hp-Finite Element and Boundary Element Methods for Elliptic, Elliptic Stochastic, Parabolic and Hyperbolic Obstacle and Contact Problems

Von der Fakultät für Mathematik und Physik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades
Doktor der Naturwissenschaften
Dr. rer. nat.

genehmigte Dissertation

von

M.Sc. Dipl.-Ök. Lothar Banz

geboren am 12.08.1985 in Hannover

2012

Abstract

The present dissertation deals with the theoretical and numerical analysis of contact and obstacle problems for different classes of differential operators and different discretization methods. In particular, two competing approaches are analyzed. The first is based on a variational inequality (VI) approach. For non-symmetric bilinear forms this approach is not equivalent to a constraint minimization problem. The second is a mixed method in which the non-penetration condition is only weakly enforced by a variational inequality.

The elliptic and parabolic obstacle problems are discretized by an hp-interior penalty discontinuous Galerkin (IPDG) method in space and an hp-time discontinuous Galerkin (TDG) method in time, respectively. The employed discrete Lagrange multiplier is a linear combination of locally constructed basis functions which are globally biorthogonal to the Gauss-Lobatto-Lagrange basis function of the primal variable. This choice allows to reduce the VIs for the sign and non-penetration condition into a simple complementarity problem for each coefficient of the solution's vectors. Using the penalized Fischer-Burmeister non-linear complementarity function, the resulting non-linear, strongly semi-smooth problem can be solved by a locally Q-quadratic converging semi-smooth Newton method. If the discrete non-penetration condition is chosen appropriately, the discrete mixed method is equivalent to the discrete VI method. For the VI method, a p-hierarchical error estimator is constructed which allows an hp-adaptive refinement strategy leading to improved, compared to uniform meshes, or even exponential experimental convergence rates. Furthermore, a priori error estimates for the VI method based on [27] are presented. For these, the continuous and discrete formulations are rewritten in the same bilinear and linear forms using the operators from [35].

The above strategy is general enough to be carried over to an exterior, elliptic stochastic contact problem using boundary integral operators. After a finite Karhunen-Loève expansion the stochastic problem becomes a deterministic but high-dimensional problem. Due to the $\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)$-conformity of the primal variable, a residual a posteriori error estimator based on the ideas in [8, 10] is constructed for hp-adaptive refinements.

Furthermore, numerical experiments indicate that the above discretization strategy, but in a $H^1(\Omega)$-conforming FE context, yields a higher order TDG method which is efficient in terms of CPU-time and artificial energy loss for solving linear elasto-dynamic frictional contact problems approximatively.

Key words: stochastic/time dependent Contact/Obstacle Problems, mixed and VI hp-FEM/BEM, a priori and a posteriori error estimates and hp-adaptivity, semi-smooth Newton, biorthogonal basis functions.

Zusammenfassung

In der vorliegenden Dissertation werden Kontakt- und Hindernisprobleme für verschiedene Klassen von Differentialoperatoren und verschiedene Diskretisierungsmethoden sowohl theoretisch wie auch numerisch analysiert. Insbesondere werden zwei konkurrierende Ansätze verfolgt. Der eine basiert auf einem Variationsungleichungsansatz (VU). Für nichtsymmetrische Bilinearformen ist dieser nicht äquivalent zu einem Minimierungsproblem unter Nebenbedingungen. Der andere ist eine gemischte Methode, in welcher die Nichteindringungsbedingung nur im schwachen Sinne durch eine Variationsungleichung erfüllt wird.

Sowohl das elliptische wie auch das parabolische Hindernisproblem werden im Ort mittels einer hp-interior penalty discontinuous Galerkin (IPDG) Methode und in der Zeit mittels einer hp-time discontinuous Galerkin (TDG) Methode diskretisiert. Der diskrete Lagrangemultiplikator ist eine Linearkombination von lokal konstruierten Basisfunktionen, die global biorthogonal zu den Gauß-Lobatto-Lagrange Basisfunktionen der primalen Variable sind. Diese Wahl erlaubt es, die VUs für die Vorzeichen- und Nichteindringungsbedingung in ein einfaches Komplementaritätsproblem für jeden Koeffizienten der Lösungsvektoren zu reduzieren. Durch Anwendung der penalized Fischer-Burmeister nichtlinearen Komplementaritätsfunktion kann das resultierende nichtlineare, stark halbglatte Problem durch ein lokal Q-quadratisch konvergierendes halbglattes Newton-Verfahren gelöst werden. Bei geeigneter Wahl der diskreten Nichteindringungsbedingung sind die diskrete gemischte Methode und die diskrete VU Methode äquivalent. Für die VU Methode wird ein p-hierarchischer Fehlerschätzer konstruiert, welcher hp-adaptive Verfeinerungsstrategien zulässt. Im Vergleich zu uniformen Gittern führen diese zu verbesserter oder sogar exponentieller Konvergenz. Desweiteren werden, basierend auf [27], a priori Fehlerabschätzungen für die VU Methode hergeleitet. Für diese werden durch Anwendung der Operatoren aus [35] die stetigen und diskreten Formulierungen umgeschrieben, so dass sie die selben Bilinear- und Linearformen verwenden.

Obiges Verfahren ist so allgemein, dass es ebenfalls für elliptisch-stochastische Aussenraumkontaktprobleme mit Randintegraloperatoren angewendet werden kann. Durch eine endliche Karhunen-Loève Entwicklung wird das stochastische Problem in ein hochdimensionales deterministisches Problem umgewandelt. Aufgrund der $\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)$-Konformität in der primalen Variable kann in Anlehnung an [8, 10] ein residualer a posteriori Fehlerschätzer für hp-adaptive Verfeinerungen konstruiert werden.

Darüberhinaus zeigen numerische Experimente, dass die obige Diskretisierungsstrategie in einem $H^1(\Omega)$-konformen FE Kontext zu einer TDG Methode höherer Ordnung führt, die beim approximativen Lösen von reibungsbehafteten linearen elasto-dynamischen Kontaktproblemen sowohl bezüglich der CPU-Zeit wie auch dem künstlichen Energieverlust effizient ist.

Schlagworte: stochastische/zeitabhängige Kontakt/Hindernis Probleme, gemischte und VU hp-FEM/BEM, a priori und a posteriori Fehlerabschätzungen, halbglattes Newton, biorthogonale Basisfunktionen

Acknowledgments

Contents

List of Figures

List of Figures

List of Tables

1 Introduction

Contact and obstacle problems have a variety of applications in the real world. Frictional dynamic contact problems appear in many engineer's applications, e.g. in gear transmissions or in belt grinding processes. Whereas parabolic obstacle problems appear in the financial markets for pricing American put basket options within the Black-Scholes model world [6]. An American put option gives the holder the right to sell the underlying, e.g. a share, to a predetermined, fixed price at any time prior to its expiration date. If the underlying is not a single entity but a portfolio, e.g. consisting of two or more shares, the option is called a basket option. Mathematically, the space dimension of the parabolic obstacle problem is increased. From an economic's point of view mathematical simulations of these problems are indispensable. However, the underlying differential inequalities in conjunction with the non-penetration condition can seldom be solved analytically. Hence, numerical methods such as the Galerkin approximation have to be applied.

The standard approach for these is a variational inequality method as analyzed in [56, 45, 9] among others. However, the discretization of the convex cone of admissible functions is by no means trivial and leads to non-optimal a priori error estimates [27, 59] even if the approximation property of the discrete cone may be optimal [5]. More recently, in [70, 71, 34, 37] and the references therein, mixed methods are employed in which a Lagrange multiplier is used to weakly enforce the non-penetration condition by a variational inequality constraint. For the lowest order h-version, optimal a priori error estimates have been shown in [34, 37]. If the primal and dual space are discretized independently, the discrete inf-sup-condition will require the discrete dual space to be sufficiently coarser than the discrete primal space. Typically, this is achieved by coarsening the mesh or decreasing the polynomial degree for the Lagrange multiplier space as in [70]. If, however, the duality of the spaces is also considered on the discrete level by using biorthogonal basis functions [76], the same mesh and polynomial degree can be used for both the primal and dual variable. The use of these basis functions for a lowest order continuous Galerkin uniform h-version in elliptic contact problems has first been studied in [37, 38] and shown to be quite successful. The construction and use of higher order biorthogonal basis functions on irregular meshes for elliptic obstacle problems is studied in [3].

In contact and obstacle problems the space/space-time/space-stochastic domain can be separated into two disjoint sets, the contact and non-contact set. In the contact set, the solution is completely determined by the rigid foundation/obstacle. In the non-contact set, the solution satisfies a simpler differential equation but typically in a very complex

domain. In general, the solution is only of reduced regularity and, thus, higher order *h*-versions are not an appropriate tool to approximate the solution. Furthermore, *hp*-methods with an a priori defined, geometrically graded mesh cannot be applied since the free boundary, the location of the singularity, is a priori unknown. Consequently, local a posteriori error indicators are required to automatically generate sequences of *hp*-adapted meshes. For elliptic problems with a $H_0^1(\Omega)$-conforming approach in the primal variable, Braess [8] provides a general scheme for constructing such a posteriori error estimators by reusing existing estimators for an auxiliary problem without contact. However, it still remains open how the consistency error of the dual variable in the $H^{-1}(\Omega)$-norm can be computed. For parabolic obstacle problems with a finite difference approximation in time, a residual error estimator based on the Galerkin functional has been derived in [63]. However, it also requires $H_0^1(\Omega)$-conformity. A completely different approach is the hierarchical h-$\frac{h}{2}$ error estimator which outsources all difficulties into the saturation assumption [28, 26]. This estimator is very costly as the refined solution must be explicitly calculated and is not approximated by a preconditioner as in [59].

Additionally, the contact and obstacle problems are highly non-linear such that the construction of efficient iterative solvers itself is challenging. If the differential operator is symmetric and coercive, the problem is equivalent to a constraint quadratic programming problem for which many efficient iterative solvers exist. In general, however, the operator is not symmetric and consequently all these solvers are not guaranteed to converge and often fail to do so in practice. For non-symmetric variational inequalities, to the best of the author's knowledge, only projection and contraction methods like [33] are guaranteed to converge. But their convergence rate depends on many user-chosen, mesh and problem dependent parameters and is in general very small.

The aim of this dissertation is to construct an approximation strategy which is general enough to be applied to different classes of differential operators for contact and obstacle problems and to different discretization methods. The focus is on the construction of an efficient iterative solver and on an a posteriori error estimator. This in conjunction with the analyticity estimator of [36] allows an *hp*-adaptive mesh refinement which leads to improved, compared to uniform meshes, or even exponential experimental convergence rates. For the obstacle problems, this is achieved by employing a mixed *hp*-FEM interior penalty discontinuous Galerkin discretization in space, and a time discontinuous Galerkin approach in time, which is equivalent to a discrete variational inequality method. By using biorthogonal basis functions for the Lagrange multiplier, the sign and non-penetration conditions are reduced to a set of complementarity problems. Employing a non-linear complementarity function, the resulting discrete problem can be solved iteratively by a locally Q-quadratic converging semi-smooth Newton method. Since the discrete variational inequality and mixed method are equivalent, a *p*-hierarchical error estimator for the discrete variational inequality allows *hp*-adaptivity. In discontinuous Galerkin, irregular meshes which naturally arises in *hp*-adaptivity imply no difficulties. Furthermore, for an a priori error estimated analogously to [27] the extension operators from [35] are used. For a symmetric, exterior, elliptic stochastic contact problem with

a boundary integral formulation a finite Karhunen-Loève expansion is used to transform the stochastic problem into a deterministic but high-dimension formulation. This problem is discretized similarly to the obstacle problems, except with a conforming, continuous Galerkin discretization. This allows to derive a residual based a posteriori error estimator.

