}\,, \tag{6.62b}$$

$$\gamma^{H} = k^{-1/2}\|\mathbf{P}^{H}\mathbf{T}_s\mathbf{P}^{H}\|^{-1/2}\,, \tag{6.62c}$$

for the standard EFIE and by

$$\widetilde{\mathbf{Q}} = \gamma^{\widetilde{\Sigma}}\,\mathbf{Q}^{\widetilde{\Sigma}} + \gamma^{\widetilde{\Lambda}}\,\mathbf{Q}^{\widetilde{\Lambda}} + \gamma^{\widetilde{H}}\mathbf{P}^{\widetilde{H}}\,, \tag{6.63}$$

with

$$\gamma^{\widetilde{\Sigma}} = jk^{1/2}\|\mathbf{Q}^{\widetilde{\Sigma}}\widetilde{\mathbf{T}}_h\mathbf{Q}^{\widetilde{\Sigma}}\|^{-1/2}\,, \tag{6.64a}$$

$$\gamma^{\widetilde{\Lambda}} = k^{-1/2}\|\mathbf{Q}^{\widetilde{\Lambda}}\widetilde{\mathbf{T}}_s\mathbf{Q}^{\widetilde{\Lambda}}\|^{-1/2}\,, \tag{6.64b}$$

$$\gamma^{\widetilde{H}} = k^{-1/2}\|\mathbf{P}^{\widetilde{H}}\widetilde{\mathbf{T}}_s\mathbf{P}^{\widetilde{H}}\|^{-1/2}\,, \tag{6.64c}$$

for the normalized EFIE. This yields the preconditioned systems

$$\mathbf{QTQ}\boldsymbol{y} = \mathbf{Q}\boldsymbol{e}\,, \tag{6.65}$$

with $\boldsymbol{j} = \mathbf{Q}\boldsymbol{y}$, and

$$\widetilde{\mathbf{Q}}\widetilde{\mathbf{T}}\widetilde{\mathbf{Q}}\widetilde{\boldsymbol{y}} = \widetilde{\mathbf{Q}}\widetilde{\boldsymbol{e}}\,, \tag{6.66}$$

with $\widetilde{\boldsymbol{j}} = \widetilde{\mathbf{Q}}\widetilde{\boldsymbol{y}}$.

The analysis conducted in [94] demonstrates that

$$\mathrm{cond}\left(\widetilde{\mathbf{Q}}\widetilde{\mathbf{T}}\widetilde{\mathbf{Q}}\right) = O(1),\ h \to 0,\ k \to 0\,, \tag{6.67}$$

which is a consequence of the choice of the scaling factors (6.60) and (6.64) and the convergence results on the eigenvalues of (6.24) and (6.31).

As it concerns the standard EFIE, cond ($\mathbf{QTQ}$) worsens with respect to cond($\widetilde{\mathbf{Q}}\widetilde{\mathbf{T}}\widetilde{\mathbf{Q}}$) because of the ill-conditioning of the Gram matrix $\mathbf{G}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}$ for increasing p and for meshes with non-uniform diameters of the cells. Nevertheless, the condition number can be bounded whenever an efficient strategy is used to precondition the Gram matrix without necessarily relying on the computation of its inverse square root [94].

6.5 Implementation related details

In this section, we discuss the techniques that will be useful to efficiently implement the methods presented here within an iterative solver.

Fast algorithms are assumed to be used for the computation of the matrices, i.e., multipole or compression-based techniques for the integral operator matrix discretizations [40, 112, 119], together with sparse algorithms for the Gram matrices [37]. The action of the linear algebra operations, such as matrix inversions and matrix square roots, are computed on the fly when multiplied with the vectors of the iterative process. They must be done with iterative methods which imply a reduced computational cost (in time an storage) due to the fast algorithms employed to store the matrices and compute their associated matrix-vector products. The matrices $\mathbf{\Lambda}^\mathsf{T}\mathbf{\Lambda}$ and $\mathbf{\Sigma}^\mathsf{T}\mathbf{\Sigma}$ are ill-conditioned for increasing discretization (in h and p) and their pseudo-inversions require suitable algorithms such as multigrid techniques [120, 84, 109, 104, 22]. These pseudo-inverses obtained via multigrid can be used as preconditioners for the conjugate gradient method when pseudo-inverting their normalized counterpart (6.24) and (6.31). Similarly, the evaluation of the spectral norms is efficiently done via power iteration methods that also benefit from matrix-vector product acceleration.

Moreover, the definition of the filters involves the computation of several SVDs, which, while offering a comprehensive theoretical framework, are computationally prohibitive. In [94], alternative filtering algorithms are introduced, and their effectiveness is shown for the computation of the zeroth-order quasi-Helmholtz filters. These algorithms can be used equally for higher orders, and we refer to the aforementioned paper for more details about their descriptions.

Finally, all the products of $\mathbf{T}_h$ / $\widetilde{\mathbf{T}}_h$ with $\mathbf{Q}^\Lambda$ / $\mathbf{Q}^{\widetilde{\Lambda}}$ and $\mathbf{P}^H$ / $\mathbf{P}^{\widetilde{H}}$ must be explicitly set to zero to avoid numerical instabilities due to the loss of significant digits appearing at low frequency. Further treatments on the right-hand side, solution vector, and computation of the scattered field are required to ensure numerical stability at very low frequency, as is commonly done in low-frequency stabilization strategies (see Chapter 5).

6.6 Numerical results

To corroborate our theoretical developments, we first display in Figure 6.1 the singular values of the Laplace-Beltrami operator Δ, for the unitary sphere, together with the approximating eigenvalues of $\widetilde{\mathbf{\Lambda}}^\mathsf{T}\widetilde{\mathbf{\Lambda}}$ and $\widetilde{\mathbf{\Sigma}}^\mathsf{T}\widetilde{\mathbf{\Sigma}}$, obtained for $h = 58$cm and $p = 2$. We observe that there is a good match between these singular values and they all grow linearly with the spectrum index, as expected by the Weyl's law (6.34) and the convergence of the discrete eigenvalues to those of the continuous operator (6.32)–(6.33). This allows to effectively filter these spectra through the use of truncated singular value decompositions. In the same figure, we display the filtered singular value spectrum of the operators with a filtering parameter $\alpha = 2$. We observe that these three spectra are flat and bounded, indicating a condition number almost equal to α. Subsequently, we display in Figure 6.2 the eigenvalues of the two matrices that we filter in the EFIE filtering approach, namely $\mathbf{P}^{\widetilde{\Lambda}}\widetilde{\mathbf{T}}_s\mathbf{P}^{\widetilde{\Lambda}}$ and $\mathbf{P}^{\widetilde{\Sigma}}\widetilde{\mathbf{T}}_h\mathbf{P}^{\widetilde{\Sigma}} = \widetilde{\mathbf{T}}_h$. We observe that the decrease and increase of the singular values, respectively, are linear with respect to the spectrum index, as they inherit the behavior of the associated Laplacian discretizations $\widetilde{\mathbf{\Lambda}}^\mathsf{T}\widetilde{\mathbf{\Lambda}}$ and $\widetilde{\mathbf{\Sigma}}^\mathsf{T}\widetilde{\mathbf{\Sigma}}$. Again, this enables effective filtering of these spectra, that we display in the same figure. Equally, we observe the the filtered spectra are flat and bounded, indicating a condition number almost equal to α.

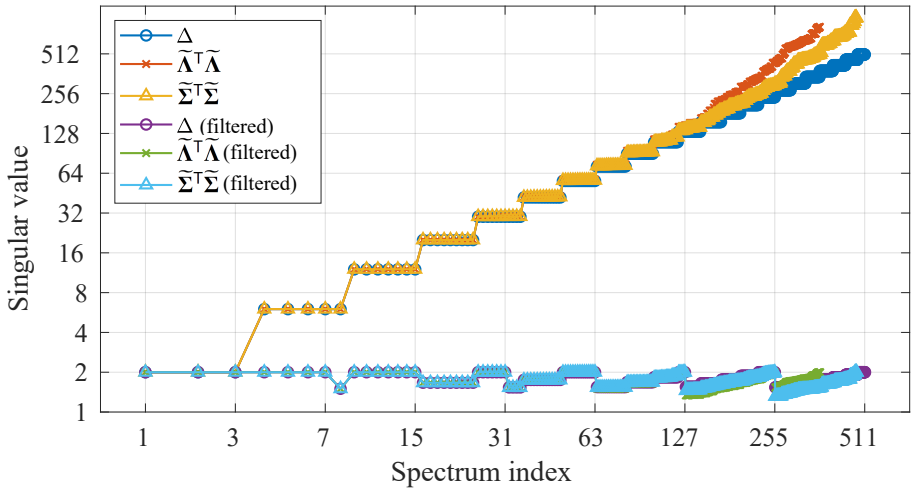


Figure 6.1: Singular value spectra of the continuous and discretized Laplacian operators along with their filtered counterparts for the sphere at order 2 with $h = 58$cm.

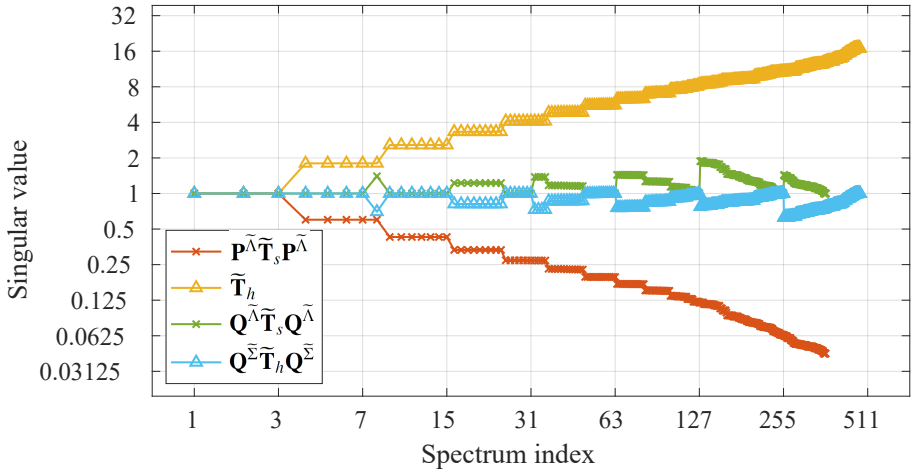


Figure 6.2: Singular value spectra of $\mathbf{P}^{\widetilde{\Lambda}}\widetilde{\mathbf{T}}_s\mathbf{P}^{\widetilde{\Lambda}}$ and $\widetilde{\mathbf{T}}_h$ and of their filtered counterparts for the sphere at order 2 with $h = 58$cm.

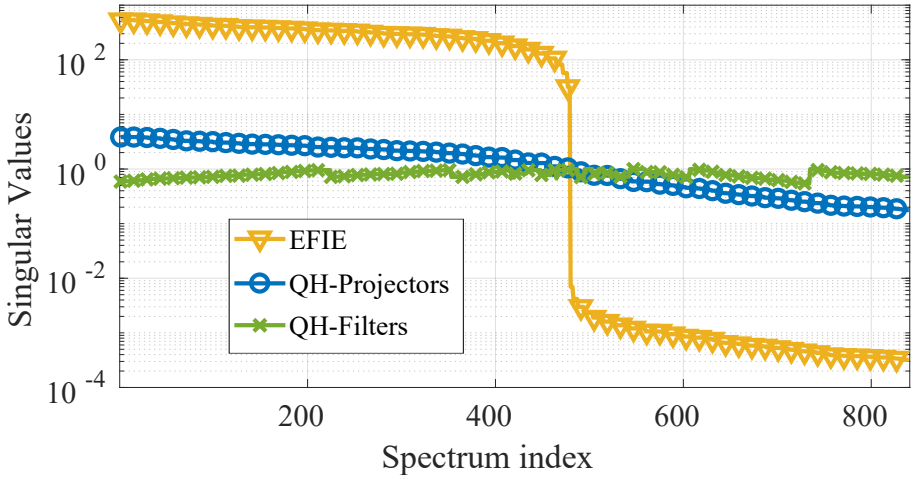


Figure 6.3: Singular value spectra of the EFIE (without preconditioning), projector-based EFIE and filter-based EFIE for the sphere at order 2 with $h = 58$cm. Frequency is $f = 10$kHz.

With the same discretization, and at frequency $f = 10\text{kHz}$, we display in Figure 6.3 the spectra of the EFIE, quasi-Helmholtz projector-based EFIE and quasi-Helmholtz filter-based EFIE (with normalization). We observe that the gap between the maximal and minimal singular values of the EFIE is substantially larger than with the two other formulations. The projector-based EFIE displays a smaller gap than the non-preconditioned EFIE. It notably reduces the distance between the two clusters of singular values associated with non-solenoidal (top-left) and solenoidal (bottom-right) singular values. However, it does not entirely flatten the spectrum. The filter-based EFIE, instead, displays a very flat spectrum, indicating a condition number almost equal to α.

To showcase the performance of our technique, the normalized filter-based EFIE (6.66) is compared with the normalized EFIE (6.1) and the normalized quasi-Helmholtz projector-based EFIE (6.19). These formulations are applied to solve electromagnetic problems on three canonical shapes: a sphere of radius 1 m uniformly meshed (all triangles of the mesh discretization are identical), a torus of outer radius 0.5 m and inner radius 0.15 m and the NASA almond (of size 25.24 cm × 9.75 cm × 3.25 cm) [143].

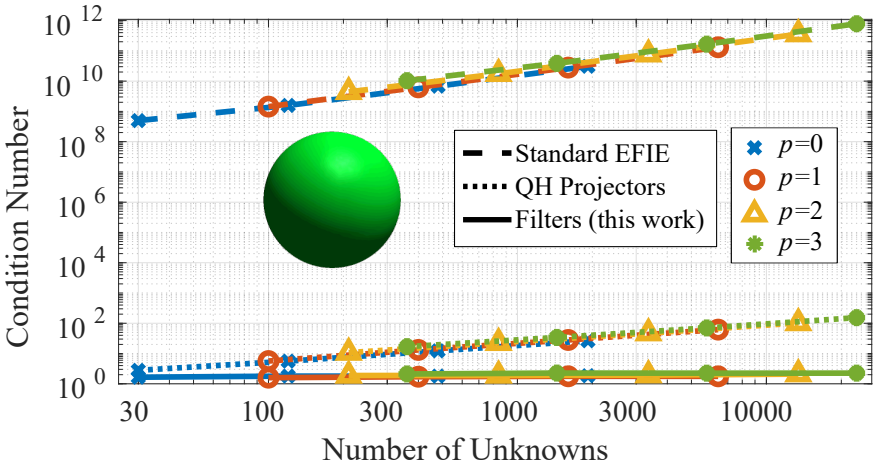


Figure 6.4: Condition number of the different formulations matrices for the Sphere of radius 1m as a function of the number of unknowns (for decreasing h). Frequency is $f = 10\text{kHz}$.

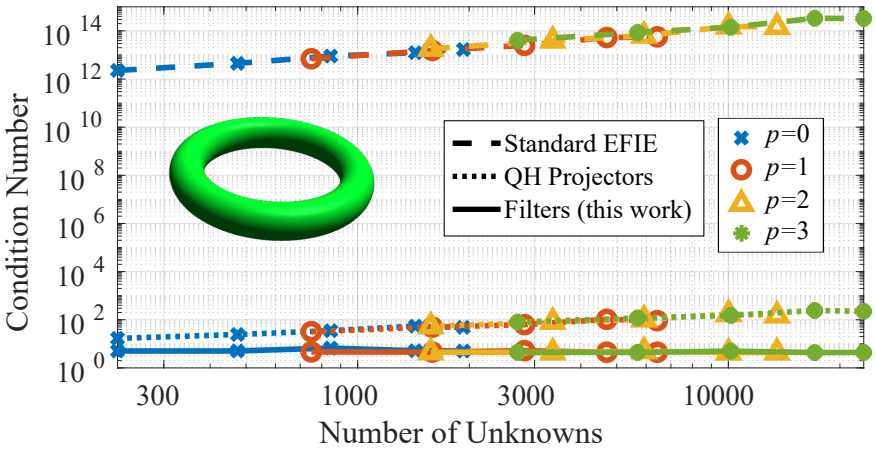


Figure 6.5: Condition number of the different formulations matrices for the Torus of outer radius 50cm and inner radius 15cm as a function of the number of unknowns (for decreasing h). Frequency is $f = 10\text{kHz}$.

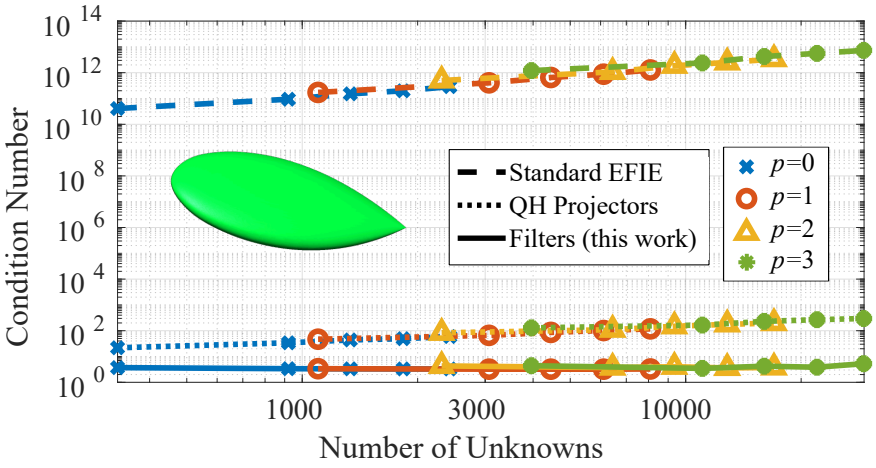


Figure 6.6: Condition number of the different formulations matrices for the NASA almond of dimensions 25.24cm × 9.75cm × 3.25cm as a function of the number of unknowns (for decreasing h). Frequency is f = 10kHz.

Figures 6.4, 6.5 and 6.6 display the condition number of the underlying matrices as a function of the number of unknowns, at fixed frequency f = 10kHz, for the system matrices obtained for simulations with the sphere, torus and NASA almond, respectively. As expected, the matrices obtained with the standard EFIE present an extremely high condition number that increases linearly with the number of unknowns (i.e., as the square of $1/h$), whereas matrices preconditioned with the quasi-Helmholtz projectors present a lower condition number which also increases for decreasing h. The quasi-Helmholtz filtered matrices, instead, present an almost constant condition number.

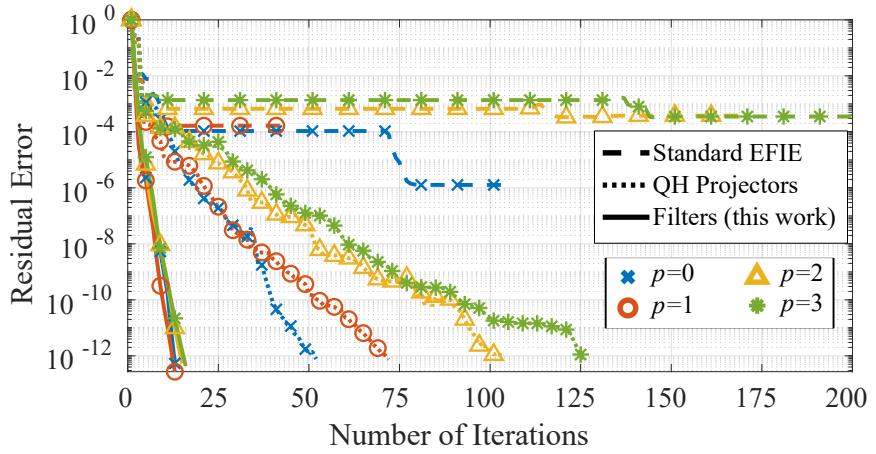


Figure 6.7: Residual error as a function of the number of unknowns for the iterative solver TFQMR for simulations of the unitary sphere discretized with h = 10cm. Frequency is f = 10 kHz.

Further, we display in Figure 6.7 the residual error as a function of the number of unknowns obtained with the iterative solver TFQMR when modelling the unitary sphere at frequency f = 10kHz, with h = 10cm. As it was expected with the computation of the condition number, we observe an almost constant number of iterations at a fixed precision with the quasi-Helmholtz filters preconditioning, while it increases for increasing order (that is with the number of unknowns) with the quasi-Helmholtz projectors preconditioning. At this frequency, TFQMR

does not convergence when solving the non-preconditioned EFIE. These observations are valid for different orders and showcase the effectiveness of our scheme in curing the dense-refinement breakdown of the EFIE and quasi-Helmholtz projectors-based EFIE.

Figures 6.8, 6.9 and 6.10 display the condition number as a function of the frequency, at a fixed mesh parameter h, for the system matrices obtained for simulations with the sphere, torus and NASA almond, respectively. As expected, the condition number of the non-preconditioned EFIE increases quadratically for decreasing frequencies. Instead, projector-based and filter-based preconditioned matrices have stable condition numbers with respect to the frequency, showing that these two techniques are both capable of curing the low-frequency breakdown. Nonetheless, the filter-based preconditioning provides significantly lower condition numbers than the projector-based preconditioning. These observations are valid for different orders and showcase the effectiveness of our scheme in curing the low-frequency breakdown of the EFIE.

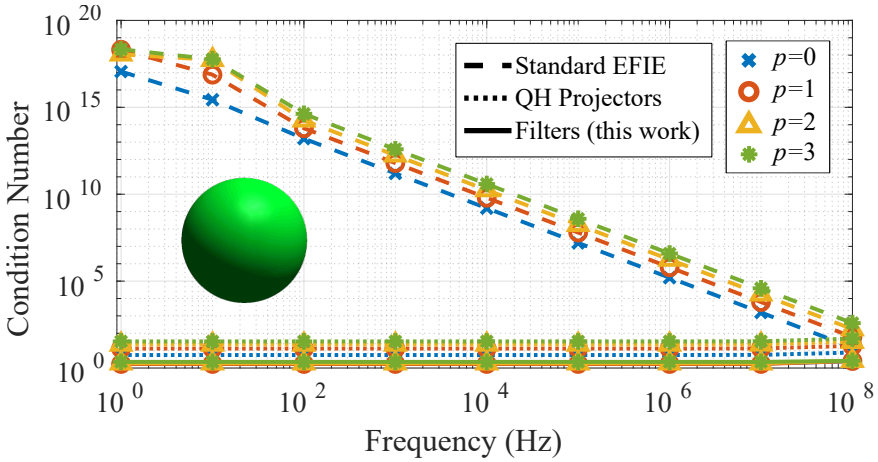


Figure 6.8: Condition number of the different formulations matrices for the sphere of radius 1 m at different frequencies. Average mesh edge size is h = 58cm.