The remainder of this work is structured as follows. Chapter 2 contains some fundamentals which are used throughout the entire work. In particular, Section 2.2 revises non-linear complementarity functions, subdifferentials and the semi-smooth Newton method. Chapter 3 is devoted to a non-symmetric elliptic obstacle problem and a parabolic obstacle problem is considered in Chapter 4. In Chapter 5 an exterior, elliptic stochastic contact problem with a boundary integral formulation is analyzed. An application of the general *hp*-FEM strategy to an elasto-dynamical frictional contact problem in Chapter 6 shows the flexibility and powerfulness of this method. Concluding remarks are given in Chapter 7.

2 Fundamentals

In this chapter basic results about boundary integral operators are recalled and a brief introduction to nonlinear complementarity functions and the semi-smooth Newton method is given. These are fundamental for the analysis in Chapter 5 and for the development of the iterative solvers for the discrete mixed formulations. Additionally, basic definitions in probability theory and an short introduction into the Karhunen-Loève expansion is presented.

2.1 Boundary Integral Operators

In boundary value problems, a differential operator acts on a function u in every point x of a domain. Once the fundamental solution is known, this operator can be replaced by a boundary integral operator. In Chapter 5 such boundary integral operators are used for an exterior Laplacian problem.

Let $\Omega \subset \mathbb{R}^2$ be a domain with piecewise Lipschitz boundary $\Gamma = \partial\Omega$, a and b real constants and let the solution u satisfy

$$-\Delta u(x) = 0 \qquad\qquad x \in \mathbb{R}^2 \setminus \overline{\Omega} \tag{2.1a}$$

$$u(x) = a \cdot \log(x) + b + o(1) \quad \text{as } \|x\| \to \infty. \tag{2.1b}$$

Here, $o(1)$ is the Landau symbol with $\lim_{\|x\| \to \infty} o(1) = 0$. Further, let $k(x, y)$ be the fundamental solution of the Laplace operator in two dimensions, i.e.

$$k(x, y) = -\frac{1}{2\pi} \log \|x - y\|.$$

Using the representation formula

$$u(x) = -\int_\Omega k(x, y)\Delta u(y)\, dx_y + \int_\Gamma k(x, y)\partial_{n_y} u(y) - \partial_{n_y} k(x, y)u(y)\, ds_y, \quad x \in \mathbb{R}^2 \setminus \overline{\Omega} \tag{2.2}$$

the exterior Calderon projector $\mathcal{C}^+$ can be defined. Taking the limit $x \to \Gamma$ and denoting $\phi = \partial_n u$ yields

$$\begin{pmatrix} u \\ \phi \end{pmatrix} = \begin{pmatrix} \frac{1}{2} + K & -V \\ -W & \frac{1}{2} - K' \end{pmatrix} \begin{pmatrix} u \\ \phi \end{pmatrix} =: \mathcal{C}^+ \begin{pmatrix} u \\ \phi \end{pmatrix}$$

with the single layer potential V, the double layer potential K, its adjoint K' and the hypersingular integral operator W.

$$V\phi(x) := \int_\Gamma k(x,y)\phi(y)\,ds_y, \qquad Wu(x) := -\frac{\partial}{\partial n_x}\int_\Gamma \frac{\partial}{\partial n_y}k(x,y)u(y)\,ds_y,$$

$$Ku(x) := \int_\Gamma \frac{\partial}{\partial n_y}k(x,y)u(y)\,ds_y, \quad K'\phi(x) := \frac{\partial}{\partial n_x}\int_\Gamma k(x,y)\phi(y)\,ds_y.$$

Lemma 2.1 (Costabel [20])**.** *Let Γ be the boundary of a Lipschitz domain. Then the integral operators*

$$V : H^{-\frac{1}{2}+s}(\Gamma) \to H^{\frac{1}{2}+s}(\Gamma), \qquad W : H^{\frac{1}{2}+s}(\Gamma) \to H^{-\frac{1}{2}+s}(\Gamma)$$

$$K : H^{\frac{1}{2}+s}(\Gamma) \to H^{\frac{1}{2}+s}(\Gamma), \qquad K' : H^{-\frac{1}{2}+s}(\Gamma) \to H^{-\frac{1}{2}+s}(\Gamma)$$

are bounded for all $s \in [-\frac{1}{2}, \frac{1}{2}]$, i.e. there exists constants C_V, C_K, $C_{K'}$, $C_W > 0$ such that

$$\|V\phi\|_{H^{\frac{1}{2}+s}(\Gamma)} \le C_V \|\phi\|_{H^{-\frac{1}{2}+s}(\Gamma)}, \quad \|Wu\|_{H^{-\frac{1}{2}+s}(\Gamma)} \le C_W \|u\|_{H^{\frac{1}{2}+s}(\Gamma)},$$

$$\|Ku\|_{H^{\frac{1}{2}+s}(\Gamma)} \le C_K \|u\|_{H^{\frac{1}{2}+s}(\Gamma)}, \quad \|K'\phi\|_{H^{-\frac{1}{2}+s}(\Gamma)} \le C_{K'} \|\phi\|_{H^{-\frac{1}{2}+s}(\Gamma)}.$$

It is well known that for $\Omega \subset \mathbb{R}^2$ with $cap(\Gamma) < 1$ the mapping V has a bounded inverse. (More general 2D cases of Γ can be treated by scaling arguments.) Hence, the symmetric, linear, positive definite Steklov-Poincaré operator $S : H^{\frac{1}{2}}(\Gamma) \to H^{-\frac{1}{2}}(\Gamma)$

$$S := W + (K' - \frac{1}{2})V^{-1}(K - \frac{1}{2}) \tag{2.3}$$

is well defined [11].

Lemma 2.2. *Let $\Gamma \subset \mathbb{R}^2$ be the boundary of a Lipschitz domain Ω, $\Gamma_0 \subset \Gamma$ and let $cap(\Gamma) < 1$. Then V is $H^{-\frac{1}{2}}(\Gamma)$-elliptic, i.e. $\exists\, c_V > 0$ s.t.*

$$\langle V\phi, \phi\rangle_\Gamma \ge c_V \|\phi\|_{H^{-\frac{1}{2}}(\Gamma)}^2 \qquad \forall\, \phi \in H^{-\frac{1}{2}}(\Gamma)$$

and S is continuous, $H^{\frac{1}{2}}(\Gamma)$-elliptic and $\tilde{H}^{\frac{1}{2}}(\Gamma_0)$-elliptic , i.e. $\exists\, c_S,\, C_S > 0$ s.t.

$$\|Su\|_{H^{-\frac{1}{2}}(\Gamma)} \le C_S \|u\|_{H^{\frac{1}{2}}(\Gamma)} \qquad \forall\, u \in H^{\frac{1}{2}}(\Gamma),$$

$$\langle Su, u\rangle_\Gamma \ge c_S \|u\|_{H^{\frac{1}{2}}(\Gamma)}^2 \qquad \forall\, u \in H^{\frac{1}{2}}(\Gamma),$$

$$\langle Su, u\rangle_{\Gamma_0} \ge c_S \|u\|_{\tilde{H}^{\frac{1}{2}}(\Gamma_0)}^2 \qquad \forall\, u \in \tilde{H}^{\frac{1}{2}}(\Gamma_0),$$

with the Sobolev space ($s \ge 0$)

$$\tilde{H}^s(\Gamma_0) := \left\{ u : \ \exists\, v \in H^s(\Gamma) \ s.t. \ u = v|_{\Gamma_0},\ \operatorname{supp} v \subset \overline{\Gamma}_0 \right\}.$$

Often the different notation $\tilde{H}^{\frac{1}{2}}(\Gamma_0) = H_{00}^{\frac{1}{2}}(\Gamma_0)$ is used.

2.2 Nonlinear Complementarity Functions and Semi-Smooth Newton Methods

Nonlinear complementarity functions (NCF) find their applications in the theory of complementarity which is used in mathematical programming but also in game theory and fixed point theory [25]. Within this research area it is well known [60, 30] that the complementarity problem

$$u \geq 0, \quad \lambda \geq 0, \quad u \cdot \lambda = 0 \tag{2.4}$$

can be transformed by a suitable $\phi : \mathbb{R}^2 \to \mathbb{R}$ into the equivalent nonlinear problem

$$\phi(u, \lambda) = 0. \tag{2.5}$$

The function ϕ is in general not linear and not even differentiable everywhere. Throughout this work the penalized Fischer-Burmeister function

$$\phi_\mu(u, \lambda) := \mu \left(u + \lambda - \sqrt{u^2 + \lambda^2} \right) + (1 - \mu) \max \{0, u\} \max \{0, \lambda\}, \quad \mu \in (0, 1), \tag{2.6}$$

introduced in [14], is used, which is a convex combination of the Fischer-Burmeister function $u + \lambda - \sqrt{u^2 + \lambda^2}$ introduced in [29] and $\max \{0, u\} \max \{0, \lambda\}$. The later term penalizes violations of the complementarity condition (both $u, \lambda \geq 0$ but $u \cdot \lambda \neq 0$) and increases the slop of ϕ in the positive quadrant. Figure 2.1(a) shows the classical Fischer-Burmeister function and Figure 2.1(b) the penalized Fischer-Burmeister function with $\mu = 0.3$. In the latter cases, the contour lines are evenly distributed which is desirable when solving (2.5) iteratively with a semi-smooth Newton method.

The semi-smooth Newton method is introduced analogously to [22, 65]. Let $\phi : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitzian, then it is differentiable almost everywhere by Rademacher's theorem and the set where ϕ is differentiable is denoted by D_ϕ. The B-subdifferential of ϕ at x is defined as

$$\partial_B \phi(x) := \left\{ H : \exists \text{ a sequence } \left\{ x^k \right\}, \ x^k \in D_\phi \text{ with } \lim_{x^k \to x} \phi'(x^k) = H \right\} \tag{2.7}$$

and the Clark subdifferential $\partial \phi(x) := co \, \partial_B \phi(x)$ of ϕ at x as the convex hull of the B-subdifferential. Qi introduced in [64] the C-subdifferential as

$$\partial_C \phi(x)^T := \partial \phi_1(x) \times \ldots \times \partial \phi_m(x) \tag{2.8}$$

where the right hand side is a set of matrices whose i^{th}-column is an arbitrary element of the Clark subdifferential of ϕ_i. By [30], there holds the following overestimation of Clark subdifferential by the C-subdifferential.

Lemma 2.3 (Lemma 1(v) in [30]). *Let $\phi : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitzian near x, then*

$$\partial \phi(x) \subseteq \partial \phi_1(x) \times \ldots \times \partial \phi_m(x) = \partial_C \phi(x)^T$$
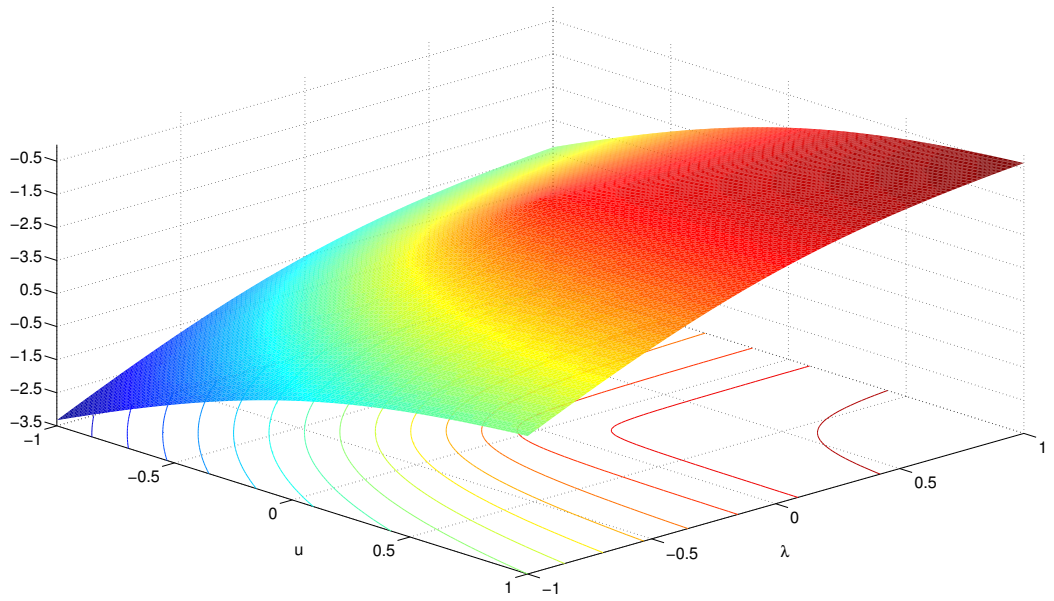
(a) Fischer-Burmeister NCF
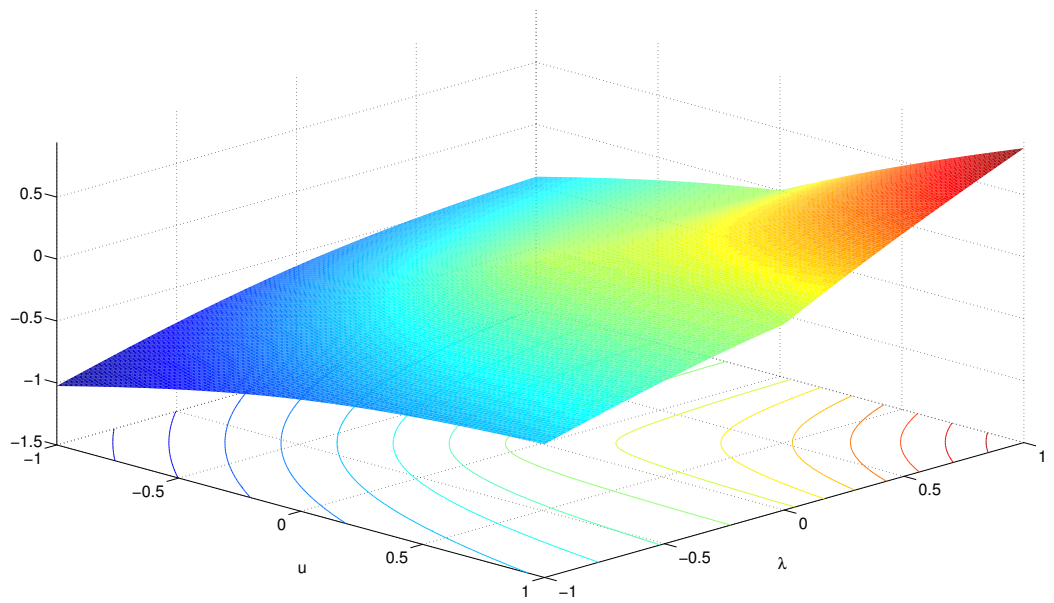
(b) Penalized Fischer-Burmeister NCF (2.6) with $\mu = 0.3$

Figure 2.1: Function values and contour lines of the Fischer-Burmeister NCFs

Definition 2.1 (Strongly semi-smooth)**.** *Let $\phi : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitzian at x. ϕ is called semi-smooth at x if*

$$\lim_{H \in \partial\phi(x+tv'), v' \to v, t \to 0^+} Hv' \tag{2.9}$$

exists for all $v \in \mathbb{R}^n$ and is called strongly semi-smooth at x if additionally for any $H \in \partial\phi(x+d)$ and any $d \to 0$

$$Hd - \phi'(x; d) = \mathcal{O}\left(\|d\|^2\right) \tag{2.10}$$

holds with $\phi'(x; d)$ the directional derivative.

Lemma 2.4 (Chen et al. [14])**.** *The penalized Fischer-Burmeister function $\phi_\mu : \mathbb{R}^2 \to \mathbb{R}$ defined in (2.6) satisfies the following properties:*

1. *ϕ_μ is a NCF.*

2. *ϕ_μ continuously differentiable on $\mathbb{R}^2 \setminus \{(u, \lambda) : \ u \geq 0, \ \lambda \geq 0, \ u \cdot \lambda = 0\}$.*

3. *ϕ_μ is strongly semi-smooth on $\mathbb{R}^2$.*

4. *The generalized gradient $\partial\phi_\mu(u, \lambda)$ is equal to the set*

$$\frac{\partial\phi_\mu}{\partial u} = \begin{cases} \mu(1 - \xi) & , \ if \ u = \lambda = 0 \\ \mu(1 - \frac{u}{\sqrt{u^2+\lambda^2}}) + (1-\mu)\max\{0, \lambda\}\,\partial\max\{0, u\} & , \ otherwise \end{cases}$$

$$\frac{\partial\phi_\mu}{\partial \lambda} = \begin{cases} \mu(1 - \zeta) & , \ if \ u = \lambda = 0 \\ \mu(1 - \frac{\lambda}{\sqrt{u^2+\lambda^2}}) + (1-\mu)\max\{0, u\}\,\partial\max\{0, \lambda\} & , \ otherwise \end{cases}$$

with $\sqrt{\xi^2 + \zeta^2} \leq 1$ and

$$\partial\max\{0, x\} = \begin{cases} 1 & , \ if \ x > 0 \\ [0, 1] & , \ if \ x = 0 \\ 0 & , \ otherwise. \end{cases}$$

Lemma 2.5. *Let $\Phi_\mu = vec\{\phi_\mu\}$ be a vector valued penalized Fischer-Burmeister NCF, then there holds:*

1. *Φ_μ is strongly semi-smooth.*

2. *For any $u, \ \lambda \in \mathbb{R}^n$ there holds the overestimation*

$$\partial\Phi_\mu \subseteq \partial_C\Phi_\mu \subseteq D_u(u, \lambda) \times D_\lambda(u, \lambda)$$

   with $D_u(u, \lambda) = diag\left\{\frac{\partial\phi_\mu(u,\lambda)}{\partial u}\right\}$ and $D_\lambda(u, \lambda) = diag\left\{\frac{\partial\phi_\mu(u,\lambda)}{\partial \lambda}\right\}$ diagonal matrices.

3. *The merit function $\Psi_\mu(u, \lambda) := \frac{1}{2}\Phi_\mu(u, \lambda)^T\Phi_\mu(u, \lambda)$ is continuously differentiable with $\nabla\Psi_\mu(u, \lambda) = H^T\Phi_\mu(u, \lambda)$ for any $H \in \partial_C\Psi_\mu(u, \lambda)$.*

Proof.   *Result 1 is [13, Theorem 2.2]. Result 2 is analogous to [13, Proposition 2.3] together with Lemma 2.3. Result 3 follows from [13, Theorem 3.2].* $\square$

Definition 2.2 (BD/CD-regular). *Let $n = m$, then ϕ is called BD-, CD-regular at x if all $H \in \partial_B \phi(x)$, $H \in \partial \phi(x)$ respectively, are invertible.*

Theorem 2.1 (Semi-smooth Newton, [65]). *Suppose that x^* is a solution of $\phi(x) = 0$, ϕ is locally Lipschitzian, semi-smooth and CD-regular at x^*. Then the iteration method*

$$x^{k+1} = x^k - H_k^{-1}\phi(x^k), \quad H_k \in \partial\phi(x^k) \tag{2.11}$$

is well defined and the sequence $\{x^k\}$ converges to x^* Q-superlinearly in a neighborhood of x^*. If in addition, ϕ is strongly semi-smooth at x^*, then this convergence is Q-quadratic.

If the merit function ψ is continuously differentiable, the iteration method (2.11) can be globalized by a strategy proposed in [22]. The globalization is achieved by a line search to minimize ψ in step 2.c) and in case of an unsuited search direction it switches to one globally converging gradient descent step in step 2.b).

Algorithm 2.1. *(Damped semi-smooth Newton algorithm [40])*

---

1. *Choose initial solution $x^0 \in \mathbb{R}^n$ and control parameters $\rho > 0$, $\beta \in (0,1)$, $\sigma \in (0, \frac{1}{2})$, $p > 2$, tol > 0*

2. *For $k = 0, 1, 2, \dots$ do*

   a) *If $\left\|\nabla\psi(x^k)\right\| <$ tol or $\left\|\psi(x^k)\right\| <$ tol then stop.*

   b) *Compute subdifferential $H_k \in \partial\phi(x^k)$ and find $d^k \in \mathbb{R}^n$ s.t.*

   $$H_k d^k = -\phi(x^k). \tag{2.12}$$

   If (2.12) not solvable or if the descent condition

   $$\nabla\psi(x^k)^T d^k \leq -\rho||d^k||^p \tag{2.13}$$

   is not satisfied, set $d^k := -\nabla\psi(x^k)$.

   c) *Compute search length $t_k := \max\left\{\beta^l : l = 0, 1, 2, \dots\right\}$ s.t.*

   $$\psi(x^k + t_k d^k) \leq \psi(x^k) + \sigma t_k \nabla\psi(x^k)d^k. \tag{2.14}$$

   d) *Update the solution vector and goto step 2.*

   $$x^{k+1} = x^k + t_k d^k.$$

The computation of the search direction can be very expensive. Therefore inexact SSN methods have been studied intensively over the last years, c.f. [40] and the references therein. If (2.12) is solved inexactly in the sense that

$$\left\| H_k d^k + \phi(x^k) \right\| \le \eta_k \left\| \phi(x^k) \right\| \tag{2.15}$$

with $\eta_k = \mathcal{O}\left(\left\| \phi(x^k) \right\|\right)$ the method still converges Q-quadratically. The monotone line search (2.14) in Algorithm 2.1 can lead to very small step sizes. This in turn can lead to reduced rate of or even no convergence [22]. For the implementation De Luca et al. recommend to replace the line search by a non-monotone one, explained in detail in [22, Section 7].

2.3 Basics of Probability Theory and the Karhunen-Loève Expansion

Definition 2.3 (Probability space). *The triple $(\Omega, \mathcal{F}, P)$ is called probability space with Ω the set outcomes, $\mathcal{F} \subset 2^\Omega$ the σ-algebra of events and $P : \mathcal{F} \to [0, 1]$ the probability measure.*

Definition 2.4 (Expected Value, c.f. [1]). *Let $Y \in L_P^1(\Omega)$ be an $\mathbb{R}^N$-valued random variable in $(\Omega, \mathcal{F}, P)$. Then, the expected value is defined by*

$$E\left[Y\right] = \int_\Omega Y(\omega)\, dP(\omega) = \int_{\mathbb{R}^N} y\, d\mu_Y(y)$$

where μ_Y is the distribution measure for Y, defined for the Borel sets $\tilde{b} \in \mathcal{B}(\mathbb{R}^N)$ by $\mu_Y(\tilde{b}) \equiv P(Y^{-1}(\tilde{b}))$.

In the case of an (a.e.-)absolute continuous μ_Y there exists a density function $\rho_Y : \mathbb{R}^N \to [0, \infty)$ such that $E\left[Y\right] = \int_{\mathbb{R}^N} y\rho_Y(y)\, dy$.