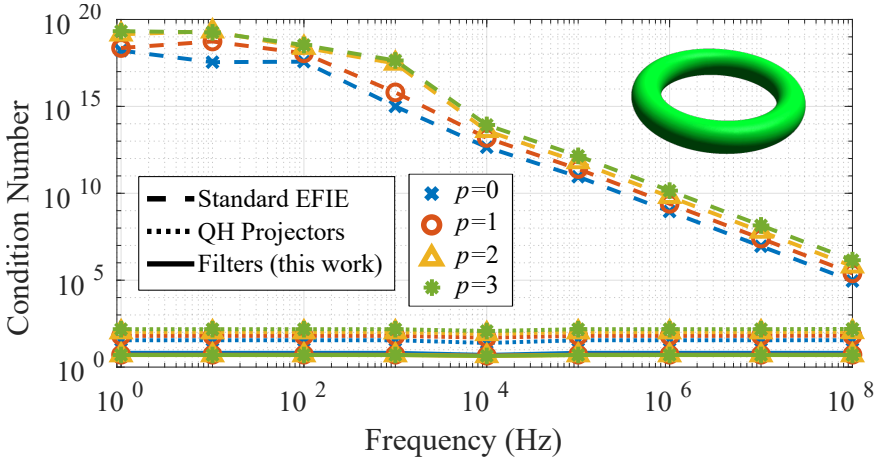


Figure 6.9: Condition number of the different formulations matrices for the torus of outer radius 50cm and inner radius 15cm at different frequencies. Average mesh edge size is h = 8cm.

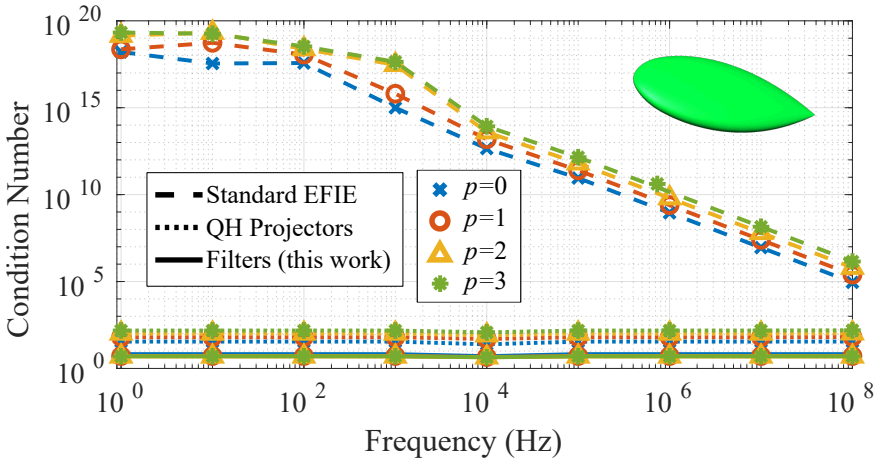


Figure 6.10: Condition number of the different formulations matrices for the NASA almond of dimensions 25.24cm × 9.75cm × 3.25cm at different frequencies. Average mesh edge size is $h = 1.3$cm.

Finally, we display in Figure 6.11 the condition number obtained with the different formulations without normalization (see (2.93), (4.6) and (6.65)). We observe that the filtering technique still allows to get an almost constant condition number with respect to the mesh parameter h, but it is not stable for increasing order of the basis functions. In fact we observed an exponential growth of the condition number for increasing p. These observations indicate however that an absence of normalization can be acceptable for relatively low orders, while an efficient technique to normalize the basis functions becomes necessary for higher orders.

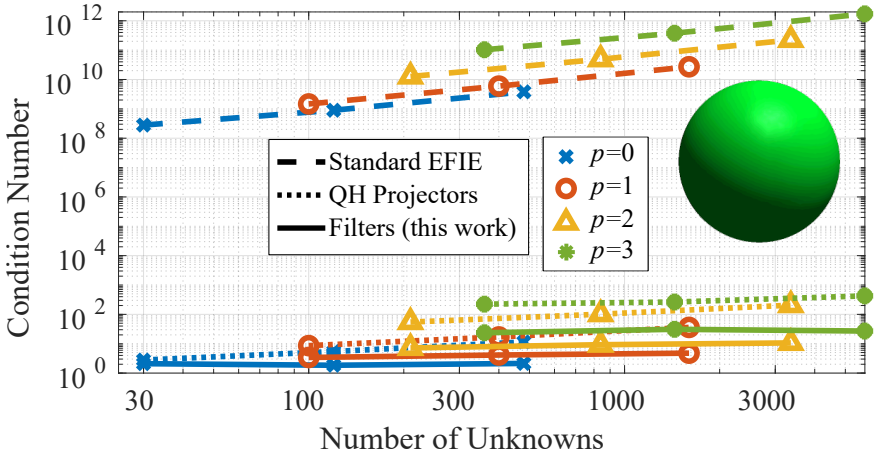


Figure 6.11: Condition number of the different formulations matrices for the Sphere of radius 1 m as a function of the number of unknowns (for decreasing h) without normalization. Frequency is $f = 10$kHz.

6.7 Conclusion

We presented an arbitrary-order preconditioning strategy capable of addressing both h-refinement and low-frequency breakdowns. These issues are met in several industrial applications where mesh densities and object size appear to be substantially smaller than the frequency of the electromagnetic problem and can compromise the precision of the solution in addition to

slow done or even prevent the convergence of iterative solvers. Furthermore, the proposed approach allows the use of basis functions of arbitrary order, leading to faster convergence toward the physical solution as the mesh is refined, which is of considerable industrial significance.

Our technique resorts to the filtering of the quasi-Helmholtz projectors, introduced in Chapter 5. It can be seen as an extension of this previous method, where the EFIE spectrum is further stabilized. Theoretical considerations and experimental validations showcased the effectiveness of the proposed strategy against other techniques. Preconditioning effect with regard to p-refinement has also been observed and is currently under theoretical investigations.

Notice that the computation of inverse square roots of Gram matrices may represent an expensive process and we may look for some alternatives. In the next section, we will also employ these matrices to normalize the EFIE. There, we will introduce an efficient alternative that consists in partially-orthonormalizing the basis functions with sparse transformation matrices. This technique can be used equally in the context of quasi-Helmholtz Laplacian filters. Moreover, we did not explore the use of other kind of basis functions. It is known for instance that hierarchical basis functions, which span the same Raviart-Thomas spaces than the GWP functions, possess some orthogonality between the different levels of basis functions, which results in a smaller conditioning of the associated discretization matrices [108].

Chapter 7

Projector-Based Calderón Preconditioners for High-Order Discretizations

Preliminary results of this work were presented at the 2024 International Conference on Electromagnetics in Advanced Applications [25].

7.1 Introduction

In this chapter, we propose a preconditioning strategy for the EFIE that is based on the use of both quasi-Helmholtz projectors and Calderón identities (7.14). Calderón preconditioners are techniques that consist in mapping the EFIE operator to a second kind Fredholm integral operator—a class of operators that are not prone to dense-refinement breakdowns—by composing it with itself [82, 42]. However, the presence of a mixed-Gram matrix—that contains the inner products between div-conforming and curl-conforming functions—arising from the discretization of the Calderón identity, requires identifying a dual space so as to actually build a well-conditioned system [42, 28, 136].

Relationships between Calderón preconditioners and quasi-Helmholtz decompositions are well-established [28], even in the high-order case [136, 42]. Roughly speaking, the reason why a straightforward discretization of the Calderón identity is not effective relates to the duality between the quasi-Helmholtz decompositions of div-conforming and curl-conforming functions. As a matter of fact, the functions from one space can be obtained by evaluating the cross product between the normal to the surface and the functions from the other space [30]. Hence, the quasi-Helmholtz decomposition of div-conforming functions—which separates them into solenoidal (divergence-free) and quasi-irrotational contributions—transforms into a quasi-Helmholtz decomposition of curl-conforming functions in terms of irrotational (curl-free) and quasi-solenoidal contributions (see Section 2.2.2). Regrettably, when the same finite element space (up to the rotation) is used in both expanding and testing the mixed-Gram matrix, there is a mismatch between the number of solenoidal/quasi-irrotational functions of the expanding space and quasi-solenoidal/irrotational functions of the testing space, which results in an unstable discretization [42, 7]. At the lowest order, it is common to use the barycentric refinement-based Buffa-Christiansen (BC) basis functions [7] to construct a space dual to that of the Rao-Wilton-Glisson (RWG) basis functions [118]; in the sense that the solenoidal and quasi-irrotational RWG-based functions are associated with the vertices and cells of the mesh, respectively, whereas the solenoidal and quasi-irrotational BC-based functions are associated with the cells and vertices, respectively. A comparable technique based on barycentric refinement has also

been proposed at higher order [136]—for instance, when expanding the surface current density with Graglia-Wilton-Peterson (GWP) basis functions [67]. There, a successful strategy to obtain a valid dual space (with a number of solenoidal and quasi-irrotational basis functions that are exchanged with respect to those of the primal space) is to project the rotated solenoidal and quasi-irrotational functions of the primal mesh onto the quasi-irrotational and solenoidal functions of the barycentric-refined mesh, respectively, while the same functional order is set to both spaces.

In the present contribution, we derive and analyze another strategy which leverages p-refinement (that is, when increasing the number of unknowns by increasing the order of the basis functions), rather than mesh-refinement. We show that the normalized version of the high-order quasi-Helmholtz projectors applied to the normalized EFIE (see Section 6.2) allows for properly building a space that is dual to the Raviart-Thomas space of a given order p. Quite similarly to what was done in [136], in the case of barycentric refinement, the rotated solenoidal and quasi-irrotational contributions of the GWP functions of order p are projected onto the quasi-irrotational and solenoidal contributions of the GWP functions of an order higher than p, respectively. Nonetheless, the formulation proposed in this chapter is substantially different than in the latter work, as it does not require the inversion of the mixed-Gram matrix. To this extent, our strategy is closer to the one presented in the seminal work of Christiansen and Nédélec [42], where it is proposed to solve a saddle-point system in order to produce such projection, that preserves the cancellation of the scalar potential applied to itself—another property that is crucial to build effective Calderón preconditioners. In that latter case, however, the dual space is built from the primal space, which introduces a nullspace in the preconditioned system, and no EFIE normalization nor orthogonal projectors are employed.

Following our strategy, we will show experimentally that a proper choice to produce a stable and well-conditioned discretization of the Calderón identity is to use a space dual to the Raviart-Thomas space of order p, that is built from an auxiliary Raviart-Thomas space of order $p+1$ complemented with curl of order $p+1$ bubble functions (a subspace of the Raviart-Thomas space of order $p+2$). Implementation details and algorithmic efficiency of the proposed method—including an alternative to the computation of exact basis functions orthonormalization—will then be discussed, while numerical experiments will showcase stable condition numbers of the new formulation in the dense-refinement and low-frequency regimes.

7.2 Background and Notations

Consider a closed and possibly multiply-connected boundary surface Γ of a PEC object. Given the incident electric field $\boldsymbol{E}^{\text{inc}}$ that impinges on Γ, the EFIE allows finding the induced surface current density $\boldsymbol{J}$ and reads

$$\mathcal{T}\boldsymbol{J} = -\boldsymbol{n} \times \boldsymbol{E}^{\text{inc}}/\eta. \tag{7.1}$$

For a given order p, we approximate the surface current density as

$$\boldsymbol{J} \approx \sum_{n=1}^{N_p} [\boldsymbol{j}_p]_n \boldsymbol{\psi}_n^{(p)}, \tag{7.2}$$

and the EFIE is tested with rotated basis functions $\{\boldsymbol{n} \times \boldsymbol{\psi}^{(p)}\}$ which leads to the matrix system of size N_p

$$\mathbf{T}_p \boldsymbol{j}_p = \boldsymbol{e}_p, \tag{7.3}$$

with

$$\mathbf{T}_p \coloneqq -jk\mathbf{T}_{s|p} + (jk)^{-1}\mathbf{T}_{h|p}\,, \tag{7.4}$$

$$[\mathbf{T}_{s|p}]_{mn} \coloneqq \langle \boldsymbol{n} \times \boldsymbol{\psi}_m^{(p)}, \mathcal{T}_s \boldsymbol{\psi}_n^{(p)} \rangle\,, \tag{7.5}$$

$$[\mathbf{T}_{h|p}]_{mn} \coloneqq \langle \boldsymbol{n} \times \boldsymbol{\psi}_m^{(p)}, \mathcal{T}_h \boldsymbol{\psi}_n^{(p)} \rangle\,, \tag{7.6}$$

$$[\boldsymbol{e}_p]_m \coloneqq -\langle \boldsymbol{n} \times \boldsymbol{\psi}_m^{(p)}, \boldsymbol{n} \times \boldsymbol{E}^{\mathrm{inc}}/\eta \rangle\,. \tag{7.7}$$

We finally consider a normalization of the EFIE that consists in orthonormalizing the basis functions with respect to the inner-product (1.3), that becomes

$$\widetilde{\mathbf{T}}_p \widetilde{\boldsymbol{j}}_p = \widetilde{\boldsymbol{e}}_p\,, \tag{7.8}$$

with

$$\widetilde{\mathbf{T}}_p \coloneqq -jk\widetilde{\mathbf{T}}_{s|p} + (jk)^{-1}\widetilde{\mathbf{T}}_{h|p}\,, \tag{7.9}$$

where

$$\widetilde{\mathbf{T}}_{s|p} \coloneqq \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}} \mathbf{T}_{s|p} \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\,, \tag{7.10}$$

$$\widetilde{\mathbf{T}}_{h|p} \coloneqq \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}} \mathbf{T}_{h|p} \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}\,, \tag{7.11}$$

$$\widetilde{\boldsymbol{j}}_p \coloneqq \mathbf{G}^{1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}} \boldsymbol{j}_p\,, \tag{7.12}$$

$$\widetilde{\boldsymbol{e}}_p \coloneqq \mathbf{G}^{-1/2}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}} \boldsymbol{e}_p\,. \tag{7.13}$$

This normalization allows to mitigate ill-conditioning related to the lack of orthogonality between the basis functions, that substantially deteriorates for non-uniform meshes and high orders [94]. In spite of this normalization, the system may remain ill-conditioned however, and in particular, it is known that the spectrum of the EFIE is unbounded and its matrix discretization (even with normalization) has eigenvalues that go to zero and to infinity (i) for fixed frequency and increasing discretization and (ii) fixed discretization and decreasing frequency [94]. Additional preconditioning techniques are thus needed to improve the effectiveness of numerical solvers in such scenarios.

The preconditioning strategy employed in this work leverages the Calderón identity (1.120)

$$\mathcal{T} \circ \mathcal{T} = -\tfrac{1}{4}\mathcal{I} + \mathcal{K} \circ \mathcal{K}\,, \tag{7.14}$$

where $\mathcal{I}$ is the identity operator and $\mathcal{K}$ is the double layer potential (1.100). For sufficiently smooth geometries, $\mathcal{K}$ is compact and it follows that the eigenvalues of $\mathcal{T} \circ \mathcal{T}$ are bounded and clustered around $-1/4$ [82].

Furthermore, we can expand (7.14) as

$$\mathcal{T} \circ \mathcal{T} = -k^2 \mathcal{T}_s \circ \mathcal{T}_s - \mathcal{T}_h \circ \mathcal{T}_s - \mathcal{T}_s \circ \mathcal{T}_h - \frac{1}{k^2}\mathcal{T}_h \circ \mathcal{T}_h\,, \tag{7.15}$$

and the property $\mathrm{div}_\Gamma(\boldsymbol{n} \times \mathbf{grad}_\Gamma) = -\mathrm{div}_\Gamma \mathbf{curl}_\Gamma = 0$ shows that

$$\mathcal{T}_h \circ \mathcal{T}_h = 0\,. \tag{7.16}$$

The cancellation of (7.16) is crucial to ensure the boundedness of the $\mathcal{T} \circ \mathcal{T}$ spectrum [7] at low-frequency, as the operator (7.14) thus tends towards the static limit

$$-\mathcal{T}_h \circ \mathcal{T}_s - \mathcal{T}_s \circ \mathcal{T}_h \tag{7.17}$$

when k goes to zero, which has a bounded spectrum for simply-connected geometries [8]. Therefore, any effective preconditioner based on the Calderón identity must preserve that cancellation.

For multiply connected geometries, however, it can be shown that (7.17) has a nullspace in statics that is spanned by the harmonic functions [8] (associated with the topological cycles around the handles of the structure). Following the developments in [8], further treatments will thus be introduced to stabilize the Calderón identity at low-frequency when topological cycles arise from multiply connected geometries.

7.2.1 Extended Raviart-Thomas Space

In the following strategy we will need to complement the Raviart-Thomas space of order p by additional cell-based solenoidal functions. We use here the space $\mathbb{B}_p$ of bubble functions of order p, that are the continuous functions of order p which vanish on the border of every cells of the mesh (see Section 2.2). We define these extended Raviart-Thomas spaces as

$$\mathbb{RT}_{p\oplus b} \coloneqq \mathbb{RT}_p \cup \mathbf{curl}_\Gamma \mathbb{B}_p \,, \tag{7.18}$$

and we have $\mathbb{RT}_p \subseteq \mathbb{RT}_{p\oplus b} \subseteq \mathbb{RT}_{p+1}$. Notice that $p(p-1)C/2$ basis functions in $\mathbf{curl}_\Gamma \mathbb{B}_p$ are already included in the set of basis functions of $\mathbb{RT}_p$ [141]. It turns out that only pC basis functions are actually added to those of $\mathbb{RT}_p$ in order to span $\mathbb{RT}_{p\oplus b}$. Hence, this space is of dimension

$$N_{p\oplus b} = N_p + pC \,. \tag{7.19}$$

When the EFIE is discretized with the extended Raviart-Thomas space, pC additional solenoidal basis functions $\{\mathbf{curl}_\Gamma \beta_n^{(p)}\}_{n=1}^{pC}$ are added to the set of GWP functions $\{\boldsymbol{\psi}_n^{(p)}\}_{n=1}^{N_p}$ in order to obtain the basis functions $\{\boldsymbol{\psi}_n^{(p\oplus b)}\}_{n=1}^{N_{p\oplus b}}$. The space of discontinuous functions is not augmented, i.e., $\mathbb{DG}_{p\oplus b} \coloneqq \mathbb{DG}_p$, while pC bubble basis functions $\{\beta_n^{(p)}\}$ are added to the set of continuous basis functions, i.e., $\mathbb{CG}_{p\oplus b} \coloneqq \mathbb{CG}_p \cup \mathbb{B}_p$, and we have $\mathbb{CG}_p \subseteq \mathbb{CG}_{p\oplus b} \subseteq \mathbb{CG}_{p+1}$ with basis functions $\{\lambda_n^{(p\oplus b)}\}_{n=1}^{L_{p\oplus b}}$, where

$$L_{p\oplus b} = L_p + pC \,. \tag{7.20}$$

Assuming an ordering of the basis functions that places at first the GWP functions and thereafter the curl of bubble functions, the quasi-Helmholtz transformation matrices become

$$\boldsymbol{\Sigma}_{p\oplus b} = \begin{pmatrix} \boldsymbol{\Sigma}_p \\ \mathbf{0} \end{pmatrix}, \quad \boldsymbol{\Lambda}_{p\oplus b} = \begin{pmatrix} \boldsymbol{\Lambda}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{7.21}$$

and the Gram matrix associated with the augmented bases are

$$\mathbf{G}_{\boldsymbol{\psi}^{(p\oplus b)},\boldsymbol{\psi}^{(p\oplus b)}} = \begin{pmatrix} \mathbf{G}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}} & \mathbf{G}_{\boldsymbol{\psi}^{(p)},\mathbf{curl}_\Gamma\beta^{(p)}} \\ \mathbf{G}_{\mathbf{curl}_\Gamma\beta^{(p)},\boldsymbol{\psi}^{(p)}} & \mathbf{G}_{\mathbf{curl}_\Gamma\beta^{(p)},\mathbf{curl}_\Gamma\beta^{(p)}} \end{pmatrix}, \tag{7.22}$$

$$\mathbf{G}_{\lambda^{(p\oplus b)},\lambda^{(p\oplus b)}} = \begin{pmatrix} \mathbf{G}_{\lambda^{(p)},\lambda^{(p)}} & \mathbf{G}_{\lambda^{(p)},\beta^{(p)}} \\ \mathbf{G}_{\beta^{(p)},\lambda^{(p)}} & \mathbf{G}_{\beta^{(p)},\beta^{(p)}} \end{pmatrix}, \tag{7.23}$$

$$\mathbf{G}_{\sigma^{(p\oplus b)},\sigma^{(p\oplus b)}} = \mathbf{G}_{\sigma^{(p)},\sigma^{(p)}} \,, \tag{7.24}$$

which allow to compute the normalized quasi-Helmholtz transformation matrices (6.7) and (6.8), and the normalized projectors as in (6.14), (6.15) and (6.16). Further, the properties (6.17)–(6.18) are preserved when the EFIE is discretized with the same ordering of $\mathbb{RT}_{p\oplus b}$ basis functions and normalized with the inverse square-root of (7.22).

7.3 The Novel Calderón Preconditioner for High-Order Discretization of the EFIE

7.3.1 Proposed Formulation

In this contribution, we propose a discretization of $\mathcal{T} \circ \mathcal{T}$ in the form

$$-\widetilde{\mathbf{\Theta}}_{r,p}^{\mathsf{T}} \widetilde{\mathbf{T}}_r \widetilde{\mathbf{\Theta}}_{r,p} \widetilde{\mathbf{T}}_p \,, \tag{7.25}$$

where r is the order of an auxiliary Raviart-Thomas space that has to be properly chosen with respect to the space of order p and where $\widetilde{\mathbf{\Theta}}_{r,p}$ is a (normalized) discretization of the operator

$$\Theta\colon \begin{array}{ccc} \mathrm{H}_{\mathrm{div}_\Gamma}^{-1/2} & \longrightarrow & \mathrm{H}_{\mathrm{curl}_\Gamma}^{-1/2} \\ \boldsymbol{\psi} & \longmapsto & \boldsymbol{n} \times \boldsymbol{\psi} \end{array} . \tag{7.26}$$

The rationale behind this expression is that the testing functions in $\widetilde{\mathbf{T}}_p$ and $\widetilde{\mathbf{T}}_r$, that are curl-conforming, need to be rotated in order to match with the entries of the EFIE discretization, that are coefficients of div-conforming functions. Similar developments have been made in [42] and we refer to this reading for complimentary justifications.