Definition 2.5 (Covariance Matrix, c.f. [1]). *Let $Y_i \in L_P^2(\Omega)$ $1 \le i \le d$, then the covariance matrix $\mathrm{Cov}\left[Y\right] \in \mathbb{R}^{d \times d}$ is defined by*

$$\mathrm{Cov}\left[Y\right](i,j) = \mathrm{Cov}(Y_i, Y_j) = E\left[(Y_i - E\left[Y_i\right])(Y_j - E\left[Y_j\right])\right] \quad 1 \le i, j \le d.$$

The stochastic Sobolev space $L_P^2(\Omega, H^s(D))$ with the norm $\|v\|_{L_P^2(\Omega, H^s(D))}^2 = E\left[\|v\|_{H^s(D)}^2\right]$ is a Hilbert space. Moreover, it is isomorph to $L_P^2(\Omega) \otimes H^s(D) \simeq L_P^2(\Omega, H^s(D))$.

Definition 2.6 (Karhunen-Loève expansion, [41]). *The Karhunen-Loève expansion of a random field $\kappa : R \times \Omega \to \mathbb{R}$ with bounded covariance is*

$$\kappa(x, \omega) = \mu(x) + \sum_{i=1}^\infty \sqrt{\lambda_i}\kappa_i(x)\xi_i(\omega) \tag{2.16}$$

where $\mu(x) = E[\kappa]$ and the random variables

$$\xi_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_R (\kappa - \mu)\kappa_i \, dx \tag{2.17}$$

are mutually uncorrelated and centered with unit variance. Here (λ_i, κ_i) is an (eigenvalue, eigenfunction) pair of the compact and self-adjoint Fredholm operator

$$T : \ L^2(R) \to L^2(R)$$
$$u(x) \mapsto \int_R \mathrm{Cov}_\kappa(x, y)u(y) \, dy$$

i.e. $T\kappa_i = \lambda_i\kappa_i$ for $\kappa_i \in L^2(R)$ and $i \in \mathbb{N}$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$.

Often the Karhunen-Loève expansion (2.16) is truncated after a finite number of terms. Then the finite Karhunen-Loève expansion is an $L^2(R \times \Omega)$-optimal linear approximation of κ [73].

3 Elliptic Obstacle Problems

In this chapter an elliptic obstacle problem is analyzed which can be viewed as a sub-problem of the parabolic obstacle problem in Chapter 4 if a finite difference approximation in time is used. Since the differential operator is not symmetric and the solution exhibits reduced regularity across the a priori unknown free boundary, a discretization optimized for an efficient iterative solver and an hp-adaptive mesh refinement are constructed.

3.1 Weak Formulations for Elliptic Obstacle Problems

Let $\Omega \subset \mathbb{R}^d$ be an open, bounded, polygonal domain with boundary Γ. For given volume force $f \in H^{-1}(\Omega)$, convection coefficient $\vec{\gamma} \in \mathbb{R}^d$, Dirichlet data $g \in H^{\frac{1}{2}}(\Gamma)$ and obstacle $\chi \in H^1(\Omega)$ with $\chi|_\Gamma = g$, the elliptic obstacle problem is to find a function $u : \Omega \to \mathbb{R}$ such that

$$-\Delta u + \vec{\gamma} \cdot \nabla u + u \geq f \quad \text{in } \Omega \tag{3.1a}$$

$$u = g \quad \text{on } \Gamma \tag{3.1b}$$

$$u \geq \chi \quad \text{in } \Omega. \tag{3.1c}$$

Remark 3.1. *In general only $\chi|_\Gamma \leq g$ is required to ensure that the set of admissible functions is not empty. The assumption $\chi|_\Gamma = g$ has been made for simplicity. Although, this is a restrictive assumption, such obstacles frequently occur in financial mathematics, e.g. in American put option pricing [63].*

For a weak formulation of the obstacle problem let

$$K := \left\{ v \in H^1(\Omega) : \ v|_\Gamma = g \text{ and } v \geq \chi \text{ a.e. in } \Omega \right\} \tag{3.2}$$

be the convex cone of admissible functions. Then, the variational inequality formulation is to find $u \in K$ such that

$$B(u, v - u) \geq \langle f, v - u \rangle \quad \forall\, v \in K \tag{3.3}$$

with $\langle \cdot, \cdot \rangle$ the duality pairing between $H_0^1(\Omega)$ and $H^{-1}(\Omega)$ and with the bilinear form

$$B(u, v) := \int_\Omega \nabla u \nabla v + \vec{\gamma} \cdot \nabla u v + u v \, dx. \tag{3.4}$$

Using partial integration, the following lemma holds trivially.

3 Elliptic Obstacle Problems

Lemma 3.1. *Every solution of* (3.1) *is a solution of* (3.3). *The converse holds in a distributional sense.*

The proof for the existence and uniqueness of a solution u requires coercivity and continuity of the bilinear form $B(\cdot, \cdot)$ in the same norm.

Lemma 3.2. *The bilinear form $B(\cdot, \cdot)$ is $H^1(\Omega)$-continuous and $H_0^1(\Omega)$-elliptic, i.e. there exists constants $\alpha > 0$ and $C_B > 0$ such that*

$$B(v, v) \geq \alpha \|v\|_{H^1(\Omega)}^2 \qquad \forall\, v \in H_0^1(\Omega) \tag{3.5}$$

$$B(w, v) \leq C_B \|w\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \quad \forall\, v, w \in H^1(\Omega). \tag{3.6}$$

Proof. Green's formula yields for the convection term [66, p.98]

$$\int_\Omega \vec{\gamma} \cdot \nabla v v \, dx = \frac{1}{2} \int_\Omega \vec{\gamma} \cdot \nabla(v^2) \, dx = \frac{1}{2} \int_\Gamma \vec{\gamma} \cdot \vec{n} v^2 \, ds = 0 \quad \forall\, v \in H_0^1(\Omega)$$

which immediately implies the $H_0^1(\Omega)$-ellipticity. The continuity follows with the Cauchy-Schwarz inequality and the boundedness of $\vec{\gamma}$. $\qquad\square$

Theorem 3.1. *There exists a unique solution to the variational inequality formulation* (3.3).

Proof. Decompose $u = u_0 + u_g$ into $u_0 \in H_0^1(\Omega)$ and a Dirichlet lift $u_g \in H^1(\Omega)$, i.e. $u_g|_\Gamma = g$. Moreover, let u_g satisfy

$$B(u_g, v) = 0 \quad \forall\, v \in H_0^1(\Omega). \tag{3.7}$$

The existence of u_g is guaranteed by the extension operator mapping from $H^{\frac{1}{2}}(\Gamma)$ onto $H^1(\Omega)$ and the ellipticity and continuity of $B(\cdot, \cdot)$ by Lemma 3.2. Therefore, $u_0 \in K_0(u_g) = \{v \in H_0^1(\Omega) : v \geq \chi - u_g \text{ a.e. in } \Omega\}$ satisfies

$$\langle f, v_0 - u_0 \rangle = \langle f, v_0 + u_g - u_0 - u_g \rangle \leq B(u, v - u) = B(u_0 + u_g, v_0 + u_g - u_0 - u_g)$$

$$= B(u_0 + u_g, v_0 - u_0) = B(u_0, v_0 - u_0) + \underbrace{B(u_g, v_0 - u_0)}_{=0,\ (3.7)}$$

$$= B(u_0, v_0 - u_0) \quad \forall v_0 \in K_0(u_g).$$

with $v = v_0 + u_g$ yielding $v - u = v_0 - u_0$. Since $\chi|_\Gamma = g = u_g|_\Gamma$, $K_0(u_g) \subset H_0^1(\Omega)$ is closed, convex and not empty and since $B(\cdot, \cdot)$ is $H_0^1(\Omega)$-elliptic and continuous, the Stampacchia theorem [45, Theorem 2.1] provides the unique existence of u_0.

It remains to show that u is unique as well. Assume $u_1, u_2 \in K$ were two solutions to (3.3). Then their difference $\delta := u_1 - u_2 \in H_0^1(\Omega)$ satisfies

$$\alpha \|\delta\|_{H^1(\Omega)}^2 \leq B(\delta, \delta) \leq 0$$

by adding (3.3) with $v_1 = u_2$ and $v_2 = u_1$, and by the $H_0^1(\Omega)$-ellipticity of $B(\cdot, \cdot)$. $\quad\square$

The variational formulation (3.3) has two severe drawbacks. Firstly, for $\vec{\gamma} \neq 0$ the bilinear form $B(\cdot, \cdot)$ is not symmetric. Consequently, a discretization of (3.3) yields an algebraic variational inequality formulation with a non-symmetric but positive definite system matrix. To the best of the author's knowledge there exists no efficient iterative solver for these and only projection-contraction methods like [33] are guaranteed to converge. However, their convergence is in general slow and depends on many user-chosen, problem and mesh dependent parameters. Secondly, the discretization of K itself is by no means trivial. Therefore, a mixed method in which the counter force λ of the obstacle is sought as an additional unknown seems to be favorable. Let

$$L^+ := \left\{ \mu \in H^{-1}(\Omega) : \ \langle \mu, v \rangle \leq 0 \ \ \forall\, v \in H_0^1(\Omega) \text{ with } v \leq 0 \right\} \tag{3.8}$$

be the set of admissible Lagrange multipliers. Then, the mixed method is to find the pair $(u, \lambda) \in H^1(\Omega) \times L^+$ such that $u|_\Gamma = g$ and

$$B(u, v) - \langle v, \lambda \rangle = \langle f, v \rangle \qquad \forall\, v \in H_0^1(\Omega) \tag{3.9a}$$

$$\langle u, \mu - \lambda \rangle \geq \langle \chi, \mu - \lambda \rangle \quad \forall\, \mu \in L^+. \tag{3.9b}$$

Theorem 3.2. *The variational inequality formulation* (3.3) *and the mixed method* (3.9) *are equivalent.*

Proof. "(3.3) $\Leftarrow$ (3.9)": Choosing $\mu = 0$ and $\mu = 2\lambda$ in (3.9b) yields

$$\langle u - \chi, \lambda \rangle = 0 \quad \text{and} \quad \langle u - \chi, \mu \rangle \geq 0 \quad \forall\, \mu \in L^+. \tag{3.10}$$

Assume there were to exist a Lebesgue measurable set $P \subset \Omega$ in which $u < \chi$. Then, for $\mu = 1_P \in L^+$ which is one in P and zero elsewhere, this yields

$$\langle u - \chi, 1_P \rangle = \int_P u - \chi \, dx < 0$$

which contradicts (3.10). Therefore, $u \geq \chi$ almost everywhere in Ω, i.e. $u \in K$. Furthermore, with $w \in K$, equation (3.10) yields

$$\langle w - u, \lambda \rangle = \langle w - \chi, \lambda \rangle \geq 0$$

since $\lambda \in L^+$. Hence, setting $v = w - u$ in (3.9a) yields (3.3).
"(3.3) $\Rightarrow$ (3.9)": Define the residual $Res \in H^{-1}(\Omega)$ of (3.3) by

$$\langle Res, v \rangle := B(u, v) - \langle f, v \rangle \quad \forall\, v \in H_0^1(\Omega). \tag{3.11}$$