This discretization leads us to the Calderón preconditioning of the normalized system (7.8)

$$\widetilde{\mathbb{T}}_{r,p} \widetilde{\mathbf{T}}_p \widetilde{\boldsymbol{j}}_p = \widetilde{\mathbb{T}}_{r,p} \widetilde{\boldsymbol{e}}_p \,, \tag{7.27}$$

with the preconditioner

$$\widetilde{\mathbb{T}}_{r,p} = -\widetilde{\mathbf{\Theta}}_{r,p}^{\mathsf{T}} \widetilde{\mathbf{T}}_r \widetilde{\mathbf{\Theta}}_{r,p} \,. \tag{7.28}$$

In the following, we investigate the construction of $\widetilde{\mathbf{\Theta}}_{r,p}$ in order to build an effective preconditioner for the normalized system. In particular, we look for a matrix that will ensure the following properties: (i) it has to preserve the cancellation of $\mathcal{T}_h \circ \mathcal{T}_h$ after discretization, i.e.,

$$\widetilde{\mathbf{T}}_{h|r} \widetilde{\mathbf{\Theta}}_{r,p} \widetilde{\mathbf{T}}_{h|p} = \mathbf{0} \,, \tag{7.29}$$

and (ii) it has to be a stable discretization of $\Theta : \boldsymbol{\psi} \mapsto \boldsymbol{n} \times \boldsymbol{\psi}$, in the sense that $\widetilde{\mathbf{\Theta}}_{r,p}$ is a matrix with a bounded spectrum.

These two properties will lead to a stable discretization of the Calderón identity and, as a consequence, the preconditioned system (7.27) will have a bounded spectrum for dense discretizations and at low-frequency in the case of simply connected geometries [42, 79]. The case of multiply connected geometries will require a slight modification of the preconditioned formulation (7.27) that will be presented thereafter.

7.3.2 Discretization of the Surface Rotation Operator

The boundary element discretization of the operator $\Theta : \boldsymbol{\psi} \mapsto \boldsymbol{n} \times \boldsymbol{\psi}$ (with orthonormalization of the basis functions) leads to the definition of the normalized mixed-Gram matrix

$$\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}} := \mathbf{G}_{\boldsymbol{\psi}^{(r)},\boldsymbol{\psi}^{(r)}}^{-1/2} \mathbf{G}_{\boldsymbol{\psi}^{(r)},\boldsymbol{n}\times\boldsymbol{\psi}^{(p)}} \mathbf{G}_{\boldsymbol{\psi}^{(p)},\boldsymbol{\psi}^{(p)}}^{-1/2} \,. \tag{7.30}$$

First, it is important to notice that

$$\mathbf{P}_r^{\widetilde{\Lambda}}\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\mathbf{P}_p^{\widetilde{\Lambda}} = \mathbf{P}_r^{\widetilde{\Lambda}}\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\mathbf{P}_p^{\widetilde{H}} = \mathbf{P}_r^{\widetilde{H}}\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\mathbf{P}_p^{\widetilde{\Lambda}} = \mathbf{0}\,, \tag{7.31}$$

which results from the property

$$\begin{aligned}\langle \boldsymbol{f}, \boldsymbol{n}\times\mathbf{curl}_\Gamma\lambda\rangle &= \langle \boldsymbol{f}, \mathbf{grad}_\Gamma\lambda\rangle \\ &= -\langle \mathrm{div}_\Gamma \boldsymbol{f}, \lambda\rangle = 0\,,\end{aligned} \tag{7.32}$$

for any solenoidal functions $\mathbf{curl}_\Gamma\lambda$, with continuous functions λ—associated with the coefficients in the columns of $\mathbf{P}_p^{\widetilde{\Lambda}}$ / $\mathbf{P}_r^{\widetilde{\Lambda}}$—and any divergence-free functions $\boldsymbol{f}$—associated with the coefficients in the columns of $\mathbf{P}_p^{\widetilde{\Lambda}}$ / $\mathbf{P}_r^{\widetilde{\Lambda}}$ and $\mathbf{P}_p^{\widetilde{H}}$ / $\mathbf{P}_r^{\widetilde{H}}$. Another property that we would like to preserve from the continuous case is

$$\begin{aligned}\langle \mathbf{grad}_\Gamma\gamma, \boldsymbol{n}\times\mathbf{grad}_\Gamma\phi\rangle &= -\langle \mathbf{grad}_\Gamma\gamma, \mathbf{curl}_\Gamma\phi\rangle \\ &= \langle \gamma, \mathrm{div}_\Gamma\mathbf{curl}_\Gamma\phi\rangle = 0\,,\end{aligned} \tag{7.33}$$

for any continuous functions ϕ and γ such that $\mathbf{grad}_\Gamma\gamma$ and $\mathbf{grad}_\Gamma\phi$ are irrotational functions. Because div-conforming spaces do not include irrotational functions, (7.33) does not have a discretized counterpart, and, as a matter of fact, the cancellation does not hold with quasi-irrotational functions (that are weak gradients of discontinuous functions)

$$\mathbf{P}_r^{\widetilde{\Sigma}}\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\mathbf{P}_p^{\widetilde{\Sigma}} \neq 0\,. \tag{7.34}$$

It turns out that the discretization of $\mathcal{T}_h \circ \mathcal{T}_h$ (7.16) using the mixed-Gram matrix (7.30) for the discretization of $\Theta : \boldsymbol{\psi} \mapsto \boldsymbol{n}\times\boldsymbol{\psi}$ is not zero, according to the following derivation

$$\begin{aligned}\widetilde{\mathbf{T}}_{h|r}\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\widetilde{\mathbf{T}}_{h|p} &= \widetilde{\mathbf{T}}_{h|r}\mathbf{P}_r^{\widetilde{\Sigma}}\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\mathbf{P}_p^{\widetilde{\Sigma}}\widetilde{\mathbf{T}}_{h|p} + \widetilde{\mathbf{T}}_{h|r}\mathbf{P}_r^{\widetilde{\Sigma}}\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\left(\mathbf{I}-\mathbf{P}_p^{\widetilde{\Sigma}}\right)\widetilde{\mathbf{T}}_{h|p} \\ &\quad + \widetilde{\mathbf{T}}_{h|r}\left(\mathbf{I}-\mathbf{P}_r^{\widetilde{\Sigma}}\right)\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\mathbf{P}_p^{\widetilde{\Sigma}}\widetilde{\mathbf{T}}_{h|p} + \widetilde{\mathbf{T}}_{h|r}\left(\mathbf{I}-\mathbf{P}_r^{\widetilde{\Sigma}}\right)\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\left(\mathbf{I}-\mathbf{P}_p^{\widetilde{\Sigma}}\right)\widetilde{\mathbf{T}}_{h|p} \\ &= \widetilde{\mathbf{T}}_{h|r}\mathbf{P}_r^{\widetilde{\Sigma}}\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\mathbf{P}_p^{\widetilde{\Sigma}}\widetilde{\mathbf{T}}_{h|p} \neq 0\,,\end{aligned} \tag{7.35}$$

where the cancellation of the three last terms in (7.35) follows from (6.17).

Instead, let us define the following modified discretization of $\Theta : \boldsymbol{\psi} \mapsto \boldsymbol{n}\times\boldsymbol{\psi}$

$$\widetilde{\boldsymbol{\Theta}}_{r,p} \coloneqq \widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}} - \mathbf{P}_r^{\widetilde{\Sigma}}\widetilde{\mathbf{G}}_{r,p}^{\mathrm{mix}}\mathbf{P}_p^{\widetilde{\Sigma}}\,, \tag{7.36}$$

that can be seen as a correction of (7.30) in order to impose the discretized counterpart of (7.33). Indeed, we now have by construction

$$\mathbf{P}_r^{\widetilde{\Sigma}}\widetilde{\boldsymbol{\Theta}}_{r,p}\mathbf{P}_p^{\widetilde{\Sigma}} = \mathbf{0}\,, \tag{7.37}$$

while the discretized counterpart of (7.32) is still preserved

$$\mathbf{P}_r^{\widetilde{\Lambda}}\widetilde{\boldsymbol{\Theta}}_{r,p}\mathbf{P}_p^{\widetilde{\Lambda}} = \mathbf{P}_r^{\widetilde{\Lambda}}\widetilde{\boldsymbol{\Theta}}_{r,p}\mathbf{P}_p^{\widetilde{H}} = \mathbf{P}_r^{\widetilde{H}}\widetilde{\boldsymbol{\Theta}}_{r,p}\mathbf{P}_p^{\widetilde{\Lambda}} = \mathbf{0}\,, \tag{7.38}$$

as a consequence of (6.17) and (7.31). We thus obtain the first desired property (7.29) when using (7.36) as a modified discretization of $\Theta : \boldsymbol{\psi} \mapsto \boldsymbol{n}\times\boldsymbol{\psi}$.

Expanding (7.36) with the projectors (i.e., by multiplying with $\mathbf{I} = \mathbf{P}_r^{\widetilde{\Sigma}} + \mathbf{P}_r^{\widetilde{\Lambda}} + \mathbf{P}_r^{\widetilde{H}}$ on the left and on the right) and imposing the cancellations (6.17), (7.37) and (7.38) give us another expression of the modified mixed-Gram matrix

$$\widetilde{\mathbf{\Theta}}_{r,p} = \left(\mathbf{P}_r^{\widetilde{\Sigma}} + \mathbf{P}_r^{\widetilde{H}}\right) \widetilde{\mathbf{G}}_{r,p}^{\text{mix}} \left(\mathbf{P}_p^{\widetilde{\Lambda}} + \mathbf{P}_p^{\widetilde{H}}\right) + \left(\mathbf{P}_r^{\widetilde{\Lambda}} + \mathbf{P}_r^{\widetilde{H}}\right) \widetilde{\mathbf{G}}_{r,p}^{\text{mix}} \mathbf{P}_p^{\widetilde{\Sigma}} \tag{7.39}$$

Specifically, when Γ is simply connected, $\mathbf{P}_p^{\widetilde{H}} = \mathbf{0}$ and $\mathbf{P}_r^{\widetilde{H}} = \mathbf{0}$, so that $\widetilde{\mathbf{\Theta}}_{r,p}$ can be characterized as the transformation matrix that maps coefficients of solenoidal functions of order p onto coefficients of quasi-irrotational functions of order r and, vice-versa, coefficients of quasi-irrotational functions of order p onto coefficients of solenoidal functions of order r. This defines a mapping of the order r basis functions into a set of dual functions; in the sense that—if $\widetilde{\mathbf{\Theta}}_{r,p}$ is full rank—the dual space is of same dimension than the primal space of order p and the numbers of solenoidal and quasi-irrotational functions are exchanged with those of the primal space [136]. However, if the dimension of the auxiliary space of order r is not high-enough with respect to the dimension of the primal space of order p, this transformation matrix may be singular, or introduce an ill-conditioning of the preconditioned formulation (7.27). In Section 7.5, we will present numerical experiments to showcase that $r \geq (p+1) \oplus b$ provides a well-conditioned formulation. Theoretical developments are still under investigation to corroborate these observations.

7.3.3 Suppression of the Static Nullspace Associated with the Quasi-Harmonic Subspace

It is known that the integral operator $\mathcal{T} \circ \mathcal{T}$ has a static nullspace associated with the harmonic functions, which comes from the static nullspace of both interior and exterior magnetic field integral operators [48]. It turns out that there are the same number of eigenvalues than the number of topological loops that tend towards zero as $O(k^2)$ [8]. This results in an ill-conditioning of the preconditioned formulation (7.27) at low-frequency.

To remove the static nullspace, we can stabilize the two EFIE discretizations $\widetilde{\mathbf{T}}_p$ and $\widetilde{\mathbf{T}}_r$ at low-frequency, with techniques similar to the usual low-frequency preconditioning of the EFIE leveraging the quasi-Helmholtz projectors [2]. This approach is similar to the one proposed in [8] and it allows to remove the overall dependency of the preconditioned formulation in the low-frequency regime. We thus propose to define the matrices (for $\ell = p, r$)

$$\widetilde{\mathbf{P}}_\ell^k = k^{-1} \left(\mathbf{I} - \mathbf{P}_\ell^{\widetilde{\Sigma}}\right) + k \mathbf{P}_\ell^{\widetilde{\Sigma}} \tag{7.40}$$

so that we have from (6.17) and (6.18)

$$\widetilde{\mathbf{P}}_p^k \widetilde{\mathbf{T}}_p = -j \left(\mathbf{I} - \mathbf{P}_p^{\widetilde{\Sigma}}\right) \widetilde{\mathbf{T}}_{s|p} - j \widetilde{\mathbf{T}}_{h|p} - j k^2 \mathbf{P}_p^{\widetilde{\Sigma}} \widetilde{\mathbf{T}}_{s|p} \,, \tag{7.41}$$

$$\widetilde{\mathbf{T}}_r \widetilde{\mathbf{P}}_r^k = -j \widetilde{\mathbf{T}}_{s|r} \left(\mathbf{I} - \mathbf{P}_r^{\widetilde{\Sigma}}\right) - j \widetilde{\mathbf{T}}_{h|r} - j k^2 \widetilde{\mathbf{T}}_{s|r} \mathbf{P}_r^{\widetilde{\Sigma}} \,, \tag{7.42}$$

which tend towards frequency-independent and full-rank matrices in the low-frequency limit. This, in turns, shows that the preconditioned system

$$\widetilde{\mathbb{T}}_{r,p}^k \widetilde{\mathbf{T}}_p^k \widetilde{\boldsymbol{j}}_p = \widetilde{\mathbb{T}}_{r,p}^k \widetilde{\boldsymbol{e}}_p \,, \tag{7.43}$$

with the rescaled EFIE matrix

$$\widetilde{\mathbf{T}}_p^k := \widetilde{\mathbf{P}}_p^k \widetilde{\mathbf{T}}_p \,, \tag{7.44}$$

and the rescaled preconditioner

$$\widetilde{\mathbb{T}}^k_{r,p} := -\widetilde{\boldsymbol{\Theta}}^{\mathsf{T}}_{r,p}\widetilde{\mathbf{T}}_r\widetilde{\mathbf{P}}^k_r\widetilde{\boldsymbol{\Theta}}_{r,p}\,, \tag{7.45}$$

is independent of the frequency in the low-frequency limit, as no frequency-dependent terms are present in $\widetilde{\boldsymbol{\Theta}}_{r,p}$.

Moreoever, the cancellation of the $\mathcal{T}_h \circ \mathcal{T}_h$ discretization is preserved under the proposed rescaling

$$\left(\widetilde{\mathbf{T}}_{h|r}\widetilde{\mathbf{P}}^k_r\right)\widetilde{\boldsymbol{\Theta}}_{r,p}\left(\widetilde{\mathbf{P}}^k_p\widetilde{\mathbf{T}}_{h|p}\right) = \mathbf{0}\,, \tag{7.46}$$

indicating that the preconditioning is also effective for dense discretizations. Another justification of this assertion is that no dependency of the spectrum with respect to the discretization is introduced by the matrices (7.40) as projectors have eigenvalues equal to one or zero.

Additionally, we can notice that

$$\begin{aligned}\widetilde{\mathbf{P}}^k_r\widetilde{\boldsymbol{\Theta}}_{r,p}\widetilde{\mathbf{P}}^k_p &= \mathbf{P}^{\widetilde{\Sigma}}_r\widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_p\right) + \left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_r\right)\widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\mathbf{P}^{\widetilde{\Sigma}}_p + k^{-2}\left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_r\right)\widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_p\right)\\ &= \widetilde{\boldsymbol{\Theta}}_{r,p} + (k^{-2}-1)\left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_r\right)\widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_p\right)\,,\end{aligned} \tag{7.47}$$

has a singularity that is associated with the quasi-harmonic functions, as we can show from (7.31) that

$$\left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_r\right)\widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_p\right) = \mathbf{P}^{\widetilde{H}}_r\widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\mathbf{P}^{\widetilde{H}}_p\,. \tag{7.48}$$

7.4 Implementation Related Details

7.4.1 Enforcing the Numerical Stability at Low-Frequency

At low-frequency, important instabilities may arise because of finite precision and round-off errors in the computation of the matrices, and re-expressions of the preconditioned equations need to be established to remove terms going to infinity as the frequency goes to zero. These terms are implicitly canceling out in the original formulations (7.27) and (7.43), but they are not truly removed within numerical finite precision. Making these cancellations explicit will provide actual numerical stability of the proposed preconditioning strategy at low-frequency.

From (7.39), we define the matrices

$$\widetilde{\mathbf{A}}_{r,p} := \left(\mathbf{P}^{\widetilde{\Sigma}}_r + \mathbf{P}^{\widetilde{H}}_r\right)\widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\left(\mathbf{P}^{\widetilde{\Lambda}}_p + \mathbf{P}^{\widetilde{H}}_p\right)\,, \tag{7.49}$$

$$\widetilde{\mathbf{B}}_{r,p} := \left(\mathbf{P}^{\widetilde{\Lambda}}_r + \mathbf{P}^{\widetilde{H}}_r\right)\widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\mathbf{P}^{\widetilde{\Sigma}}_p\,, \tag{7.50}$$

such that $\widetilde{\boldsymbol{\Theta}}_{r,p} = \widetilde{\mathbf{A}}_{r,p} + \widetilde{\mathbf{B}}_{r,p}$. In order to avoid computing the projectors on the solenoidal and quasi-harmonic subspaces—which, we recall, may be cumbersome and computationally expensive—(6.16) and (7.31) are used to rewrite

$$\widetilde{\mathbf{A}}_{r,p} = \widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_p\right)\,, \tag{7.51}$$

$$\widetilde{\mathbf{B}}_{r,p} = \left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_r\right)\widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\mathbf{P}^{\widetilde{\Sigma}}_p\,. \tag{7.52}$$

We also define

$$\widetilde{\mathbf{A}}^k_{r,p} := \left(\mathbf{I}-(1-k^2)\mathbf{P}^{\widetilde{\Sigma}}_r\right)\widetilde{\mathbf{G}}^{\text{mix}}_{r,p}\left(\mathbf{I}-\mathbf{P}^{\widetilde{\Sigma}}_p\right)\,. \tag{7.53}$$

such that $\widetilde{\mathbf{P}}_r^k \widetilde{\mathbf{\Theta}}_{r,p} \widetilde{\mathbf{P}}_p^k = k^{-2} \widetilde{\mathbf{A}}_{r,p}^k + \widetilde{\mathbf{B}}_{r,p}$.

We can now use (7.37) and (7.38) to rewrite the preconditioned systems (7.27) and (7.43) as

$$\widetilde{\mathbb{T}}_{r,p} \widetilde{\mathbf{T}}_p = k^2 \widetilde{\mathbf{\Theta}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{s|r} \widetilde{\mathbf{\Theta}}_{r,p} \widetilde{\mathbf{T}}_{s|p} + \widetilde{\mathbf{A}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{h|r} \widetilde{\mathbf{A}}_{r,p} \widetilde{\mathbf{T}}_{s|p} + \widetilde{\mathbf{\Theta}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{s|r} \widetilde{\mathbf{B}}_{r,p} \widetilde{\mathbf{T}}_{h|p} \,, \tag{7.54}$$

$$\widetilde{\mathbb{T}}_{r,p}^k \widetilde{\mathbf{T}}_p = \widetilde{\mathbf{\Theta}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{s|r} \left(\widetilde{\mathbf{A}}_{r,p}^k + k^2 \widetilde{\mathbf{B}}_{r,p} \right) \widetilde{\mathbf{T}}_{s|p} + \widetilde{\mathbf{A}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{h|r} \widetilde{\mathbf{A}}_{r,p} \widetilde{\mathbf{T}}_{s|p} + \widetilde{\mathbf{\Theta}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{s|r} \widetilde{\mathbf{B}}_{r,p} \widetilde{\mathbf{T}}_{h|p} \,, \tag{7.55}$$

respectively. These expressions do not suffer from numerical instability at low-frequency and should therefore be used when going to that regime.

Additionally, subtraction of the static (divergence-free) part of the righ-hand side vector $\widetilde{\boldsymbol{e}}_p$ needs to be carried out to enforce the same cancellations than for the system matrix. We denote the frequency-scaled static component of the right-hand side as $\widetilde{\boldsymbol{e}}_p^{\text{sub}}$, which is found so as to verify

$$\widetilde{\boldsymbol{e}}_p = \mathbf{P}_p^{\widetilde{\Sigma}} \widetilde{\boldsymbol{e}}_p + k \left(\mathbf{I} - \mathbf{P}_p^{\widetilde{\Sigma}} \right) \widetilde{\boldsymbol{e}}_p^{\text{sub}} \,. \tag{7.56}$$

Therefore, the right-hand side vectors of the preconditioned systems (7.27) and (7.43) are reformulated, respectively as

$$\widetilde{\mathbb{T}}_{r,p} \widetilde{\boldsymbol{e}}_p = j \widetilde{\mathbf{\Theta}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{s|r} \left(k^2 \widetilde{\mathbf{A}}_{r,p} \widetilde{\boldsymbol{e}}_p^{\text{sub}} + k \widetilde{\mathbf{B}}_{r,p} \widetilde{\boldsymbol{e}}_p \right) + j \widetilde{\mathbf{A}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{h|r} \widetilde{\mathbf{A}}_{r,p} \widetilde{\boldsymbol{e}}_p^{\text{sub}} + j k \widetilde{\mathbf{\Theta}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{s|r} \widetilde{\mathbf{B}}_{r,p} \widetilde{\boldsymbol{e}}_p \,, \tag{7.57}$$

$$\widetilde{\mathbb{T}}_{r,p}^k \widetilde{\boldsymbol{e}}_p = j \widetilde{\mathbf{\Theta}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{s|r} \left(\widetilde{\mathbf{A}}_{r,p}^k \widetilde{\boldsymbol{e}}_p^{\text{sub}} + k \widetilde{\mathbf{B}}_{r,p} \widetilde{\boldsymbol{e}}_p \right) + j \widetilde{\mathbf{A}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{h|r} \widetilde{\mathbf{A}}_{r,p} \widetilde{\boldsymbol{e}}_p^{\text{sub}} + j k \widetilde{\mathbf{\Theta}}_{r,p}^\mathsf{T} \widetilde{\mathbf{T}}_{s|r} \widetilde{\mathbf{B}}_{r,p} \widetilde{\boldsymbol{e}}_p \,. \tag{7.58}$$

These expressions do not suffer from numerical instability at low-frequency if we assume that $k\, \widetilde{\boldsymbol{e}}_p$ and $\widetilde{\boldsymbol{e}}_p^{\text{sub}}$ are bounded from above when the frequency goes to zero. This is in particular the case for plane-wave excitations [2]. The case of other excitation kinds is not studied here, but similar developments with the quasi-Helmholtz projectors can be found in [64, 81].

7.4.2 Efficiency of the Algorithm within Fast Iterative Frameworks

The proposed preconditioned systems (7.27) and (7.43) are fully compatible with acceleration strategies for iterative solvers. Fast matrix-vector products can indeed be performed by applying the system matrix's action to the solver's iterate stepwisely, from the rightmost to the leftmost products in the formulation, and by leveraging fast matrix-vector products associated with the matrices that compose the underlying systems. In absence of normalization (i.e., if we replace the inverse square-roots of Gram matrices by the identity), these fast matrix-vector products can all be carried out in a log-linear complexity with respect to the dimensions of the matrices, N_p and N_r:

- For matrices composing the non-normalized EFIE matrices ($\mathbf{T}_{s|p}$, $\mathbf{T}_{h|p}$, $\mathbf{T}_{s|r}$ and $\mathbf{T}_{h|r}$) fast multipole method (FMM) [40] and compression techniques [20] can be achieved with log-linear complexity.
- The quasi-irrotational transformation matrices $\mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_r$ as well as the mixed-Gram matrix $\mathbf{G}_{\boldsymbol{\psi}^{(r)}, \boldsymbol{n} \times \boldsymbol{\psi}^{(p)}}$ are sparse, so that their associated matrix-vector products are done in linear complexity, as well as the explicit computation of $\mathbf{\Sigma}_p^\mathsf{T} \mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_r^\mathsf{T} \mathbf{\Sigma}_r$ that still possess a linear number of non-zero entries.
- The action of the pseudo-inversions $(\mathbf{\Sigma}_p^\mathsf{T} \mathbf{\Sigma}_p)^+$ and $(\mathbf{\Sigma}_r^\mathsf{T} \mathbf{\Sigma}_r)^+$ to the solver's iterate involved in the multiplication by the quasi-Helmholtz projectors can be carried out with algebraic-multigrid (AMG) solvers [104, 78, 134], in a log-linear complexity.

Table 7.1: Ratio between the dimensions of the auxiliary space (of order r) and of the primal space (of order p) for $C = 2E/3$

Order p	0	1	2	3	4	5	6
$N_{(p+1)\oplus b}/N_p$	4.00	2.50	2.00	1.75	1.60	1.50	1.43
N_{p+2}/N_p	7.00	3.60	2.62	2.17	1.91	1.74	1.63

The normalization of the system may be more expensive as it requires fast iterative solvers for computing the multiplication by inverse square-roots of Gram matrices. This can be done efficiently with Tchebychev polynomial expansions or Padé approximants [37]. Even though, this technique may imply a large number of inner iterations if the Gram matrices are ill-conditioned. Additionally, the matrix to pseudo-invert in the projectors (6.14) is not sparse anymore and AMG cannot be directly used. A strategy recently proposed [94] is to use algebraic-multigrid applied to $\mathbf{\Sigma}_p^\intercal \mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_r^\intercal \mathbf{\Sigma}_r$ as preconditioners for the conjugate gradient (CG) solver applied to the iterative inversion of $\widetilde{\mathbf{\Sigma}}_p^\intercal \widetilde{\mathbf{\Sigma}}_p$ and $\widetilde{\mathbf{\Sigma}}_r^\intercal \widetilde{\mathbf{\Sigma}}_r$, respectively, where the matrix-vector products are done stepwisely too. In the following subsection, we will investigate experimentally the use of a substitution matrix, that allows to form a partially-orthonormal basis (leading to an effective preconditioning of the system) without the need for any inverse square-roots and that preserves the sparsity of the normalized matrices (enabling the direct use of AMG for the computation of the projectors).

Regarding the computational cost resulting from the use of an auxiliary space (of order r) of higher dimension than the primal space (of order p), we can have a look to the ratio between their dimensions. For the auxiliary space of order $r = (p+1) \oplus b$, we find

$$N_{(p+1)\oplus b}/N_p = 1 + \frac{E + 3(p+1)C}{(p+1)E + p(p+1)C}\,. \tag{7.59}$$

If the reader prefers rather using complete order Raviart-Thomas spaces, one possible choice is to use $r = p + 2$ instead, for which we find

$$N_{p+2}/N_p = 1 + \frac{2E + (4p+6)C}{(p+1)E + p(p+1)C}\,. \tag{7.60}$$

These two ratios tend towards 1 as p increases. In Table 7.1 we display their values for p going from zero to six and for $C = 2E/3$—which corresponds to the case of simply-connected geometries, and that is a fair approximation of the mesh elements distribution for closed surfaces with a small amount of topological loops. This shows that for $r = (p+1) \oplus b$, the auxiliary space dimension is never higher than four times the primal space dimension, and rapidly goes under twice that size (for $p \geq 2$). Instead, standard strategies based on barycentric refinement, such as in [136], always require the use of an auxiliary space that is six times bigger than the primal space, as both numbers of edges and cells are multiplied by six. These considerations show that N_r is of the same order of magnitude than N_p and that the overall multiplication with the system matrix is done in a log-linear complexity with respect to the dimension of the original system (2.93).

Table 7.2: Number of instances of each kind of matrix-vector product per iteration.

Formulation	EFIO	Sparse	AMG (+ CG)	Inv-sqrt
(7.27)	2	12	4	26
(7.43)	2	14	5	30
(7.54)	5	25	9	60
(7.55)	5	30	11	64

To conclude this subsection, we summarize in Table 7.2 the number of instances carried out for each kind of matrix-vector product at one step of the iterative inversion of the preconditioned systems (7.27) and (7.43)—corresponding to the simply- and multiply-connected cases, respectively—as well as for their expanded counterparts (7.54) and (7.55)—needed in the low-frequency regime. "EFIO" stands for fast multiplications with the EFIE operators (2.94) (for orders r and p) in the case of formulations (7.27) and (7.43), or with the underlying matrices (7.5) and (7.6) in the case of formulations (7.54) and (7.55). "Sparse" stands for sparse multiplications with the matrices $\boldsymbol{\Sigma}_p$, $\boldsymbol{\Sigma}_r$ and $\mathbf{G}_{\boldsymbol{\psi}^{(r)},\boldsymbol{n}\times\boldsymbol{\psi}^{(p)}}$. "AMG (+ CG)" stands for the iterative pseudo-inversion of $\widetilde{\boldsymbol{\Sigma}}_p^\intercal\widetilde{\boldsymbol{\Sigma}}_p$ and $\widetilde{\boldsymbol{\Sigma}}_r^\intercal\widetilde{\boldsymbol{\Sigma}}_r$ and hides additional products with the normalization matrices, several iterations needed to solve the underlying systems, and a setup phase done once for all at the beginning of the program. "Inv-sqrt" stands for the additional multiplications with inverse square-roots of Gram-matrices used for the normalization. We remind the reader that the inverse square-roots of Gram matrices can be replaced by a product with a sparse matrix, that still enables an effective preconditioning of the system, as it will be discussed in further details now.

7.4.3 Replacing the Inverse Square-Roots by Sparse Transformation Matrices

Only two types of inverse square-roots of Gram matrices are used in the proposed preconditioning technique. The inverse square-roots of the Gram matrix $\mathbf{G}_{\sigma,\sigma}$ between $\mathbb{DG}$ functions, and of the Gram matrix $\mathbf{G}_{\boldsymbol{\psi},\boldsymbol{\psi}}$ between $\mathbb{RT}_p$ functions, where we have removed the annotations for the order (superscripts over the basis functions and subscripts under the functional spaces), as this strategy can be applied equally to any order.

On the one hand, the computation of the inverse square-root of $\mathbf{G}_{\sigma,\sigma}$ can be done in a linear complexity, if we assume the discontinuous basis functions to be ordered cell-by-cell, as it results in a block diagonal matrix, for which the matrix operations can be done block-wisely. If the basis functions are not grouped properly, this should be enforced with a permutation matrix (see Step 0 and Step 1 of the algorithm for $\mathbf{G}_{\boldsymbol{\psi},\boldsymbol{\psi}}$).

On the other hand, however, the computation of the inverse square-root of $\mathbf{G}_{\boldsymbol{\psi},\boldsymbol{\psi}}$ is more complicated, as the presence of edge-based functions implies a global orthonormalization of the basis functions. This results in an inverse square-root of the Gram matrix that is a dense matrix—and that is computed in a cubic complexity with a direct algorithm. The algorithm we present now allows for computing a sparse transformation matrix that partially-orthonormalizes the basis functions and that leads to Gram matrices and preconditioned systems that are well-conditioned with respect to the mesh refinement and the basis functions order.

Step 0: We first impose that the basis functions are ordered as follows:

1. We group the edge-based functions together.

2. Within this group, we give contiguous indices to edge-based functions associated with the same edge.

3. We group the cell-based functions together.

4. Within this group, we give contiguous indices to cell-based functions associated with the same cell.

Notice that among the cell-based functions are included the curl of bubble functions, and that this defines an ordering different from the one used in subsection 7.2.1. If this ordering is not naturally imposed, we can assemble a permutation matrix $\mathbf{Q}_0$ such that

$$\mathbf{Q}_0^{\mathsf{T}} \mathbf{G}_{\boldsymbol{\psi},\boldsymbol{\psi}} \mathbf{Q}_0 := \begin{pmatrix} \mathbf{G}_{e,e} & \mathbf{G}_{e,c} \\ \mathbf{G}_{c,e} & \mathbf{G}_{c,c} \end{pmatrix}, \tag{7.61}$$

which contains the inner products between edge-based / edge-based, edge-based / cell-based and cell-based / cell-based basis functions, grouped by common edges $\{e_i\}_{i=1}^{N}$ and common cells $\{c_i\}_{i=1}^{C}$, i.e.,

$$\mathbf{G}_{e,e} := \begin{pmatrix} \mathbf{G}_{e_1,e_1} & \mathbf{G}_{e_1,e_2} & \cdots & \mathbf{G}_{e_1,e_E} \\ \mathbf{G}_{e_2,e_1} & \mathbf{G}_{e_2,e_2} & \cdots & \mathbf{G}_{e_2,e_E} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{e_1,e_1} & \mathbf{G}_{e_1,e_2} & \cdots & \mathbf{G}_{e_E,e_E} \end{pmatrix} \tag{7.62}$$

$$\mathbf{G}_{e,c} := \begin{pmatrix} \mathbf{G}_{e_1,c_1} & \mathbf{G}_{e_1,c_2} & \cdots & \mathbf{G}_{e_1,c_C} \\ \mathbf{G}_{e_2,c_1} & \mathbf{G}_{e_2,c_2} & \cdots & \mathbf{G}_{e_2,c_C} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{e_1,c_1} & \mathbf{G}_{e_1,c_2} & \cdots & \mathbf{G}_{e_E,c_C} \end{pmatrix}, \tag{7.63}$$

$$\mathbf{G}_{c,c} := \begin{pmatrix} \mathbf{G}_{c_1,c_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{c_2,c_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{G}_{c_C,c_C} \end{pmatrix}, \tag{7.64}$$

and $\mathbf{G}_{c,e} = \mathbf{G}_{e,c}^{\mathsf{T}}$. Notice that $\mathbf{G}_{e_i,e_j} \neq 0$ if and only if edges e_i and e_j belong to the same triangle, while $\mathbf{G}_{e_i,c_j} \neq 0$ if and only if e_i is an edge of triangle c_j.

Step 1: We make the cell-based functions orthonormal to each other. Because cell-based functions defined on different cells are already orthogonal to each other, we only need to orthonormalize the cell-based functions cell-by-cell. As a matter of fact, $\mathbf{G}_{c,c}$ is block-diagonal, so that its inverse square-root is computed block-wisely, in a linear complexity

$$\mathbf{G}_{c,c}^{-1/2} = \begin{pmatrix} \mathbf{G}_{c_1,c_1}^{-1/2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{c_2,c_2}^{-1/2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{G}_{c_C,c_C}^{-1/2} \end{pmatrix}. \tag{7.65}$$

We now set

$$\mathbf{Q}_1 := \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{c,c}^{-1/2} \end{pmatrix}, \tag{7.66}$$

such that

$$\mathbf{Q}_1^\mathsf{T}\mathbf{Q}_0^\mathsf{T}\mathbf{G}_{\psi,\psi}\mathbf{Q}_0\mathbf{Q}_1 = \begin{pmatrix} \mathbf{G}_{e,e} & \mathbf{G}_{e,c}\mathbf{G}_{c,c}^{-1/2} \\ \mathbf{G}_{c,c}^{-1/2}\mathbf{G}_{c,e} & \mathbf{I} \end{pmatrix}. \tag{7.67}$$

Step 2: We make the edge-based functions orthogonal to the cell-based functions. This is done by setting

$$\mathbf{Q}_2 := \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{c,c}^{-1/2}\mathbf{G}_{c,e} & \mathbf{I} \end{pmatrix}, \tag{7.68}$$

which yields

$$\mathbf{Q}_2^\mathsf{T}\mathbf{Q}_1^\mathsf{T}\mathbf{Q}_0^\mathsf{T}\mathbf{G}_{\psi,\psi}\mathbf{Q}_0\mathbf{Q}_1\mathbf{Q}_2 = \begin{pmatrix} \mathbf{B}_{e,e} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{7.69}$$

with $\mathbf{B}_{e,e} := \mathbf{G}_{e,e} - \mathbf{G}_{e,c}\mathbf{G}_{c,c}^{-1}\mathbf{G}_{c,e}$.

Step 3: We *partially-orthonormalize* the edge-based basis functions. We first write

$$\mathbf{B}_{e,e} := \begin{pmatrix} \mathbf{B}_{e_1,e_1} & \mathbf{B}_{e_1,e_2} & \cdots & \mathbf{B}_{e_1,e_E} \\ \mathbf{B}_{e_2,e_1} & \mathbf{B}_{e_2,e_2} & \cdots & \mathbf{B}_{e_2,e_E} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{e_1,e_1} & \mathbf{B}_{e_1,e_2} & \cdots & \mathbf{B}_{e_E,e_E} \end{pmatrix}, \tag{7.70}$$

where each block corresponds to the grouping of the basis functions edge-by-edge, with $\mathbf{B}_{e_i,e_j} \neq 0$ if and only if e_i and e_j belong to the same triangle. In order to fully-orthonormalize these functions, the correct strategy would laverage the computation of the inverse square-root of $\mathbf{B}_{e,e}$. We want to avoid that as it would result in a dense matrix for which the direct computation is done in a cubic complexity. Instead, we solely compute the inverse square-root of the block diagonal,

$$\mathbf{Q}_3 := \begin{pmatrix} \mathbf{B}_{e_1,e_1}^{-1/2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{e_2,e_2}^{-1/2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}_{e_E,e_E}^{-1/2} \end{pmatrix}, \tag{7.71}$$

and we finally set $\mathbf{Q} := \mathbf{Q}_0\mathbf{Q}_1\mathbf{Q}_2\mathbf{Q}_3$ such that the block diagonal of $\mathbf{Q}^\mathsf{T}\mathbf{G}_{\psi,\psi}\mathbf{Q}$ is the identity matrix and its off-diagonal blocks are non-zero if and only if they are associated with edges belonging to the same triangle.

Overall, this transformation maps the GWP basis functions into another set of basis functions spanning the same Raviart-Thomas space, for which:

- Cell-based basis functions are orthonormal to every basis function.
- Edge-based basis functions are orthonormal to every cell-based basis function, edge-based basis function defined on the same edge, and edge-based basis function that do not share the same domain.
- Only edge-based basis functions defined on different edges of the same triangle are not orthonormal to each other.

The transformation matrix $\mathbf{Q}$ is sparse and can be computed in a linear complexity as it only involves inverse square-roots of diagonal blocks and multiplications between sparse matrices. Numerical experiments have shown that for quasi-uniform meshes, this partial-orthonormalization of the basis functions yields a well conditioned Gram matrix $\mathbf{Q}^\mathsf{T}\mathbf{G}_{\psi,\psi}\mathbf{Q}$, with a condition number that is almost constant when refining the mesh and/or increasing the basis functions order (see Table 7.3 in the numerical experiments). It turns out that $\mathbf{Q}$ can replace $\mathbf{G}_{\psi,\psi}^{-1/2}$ in the preconditioned formulations (7.27), (7.43), (7.54) and (7.55) and still allows to get a bounded spectrum independent of the mesh refinement. With this strategy, no direct nor iterative computation of inverse square-root of Gram matrices is needed, and these operations are replaced by sparse multiplications with the matrix $\mathbf{Q}$ (in Table 7.2, the values in the column "Inv-sqrt" are set to zero and added to the column "Sparse"). Additionally, the partial-orthonormalization of the quasi-irrotational transformation matrices $\mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_r$ can be computed directly (with sparse matrix-matrix products) so that AMG can be applied directly to compute the iterative pseudo-inversion in (6.14).

7.5 Numerical Validations

7.5.1 Validation of the Simply-Connected Formulation

To illustrate the effects of the choice of r on the spectrum of the preconditioned formulation (7.27), we set

$$\widetilde{\mathbb{T}}_{r,p}\widetilde{\mathbf{T}}_p = -\tfrac{1}{4}\mathbf{L}_{r,p} + \mathbf{C}_{r,p}\,, \tag{7.72}$$

with

$$\mathbf{L}_{r,p} \coloneqq \left(\widetilde{\mathbf{\Theta}}_{r,p}^\mathsf{T}\widetilde{\mathbf{G}}_{r,p}^{\text{mix}}\right)^\mathsf{T}\left(\widetilde{\mathbf{\Theta}}_{r,p}^\mathsf{T}\widetilde{\mathbf{G}}_{r,p}^{\text{mix}}\right)\,, \tag{7.73}$$

and

$$\mathbf{C}_{r,p} \coloneqq \widetilde{\mathbb{T}}_{r,p}\widetilde{\mathbf{T}}_p + \tfrac{1}{4}\mathbf{L}_{r,p}\,, \tag{7.74}$$

and we display in Figure 7.1 the moduli of the eigenvalues of the matrices involved in the formulation (7.27) for various values of r when modelling the unitary sphere at 1MHz with $p = 2$ and $h = 58$cm. We additionally display the spectrum of the matrix

$$\mathbf{L}_{r,p}^{-1}\left(\widetilde{\mathbb{T}}_{r,p}\widetilde{\mathbf{T}}_p\right) \tag{7.75}$$

to illustrate the spectral equivalence between $\mathbf{L}_{r,p}^{-1}$ and the preconditioned matrix $\widetilde{\mathbb{T}}_{r,p}\widetilde{\mathbf{T}}_p$.

We firstly observe that the EFIE matrices $\widetilde{\mathbf{T}}_p$ and $\widetilde{\mathbb{T}}_{r,p}$ exhibit analogous spectra characterized by an interchange between solenoidal (small eigenvalues) and non-solenoidal (high eigenvalues) contributions. For $r = p$ and $r = p + 1$, however, there is an ill-conditioning of $\widetilde{\mathbb{T}}_{r,p}$, $\mathbf{L}_{r,p}$ and $\widetilde{\mathbb{T}}_{r,p}\widetilde{\mathbf{T}}_p$, with eigenvalues equal to zero for $r = p$ and going to zero for $r = p + 1$. Instead, these matrices are well-conditioned for $r = (p + 1) \oplus b$ and $r = p + 2$, with almost flat spectra. In comparison, the eigenvalues moduli of $\mathbf{C}_{r,p}$ are substantially smaller than those of $\mathbf{L}_{r,p}$ and go to zero with the spectrum index, indicating that it corresponds to a compact perturbation in the spectrum of $\widetilde{\mathbb{T}}_{r,p}\widetilde{\mathbf{T}}_p$. We also observe that the eigenvalues moduli of (7.75) are very close to $1/4$, indicating that $-\frac{1}{4}\mathbf{L}_{r,p}$ and $\widetilde{\mathbb{T}}_{r,p}\widetilde{\mathbf{T}}_p$ are spectrally equivalent. To further confirm these observations, we display the eigenvalues of these matrices in the complex plane for $r = (p+1)\oplus b$ in Figure 7.2, showing that the eigenvalues of the preconditioned matrix actually cluster around -1/4—and are not scattered along the circle of radius 1/4.