Hence, with u solving (3.3) there holds additionally: Find $u \in H^1(\Omega)$ s.t. $u|_{\partial\Omega} = g_D$ and

$$B(u, w) - \langle Res, w \rangle = \langle f, w \rangle \quad \forall\, w \in H_0^1(\Omega)$$

and also by (3.3)

$$\langle Res, v - u \rangle \geq 0 \quad \forall\, v \in K. \tag{3.12}$$

Choosing $v = \chi$ and $v = 2u - \chi$ yields

$$\langle Res, u - \chi \rangle \leq 0, \quad \langle Res, u - \chi \rangle \geq 0 \quad \Rightarrow \quad \langle u - \chi, Res \rangle = 0.$$

Recall that $u \in K$. Then for $\mu \in L^+$, $\langle u - \chi, \mu \rangle \geq 0$ and together with the above result yields (3.9b) when setting $\lambda = Res$. Choosing $v = w + u$ with $w \geq 0$ in (3.12) yields $Res \in L^+$ which completes the proof. $\qquad \square$

Lemma 3.3. *There exists a unique pair (u, λ) solving (3.9).*

Proof. By Theorem 3.1 and Theorem 3.2 the unique existence of u and the existence of λ is guaranteed. Assume both (u, λ_1) and (u, λ_2) solve (3.9). Then their difference satisfies

$$\langle v, \lambda_1 - \lambda_2 \rangle = 0 \quad \forall\, v \in H_0^1(\Omega)$$

by subtracting the two (3.9a) from each other. With the definition of the dual norm

$$\|\lambda_1 - \lambda_2\|_{H^{-1}(\Omega)} := \sup_{0 \neq v \in H_0^1(\Omega)} \frac{\langle v, \lambda_1 - \lambda_2 \rangle}{\|v\|_{H_0^1(\Omega)}} = 0$$

follows the assertion. $\qquad \square$

Lemma 3.4. *The inf-sup-condition (3.13) is satisfied with $\beta = 1$.*

$$\inf_{0 \neq \mu \in L^+} \sup_{0 \neq v \in H_0^1(\Omega)} \frac{-\langle v, \mu \rangle}{\|v\|_{H_0^1(\Omega)} \|\mu\|_{H^{-1}(\Omega)}} \geq \beta > 0 \tag{3.13}$$

Proof. Recall the definition of the dual norm

$$\|\mu\|_{H^{-1}(\Omega)} := \sup_{0 \neq v \in H_0^1(\Omega)} \frac{\langle v, \mu \rangle}{\|v\|_{H_0^1(\Omega)}} = \sup_{0 \neq v \in H_0^1(\Omega)} \frac{-\langle v, \mu \rangle}{\|v\|_{H_0^1(\Omega)}}.$$

Hence,

$$\inf_{0 \neq \mu \in L^+} \sup_{0 \neq v \in H_0^1(\Omega)} \frac{-\langle v, \mu \rangle}{\|v\|_{H_0^1(\Omega)} \|\mu\|_{H^{-1}(\Omega)}} = \inf_{\mu \in L^+} \frac{\|\mu\|_{H^{-1}(\Omega)}}{\|\mu\|_{H^{-1}(\Omega)}} = 1$$

which completes the proof. $\qquad \square$

3.2 hp-**IPDG Discretization for Elliptic Obstacle Problems**

There are several ways to discretize (3.9). First, a $H^1(\Omega)$-conforming continuous Galerkin approach with possible hanging nodes which is dealt with in the work [3]. Second, an interior penalty discontinuous Galerkin (IPDG) method in which the continuity constraint is only weakly enforced. For the IPDG scheme let $\mathcal{E}_h$ be a subdivision of Ω into rectangulars, Γ_h the set of interior edges and Γ_D the set of Dirichlet edges. With each edge in $\Gamma_h \cup \Gamma_D$ a unit normal $\mathbf{n}_e$ is associated. The orientation of $\mathbf{n}_e$ is

outwards on Γ_D and on Γ_h from E_1^e to E_2^e if $e = \partial E_1^e \cap \partial E_2^e$. However, the numbering of the elements E themselves is arbitrary. Therewith, the jump and average across the edge can be defined as in [66, p. 28] by

$$[v] := v|_{E_1^e} - v|_{E_2^e}, \quad \{v\} := \frac{1}{2}v|_{E_1^e} + \frac{1}{2}v|_{E_2^e} \quad \forall\, e = \partial E_1^e \cap \partial E_2^e$$

on the interior edges and on the Dirichlet edges by

$$[v] := v|_{E_1^e}, \quad \{v\} := v|_{E_1^e} \quad \forall\, e = \partial E_1^e \cap \Gamma.$$

3.2.1 FE-IPDG Discretization for Mixed Formulation

The subdivision $\mathcal{E}_h$ together with a polynomial degree vector p defines the FE-sets for both the primal and dual variables.

$$V_{hp} := \left\{ v \in L^2(\Omega) : \ v|_Q \in \mathbb{P}_{p_Q}(Q) \ \ \forall\, Q \in \mathcal{E}_h \right\} = \mathrm{span}\,\{\phi_j\}_{j=1}^{\dim V_{hp}} \tag{3.14}$$

$$M_{hp}^+ := \left\{ \mu \in \mathrm{span}\,\{\psi_j\}_{j=1}^{\dim V_{hp}} : \int_\Omega \mu v \, dx \leq 0 \ \ \forall\, v = \sum_{i=1}^{\dim V_{hp}} v_i \phi_i \in V_{hp} \text{ with } v_i \leq 0 \right\} \tag{3.15}$$

Here, ϕ_j are affinely transformed Gauss-Lobatto-Lagrange (GLL) basis functions defined on the reference square $[-1,1]^2$ using a tensor product of the 1D-GLL functions. The dual basis functions ψ_j are globally biorthogonal to ϕ_i, i.e.

$$\int_\Omega \psi_j \phi_i \, dx = \delta_{ij} \int_\Omega \phi_i \, dx \quad 1 \leq i,j \leq \dim V_{hp}.$$

These basis functions are studied thoroughly by Wohlmuth et al. in [77, 76, 38] for the lowest order h-version. In [52] the construction of higher order biorthogonal basis functions are studied. In contact problems their analysis is always restricted to H^1-conforming methods with regular meshes. In domain decomposition methods their analysis is restricted to approaches which are conforming in the subdomains which are independent of the mesh size. For obstacle problems with an H^1-conforming approach, the use of biorthogonal basis functions on irregular meshes with an arbitrary number of hanging nodes is studied in [3]. Especially, it explains that the same connectivity informations which are used for the assembly of the primal basis functions can also be used for the biorthogonal ones. In the case of IPDG, affinely transformed local biorthogonal basis functions are also globally biorthogonal. Moreover, there holds the following lemma which implies that the discrete variational inequality constraint (3.16b) is nothing more than an L^2-projection problem.

Lemma 3.5. *The primal and dual basis functions span the same set, i.e.*

$$\mathrm{span}\,\{\phi_j\}_{j=1}^{\dim V_{hp}} = \mathrm{span}\,\{\psi_j\}_{j=1}^{\dim V_{hp}}.$$

Proof. Same arguments as in Lemma 4.7. □

3 Elliptic Obstacle Problems

Therewith, the IPDG method is: Find $u_h \in V_{hp}$ and $\lambda_h \in M_{hp}^+$ such that

$$a_\epsilon(u_h, v) + b_{\vec{\gamma}}(u_h, v) - \langle v, \lambda_h \rangle = F_\epsilon(v) \qquad \forall\, v \in V_{hp} \qquad (3.16a)$$

$$\langle u_h, \mu - \lambda_h \rangle \geq \langle \chi, \mu - \lambda_h \rangle \quad \forall\, \mu \in M_{hp}^+ \qquad (3.16b)$$

with the bilinear and linear forms

$$a_\epsilon(u, v) := \sum_{E \in \mathcal{E}_h} \int_E \nabla u \nabla v + uv \, dx + \sum_{e \in \Gamma_h \cup \Gamma_D} \int_e - \left\{ \frac{\partial u}{\partial \mathbf{n}_e} \right\} [v] + \epsilon \left\{ \frac{\partial v}{\partial \mathbf{n}_e} \right\} [u] + \frac{\sigma_e p_e^2}{|e|^\beta} [u]\,[v] \, ds$$

$$(3.17)$$

$$b_{\vec{\gamma}}(u, v) := -\sum_{E \in \mathcal{E}_h} \int_E \vec{\gamma} u \cdot \nabla v \, dx + \sum_{e \in \Gamma_h} \int_e \vec{\gamma} \cdot \mathbf{n}_e u^{\mathrm{up}} [v] \, ds \qquad (3.18)$$

$$F_\epsilon(v) := \int_\Omega f v \, dx + \sum_{e \in \Gamma_D} \int_e \left(\epsilon \frac{\partial v}{\partial \mathbf{n}_e} + \left(\frac{\sigma_e p_e^2}{|e|^\beta} - \vec{\gamma} \cdot \mathbf{n}_e \right) v \right) g \, ds, \qquad (3.19)$$

respectively. Here, an upwind discretization for the convection term is used, i.e.

$$v^{\mathrm{up}} := \begin{cases} v|_{E_1^e}, & \text{if } \vec{\gamma} \cdot \mathbf{n}_e \geq 0 \\ v|_{E_2^e}, & \text{if } \vec{\gamma} \cdot \mathbf{n}_e < 0. \end{cases}$$

The choice of the parameter $\epsilon \in \{-1, 0, 1\}$ determines which particular IPDG method is used, e.g. for $\epsilon = 1$ and $\sigma_e = 1$ it is called the non-symmetric interior penalty Galerkin (NIPG) method and for $\epsilon = 0$ incomplete interior penalty Galerkin (IIPG), c.f. [67, 66] among others. The penalty parameter σ_e is always non-negative but may vary for different edges [68]. The exponent β is a positive constant depending on the dimension d of Ω such that $\beta(d-1) \geq 1$ and p_e is the maximum of the two polynomial degrees on the edge e. This penalty term guarantees the convergence towards a $H^1(\Omega)$-function and the coercivity of the bilinear form $a_\epsilon(\cdot, \cdot)$ what gives rise to the mesh dependent norm

$$\|v\|_{1,hp}^2 := \sum_{E \in \mathcal{E}_h} \int_E (\nabla v)^2 + v^2 \, dx + \sum_{e \in \Gamma_h \cup \Gamma_D} \int_e \frac{\sigma_e p_e^2}{|e|^\beta} [v]^2 \, ds. \qquad (3.20)$$

Lemma 3.6. *If σ_e is sufficiently large, there exists a constant $\alpha > 0$ such that for all $v \in V_{hp}$*

$$a_\epsilon(v, v) + b_{\vec{\gamma}}(v, v) \geq \alpha \|v\|_{1,hp}^2. \qquad (3.21)$$

Proof. By [66, p. 38 and p. 99] $a_\epsilon(v, v) \geq \alpha \|v\|_{1,hp}^2$ and $b_{\vec{\gamma}}(v, v) \geq 0$. $\qquad \square$

In the forthcoming σ_e is always assumed to be sufficiently large such that Lemma 3.6 can be applied. For the proof of existence and uniqueness of (u_h, λ_h) and for the construction of the iterative solver, the following lemma is of central importance.

Lemma 3.7. *There holds for the integral value of the primal and dual basis functions* $\int_\Omega \psi_i \, dx = \int_\Omega \phi_i \, dx =: D_i > 0.$

Proof. Follows from the biorthogonality relationship, a partition of unity, and the same arguments as in Lemma 5.11. □

Lemma 3.8. *If $\mu_h \in M_{hp}$ satisfies*

$$\int_\Omega \mu_h v_h \, dx = 0 \quad \forall \, v_h \in V_{hp}$$

then μ_h is zero.