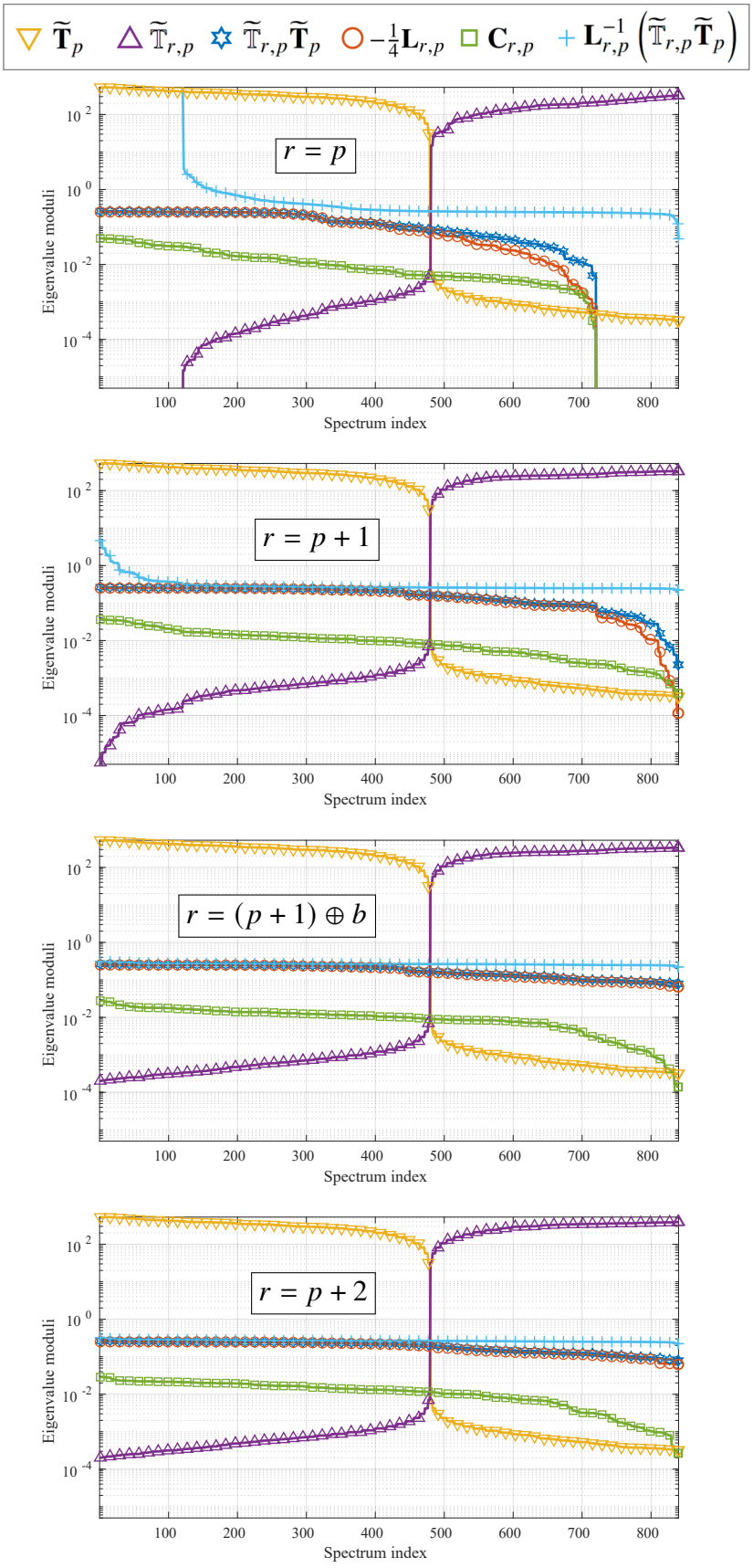


Figure 7.1: Moduli of the eigenvalues of the matrices involved in the preconditioned formulation (7.27) for simulations of the sphere at frequency $f = 1$MHz, with $p = 2$ and $h = 0.58$cm.

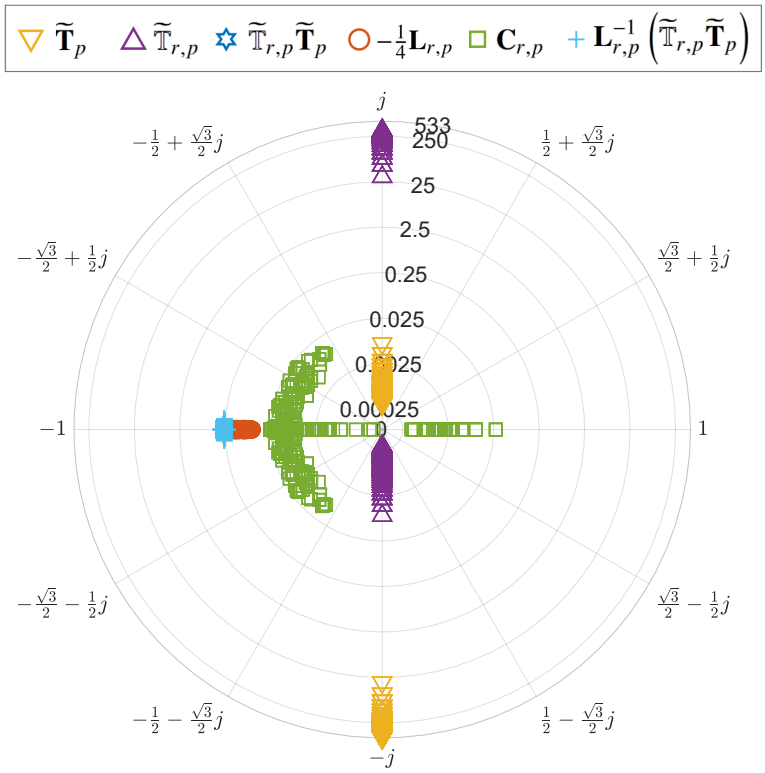


Figure 7.2: Eigenvalues of the matrices involved in the preconditioned formulation (7.27) for simulations of the sphere at frequency $f = 1\text{MHz}$, with $p = 2$ and $h = 0.58\text{cm}$.

From now, and in the following experiments, the order of the auxiliary space is always set to $r = (p+1) \oplus b$.

To further validate the preconditioned formulation (7.27), we compare the ratio between the maximal and minimal moduli of the eigenvalues of the underlying matrix with that of the normalized EFIE stabilized using quasi-Helmholtz projectors (6.19) and with that of the normalized EFIE without preconditioning (6.1). To start, we display in Figure 7.3 the eigenvalue moduli ratios as a function of the number of unknowns, for order $p = 0,1,2$, and for $r = (p+1)\oplus b$, that are obtained for simulation of the sphere of radius 1m at frequency $f = 1\text{MHz}$, with $h = 0.58\text{cm}$. We observe that the eigenvalues obtained with the Calderón preconditioning strategy have moduli that are relatively close to each other, indicating that they have a bounded eigenvalue spectrum independently of the discretization; while both projectors-based and non-preconditioned EFIE relate system matrices with unbounded spectra. To showcase the stability at low-frequency, we display in Figure 7.4 the eigenvalue moduli ratio of the expanded version of the preconditioned formulation (7.54) as well as that of the expanded version of the projectors-based EFIE (as in (5.35) but with normalization), in addition to that of the standard (normalized) EFIE. The system matrices are obtained for simulations of the sphere of radius 1m, with $h = 0.58\text{cm}$. Again, the eigenvalues of the Calderón preconditioned matrix are bounded when the frequency decreases, and it is also the case for the projectors-based EFIE—that display a higher ratio nonetheless. Instead, the non-preconditioned EFIE displays an unbounded spectrum as the frequency goes to zero. Finally, we show in Figure 7.5 the effects of the preconditioning on the number of iterations carried out by the iterative solver TFQMR. We observe that for a given residual error, an almost constant number of iterations is needed when solving the Calderón formulation, for increasing order, that is with an increasing number of unknowns. Instead,

TFQMR suffers from an increase of the number of iterations when solving the projectors-based EFIE formulation, while it does not converge when solving the standard EFIE formulation.

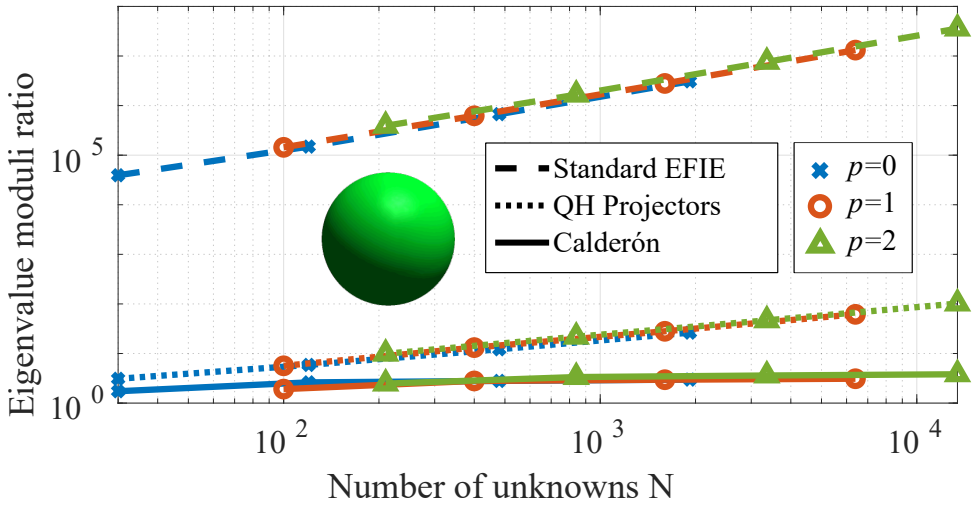


Figure 7.3: Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the sphere (f = 1MHz).

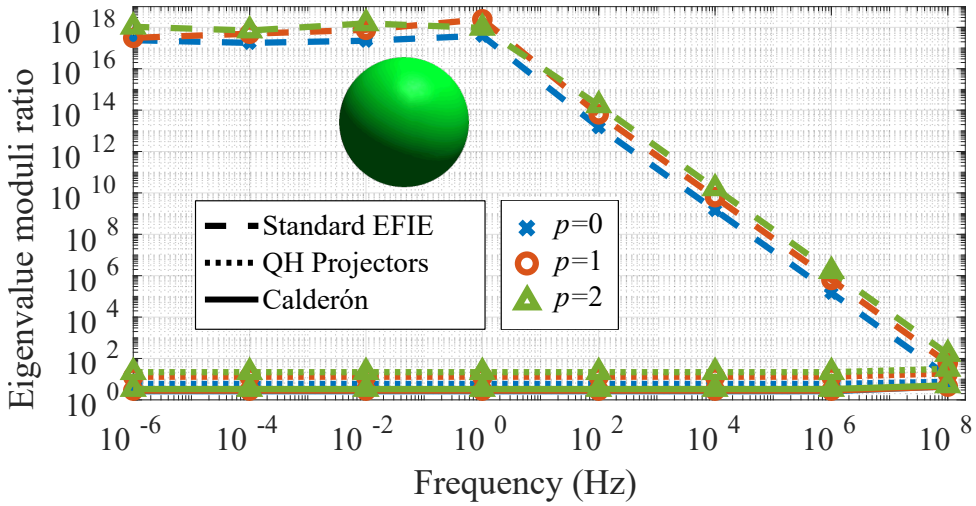


Figure 7.4: Ratio between the maximal and minimal eigenvalue moduli as a function of the frequency for simulations of the sphere (h = 58cm).

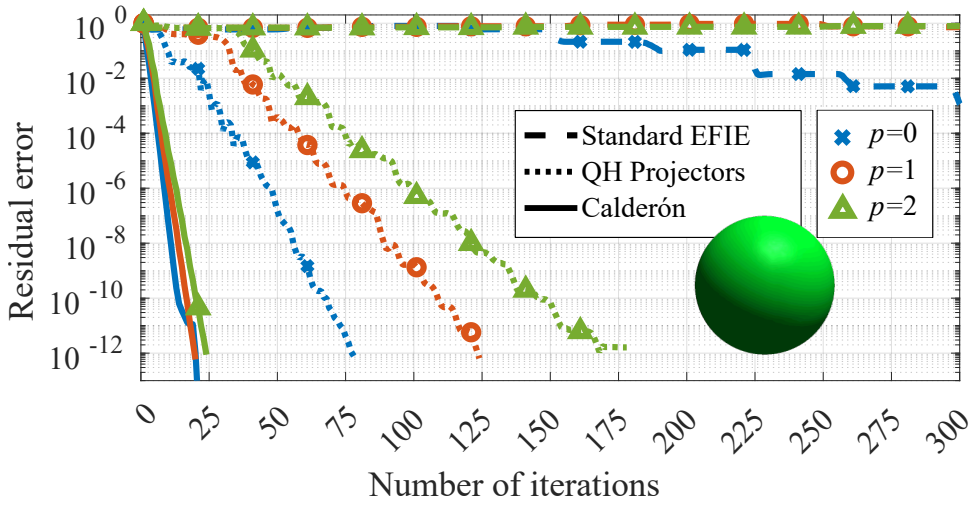


Figure 7.5: Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the sphere (f = 1MHz, h = 0.58cm).

7.5.2 Validation of the Multiply-Connected Formulation

To illustrate the effects of the low-frequency rescaling on the spectrum of the rescaled formulation (7.43) for multiply connected geometries, we set

$$\widetilde{\mathbb{T}}^k_{r,p}\widetilde{\mathbf{T}}^k_p = -\tfrac{1}{4}\mathbf{L}_{r,p} + \mathbf{C}^k_{r,p}\,, \tag{7.76}$$

with $\mathbf{L}_{r,p}$ defined as in (7.73) and with

$$\mathbf{C}^k_{r,p} := \widetilde{\mathbb{T}}^k_{r,p}\widetilde{\mathbf{T}}^k_p + \tfrac{1}{4}\mathbf{L}_{r,p}\,, \tag{7.77}$$

and we display in Figure 7.6 the eigenvalue moduli of the matrices involved in (7.27) and (7.43) obtained for simulations of a torus at 1Hz. We additionally display the spectrum of the matrix

$$\mathbf{L}^{-1}_{r,p}\left(\widetilde{\mathbb{T}}^k_{r,p}\widetilde{\mathbf{T}}^k_p\right) \tag{7.78}$$

to illustrate the spectral equivalence between $\mathbf{L}^{-1}_{r,p}$ and the preconditioned matrix $\widetilde{\mathbb{T}}^k_{r,p}\widetilde{\mathbf{T}}^k_p$. We first observe that without rescaling, the original preconditioned matrix (7.27) has two singular values that go to zero, associated with the two topological loops of the geometry. With the rescaling in (7.43), $\widetilde{\mathbf{T}}^k_p$ and $\widetilde{\mathbb{T}}^k_{r,p}$ have flatter spectra than their non-rescaled counterparts $\widetilde{\mathbf{T}}_p$ and $\widetilde{\mathbb{T}}_{r,p}$, as their divergence-free and quasi-irrotational components have been properly rescaled for the low-frequency regime. It turns out that the two singular values going to zero are removed from the spectrum of $\widetilde{\mathbb{T}}^k_{r,p}\widetilde{\mathbf{T}}^k_p$ and we still observe that the singular values of $\mathbf{C}^k_{r,p}$ correspond to a compact perturbation of $-\frac{1}{4}\mathbf{L}_{r,p}$ in the spectrum of the preconditioned matrix. We also observe that the eigenvalues of (7.78) have moduli very close to $1/4$, indicating that $-\frac{1}{4}\mathbf{L}_{r,p}$ and $\widetilde{\mathbb{T}}^k_{r,p}\widetilde{\mathbf{T}}^k_p$ are spectrally equivalent. To further confirm these observations, we display the eigenvalues of these matrices in the complex plane for the rescaled formulation (7.43) in Figure 7.7, showing that the eigenvalues of the preconditioned matrix actually cluster around $-1/4$—and are not scattered around the circle of radius $1/4$.

To further validate the preconditioned formulation (7.43), we display in Figure 7.8 the eigenvalue moduli ratio of the underlying system matrices as a function of the number of unknowns, for order p = 0,1,2, and for $r = (p+1)\oplus b$, that are obtained for simulation of the torus of outer

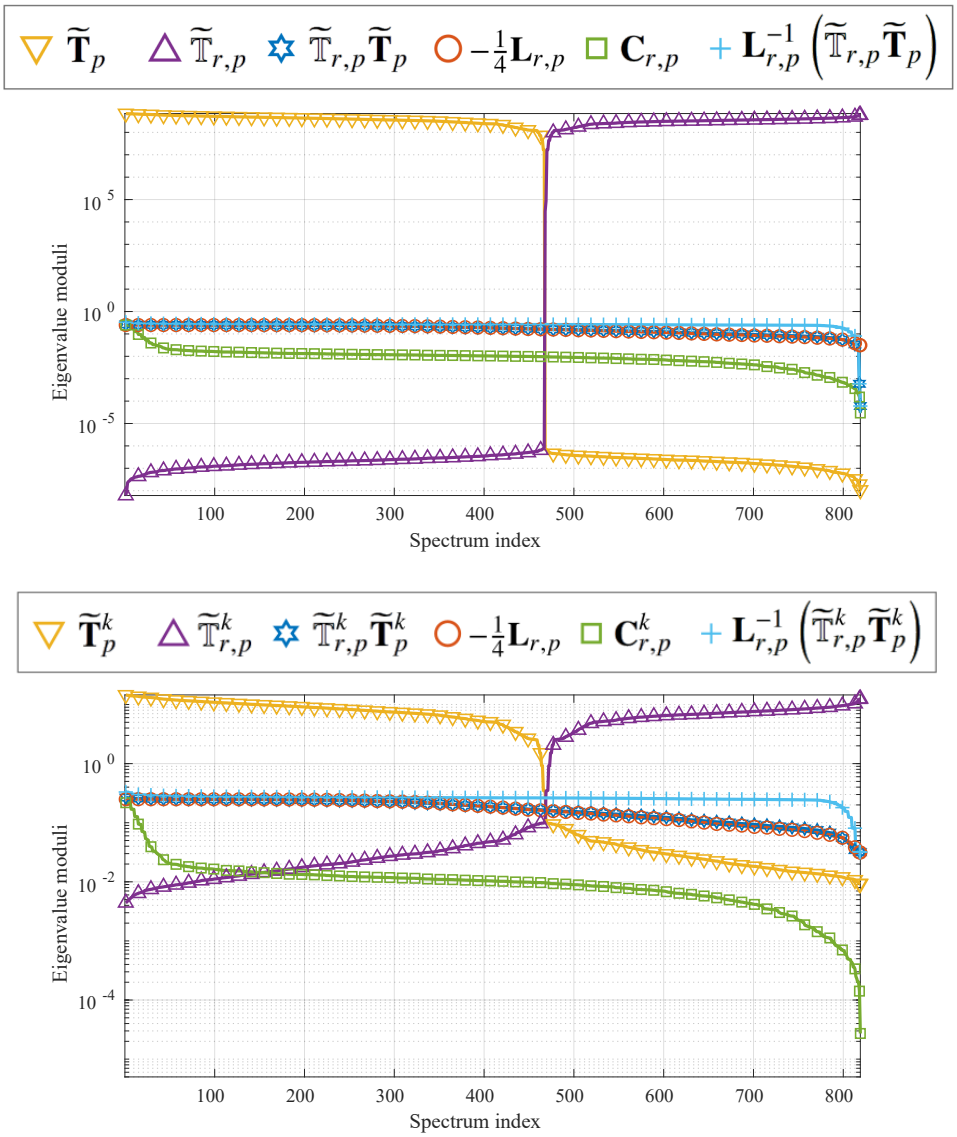


Figure 7.6: Moduli of the eigenvalues of the matrices involved in the preconditioned formulation (7.43) for simulations of the torus of outer radius 0.5 m and inner radius 0.15 m at frequency $f = 1$Hz, with $p = 2$ and $h = 8$cm.

radius 0.5 m and inner radius 0.15 m at frequency $f = 10$MHz, with the different formulations. We observe that the eigenvalues of the matrix obtained with the Calderón formulation is bounded when the number of unknowns increases; while both projectors-based and non-preconditioned EFIE yield matrices with unbounded spectra as the number of unknowns increases. To showcase the stability at low-frequency, we display in Figure 7.8 the eigenvalue moduli ratio of the expanded version of the Calderón formulation (7.55) and those of the expanded version of the projectors-based EFIE and of the standard (normalized) EFIE. The system matrices are obtained for simulation of the torus for a mesh discretization corresponding to $h = 8$cm. Again, the eigenvalues obtained with the Calderón formulation are bounded when the frequency decreases, and it is also the case for the projectors-based EFIE—that display a higher ratio nonetheless. Instead, the non-preconditioned EFIE displays an unbounded spectrum as the frequency goes to zero. Finally, we show in Figure 7.10 the effects of the preconditioning on the number of iterations of the iterative solver TFQMR. We observe that for a given residual error, an almost constant number of iterations is needed when solving the Calderón formulation, for increasing

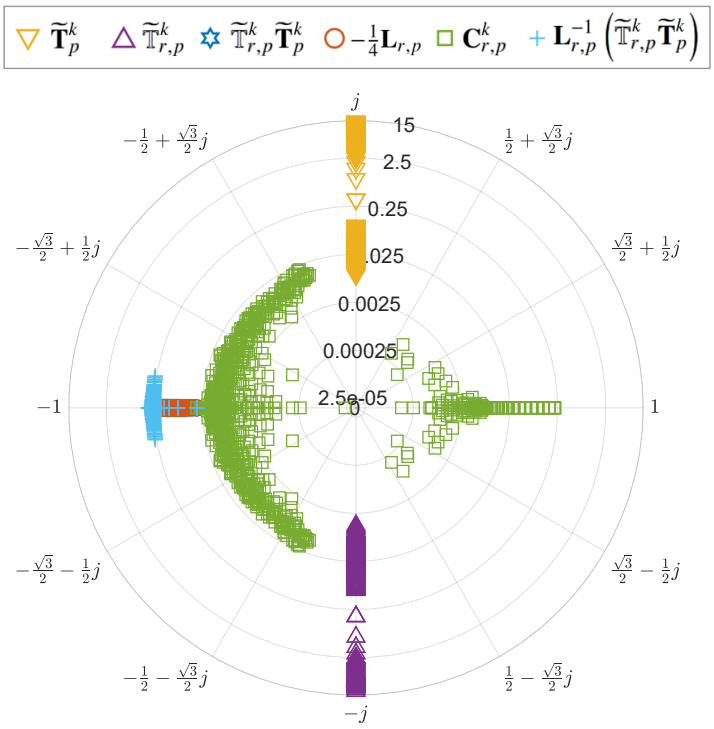


Figure 7.7: Eigenvalues of the matrices involved in the preconditioned formulation (7.43) for simulations of the torus of outer radius 0.5 m and inner radius 0.15 m at frequency f = 1Hz, with $p = 2$ and h = 8cm.

order, that is with an increasing number unknowns. Instead, TFQMR suffers from an increase of the number of iterations when solving the projectors-based EFIE formulation, while it does not converge when solving the standard EFIE formulation.