Proof. Since $\mu_h \in M_{hp}$ and $v_h \in V_{hp}$, they can be written as a linear combination of ψ_j and ϕ_j respectively, i.e.

$$\mu_h(x) = \sum_{j=1}^{\dim V_{hp}} \mu_j \psi_j(x) \quad \text{and} \quad v_h(x) = \sum_{j=1}^{\dim V_{hp}} v_j \phi_j(x).$$

Then, by the assumption of this lemma and by the biorthogonality of ψ_j to ϕ_i

$$0 = \int_\Omega \mu_h v_h \, dx = \sum_{j=1}^{\dim V_{hp}} \mu_j v_j D_j. \tag{3.22}$$

Since $v_h \in V_{hp}$ is arbitrary, choosing $v_i \neq 0$ and $v_j = 0$ for $j \neq i$ yields with $D_i > 0$ by Lemma 3.7 that $\mu_i v_i = 0$ for $1 \leq i \leq \dim V_{hp}$. Hence, $\mu_i = 0$, i.e. $\mu \equiv 0$. □

The previous three lemmas are sufficient to proof unique existence of a discrete solution.

Theorem 3.3. *There exists a unique solution pair $(u_h, \lambda_h) \in V_{hp} \times M_{hp}^+$ to (3.16).*

Proof. **Uniqueness:** Let both (u_1, λ_1) and (u_2, λ_2) solve (3.16). Then subtracting the two (3.16a) from each other yields

$$a_\epsilon(u_1 - u_2, v) + b_{\vec{\gamma}}(u_1 - u_2, v) - \langle v, \lambda_1 - \lambda_2 \rangle = 0 \quad \forall \, v \in V_{hp}. \tag{3.23}$$

Choosing $v = u_1 - u_2$ in (3.23) and using the ellipticity result of Lemma 3.6 yields

$$\alpha \|u_1 - u_2\|_{1,hp}^2 - \langle u_1 - u_2, \lambda_1 - \lambda_2 \rangle \leq 0. \tag{3.24}$$

Further, choosing $\mu = \lambda_2$ and $\mu = \lambda_1$, in (3.16b) and adding these inequalities yields

$$\langle u_1 - u_2, \lambda_1 - \lambda_2 \rangle \leq 0. \tag{3.25}$$

Inserting (3.25) into (3.24) implies $\|u_1 - u_2\|_{1,hp}^2 = 0$ and therewith (3.23) reduces to

$$\langle v, \lambda_1 - \lambda_2 \rangle = 0 \quad \forall \, v \in V_{hp}. \tag{3.26}$$

for which Lemma 3.8 yields $\lambda_1 - \lambda_2 = 0$.
Existence: Due to Lemma 3.5, it is well known that the problem (3.16b) can be written as the projection equation [51]

$$\lambda = \mathcal{P}_{M_{hp}^+} \left(\lambda + r(\chi - u) \right) \tag{3.27}$$

where $\mathcal{P}_{M_{hp}^+}$ is the L^2-projection operator mapping onto M_{hp}^+ and $r > 0$ is an arbitrary constant. For a given λ_h, the equation (3.16a) reduces to an elliptic, finite dimensional linear problem. Hence, the uniqueness result of u_h ensured by (3.24) also implies the existence of a $u_h(\lambda_h)$. Let

$$T : \ M_{hp}^+ \to M_{hp}^+$$
$$\lambda \mapsto \mathcal{P}_{M_{hp}^+}\left(\lambda + r(\chi - u)\right).$$

be the operator to describe the fixed point iteration $\lambda^{(k+1)} = T\lambda^{(k)}$. If T is a contraction then by the Banach fixed point theorem, there exists a λ which satisfies (3.27) and, hence, also a corresponding u solving (3.16). For the ease of presentation denote $\delta\lambda = \lambda_1 - \lambda_2$, $\delta u = u_1 - u_2$ and $\|\cdot\| = \|\cdot\|_{L^2(\Omega)}$.

$$
\begin{aligned}
\|T\lambda_1 - T\lambda_2\|^2 &= \left\| \mathcal{P}_{M_{hp}^+}\left(\lambda_1 + r(\chi - u_1)\right) - \mathcal{P}_{M_{hp}^+}\left(\lambda_2 + r(\chi - u_2)\right) \right\|^2 \\
&\leq \|\delta\lambda - r\delta u\|^2 \\
&= \|\delta\lambda\|^2 - 2r\langle\delta\lambda, \delta u\rangle + r^2\|\delta u\|^2 \\
&= \|\delta\lambda\|^2 - 2r\left[a_\epsilon(\delta u, \delta u) + b_{\vec{\gamma}}(\delta u, \delta u)\right] + r^2\|\delta u\|^2 \\
&\leq \|\delta\lambda\|^2 - 2\alpha r\|\delta u\|^2 + r^2\|\delta u\|^2 \\
&= \|\delta\lambda\|^2\left(1 - 2\alpha r\gamma^2 + r^2\gamma^2\right)
\end{aligned}
$$

with $\gamma = \frac{\|\delta u\|_{L^2(\Omega)}}{\|\delta\lambda\|_{L^2(\Omega)}}$. The second line is the standard projection result, the fourth line results form (3.23) and the fifth line from Lemma 3.6. Hence, for $0 < r < 2\alpha$, T is a strict contraction which completes the proof. $\qquad\square$

Theorem 3.4. *The condition* (3.16b) *is equivalent to the system*

$$u_i \geq g_i := \frac{1}{D_i}\int_\Omega \chi\psi_i(x)\ dx \qquad\qquad (3.28a)$$

$$\lambda_i \geq 0 \qquad\qquad (3.28b)$$

$$\lambda_i\left(u_i - g_i\right) = 0 \qquad\qquad (3.28c)$$

for $1 \leq i \leq \dim V_{hp}$. Here u_i and λ_i are the expansion coefficients of u_h and λ_h, respectively.

Proof. Same arguments as in the proof of Theorem 4.4 except without the additional time integration. $\qquad\square$

3.2.2 FE-IPDG Discretization for VI Formulation

As on the continuous level, there exists a discrete variational inequality formulation equivalent to the discrete mixed formulation (3.16). If the discrete non-penetration

condition is incorporated into the convex cone of admissible ansatz and test functions, the Lagrange multiplier can be eliminated. More precisely,

$$K_{hp} := \{v \in V_{hp} : v_i \geq \chi_i\} \tag{3.29}$$

is the convex cone of admissible functions where χ_i are one-sided box constraints on the solution coefficients. Then, the discrete variation inequality problem is:
Find $u_h \in K_{hp}$ such that

$$a_\epsilon(u_h, v - u_h) + b_{\vec{\gamma}}(u_h, v - u_h) \geq F_\epsilon(v - u_h) \quad \forall v \in K_{hp}. \tag{3.30}$$

Theorem 3.5. *There exists exactly one solution to the discrete variational inequality formulation* (3.30).

Proof. Same arguments as in the proof of Theorem 4.5. □

Theorem 3.6. *The discrete problems* (3.16) *and* (3.30) *are equivalent if* $\chi_i = g_i$ *of Theorem 3.4.*

Proof. Let (u, λ) solve (3.16). Then, Theorem 3.4 implies $u_i \geq g_i$, i.e. $u \in K_{hp}$. Furthermore, the biorthogonality and (3.28c) of Theorem 3.4 yield for all $w \in K_{hp}$

$$- \int_\Omega (w - u)\lambda \, dx = - \sum_{i=1}^{\dim V_{hp}} (w_i - u_i)\lambda_i D_i = - \sum_{i=1}^{\dim V_{hp}} (w_i - g_i)\lambda_i D_i \leq 0.$$

Hence, choosing $v = w - u$ with $w \in K_{hp}$ in (3.16) yields (3.30).
For the opposite direction let $u \in K_{hp}$ solve (3.30), i.e. $u_i \geq g_i$. Define the Lagrange multiplier by $\lambda = \sum_i \lambda_i \psi_i(x)$ with

$$\lambda_i = \frac{a_\epsilon(u, \phi_i) + b_{\vec{\gamma}}(u, \phi_i) - F_\epsilon(\phi_i)}{D_i} \qquad \text{for } 1 \leq i \leq \dim V_{hp}, \tag{3.31}$$

i.e. $\langle \lambda, v \rangle$ is the residual of the discrete variational inequality for all $v \in V_{hp}$. Choosing $v = u + \phi_i$, i.e. $v \in K_{hp}$, in (3.30) yields

$$0 \leq a_\epsilon(u, \phi_i) + b_{\vec{\gamma}}(u, \phi_i) - F_\epsilon(\phi_i) = \lambda_i D_i \quad \Rightarrow \quad \lambda_i \geq 0.$$

Finally, choose $v \in V_{hp}$ such that $v_i = g_i$, $v_i = 2u_i - g_i$ respectively, in (3.30) to obtain

$$0 = \sum_{i=1}^{\dim V_{hp}} \left[a_\epsilon(u, \phi_i) + b_{\vec{\gamma}}(u, \phi_i) - F_\epsilon(\phi_i) \right] (u_i - g_i) = \sum_{i=1}^{\dim V_{hp}} \lambda_i (u_i - g_i) D_i = \langle u - \chi, \lambda \rangle.$$

The assertion follows with Theorem 3.4. □

Remark 3.2. *For* $\chi_i = g_i$ *the convex cone is the same as*

$$K_{hp} := \left\{ v \in V_{hp} : \int_\Omega (v - \chi)\mu \, dx \geq 0 \quad \forall \mu \in M_{hp}^+ \right\}. \tag{3.32}$$

3.2.3 An Efficient Iterative Solver for the Discrete Mixed Formulation

The convergence proof of all efficient iterative solvers for variational inequality formulations are based on equivalent minimization problems. Therefore, these are not guaranteed to converge if the bilinear form is not symmetric. In the following a fast iterative solver is constructed for the discrete mixed formulation and by Theorem 3.6 also for the discrete variational inequality formulation.

Lemma 3.9. *The discrete problem* (3.16) *is equivalent to solving*

$$0 \overset{!}{=} F(\vec{u}, \vec{\lambda}) := \begin{pmatrix} A\vec{u} - D\vec{\lambda} - \vec{f} \\ \varphi_\eta(\vec{u}, \vec{\lambda}) \end{pmatrix} \tag{3.33}$$

where $A\vec{u} - D\vec{\lambda} - \vec{f} = 0$ is the matrix representation of (3.16a) *and $\varphi_\eta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is the vector-valued penalized Fischer-Burmeister non-linear complementarity function (NCF) defined by*

$$\varphi_\eta(\vec{u}, \vec{\lambda}) = \eta \left(\vec{\lambda} + (\vec{u} - \vec{g}) - \sqrt{\vec{\lambda}^2 + (\vec{u} - \vec{g})^2} \right) + (1 - \eta) \max\left\{0, \vec{\lambda}\right\} \max\{0, \vec{u} - \vec{g}\}$$

with $\eta \in (0, 1]$ and a componentwise understood right hand side.