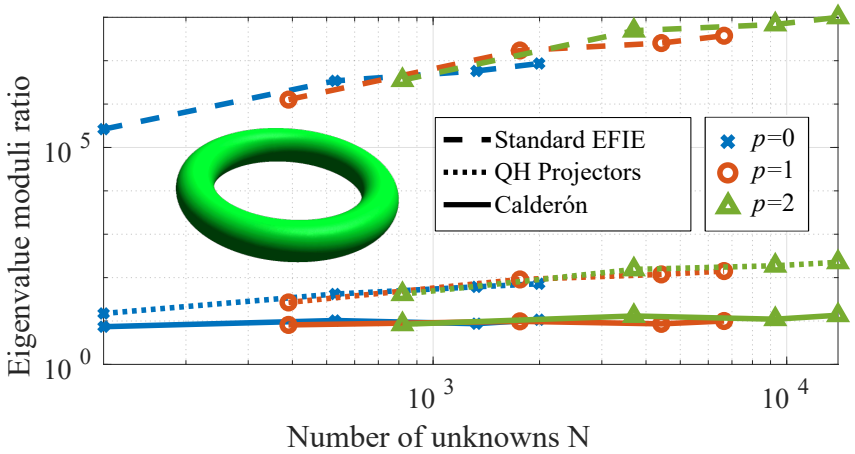


Figure 7.8: Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the torus (f = 1MHz).

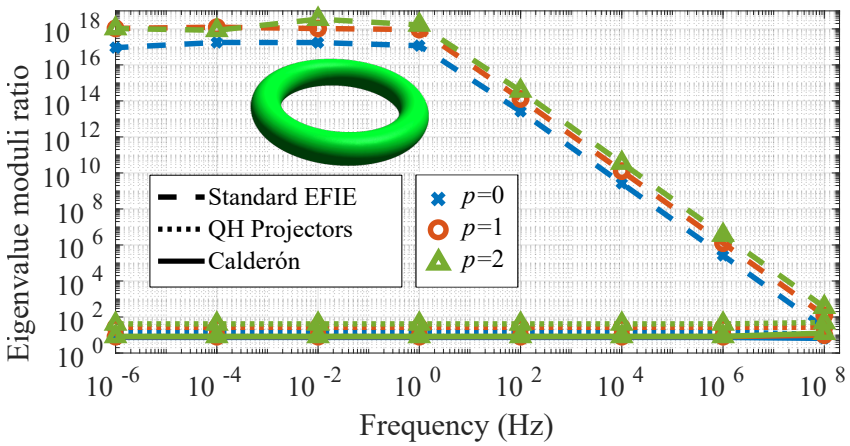


Figure 7.9: Ratio between the maximal and minimal eigenvalue moduli as a function of the frequency for simulations of the torus (h = 8cm).

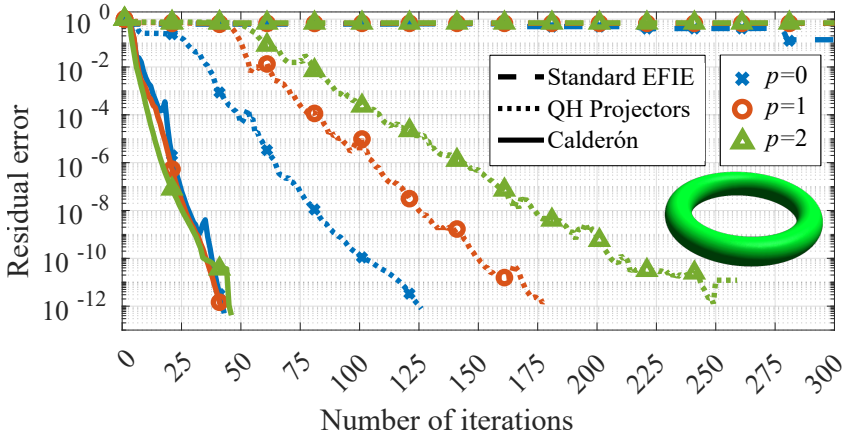


Figure 7.10: Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the torus (f = 1MHz and h = 8cm).

7.5.3 Validation for Less Regular Geometries

For smooth geometries, the Calderón identity allows to get a preconditioned EFIE formulation for which the eigenvalues cluster around -1/4. In that section, we show that well-conditioned matrices can also be obtained for less regular geometries. We do the same tests than for the torus and for the sphere with the NASA Almond of dimensions 25.24cm × 9.75cm × 3.25cm. In Figure 7.11 and Figure 7.12, we display the ratios between the moduli of the eigenvalues as a function of the number of unknowns (for f = 1MHz) and as a function of the frequency (for h = 1.3cm), respectively. In Figure 7.13, we display the residual error as a function of the number of iterations. The same observations than for the sphere and the torus are obtained with that geometry, indicating that the proposed Calderón preconditioning strategy is also effective for non-smooth geometries.

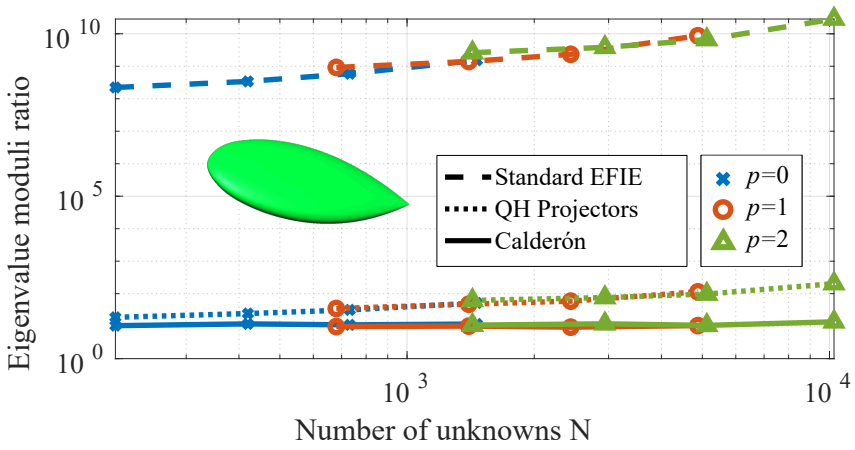


Figure 7.11: Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the NASA almond (f = 1MHz).

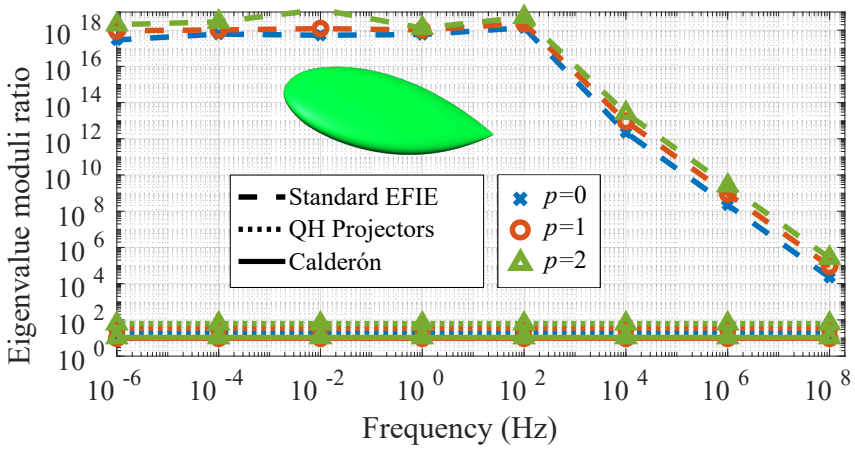


Figure 7.12: Ratio between the maximal and minimal eigenvalue moduli as a function of the frequency for simulations of the NASA almond (h = 1.3cm).

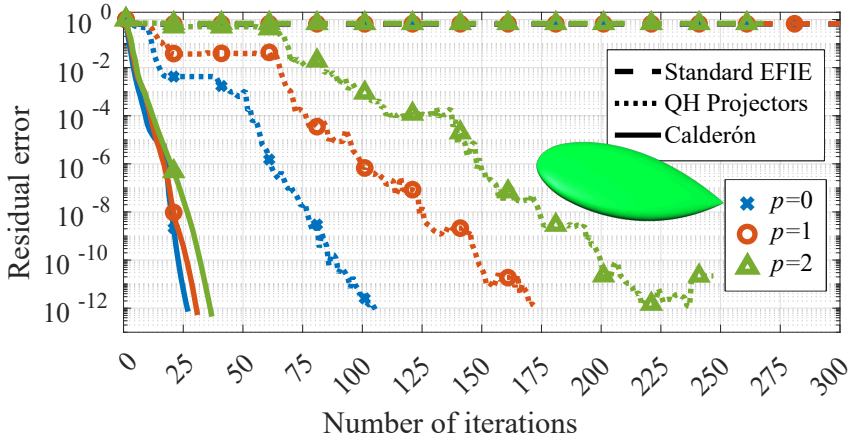


Figure 7.13: Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the NASA almond (f = 1MHz and h = 1.3cm).

7.5.4 On the Use of Partial-Orthonormalization

In Section 7.4.3, we described an algorithm that provides a partial-orthonormalization of the basis functions carried out via the computation of a sparse transformation matrix. To end this experimental section, we showcase the effectiveness of the preconditioning when using partial-orthonormalization instead of the full orthonormalization provided by the expensive inverse square-root computations.

First, Table 7.3 displays the condition number of the partially-normalized Gram matrices $\mathbf{Q}^\mathsf{T}\mathbf{G}_{\psi,\psi}\mathbf{Q}^\mathsf{T}$ for basis functions order from 0 to 6 that are obtained for simulations of the unitary sphere with $h = 105cm$ and $h = 30$cm. This shows that an almost constant condition number of the Gram matrix can be obtained with this mapping, and that the partially-orthonormalized Gram matrix is spectrally equivalent to the identity matrix (that is, the Gram matrix associated with fully-orthonormalized basis functions). As a comparison, the condition number of the non-normalized Gram matrix is shown for the same orders and we observe that it severely increases with the order and the mesh refinement. Notice that this condition number equals one with the full-orthonormalization, by definition of the inverse square-root $\mathbf{G}_{\psi,\psi}^{-1/2}\mathbf{G}_{\psi,\psi}\mathbf{G}_{\psi,\psi}^{-1/2} = \mathbf{I}$.

Table 7.3: Comparision of the condition number betweem the Gram matrix associated with the GWP functions and its partially-orthonormalized counterpart, for different orders and mesh discretization of the unitary sphere.

h (cm)		$p=0$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$
105	N_p	100	210	360	550	780	1050	1360
	cond($\mathbf{G}_{\psi,\psi}$)	1.07E1	1.11E2	6.71E2	3.63E3	1.76E4	9.59E4	4.45E5
	cond($\mathbf{Q}^\mathsf{T}\mathbf{G}_{\psi,\psi}\mathbf{Q}$)	1.98	1.94	2.01	2.02	2.11	2.12	2.19
26.3	N_p	1600	3360	5760	8800	12480	16800	21760
	cond($\mathbf{G}_{\psi,\psi}$)	1.35E1	1.36E2	8.47E2	4.55E3	2.22E4	1.22E5	5.66E5
	cond($\mathbf{Q}^\mathsf{T}\mathbf{G}_{\psi,\psi}\mathbf{Q}$)	2.20	2.07	2.21	2.19	2.31	2.32	2.40

In Figures 7.14, 7.15 and 7.16, we display the eigenvalue moduli ratios of the Calderón preconditioned system along with that of the projectors-based EFIE and the standard EFIE, obtained with partial-orthonormalization (where $\mathbf{G}_{\psi,\psi}^{-1/2}$ is replaced by $\mathbf{Q}$), at frequency $f = 1$MHz, for the sphere, the torus, and the NASA almond, respectively. In Figures 7.17, 7.18 and 7.19, we display the associated effects on the number of iterations to achieve a given residual error when solving these formulations with the iterative solver TFQMR. These experiments show that similar results can be obtained when replacing the full-orthonormalization by partial-orthonormalization, and, in particular, we get almost the same number of iterations for the Calderón preconditioned formulation in both cases. This indicates that an effective preconditioning scheme can be obtained at a smaller cost, without the need for computing inverse square-roots of Gram matrices.

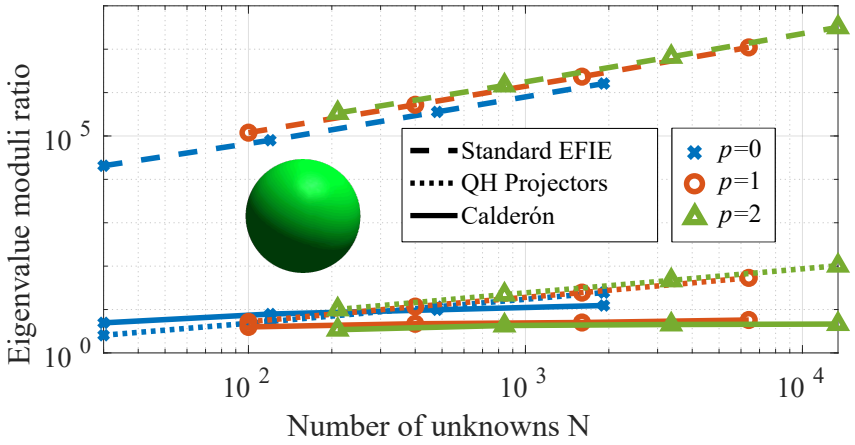


Figure 7.14: Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the sphere, with partial orthonormalization (f = 1MHz).

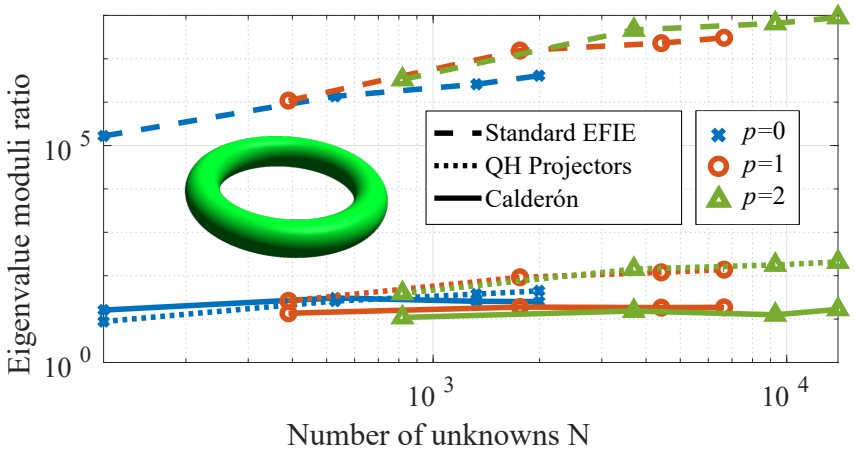


Figure 7.15: Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the torus, with partial orthonormalization (f = 1MHz).

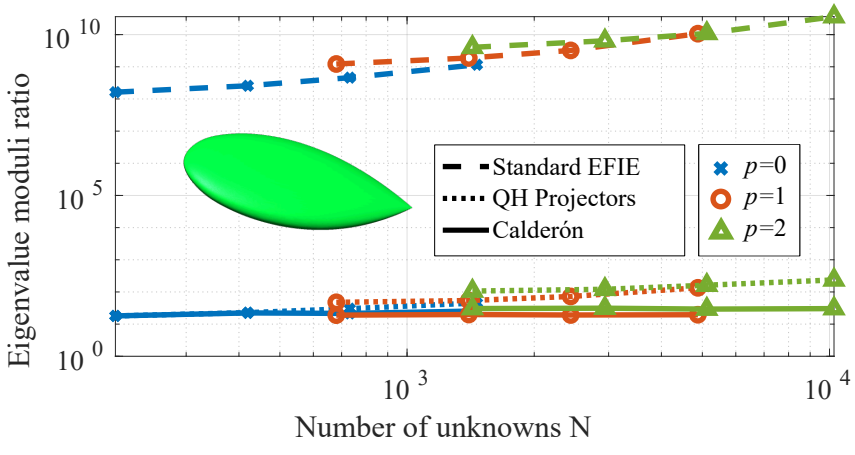


Figure 7.16: Ratio between the maximal and minimal eigenvalue moduli as a function of the number of unknowns for simulations of the NASA almond, with partial orthonormalization (f = 1MHz).

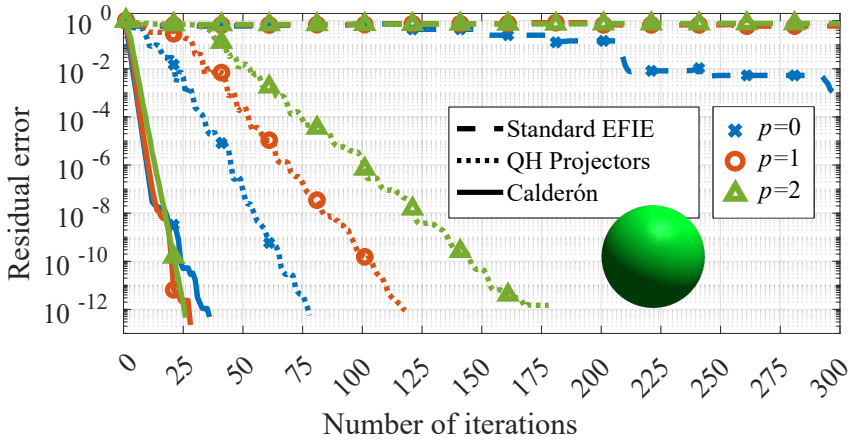


Figure 7.17: Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the sphere, with partial orthonormalization (f = 1MHz, h = 58cm).

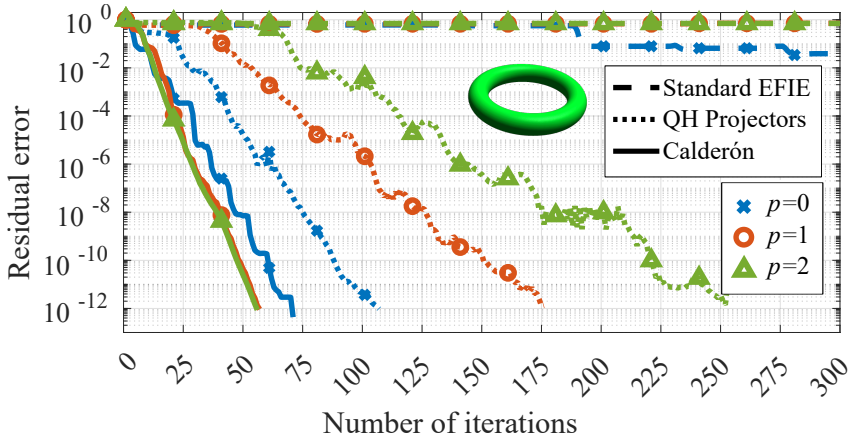


Figure 7.18: Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the torus, with partial orthonormalization (f = 1MHz, h = 8cm).

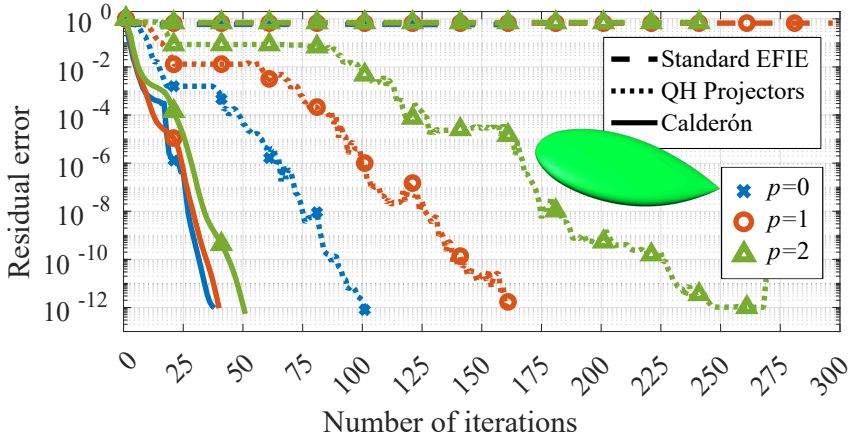


Figure 7.19: Error residual as a function of the number of iterations carried out by the solver TFQMR with the different formulations for simulations of the NASA Almond, with partial orthonormalization (f = 1MHz, h = 1.3cm).

7.6 Conclusion

In this chapter, we proposed a novel Calderón preconditioner for the high-order boundary element discretization of the electric field integral equation. This new technique leverages the high-order quasi-Helmholtz projectors and is based on the definition of a dual space that is built from a space of higher order. The preconditioned system is also stable at low-frequency, and we presented an extension of the scheme for multiply-connected geometries, for which a null-space associated with the topological loops of the geometry was originally present. We gave additional details about the implementation of the method and we introduced a technique to partially-orthonormalize the basis functions, which allows to stabilize the system with respect to the choice of the basis functions without the need for any expensive computation of inverse square-root matrices. Final experiments allowed to corroborate our theory and to showcase the effectiveness of our approach in providing well-conditioned systems, solvable in a small and stable number of iterations.

Chapter 8

Conclusion and Perspectives

In this thesis, we presented new preconditioning techniques allowing to tackle the low-frequency and dense-refinement breakdowns of the electric field integral equation (EFIE). These strategies enable the computation of well-conditioned linear systems that can subsequently be solved in few iterations when resorting to iterative solvers.

In Chapters 1 and 2, we recapped the derivation of some surface integral equations and their discretization via the boundary element method (BEM) allowing us to define the linear systems to be solved in the subsequent chapters. This includes the introduction of high-order boundary elements which were the focus of an important part of that PhD. In Chapter 3, we introduced the background in linear algebra and the key concepts to tailor efficient fast iterative solvers, including the implementation of fast matrix-vector product algorithms in the context of integral equations and their use in conjunction with preconditioning techniques.

In Chapter 4, we focused on the generalization of the quasi-Helmholtz projectors to precondition the EFIE when modelling PEC boundaries with junctions. To our knowledge, this is the first time a quasi-Helmholtz decomposition approach is applied to the preconditioning of such structures without any modification of the degrees of freedom setting. We then showed that the use of these projectors allows to properly rescale the solenoidal and non-solenoidal contributions of the EFIE—which provides a matrix system with a bounded condition number at low-frequency—and we verified that algebraic multigrid can be used efficiently to compute the projectors in the presence of junctions.

In Chapter 5, we introduced the first ever proposed definition of the quasi-Helmholtz projectors within the framework of high-order boundary element discretization. Notably, we proved their uniqueness given a particular choice and ordering of basis functions, and their well-posedness in providing a quasi-Helmholtz decomposition of high-order Raviart-Thomas elements. We subsequently showed the effectiveness of the quasi-Helmholtz projectors stabilization technique to precondition the EFIE at low-frequency when discretized with high-order basis functions.

The definition of these high-order quasi-Helmholtz projectors has been the key contribution of this thesis. In particular, it allowed us to define two other high-order compatible preconditioning techniques, that are effective to cure both low-frequency and dense-refinement breakdowns of the EFIE.

The first technique to cure the dense-refinement breakdown was presented in Chapter 6. It consists in filtering the discretized Laplacian operator matrices appearing in the definition of the quasi-Helmholtz projectors, which, combined with a wavelet approach, enables the construction of a successful preconditioner for the EFIE. In particular, we showed that it allows to get a bounded spectrum and an almost constant number of iterations at a given precision threshold

for increasingly dense discretization as well as for decreasing frequency.

The second presented technique that cures the dense refinement breakdown belongs to the class of Calderón preconditioners and was introduced in Chapter 7. Thanks to the use of the quasi-Helmholtz projectors, we have constructed for the first time a dual space that is based on p-refinement rather than mesh-refinement, and that is therefore well-suited for high-order frameworks. Several numerical experiments were given to corroborate that theory.

Moreover, these two last techniques introduced the need for an orthonormalization of the basis functions, which relies on the iterative computation of inverse square-roots of Gram matrices—the matrices that evaluate the inner products between the basis functions. In Chapter 7, we also described an algorithm to replace these operations by partial-orthonormalization of the basis functions, which relies on the cheaper computation of sparse matrices. While these developments are as of now empirical, we have shown numerically that the well-conditioning of the Calderón formulation is relatively preserved when resorting to partial- instead of full-orthonormalization.

Promising extensions of this work are currently investigated:

- While in each of these chapters we have discussed the algorithmic efficiency of these approaches, some implementation efforts remain to be done in order to get an overall fast solver. Notably, the hybridization of these techniques with fast matrix-vector products algorithms and high-performance computing paradigms are still under developments and have not been thoroughly studied in this thesis.

- The extension of these preconditioning techniques to the other formulations introduced in Chapter 1 are under investigation. The reason why we focused on the EFIE is that the integral operator appearing in that formulation is also part of the PMCHWT and CFIE formulations, and is the main cause of their ill-conditioning. Previous works at order zero have shown that these equations can be preconditioned with few modifications of the EFIE preconditioning strategies [2], and the same adaptations should be effective with our high-order schemes. Additionally, the definition of a new dual space based on p-refinement can find other applications than the discretization of the Calderón identity. We are currently investigating its use to discretize the MFIE operator, accordingly to the applications that have been found at order zero with Buffa-Christiansen basis functions [46].

- In Chapter 1, we have discussed the difficulties of preconditioning problems arising from the single-trace modelling of dielectric objects when there are junctions between multiple interfaces. As it has been done in Chapter 4, in the context of metallic junctions, an extension of the quasi-Helmholtz projectors to the context of dielectric junctions could be investigated, that would allow to define low-frequency and dense-refinement preconditioning strategies for very arbitrary problems. Alternatively, the extension of our methods to the context of multi-trace modelling could rely on existing developments done at order zero [43, 47].

Appendix A

Physics Quantities

Light-speed in the vacuum: $c = 299792458\ \mathrm{m{\cdot}s^{-1}}$
Magnetic permeability of the vacuum: $\mu_0 = 4\pi \times 10^{-7} \approx 1.2566370614 \times 10^{-6}\ \mathrm{H{\cdot}m^{-1}}$
Electric permittivity of the vacuum: $\varepsilon_0 = (\mu_0 c^2)^{-1} \approx 8.8541878176 \times 10^{-12}\ \mathrm{F{\cdot}m^{-1}}$

Magnetic permeability of the material: $\mu = \mu_0 \mu_r\ (\mathrm{H{\cdot}m^{-1}})$ $[\mu_r \in \mathbb{C}$ in frequency-domain$]$
Electric permittivity of the material: $\varepsilon = \varepsilon_0 \varepsilon_r\ (\mathrm{F{\cdot}m^{-1}})$ $[\varepsilon_r \in \mathbb{C}$ in frequency-domain$]$
Characteristic impedance of the material: $\eta = \sqrt{\mu/\varepsilon}\ (\Omega)$

Frequency: f (Hz)
Angular frequency: $\omega = 2\pi f\ (\mathrm{rad{\cdot}s^{-1}})$
Wavenumber: $k = \omega\sqrt{\mu\varepsilon}\ (\mathrm{rad{\cdot}m^{-1}})$
Wavelength: $\lambda = 2\pi/k$ (m)

Electric field: $\boldsymbol{E}\ (\mathrm{V{\cdot}m^{-1}})$
Magnetic field: $\boldsymbol{H}\ (\mathrm{A{\cdot}m^{-1}})$

Electric induction: $\boldsymbol{D}\ (\mathrm{C{\cdot}m^{-2}})$
Magnetic induction: $\boldsymbol{B}\ (\mathrm{Wb{\cdot}m^{-2}} = \mathrm{T})$

Electric surface current density: $\boldsymbol{J}\ (\mathrm{A{\cdot}m^{-1}})$
Magnetic surface current density: $\boldsymbol{M}\ (\mathrm{V{\cdot}m^{-1}})$

Electric surface charge density: $\rho_e\ (\mathrm{C{\cdot}m^{-2}})$
Magnetic surface charge density: $\rho_h\ (\mathrm{Wb{\cdot}m^{-2}} = \mathrm{T})$

Bibliography

[1] Simon B Adrian, Francesco Paolo Andriulli, and Thomas F Eibert. "On a refinement-free Calderón multiplicative preconditioner for the electric field integral equation". In: *Journal of Computational Physics* 376 (2019), pp. 1232–1252.

[2] Simon B Adrian et al. "Electromagnetic integral equations: Insights in conditioning and preconditioning". In: *IEEE Open Journal of Antennas and Propagation* 2 (2021), pp. 1143–1174.

[3] Simon B. Adrian, Francesco P. Andriulli, and Thomas F. Eibert. "Hierarchical Bases Preconditioners for the Electric Field Integral Equation on Multiply Connected Geometries". In: *IEEE Transactions on Antennas and Propagation* 62.11 (2014), pp. 5856–5861. DOI: 10.1109/TAP.2014.2347392.

[4] F. P. Andriulli, A. Tabacco, and G. Vecchi. "A Multiresolution Approach to the Electric Field Integral Equation in Antenna Problems". In: *SIAM Journal on Scientific Computing* 29.1 (2007), pp. 1–21. DOI: 10.1137/050634943. eprint: https://doi.org/10.1137/050634943. URL: https://doi.org/10.1137/050634943.

[5] Francesco Andriulli et al. "Projection Based Quasi-Helmholtz Decompositions for Well-Conditioned Integral Equations Without the Search for Global Loops". In: *ACES 2013 : the 29th International Review of Progress in Applied Computational Electromagnetics*. Monterey, United States, Apr. 2013.

[6] Francesco P. Andriulli. "Loop-Star and Loop-Tree Decompositions: Analysis and Efficient Algorithms". In: *IEEE Transactions on Antennas and Propagation* 60.5 (2012), pp. 2347–2356. DOI: 10.1109/TAP.2012.2189723.

[7] Francesco P. Andriulli et al. "A Multiplicative Calderón Preconditioner for the Electric Field Integral Equation". In: *IEEE Transactions on Antennas and Propagation* 56.8 (2008), pp. 2398–2412. DOI: 10.1109/TAP.2008.926788.

[8] Francesco P. Andriulli et al. "On a Well-Conditioned Electric Field Integral Operator for Multiply Connected Geometries". In: *IEEE Transactions on Antennas and Propagation* 61.4 (2013), pp. 2077–2087. DOI: 10.1109/TAP.2012.2234072.

[9] Xavier Antoine and Marion Darbas. "An Introduction to Operator Preconditioning for the Fast Iterative Integral Equation Solution of Time-Harmonic Scattering Problems". In: *Multiscale Science and Engineering* 3 (Feb. 2021). DOI: 10.1007/s42493-021-00057-6.

[10] Wolfgang Arendt et al. "Weyl's Law: Spectral Properties of the Laplacian in Mathematics and Physics". In: 2009.

[11] Owe Axelsson. *Iterative solution methods*. Cambridge university press, 1996.

[12] Mario Bebendorf. "Approximation of boundary element matrices". In: *Numerische Mathematik* 86 (2000), pp. 565–589. URL: https://api.semanticscholar.org/CorpusID:206858339.

[13] Yves Beghein et al. "On a Low-Frequency and Refinement Stable PMCHWT Integral Equation Leveraging the Quasi-Helmholtz Projectors". en. In: *IEEE Transactions on Antennas and Propagation* 65.10 (Oct. 2017), pp. 5365–5375. ISSN: 0018-926X, 1558-2221. DOI: 10.1109/TAP.2017.2738061. (Visited on 06/08/2022).

[14] Adi Ben-Israel and Thomas NE Greville. *Generalized inverses: theory and applications*. Vol. 15. Springer Science & Business Media, 2003.

[15] Jean-Pierre Berenger. "A perfectly matched layer for the absorption of electromagnetic waves". In: *Journal of Computational Physics* 114.2 (1994), pp. 185–200. ISSN: 0021-9991. DOI: https://doi.org/10.1006/jcph.1994.1159. URL: https://www.sciencedirect.com/science/article/pii/S0021999184711594.

[16] Sidi Biha. "Formalisation des mathématiques : une preuve du théorème de Cayley-Hamilton". In: (Jan. 2008).

[17] Daniele Boffi. "Finite element approximation of eigenvalue problems". In: *Acta Numerica* 19 (2010), pp. 1–120.

[18] Daniele Boffi, Franco Brezzi, and Michel Fortin. *Mixed Finite Element Methods and Applications*. Springer Berlin, Heidelberg, 2013. ISBN: 978-3-642-36519-5. DOI: doi.org/10.1007/978-3-642-36519-5.

[19] Matthias Bolten et al. "Algebraic Multigrid Methods for Laplacians of Graphs". In: *Linear Algebra and Its Applications - LINEAR ALGEBRA APPL* 434 (Nov. 2009). DOI: 10.1016/j.laa.2010.11.008.

[20] Steffen Börm, Lars Grasedyck, and Wolfgang Hackbusch. "Introduction to hierarchical matrices with applications". In: *Engineering Analysis with Boundary Elements* 27.5 (2003). Large scale problems using BEM, pp. 405–422. ISSN: 0955-7997. DOI: https://doi.org/10.1016/S0955-7997(02)00152-2. URL: https://www.sciencedirect.com/science/article/pii/S0955799702001522.

[21] J. Bourhis, A. Merlini, and F.P. Andriulli. "High-order quasi-Helmholtz Projectors: Definition, Analyses, Algorithms". In: *XXXVth General Assembly and Scientific Symposium of the International Union of Radio Science*. 2023.

[22] Johann Bourhis, Adrien Merlini, and Francesco P Andriulli. "High-order quasi-Helmholtz Projectors: Definition, Analyses, Algorithms". In: *IEEE Transactions on Antennas and Propagation* (2024).

[23] Johann Bourhis, Adrien Merlini, and Francesco P. Andriulli. "On the Use of Quasi-Helmholtz Projectors for Modeling Structures With Junctions at Arbitrarily Low Frequencies". In: *IEEE Transactions on Antennas and Propagation* 72.10 (2024), pp. 8097–8102. DOI: 10.1109/TAP.2024.3439772.

[24] Johann Bourhis, Andrien Merlini, and Francesco P. Andriulli. "On the Effectiveness of Quasi-Helmholtz Projectors in Preconditioning Problems with Junctions". In: *2022 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (AP-S/URSI)*. 2022, pp. 1944–1945. DOI: 10.1109/AP-S/USNC-URSI47032.2022.9886184.

[25] Johann Bourhis et al. “On a Novel Calderón Preconditioning Strategy Based on High-Order Quasi-Helmholtz Projectors”. In: *2024 International Conference on Electromagnetics in Advanced Applications (ICEAA)*. 2024, pp. 700–703. DOI: 10.1109/ICEAA61917.2024.10701907.

[26] Johann Bourhis et al. “On Operator Filtering for Integral Equations: The High-Order Case”. In: *2024 IEEE International Symposium on Antennas and Propagation and INC/USNC-URSI Radio Science Meeting (AP-S/INC-USNC-URSI)*. 2024, pp. 221–222. DOI: 10.1109/AP-S/INC-USNC-URSI52054.2024.10686349.

[27] Annalisa Buffa and Snorre Christiansen. “The electric field integral equation on Lipschitz screens: Definitions and numerical approximation”. In: *Numerische Mathematik* 94 (Jan. 2003), pp. 229–267. DOI: 10.1007/s00211-002-0422-0.

[28] Annalisa Buffa and Snorre H. Christiansen. “A dual finite element complex on the barycentric refinement”. In: *Comptes Rendus Mathematique* 340.6 (2005), pp. 461–464. ISSN: 1631-073X. DOI: https://doi.org/10.1016/j.crma.2004.12.022.

[29] Annalisa Buffa and Patrick Ciarlet. “On traces for functional spaces related to Maxwell’s equations Part II: Hodge decompositions on the boundary of Lipschitz polyhedra and applications”. In: *Mathematical Methods in the Applied Sciences* 24.1 (2001), pp. 31–48. DOI: 10.1002/1099-1476(20010110)24:1(31::aid-mma193)3.0.co;2-x. URL: https://ensta-paris.hal.science/hal-00878224.

[30] Annalisa Buffa, Martin Costabel, and Dongwoo Sheen. “On traces for H(curl,Ω) in Lipschitz domains”. In: *Journal of Mathematical Analysis and Applications* 276 (2002), pp. 845–867.

[31] Annalisa Buffa et al. “Multipatch approximation of the de Rham sequence and its traces in isogeometric analysis”. In: *Numerische Mathematik* 144 (Jan. 2020). DOI: 10.1007/s00211-019-01079-x.

[32] M Burton and S Kashyap. “A study of a recent, moment-method algorithm that is accurate to very low frequencies”. In: *Applied Computational Electromagnetics Society Journal* 10 (1995), pp. 58–68.

[33] Bruno Carpentieri et al. *Combining Fast Multipole Techniques and an Approximate Inverse Preconditioner for Large Electromagnetism Calculations*. Jan. 2005.

[34] M. Carr, E. Topsakal, and J.L. Volakis. “A procedure for modeling material junctions in 3-D surface integral equation approaches”. In: *IEEE Transactions on Antennas and Propagation* 52.5 (2004), pp. 1374–1378. DOI: 10.1109/TAP.2004.827247.

[35] Michel Cessenat. *Mathematical methods in electromagnetism problems. North-Holland*. World Scientific, 1986.

[36] Stéphanie Chaillat, Luca Desiderio, and Patrick Ciarlet. “Theory and implementation of HH-matrix based iterative and direct solvers for Helmholtz and elastodynamic oscillatory kernels”. In: *Journal of Computational Physics* 351 (Dec. 2017), pp. 165–186. DOI: 10.1016/j.jcp.2017.09.013.

[37] Rui Chen, Adrien Merlini, and Francesco Andriulli. *On the Computation of Square Roots and Inverse Square Roots of Gram Matrices for Surface Integral Equations in Electromagnetics*. May 2024.

[38] Wen-Jing Chen et al. “Improved A-EFIE System for Electromagnetic Simulation in Low Frequency Regime”. In: *IEEE Antennas and Wireless Propagation Letters* 21.9 (2022), pp. 1752–1756. DOI: 10.1109/LAWP.2022.3179270.

[39] Weng Cho Chew, Mei Song Tong, and Bin Hu. "Integral equation methods for electromagnetic and elastic waves". In: *Synthesis Lectures on Computational Electromagnetics* 3.1 (2008), pp. 1–241.

[40] Weng Cho Chew et al. *Fast and Efficient Algorithms in Computational Electromagnetics*. Artech House, 2001.

[41] Snorre Christiansen. "Discrete Fredholm properties and convergence estimates for the Electric Field Integral Equation". In: *Math. Comput.* 73 (Jan. 2004), pp. 143–167. DOI: 10.1090/S0025-5718-03-01581-3.

[42] Snorre H Christiansen and Jean-Claude Nédélec. "A preconditioner for the electric field integral equation based on Calderon formulas". In: *SIAM journal on numerical analysis* 40.3 (2002), pp. 1100–1135.

[43] Xavier Claeys et al. "Multi-trace boundary integral equations". In: *Radon Ser. Comput. Appl. Math.* 14 (Jan. 2013), pp. 51–100.

[44] Harry Contopanagos et al. "Well-conditioned boundary integral equations for three-dimensional electromagnetic scattering". In: *IEEE Transactions on Antennas and Propagation* 50.12 (2002), pp. 1824–1830.

[45] K. Cools, F.P. Andriulli, and F. Olyslager. "A Calderón preconditioned PMCHWT equation". In: *2009 International Conference on Electromagnetics in Advanced Applications*. 2009, pp. 521–524. DOI: 10.1109/ICEAA.2009.5297380.

[46] K. Cools et al. "Accurate and Conforming Mixed Discretization of the MFIE". In: *IEEE Antennas and Wireless Propagation Letters* 10 (2011), pp. 528–531. DOI: 10.1109/LAWP.2011.2155022.

[47] Kristof Cools and Carolina Urzúa-Torres. "Block diagonal Calderón preconditioning for scattering at multi-screens". In: *BIT Numerical Mathematics* 64 (Sept. 2024). DOI: 10.1007/s10543-024-01034-9.

[48] Kristof Cools et al. "Nullspaces of MFIE and Calderón Preconditioned EFIE Operators Applied to Toroidal Surfaces". In: *IEEE Transactions on Antennas and Propagation* 57.10 (2009), pp. 3205–3215. DOI: 10.1109/TAP.2009.2028669.

[49] Nair Maria Maia de Abreu. "Old and new results on algebraic connectivity of graphs". In: *Linear Algebra and its Applications* 423.1 (2007). Special Issue devoted to papers presented at the Aveiro Workshop on Graph Spectra, pp. 53–73. ISSN: 0024-3795. DOI: https://doi.org/10.1016/j.laa.2006.08.017.

[50] M. Djordjevic and B.M. Notaros. "Double higher order method of moments for surface integral equation modeling of metallic and dielectric antennas and scatterers". In: *IEEE Transactions on Antennas and Propagation* 52.8 (Aug. 2004). Conference Name: IEEE Transactions on Antennas and Propagation, pp. 2118–2129. ISSN: 1558-2221. DOI: 10.1109/TAP.2004.833175.

[51] C.T. Dodson and P.E. Parker. *A User's Guide to Algebraic Topology*. Springer New York, NY, 1997. ISBN: 978-0-7923-4293-9.

[52] Francesca Vipiana Donald R. Wilton and William A. Johnson. "Evaluating Singular, Near-Singular, and Non-Singular Integrals on Curvilinear Elements". In: *Electromagnetics* 34.3-4 (2014), pp. 307–327. DOI: 10.1080/02726343.2014.877775. eprint: https://doi.org/10.1080/02726343.2014.877775. URL: https://doi.org/10.1080/02726343.2014.877775.

[53] TF Eibert. "Iterative-solver convergence for loop-star and loop-tree decompositions in method-of-moments solutions of the electric-field integral equation". In: *IEEE Antennas and Propagation Magazine* 46.3 (2004), pp. 80–85.

[54] N. Engheta et al. "The fast multipole method (FMM) for electromagnetic scattering problems". In: *IEEE Transactions on Antennas and Propagation* 40.6 (1992), pp. 634–641. DOI: 10.1109/8.144597.

[55] Charles L Epstein and Leslie Greengard. "Debye sources and the numerical solution of the time harmonic Maxwell equations". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 63.4 (2010), pp. 413–463.

[56] Ignacia Fierro-Piccardo and Timo Betcke. "An OSRC Preconditioner for the EFIE". In: *IEEE Transactions on Antennas and Propagation* 71.4 (2023), pp. 3408–3417. DOI: 10.1109/TAP.2023.3236762.

[57] Roland W. Freund. "Transpose-Free Quasi-Minimal Residual Methods for Non-Hermitian Linear Systems". In: *Recent Advances in Iterative Methods*. Ed. by Gene Golub, Mitchell Luskin, and Anne Greenbaum. New York, NY: Springer New York, 1994, pp. 69–94. ISBN: 978-1-4613-9353-5.

[58] Roland W. Freund and N.M. Nachtigall. "QMR: a quasi-minimal residual method for non-Hermitian linear systems." In: *Numerische Mathematik* 60.3 (1992), pp. 315–340. URL: http://eudml.org/doc/133596.

[59] Xin Fu et al. "Generalized Debye Sources-Based EFIE Solver on Subdivision Surfaces". In: *IEEE Transactions on Antennas and Propagation* 65.10 (2017), pp. 5376–5386. DOI: 10.1109/TAP.2017.2740976.

[60] Christophe Geuzaine and Jean-Francois Remacle. *Gmsh*. Version 4.6.0. June 22, 2020. URL: http://http://gmsh.info/.

[61] Walton C. Gibson. *The method of moments in Electromagnetics*. Champan & Hall/CRC Press, 2021.

[62] José Ma Gil. "An Efficient Framework for Using Higher-Order Curved Div-Conforming Elements With Cancellation of Weak Singularities in Local Space for Surface Integral Equation". In: *IEEE Transactions on Antennas and Propagation* 60.8 (Aug. 2012). Conference Name: IEEE Transactions on Antennas and Propagation, pp. 3736–3743. ISSN: 1558-2221. DOI: 10.1109/TAP.2012.2201084.

[63] José Ma Gil. "Higher-order curved div-conforming elements with loop-star subspaces decomposition for Surface Integral Equation". In: *The 8th European Conference on Antennas and Propagation (EuCAP 2014)*. ISSN: 2164-3342. Apr. 2014, pp. 1139–1141. DOI: 10.1109/EuCAP.2014.6901972.

[64] Viviana Giunzioni et al. *Low-Frequency Stabilizations of the PMCHWT Equation for Dielectric and Conductive Media: On a Full-Wave Alternative to Eddy-Current Solvers*. 2024. arXiv: 2408.01321 [math.NA].

[65] Gene Golub and Charles Loan. *Matrix Computations*. Johns Hopkins University Press, Oct. 1996. ISBN: 0801854148.

[66] Michiel Gossye, Dries Vande Ginste, and Hendrik Rogier. "Electromagnetic modeling of high magnetic contrast media using Calderón preconditioning". In: *Computers & Mathematics with Applications* 77.6 (2019). 7th International Conference on Advanced Computational Methods in Engineering (ACOMEN 2017), pp. 1626–1638. ISSN: 0898-1221. DOI: https://doi.org/10.1016/j.camwa.2018.08.045. URL: https://www.sciencedirect.com/science/article/pii/S0898122118304711.