Proof. By Lemma 2.4 φ_η is a NCF, i.e.

$$\varphi_\eta(\vec{u}, \vec{\lambda}) = 0 \iff (u - g)_i \geq 0, \ \lambda_i \geq 0, \ \lambda_i \cdot (u - g)_i = 0 \quad (i = 1, \dots, n).$$

Hence, $\varphi_\eta(\vec{u}, \vec{\lambda}) = 0$ is equivalent to (3.28) which in turn is equivalent to (3.16b) by Theorem 3.4. $\qquad\square$

Lemma 3.10. *The matrix A is (in general) non-symmetric but sparse and all eigenvalues have positive real part if σ_e is sufficiently large. The matrix D is positive definite and diagonal. The function φ_μ is strongly semi-smooth and Lipschitzian.*

Proof. By Lemma 3.6 all eigenvalues of A have positive real part. The positive definiteness and diagonal property of D follows directly from the biorthogonality and the use of Gauss-Lobatto-Lagrange basis functions (c.f. Lemma 3.7). By Lemma 2.4 φ_η is strongly semi-smooth everywhere and with the definition of strongly semi-smoothness also Lipschitzian. $\qquad\square$

Theorem 3.7. *The reduced semi-smooth Newton algorithm*

$$\left(u^{k+1}, \lambda^{k+1}\right)^T = \left(u^k, \lambda^k\right)^T - H_k^{-1} F(u^k, \lambda^k) \tag{3.34}$$

where H_k is a Clarke subdifferential of F at $\left(u^k, \lambda^k\right)^T$ converges locally Q-quadratic.

Proof. The same as in Theorem 4.7 $\qquad\square$

The reduced semi-smooth Newton algorithm can be globalized using Algorithm 4.1, in which only the definition of $F(u, \lambda)$ needs to be changed. Further improvements are also discussed in Section 4.2.3.

3.3 A Hierarchical a Posteriori Error Estimator and hp-Adaptivity

As it is typical for hierarchical error estimators the proof heavily relies on the saturation assumption. Dörfler and Nochetto [24] have proven the saturation assumption for conforming FEM with piecewise linear to piecewise quadratic elements under the assumption of small data oscillation.

Assumption 3.1 (Saturation Assumption). *Let $u \in K$ solve (3.3) and $u_{\widetilde{hp}} \in K_{\widetilde{hp}}$, $u_{hp} \in K_{hp}$ respectively, solve (3.30) with $V_{hp} \subset V_{\widetilde{hp}}$. Then, the saturation assumption assumes the existence of a constant $q_S \in (0,1)$ uniformly in $hp - \widetilde{hp}$ such that*

$$\|u - u_{\widetilde{hp}}\|_{1,\widetilde{hp}} \leq q_S \|u - u_{hp}\|_{1,hp}. \tag{3.35}$$

The additional condition that the pure FE-spaces $V_{hp} \subset V_{\widetilde{hp}}$ are nested is necessary for Lemma 3.11 as the estimate (3.35) must use different norms on the left and right hand side since in general the expression $\|u - u_{\widetilde{hp}}\|_{1,hp}$ makes no sense due to the jumps of $u_{\widetilde{hp}}$ within the coarse elements $E \in \mathcal{E}_h$.

Lemma 3.11. *For $v \in H^1(\Omega)$ and $v_h \in V_{hp} \subset V_{\widetilde{hp}}$ there holds*

$$\|v - v_h\|^2_{1,hp} \leq \|v - v_h\|^2_{1,\widetilde{hp}} \tag{3.36}$$

$$\|v - v_h\|^2_{1,\widetilde{hp}} \leq C(h, p, \widetilde{h}, \widetilde{p}) \|v - v_h\|^2_{1,hp} \tag{3.37}$$

with $C(h, p, \widetilde{h}, \widetilde{p})$ defined in (3.38) if $\sigma_e = \sigma_{\widetilde{e}}$ for $\widetilde{e}$ a child edge of σ_e and $\beta \geq 1$.

Proof. Since $V_{hp} \subset V_{\widetilde{hp}}$ every coarse element $E \in \mathcal{E}_h$ is a finite sum of refined elements $\widetilde{E} \in \mathcal{E}_{\widetilde{h}}$. Hence, the mass and discrete gradient part of the two norms are identical and only the jump part differs. The h-refinement generates completely new edges which are interior edges and lie within a coarse element E. In particular $v \in H^1(\Omega)$ and v_h is a polynomial on E and, therefore, the jump of $v - v_h$ across these new edges is always zero by [35, Lemma 4.3]. Further, every edge $e \in \Gamma_h \cup \Gamma_D$ can be written as a union of refined edges $\widetilde{e}$, i.e. $e = \bigcup_{i=1}^{Ne} \widetilde{e}_{i,e}$. Then, $p_e \leq p_{\widetilde{e}_{i,e}}$ and $|e| \geq |\widetilde{e}_{i,e}|$ what implies

$$\sum_{e \in \Gamma_h \cup \Gamma_D} \int_e \frac{\sigma_e p_e^2}{|e|^\beta} [v - v_h]^2 \, ds = \sum_{e \in \Gamma_h \cup \Gamma_D} \sum_{i=1}^{Ne} \int_{\widetilde{e}_{i,e}} \frac{\sigma_e p_e^2}{|e|^\beta} [v - v_h]^2 \, ds$$

$$\leq \sum_{e \in \Gamma_h \cup \Gamma_D} \sum_{i=1}^{Ne} \int_{\widetilde{e}_{i,e}} \frac{\sigma_{\widetilde{e}} p_{\widetilde{e}_{i,e}}^2}{|\widetilde{e}_{i,e}|^\beta} [v - v_h]^2 \, ds = \sum_{\widetilde{e} \in \widetilde{\Gamma_h} \cup \widetilde{\Gamma_D}} \int_{\widetilde{e}} \frac{\sigma_{\widetilde{e}} p_{\widetilde{e}}^2}{|\widetilde{e}|^\beta} [v - v_h]^2 \, ds.$$

and hence (3.36). For the second estimate note that

$$\int_e \frac{\sigma_e p_e^2}{|e|^\beta} [v - v_h]^2 \, ds = \sum_{i=1}^{Ne} \int_{\widetilde{e}_{i,e}} \frac{\sigma_e p_e^2}{|e|^\beta} [v - v_h]^2 \, ds$$

$$= \sum_{i=1}^{Ne} \int_{\widetilde{e}_{i,e}} \frac{\sigma_e p_e^2}{|e|^\beta} \frac{\sigma_{\widetilde{e}_{i,e}} p_{\widetilde{e}_{i,e}}^2}{|\widetilde{e}_{i,e}|^\beta} \frac{|\widetilde{e}_{i,e}|^\beta}{\sigma_{\widetilde{e}_{i,e}} p_{\widetilde{e}_{i,e}}^2} [v - v_h]^2 \, ds = \sum_{i=1}^{Ne} \frac{p_e^2}{p_{\widetilde{e}_{i,e}}^2} \frac{|\widetilde{e}_{i,e}|^\beta}{|e|^\beta} \int_{\widetilde{e}_{i,e}} \frac{\sigma_{\widetilde{e}_{i,e}} p_{\widetilde{e}_{i,e}}^2}{|\widetilde{e}_{i,e}|^\beta} [v - v_h]^2 \, ds$$

$$\geq \min_{i=1,\ldots,N_e} \left\{ \frac{p_e^2}{p_{\widetilde{e}_{i,e}}^2} \frac{|\widetilde{e}_{i,e}|^\beta}{|e|^\beta} \right\} \int_{\widetilde{e}_{i,e}} \frac{\sigma_{\widetilde{e}_{i,e}} p_{\widetilde{e}_{i,e}}^2}{|\widetilde{e}_{i,e}|^\beta} [v - v_h]^2 \, ds.$$

Defining the factor $C(h, p, \widetilde{h}, \widetilde{p})$ by

$$0 \leq \frac{1}{C(h, p, \widetilde{h}, \widetilde{p})} := \min_{e \in \Gamma_h \cup \Gamma_D} \min_{i=1,\ldots,N_e} \left\{ \frac{p_e^2}{p_{\widetilde{e}_{i,e}}^2} \frac{|\widetilde{e}_{i,e}|^\beta}{|e|^\beta} \right\} \leq 1 \tag{3.38}$$

and summing over all edges $e \in \Gamma_h \cup \Gamma_D$ yields (3.37). $\qquad\square$

For the special case $h = 2\widetilde{h}$ and $p = \widetilde{p}$ the constant is $C(h, p, \widetilde{h}, \widetilde{p}) = 2^\beta$ and for $h = \widetilde{h}$, $p + 1 = \widetilde{p}$ the constant is bounded $C(h, p, \widetilde{h}, \widetilde{p}) \leq 4$ and $C(h, p, \widetilde{h}, \widetilde{p}) \overset{p \to \infty}{\to} 1$.

Theorem 3.8. *Under the saturation assumption 3.1 and the assumption of Lemma 3.11 there holds*

$$\frac{1}{\sqrt{C(h, p, \widetilde{h}, \widetilde{p})} + q_S} \widetilde{\eta} \leq \|u - u_{hp}\|_{1,hp} \leq \frac{1}{1 - q_S} \widetilde{\eta} \tag{3.39}$$

with the error estimator $\widetilde{\eta} := \|u_{\widetilde{hp}} - u_{hp}\|_{1,\widetilde{hp}}$.

Proof. Triangle inequality and equation (3.36) of Lemma 3.11 yield the upper bound

$$\|u - u_{hp}\|_{1,hp} \leq \|u - u_{hp}\|_{1,\widetilde{hp}} = \left\|u - u_{\widetilde{hp}} + u_{\widetilde{hp}} - u_{hp}\right\|_{1,\widetilde{hp}} \leq \|u - u_{\widetilde{hp}}\|_{1,\widetilde{hp}} + \widetilde{\eta}$$

$$\leq q_S \|u - u_{hp}\|_{1,hp} + \widetilde{\eta}.$$

Also triangle inequality but equation (3.37) of Lemma 3.11 yield the lower bound

$$\widetilde{\eta} = \|u_{\widetilde{hp}} - u + u - u_{hp}\|_{1,\widetilde{hp}} \leq \|u - u_{\widetilde{hp}}\|_{1,\widetilde{hp}} + \|u - u_{hp}\|_{1,\widetilde{hp}}$$

$$\leq q_S \|u - u_{hp}\|_{1,hp} + \|u - u_{hp}\|_{1,\widetilde{hp}} \leq \left(q_S + \sqrt{C(h, p, \widetilde{h}, \widetilde{p})}\right) \|u - u_{hp}\|_{1,hp}.$$

$\qquad\square$

The difficulty in the above theorem is only induced from the potential h-refinement to enrich the FE-space. For the specific $p - (p+1)$-error estimator this estimate can be sharpened since $\|u_{hp+1} - u_{hp}\|_{1,hp}$ now makes sense.

Theorem 3.9. *Under the saturation assumption 3.1 there holds for the $p-(p+1)$-error estimator*

$$\frac{1}{1+q_S}\eta \le \|u-u_{hp}\|_{1,hp} \le \frac{1}{1-q_S}\eta \le \frac{1}{1-q_S}\widetilde{\eta} \tag{3.40}$$

with $\eta := \|u_{hp+1}-u_{hp}\|_{1,hp}$.

Proof. By triangle inequality and an increase in the penalty factor from p^2 to $(p+1)^2$ there holds

$$\|u-u_{hp}\|_{1,hp} \le \|u-u_{hp+1}\|_{1,hp}+\eta \le \|u-u_{hp+1}\|_{1,hp+1}+\eta \le q_S\|u-u_{hp}\|_{1,hp}+\eta$$

and

$$\eta \le \|u_{hp+1}-u\|_{1,hp}+\|u-u_{hp}\|_{1,hp} \le \|u_{hp+1}-u\|_{1,hp+1}+\|u-u_{hp}\|_{1,hp}$$
$$\le (q_S+1)\|u-u_{hp}\|_{1,hp}.$$

The last inequality also holds from the fact that the only difference of the mesh dependent norms is that the penalty factor has increased, c.f. (3.20). □

Remark 3.3.   *1. The sharper estimate of Theorem 3.9 holds for all FE-space enrichment based only on p-refinement, e.g. $p-(p+2)$ or $p-\lceil\sqrt{2}p\rceil$.*

2. Both $\widetilde{\eta}$ and η can be written as a sum of local contributions, but the implementation of η is easier and cheaper than of $\widetilde{\eta}$ due to the gradient contribution in the norm and exploitation of the sum factorization. For $\widetilde{\eta}$, an additional mesh refinement with parent-child relations for the elements and edges must be stored and searched through.