[67] R.D. Graglia, D.R. Wilton, and A.F. Peterson. "Higher order interpolatory vector bases for computational electromagnetics". In: *IEEE Transactions on Antennas and Propagation* 45.3 (Mar. 1997). Conference Name: IEEE Transactions on Antennas and Propagation, pp. 329–342. ISSN: 1558-2221. DOI: 10.1109/8.558649.

[68] Roberto D. Graglia and Guido Lombardi. "Singular Higher Order Divergence-Conforming Bases of Additive Kind and Moments Method Applications to 3D Sharp-Wedge Structures". In: *IEEE Transactions on Antennas and Propagation* 56.12 (Dec. 2008). Conference Name: IEEE Transactions on Antennas and Propagation, pp. 3768–3788. ISSN: 1558-2221. DOI: 10.1109/TAP.2008.2007390.

[69] Roberto D. Graglia and Andrew F. Peterson. *Higher-Order Techniques in Computational Electromagnetics*. English. Edison, NJ: Scitech Publishing, Nov. 2015. ISBN: 978-1-61353-016-0.

[70] Anne Greenbaum. "Open Problems in the Analysis of Krylov Subspace Methods". In: *Notices of the American Mathematical Society* 70 (May 2023), p. 1. DOI: 10.1090/noti2683.

[71] Jonathan L. Gross and Jay Yellen. *Graph Theory and Its Applications*. Chapman and Hall/CRC, 2005.

[72] Janitha Gunatilake. "An improved multigrid solver for the p-hierarchical basis finite element method using a space decomposition smoother". In: *Computers & Mathematics with Applications* 124 (Oct. 2022), pp. 52–62. DOI: 10.1016/j.camwa.2022.08.019.

[73] Martin Gutknecht. *A Brief Introduction to Krylov Space Methods for Solving Linear Systems*. Jan. 2007, pp. 53–62. ISBN: 978-3-540-46373-3. DOI: 10.1007/978-3-540-46375-7_5.

[74] Ilari Hanninen, Matti Taskinen, and Jukka Sarvas. "Singularity subtraction integral formulae for surface integral equations with RWG, rooftop and hybrid basis functions". In: *Progress in Electromagnetics Research-pier - PROG ELECTROMAGN RES* 63 (Jan. 2006), pp. 243–278. DOI: 10.2528/PIER06051901.

[75] Roger F. Harrington. "The Method of Moments in Electromagnetics". In: *Journal of Electromagnetic Waves and Applications* 1 (1987), pp. 181–200. URL: https://api.semanticscholar.org/CorpusID:120955246.

[76] Brian Helenbrook, Dimitri Mavriplis, and Harold Atkins. "Analysis of "p"-Multigrid for Continuous and Discontinuous Finite Element Discretizations". In: *16th AIAA Computational Fluid Dynamics Conference* (June 2003). DOI: 10.2514/6.2003-3989.

[77] Magnus R. Hestenes and Eduard Stiefel. "Methods of conjugate gradients for solving linear systems". In: *Journal of research of the National Bureau of Standards* 49 (1952), pp. 409–435. URL: https://api.semanticscholar.org/CorpusID:2207234.

[78] J.J. Heys et al. "Algebraic multigrid for higher-order finite elements". In: *Journal of Computational Physics* 204.2 (2005), pp. 520–532. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2004.10.021.

[79] R. Hiptmair. "Operator Preconditioning". In: *Computers & Mathematics with Applications* 52.5 (2006). Hot Topics in Applied and Industrial Mathematics, pp. 699–706. ISSN: 0898-1221. DOI: https://doi.org/10.1016/j.camwa.2006.10.008.

[80] Ralf Hiptmair et al. "Domain decomposition for boundary integral equations via local multi-trace formulations". English (US). In: *Domain Decomposition Methods in Science and Engineering XXI*. Ed. by Taoufik Sassi et al. Lecture Notes in Computational Science and Engineering. Germany: Springer, 2014, pp. 43–57. DOI: 10.1007/978-3-319-05789-7_4.

[81] Bernd Hofmann et al. "An Excitation-Aware and Self-Adaptive Frequency Normalization for Low-Frequency Stabilized Electric Field Integral Equation Formulations". In: *IEEE Transactions on Antennas and Propagation* 71.5 (2023), pp. 4301–4314. DOI: 10.1109/TAP.2023.3247896.

[82] G.C. Hsiao and R.E. Kleinman. "Mathematical foundations for error estimation in numerical solutions of integral equations in electromagnetics". In: *IEEE Transactions on Antennas and Propagation* 45.3 (1997), pp. 316–328. DOI: 10.1109/8.558648.

[83] Immo Huismann, Jörg Stiller, and Jochen Fröhlich. "Scaling to the stars – a linearly scaling elliptic solver for p-multigrid". In: *Journal of Computational Physics* 398 (2019), p. 108868. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2019.108868.

[84] Chihiro Iwamura et al. "An efficient algebraic multigrid preconditioned conjugate gradient solver". In: *Computer Methods in Applied Mechanics and Engineering* 192 (2003), pp. 2299–2318. ISSN: 0045-7825. DOI: 10.1016/S0045-7825(02)00378-X.

[85] S. Jarvenpaa, M. Taskinen, and P. Yla-Oijala. "Singularity subtraction technique for high-order polynomial vector basis functions on planar triangles". In: *IEEE Transactions on Antennas and Propagation* 54.1 (Jan. 2006). Conference Name: IEEE Transactions on Antennas and Propagation, pp. 42–49. ISSN: 1558-2221. DOI: 10.1109/TAP.2005.861556.

[86] W. J. Kammerer and M. Z. Nashed. "On the Convergence of the Conjugate Gradient Method for Singular Linear Operator Equations". In: *SIAM Journal on Numerical Analysis* 9.1 (1972), pp. 165–181. ISSN: 00361429.

[87] Jin-Fa Lee, Robert Lee, and Robert J Burkholder. "Loop star basis functions and a robust preconditioner for EFIE scattering problems". In: *IEEE Transactions on Antennas and Propagation* 51.8 (2003), pp. 1855–1863.

[88] Marc Lenoir. *Équations intégrales et problèmes de diffraction*. ENSTA, Unité de Mathématiques Appliquées. 2011.

[89] Guido Lombardi and Roberto D. Graglia. "Modeling Junctions in Sharp Edge Conducting Structures With Higher Order Method of Moments". In: *IEEE Transactions on Antennas and Propagation* 62.11 (2014), pp. 5723–5731. DOI: 10.1109/TAP.2014.2355855.

[90] Johannes F. Lutzeyer and Andrew T. Walden. "Comparing Graph Spectra of Adjacency and Laplacian Matrices". In: *arXiv: Methodology* (2017).

[91] Anne Marsden. "Eigenvalues of the laplacian and their relationship to the connectedness of a graph". In: *University of Chicago, REU* (2013).

[92] B. Mercier et al. "Eigenvalue Approximation by Mixed and Hybrid Methods". In: *Mathematics of Computation* 36.154 (1981), pp. 427–453. ISSN: 00255718, 10886842. (Visited on 01/02/2024).

[93] Adrien Merlini. "Unified computational frameworks bridging low to high frequency simulations: fast and high fidelity modelling from brain to radio-frequency scenarios". PhD thesis. Ecole nationale supérieure Mines-Télécom Atlantique, 2019.

[94] Adrien Merlini et al. "Laplacian filtered loop-star decompositions and quasi-helmholtz filters: Definitions, analysis, and efficient algorithms". In: *IEEE Transactions on Antennas and Propagation* (2023).

[95] Russell Merris. "Laplacian matrices of graphs: a survey". In: *Linear algebra and its applications* 197 (1994), pp. 143–176.

[96] Carl Meyer. *Matrix Analysis and Applied Linear Algebra Second Edition*. SIAM, May 2023. ISBN: 1611977703.

[97] Eric Michielssen and Amir Boag. "A multilevel matrix decomposition algorithm for analyzing scattering from large structures". In: *IEEE Transactions on Antennas and Propagation* 44 (1996), pp. 1086–1093. URL: https://api.semanticscholar.org/CorpusID:121979910.

[98] Bojan Mohar. "Eigenvalues, diameter, and mean distance in graphs". In: *Graphs and Combinatorics* 7 (1991), pp. 53–64.

[99] Stephan Mohr, Yuji Nakatsukasa, and Carolina Urzúa-Torres. "Full operator preconditioning and the accuracy of solving linear systems". In: *IMA Journal of Numerical Analysis* (Jan. 2024). DOI: 10.1093/imanum/drad104.

[100] Peter Monk. *Finite Element Methods for Maxwell's Equations*. Oxford University Press, Apr. 2003. ISBN: 9780198508885. DOI: 10.1093/acprof:oso/9780198508885.001.0001. URL: https://doi.org/10.1093/acprof:oso/9780198508885.001.0001.

[101] Hadrien Montanelli, Matthieu Aussal, and Houssem Haddar. "Computing Weakly Singular and Near-Singular Integrals Over Curved Boundary Elements". In: *SIAM Journal on Scientific Computing* 44.6 (2022), A3728–A3753. DOI: 10.1137/21M1462027. eprint: https://doi.org/10.1137/21M1462027. URL: https://doi.org/10.1137/21M1462027.

[102] N. V. Nair et al. "A singularity cancellation technique on arbitrary higher order patch descriptions". In: *Proceedings of the 2012 IEEE International Symposium on Antennas and Propagation*. 2012, pp. 1–2. DOI: 10.1109/APS.2012.6348615.

[103] Artem Napov and Yvan Notay. "Algebraic Multigrid for Moderate Order Finite Elements". In: *SIAM Journal on Scientific Computing* 36.4 (2014), A1678–A1707. DOI: 10.1137/130922616. eprint: https://doi.org/10.1137/130922616.

[104] Artem Napov and Yvan Notay. "An Algebraic Multigrid Method with Guaranteed Convergence Rate". In: *SIAM Journal on Scientific Computing* 34.2 (2012), A1079–A1109. DOI: 10.1137/100818509. eprint: https://doi.org/10.1137/100818509.

[105] J.C. Nédélec. "Mixed Finite Elements in IR3 ." In: *Numerische Mathematik* 35 (1980), pp. 315–342.

[106] Jean-Claude Nédélec. *Acoustic and electromagnetic equations: integral representations for harmonic problems*. Vol. 144. Springer, 2001.

[107] Jorge Nocedal and Stephen J. Wright. "Numerical Optimization". In: *Fundamental Statistical Inference*. 2018. URL: https://api.semanticscholar.org/CorpusID:189864167.

[108] Branislav M. Notaros. "Higher Order Frequency-Domain Computational Electromagnetics". en. In: *IEEE Transactions on Antennas and Propagation* 56.8 (Aug. 2008), pp. 2251–2276. ISSN: 0018-926X. DOI: 10.1109/TAP.2008.926784. (Visited on 06/27/2022).

[109] Yvan Notay. "An aggregation-based algebraic multigrid method". In: *Electron. Trans. Numer. Anal.* 37 (2010), pp. 123–146.

[110] Günther Of, Olaf Steinbach, and Wolfgang Wendland. "Boundary Element Tearing and Interconnecting Domain Decomposition Methods". In: vol. 2006. Nov. 2006, pp. 461–490. ISBN: 978-3-540-34959-4. DOI: 10.1007/978-3-540-34961-7_14.

[111] Soumaya Oueslati et al. "Analytical method for the evaluation of singular integrals arising from boundary element method in electromagnetism". In: *International Journal of Numerical Modelling Electronic Networks Devices and Fields* 34 (Aug. 2020). DOI: 10.1002/jnm.2792.

[112] Xiao-Min Pan, Lu Cai, and Xin-Qing Sheng. "An efficient high order multilevel fast multipole algorithm for electromagnetic scattering analysis". en. In: *Progress In Electromagnetics Research* 126 (2012), pp. 85–100. ISSN: 1559-8985. DOI: 10.2528/PIER12020203. (Visited on 06/27/2022).

[113] John W. Pearson and Jennifer Pestana. "Preconditioners for Krylov subspace methods: an overview". English. In: *GAMM-Mitteilungen / GAMM-Reports* 43.4 (Nov. 2020). ISSN: 0936-7195. DOI: 10.1002/gamm.202000015.

[114] A.F. Peterson and M.M. Bibby. "Error trends in higher-order discretizations of the EFIE and MFIE". In: *2005 IEEE Antennas and Propagation Society International Symposium*. Vol. 3A. 2005, 52–55 vol. 3A. DOI: 10.1109/APS.2005.1552171.

[115] Andrew Peterson. "Mapped Vector Basis Functions for Electromagnetic Integral Equations". In: *Synthesis Lectures on Computational Electromagnetics* 1 (Dec. 2005), pp. 1–124. DOI: 10.2200/S00008ED1V01Y200508CEM001.

[116] Athanasios G. Polimeridis et al. "DIRECTFN: Fully Numerical Algorithms for High Precision Computation of Singular Integrals in Galerkin SIE Methods". In: *IEEE Transactions on Antennas and Propagation* 61.6 (June 2013). Conference Name: IEEE Transactions on Antennas and Propagation, pp. 3112–3122. ISSN: 1558-2221. DOI: 10.1109/TAP.2013.2246854.

[117] Zhi-Guo Qian and Weng Cho Chew. "Enhanced A-EFIE With Perturbation Method". In: *IEEE Transactions on Antennas and Propagation* 58.10 (2010), pp. 3256–3264. DOI: 10.1109/TAP.2010.2055795.

[118] Sadasiva M. Rao, Donald R. Wilton, and Allen W. Glisson. "Electromagnetic scattering by surfaces of arbitrary shape". In: *IEEE Transactions on Antennas and Propagation* 30.3 (1982), pp. 409–418. DOI: 10.1109/TAP.1982.1142818.

[119] Sergej Rjasanow and Lucy Weggler. "ACA accelerated high order BEM for Maxwell problems". In: *Computational Mechanics* 51 (Apr. 2013). DOI: 10.1007/s00466-012-0731-9.

[120] John Ruge and Klaus Stüben. *Efficient solution of finite difference and finite element equations by algebraic multigrid AMG*. Gesellschaft f. Mathematik u. Datenverarbeitung, 1984.

[121] Youcef Saad and Martin H. Schultz. "GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems". In: *SIAM Journal on Scientific and Statistical Computing* 7.3 (1986), pp. 856–869. DOI: 10.1137/0907058. eprint: https://doi.org/10.1137/0907058. URL: https://doi.org/10.1137/0907058.

[122] Yousef Saad. *Iterative Methods for Sparse Linear Systems*. Jan. 2003. DOI: 10.1137/1.9780898718003.ch4.

[123] S.A. Sauter and C. Schwab. *Boundary Element Methods*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2010. ISBN: 9783540680932.

[124] Joachim Schöberl and Sabine Zaglmayr. "High order Nédélec elements with local complete sequence properties". In: *COMPEL - The international journal for computation and mathematics in electrical and electronic engineering* 24.2 (Jan. 2005). Publisher: Emerald Group Publishing Limited, pp. 374–384. ISSN: 0332-1649. DOI: 10.1108/03321640510586015. (Visited on 09/16/2022).

[125] Diego M. Solís et al. "Multiresolution Preconditioners for Solving Realistic Multi-Scale Complex Problems". In: *IEEE Access* 10 (2022), pp. 22038–22048. DOI: 10.1109/ACCESS.2022.3153034.

[126] Jiming Song, Caicheng Lu, and Weng Cho Chew. "Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects". In: *IEEE Transactions on Antennas and Propagation* 45 (1997), pp. 1488–1493. URL: https://api.semanticscholar.org/CorpusID:122114463.

[127] Olaf Steinbach and Wolfgang L Wendland. "The construction of some efficient preconditioners in the boundary element method". In: *Advances in Computational Mathematics* 9 (1998), pp. 191–216.

[128] Matthew B Stephanson and Jin-Fa Lee. "Preconditioned electric field integral equation using Calderon identities and dual loop/star basis functions". In: *IEEE Transactions on Antennas and Propagation* 57.4 (2009), pp. 1274–1279.

[129] G. W. Stewart. "A Krylov–Schur Algorithm for Large Eigenproblems". In: *SIAM Journal on Matrix Analysis and Applications* 23.3 (2002), pp. 601–614. DOI: 10.1137/S0895479800371529. eprint: https://doi.org/10.1137/S0895479800371529.

[130] Gilbert Strang. *Linear algebra and its applications*. Belmont, CA: Thomson, Brooks/Cole, 2006. URL: http://www.amazon.com/Linear-Algebra-Its-Applications-Edition/dp/0030105676.

[131] Gilbert Strang and George Fix. *An Analysis of the Finite Element Methods*. Philadelphia, PA: Wellesley-Cambridge Press, 2008. DOI: 10.1137/1.9780980232707.

[132] Jose Taboada et al. "MLFMA-FFT Parallel Algorithm for the Solution of Large-Scale Problems in Electromagnetics". In: *Progress in Electromagnetics Research-Pier* 105 (Jan. 2010), pp. 15–30.

[133] Allen Taflove and S. Hagness. *Computational electrodynamics: the finite-difference time-domain method. 2nd ed*. Vol. 67–106. June 2000. ISBN: 1-58053-076-1.

[134] R. Tielen et al. "p-multigrid methods and their comparison to h-multigrid methods within Isogeometric Analysis". In: *Computer Methods in Applied Mechanics and Engineering* 372 (2020), p. 113347. ISSN: 0045-7825. DOI: https://doi.org/10.1016/j.cma.2020.113347.

[135] Felipe Valdés et al. "High-Order Calderón Preconditioned Time Domain Integral Equation Solvers". In: 61 (May 2013). Conference Name: IEEE Transactions on Antennas and Propagation, pp. 2570–2588. ISSN: 1558-2221. DOI: 10.1109/TAP.2013.2238496.

[136] Felipe Valdés et al. "High-order Div- and Quasi Curl-Conforming Basis Functions for Calderón Multiplicative Preconditioning of the EFIE". In: *IEEE Transactions on Antennas and Propagation* 59.4 (Apr. 2011), pp. 1321–1337. ISSN: 1558-2221. DOI: 10.1109/TAP.2011.2109692.

[137] L. Valle, F. Rivas, and M.F. Catedra. "Combining the moment method with geometrical modelling by NURBS surfaces and Bezier patches". In: *IEEE Transactions on Antennas and Propagation* 42.3 (1994), pp. 373–381. DOI: 10.1109/8.280724.

[138] G. Vecchi. "Loop-star decomposition of basis functions in the discretization of the EFIE". In: *IEEE Transactions on Antennas and Propagation* 47.2 (1999), pp. 339–346. DOI: 10.1109/8.761074.

[139] Francesca Vipiana, Donald R. Wilton, and William A. Johnson. "Advanced Numerical Schemes for the Accurate Evaluation of 4-D Reaction Integrals in the Method of Moments". In: *IEEE Transactions on Antennas and Propagation* 61.11 (Nov. 2013). Conference Name: IEEE Transactions on Antennas and Propagation, pp. 5559–5566. ISSN: 1558-2221. DOI: 10.1109/TAP.2013.2277864.

[140] T. C. Warburton, S. J. Sherwin, and G. E. Karniadakis. "Basis Functions for Triangular and Quadrilateral High-Order Elements". en. In: *SIAM Journal on Scientific Computing* 20.5 (Jan. 1999), pp. 1671–1695. ISSN: 1064-8275, 1095-7197. DOI: 10.1137/S1064827597315716. (Visited on 09/16/2022).

[141] R.A. Wildman and D.S. Weile. "An accurate broad-band method of moments using higher order basis functions and tree-loop decomposition". In: *IEEE Transactions on Antennas and Propagation* 52.11 (Nov. 2004). Conference Name: IEEE Transactions on Antennas and Propagation, pp. 3005–3011. ISSN: 1558-2221. DOI: 10.1109/TAP.2004.835139.

[142] Donald R. Wilton. "Topological consideration in surface patch and volume cell modeling of electromagnetic scatterers". In: *Proc. URSI Int. Symp. Electromagn. Theory*. 1983, pp. 65–68.

[143] Alex C Woo et al. "EM programmer's notebook-benchmark radar targets for the validation of computational electromagnetics programs". In: *IEEE Antennas and propagation Magazine* 35.1 (1993), pp. 84–89.

[144] Hong Xiao and Zydrunas Gimbutas. "A numerical algorithm for the construction of efficient quadrature rules in two and higher dimensions". en. In: *Computers & Mathematics with Applications* 59.2 (Jan. 2010), pp. 663–676. ISSN: 0898-1221. DOI: 10.1016/j.camwa.2009.10.027. (Visited on 06/21/2022).

[145] Kane Yee. "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media". In: *IEEE Transactions on Antennas and Propagation* 14.3 (1966), pp. 302–307. DOI: 10.1109/TAP.1966.1138693.

[146] Pasi Ylä-Oijala, Matti Taskinen, and Jukka Sarvas. "Surface Integral Equation Method for General Composite Metallic and Dielectric Structures with Junctions". en. In: *Progress In Electromagnetics Research* 52 (2005), pp. 81–108. ISSN: 1559-8985. DOI: 10.2528/PIER04071301.

[147] Xiao-Dong Zhang. *The Laplacian eigenvalues of graphs: a survey*. 2011. arXiv: 1111.2897 [math.CO].

[148] Zhimin Zhang. "How many numerical eigenvalues can we trust?" In: *Journal of Scientific Computing* 65.2 (2015), pp. 455–466.

[149] Jun-Sheng Zhao and Weng Cho Chew. "Integral equation solution of Maxwell's equations from zero frequency to microwave frequencies". In: *IEEE Transactions on Antennas and Propagation* 48.10 (2000), pp. 1635–1645. DOI: 10.1109/8.899680.

[150] Kezhong Zhao, M.N. Vouvakis, and Jin-Fa Lee. "The adaptive cross approximation algorithm for accelerated method of moments computations of EMC problems". In: *IEEE Transactions on Electromagnetic Compatibility* 47.4 (2005), pp. 763–773. DOI: 10.1109/TEMC.2005.857898.

[151] Ran Zhao et al. "EFIE-PMCHWT-Based Domain Decomposition Method for Solving Electromagnetic Scattering From Complex Dielectric/Metallic Composite Objects". In: *IEEE Antennas and Wireless Propagation Letters* 16 (2017), pp. 1293–1296. DOI: 10.1109/LAWP.2016.2633061.

This Ph.D. thesis has been typeset by means of the TEX-system facilities. The typesetting engine was pdfLATEX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TEX-system installation.