3. The saturation assumptions implies that the FE-space $V_{\widetilde{hp}}$ must be sufficiently larger than V_{hp}. Using the $p-(p+1)$ error estimator the relative FE-space enrichment $\frac{2p+3}{(p+1)^2}$ is constant for h-versions and tends towards zero for p-versions.

4. Using the $p-\lceil\sqrt{2}p\rceil$ estimator, the relative FE-space enrichment $\approx \frac{p^2+2(\sqrt{2}-1)p}{p^2+2p+1}$ tends towards one for the uniform p-version.

5. Due to the nestedness of the FE-spaces, there exists a simple basis transformation matrix I from $V_{\widetilde{hp}}$ to V_{hp} based on interpolation. Hence, all the matrices and vectors defined on the coarse mesh which do not involve the penalty factor can be computed by multiplying I with the corresponding part from the fine mesh. Further the solution from the coarse mesh can be used as an initial solution for the iterative solver for the refined solution.

Remark 3.4. *The error estimator requires the computation of both u_{hp} and $u_{\widetilde{hp}}$ which is very cost intensive compared to residual error estimators. These use an auxiliary problem to split the error contribution from the contact and the discretization of the differential operator as presented in [8]. A residual error estimator for IPDG exists*

and has been presented by Houston et al. in [35]. However, the splitting of the error requires $H^1(\Omega)$-conformity of the primal basis function which in IPDG methods is not given. Additionally, it is not clear how the arising consistency error of the discrete Lagrange multiplier in the $H^{-1}(\Omega)$-norm can be realized.

Since the $1, hp$-norm is local, the error estimators $\widetilde{\eta}$ and η can be written as a sum of local indicators.

Lemma 3.12. *The error indicator η satisfies*

$$\eta^2 = \sum_{E \in \mathcal{E}_h} \eta_{loc}^2(E) \tag{3.41}$$

with the local error indicators

$$\eta_{loc}^2(E) = \int_E \left(\nabla(u_{p+1} - u_p)\right)^2 + (u_{p+1} - u_p)^2 \ dx \tag{3.42}$$

$$+ \sum_{e \in \partial E \cap \Gamma_h} \int_e \frac{\sigma_e p_e^2}{2|e|^\beta} [u_{p+1} - u_p]^2 \ ds + \sum_{e \in \partial E \cap \Gamma_D} \int_e \frac{\sigma_e p_e^2}{|e|^\beta} (u_{p+1} - u_p)^2 \ ds.$$

Proof. Writing down the norm explicitly and spreading the jump contributions equally over the adjacent elements. $\qquad\square$

The result for $\widetilde{\eta}$ is analogical but also the parent-child relation must be used. The local error indicators (3.42) allow an hp-adaptive mesh refinement. In this strategy all elements $E \in M := \left\{ \widetilde{E} \in \mathcal{E}_h : \ \eta_{loc}^2(\widetilde{E}) \geq \theta \cdot \max \eta_{loc}^2 \right\}$ are isotropically refined for $\theta \in (0, 1)$. For other marking strategies, refer to [23, 62, 74] among others. Elementary for an hp-adaptive strategy is the decision criterion whether an element should be h-refined and when p-refined. Houston and Süli explain in [36] such a strategy based on an analyticity estimate which leads to exponential convergence in their 1D model problem. In 2D, the discrete solution $u_{hp}|_E$ in the marked element E is expanded into Legendre polynomials, i.e.

$$u_{hp}|_E(\mathcal{F}_E(x, y)) = \sum_{i=0}^{p} \sum_{j=0}^{p} a_{i,j} L_i(x) L_j(y)$$

with

$$a_{i,j} = \frac{2i+1}{2} \frac{2j+1}{2} \int_{-1}^{1} \int_{-1}^{1} u_{hp}|_E(\mathcal{F}_E(x, y)) L_i(x) L_j(y) \ dx \ dy.$$

To estimate the analyticity of $u|_E$, assume $\log |a_{i,j}| \sim (i + j) \log(\frac{1}{\rho})$ for some radius $\frac{1}{\rho}$. To approximate the radius, a least squares approach is used to compute the slope m of $|\log |a_{i,j}|| = (i + j)m + b$. If $e^{-m} \leq \delta$ then the element is p-refined, else an h-refinement is performed.

3.4 First Results on an a Priori Error Estimate

In this section an a priori error estimate for the variational inequality (3.30) based on [27, 35] is derived, yet without convergence rates. For simplicity the convection coefficient is set to zero.

To work with minimal regularity requirements on u, first the extension operators in [35, Section 4.1] are recalled. Let $V(h) := V_{hp} \cup H^1(\Omega)$ and let the operator $L : V(h) \to [V_{hp}]^2$ be defined by

$$\int_\Omega L(v) \cdot \vec{q} \, dx = \sum_{e \in \Gamma_h \cup \Gamma_D} \int_e [v] \{\vec{q} \cdot \mathbf{n}_e\} \ ds \quad \forall \vec{q} \in [V_{hp}]^2. \tag{3.43}$$

Furthermore, let $U_{\tilde{g}} \in [V_{hp}]^2$ be a lifting of $\tilde{g} \in H^{\frac{1}{2}}(\Gamma)$ defined by

$$\int_\Omega U_{\tilde{g}} \cdot \vec{q} \, dx = \sum_{e \in \Gamma_D} \int_e \tilde{g} \vec{q} \cdot \mathbf{n}_e \ ds \quad \forall \vec{q} \in [V_{hp}]^2. \tag{3.44}$$

Therewith the new bilinear and linear forms

$$\widetilde{a}_\epsilon(u,v) := \sum_{E \in \mathcal{E}_h} \int_E \nabla u \nabla v + uv - L(v)\nabla u + \epsilon L(u)\nabla v \, dx + \sum_{e \in \Gamma_h \cup \Gamma_D} \int_e \frac{\sigma_e p_e^2}{|e|^\beta} [u] [v] \ ds \tag{3.45}$$

$$\widetilde{F}_\epsilon(v) := \int_\Omega fv \, dx + \epsilon \sum_{E \in \mathcal{E}_h} \int_E U_g \nabla v \, dx + \sum_{e \in \Gamma_D} \int_e \frac{\sigma_e p_e^2}{|e|^\beta} vg \ ds. \tag{3.46}$$

can be defined. Since

$$\widetilde{a}_\epsilon(u,v) = a_\epsilon(u,v) \text{ on } V_{hp} \times V_{hp} \quad \text{and} \quad \widetilde{F}_\epsilon(v) = F_\epsilon(v) \text{ on } V_{hp},$$

the discrete variational inequality formulation (3.30) is equivalent to:

$$\text{Find } u_h \in K_{hp} : \quad \widetilde{a}_\epsilon(u_h, v_h - u_h) \geq \widetilde{F}_\epsilon(v_h - u_h) \quad \forall v_h \in K_{hp}. \tag{3.47}$$

Furthermore, since $u - v \in H_0^1(\Omega)$ for $u, v \in K$ and $L(u) = U_g$, $L(v-u) = \vec{0}$, simple algebra yields that $u \in K$ solving (3.3) also satisfies

$$\widetilde{a}_\epsilon(u, v-u) \geq \widetilde{F}_\epsilon(v-u) \quad \forall v \in K. \tag{3.48}$$

Therewith, the discrete and continuous variational inequality formulations are formulated in the same bilinear and linear form. The following lemmas provide the continuity and coercivity of the new bilinear form which is needed for Theorem 3.10.

Lemma 3.13 ([35], Lemma 4.1). *For $\sigma_e \equiv \sigma$ and $\beta = 1$, there exists a constant $C_L > 0$ independent of h, p and σ such that*

$$\|L(v)\|_{L^2(\Omega)}^2 \leq C_L \sigma^{-1} \sum_{e \in \Gamma_h \cup \Gamma_D} \int_e \frac{\sigma p^2}{|e|} | [v] |^2 \ ds \quad \forall v \in V(h)$$

$$\|U_{g_1} - U_{g_1}\|_{L^2(\Omega)}^2 \leq C_L \sigma^{-1} \sum_{e \in \Gamma_D} \int_e \frac{\sigma p^2}{|e|} |g_1 - g_2| \ ds \quad \forall g_1, g_2 \in H^{\frac{1}{2}}(\Gamma)$$

3 Elliptic Obstacle Problems

Lemma 3.14 ([35], Lemma 4.2). *For all u, $v \in V(h)$ there holds*

$$|\widetilde{a}_\epsilon(u, v)| \leq C_C \, \|u\|_{1,hp} \, \|v\|_{1,hp}$$

with $C_C = \max\left\{2, 1 + C_L \sigma^{-1}\right\}$ and C_L the constant in Lemma 3.13.

Lemma 3.15 ([35], Lemma 4.3). *For all $v \in H_0^1(\Omega)$ there holds*

$$\widetilde{a}_\epsilon(v, v) = \|v\|_{1,hp}^2.$$

By Lemma 3.14, Lemma 3.15, and Lemma 3.6 the bilinear form $\widetilde{a}_\epsilon(\cdot, \cdot)$ is continuous and elliptic. Therefore, the abstract a priori error estimate by Falk [27] can be applied.

Theorem 3.10. *Let $u \in K$ solve (3.3) and $u_h \in K_{hp}$ solve (3.30), then there holds for arbitrary $v \in K$ and $v_h \in K_{hp}$*

$$\frac{\alpha}{2} \|u - u_h\|_{1,hp}^2 \leq \left(C_F + C_C \|u\|_{1,hp}\right) \left[\|u - v_h\|_{1,hp} + \|u_h - v\|_{1,hp}\right] + \frac{C_C^2}{2\alpha} \|u - v_h\|_{1,hp}^2$$

if $f \in L^2(\Omega)$, $\vec{\gamma} = \vec{0}$, and $\beta = 1$ with C_C defined in Lemma 3.14 and C_F the continuity constant of $\widetilde{F}_\epsilon(\cdot)$.

Proof. Adding (3.48) and (3.47) and subtracting $\widetilde{a}_\epsilon(u, u_h) + \widetilde{a}_\epsilon(u_h, u)$ yields with the (bi)-linearity

$$\widetilde{a}_\epsilon(u - u_h, u - u_h) \leq \widetilde{F}_\epsilon(u - v_h) + \widetilde{F}_\epsilon(u_h - v) - \widetilde{a}_\epsilon(u, u - v_h) - \widetilde{a}_\epsilon(u, u_h - v) \\ + \widetilde{a}_\epsilon(u - u_h, u - v_h) \qquad (3.49)$$

By the coercivity and continuity there follows

$$\alpha \|u - u_h\|_{1,hp}^2 \leq \left(C_F + C_C \|u\|_{1,hp}\right) \left[\|u - v_h\|_{1,hp} + \|u_h - v\|_{1,hp}\right] \\ + C_C \|u - u_h\|_{1,hp} \|u - v_h\|_{1,hp}.$$

Applying Young's inequality yields the assertion. $\qquad\square$

This estimate is not optimal due to the additional square root in the estimate. By Theorem 3.6 this also yields an a priori error estimate for the mixed method (3.16), yet only for the primal variable. An error estimate for the dual variable requires the discrete inf-sup-condition which for the p-version in the context of IPDG with biorthogonal basis functions is still an open question.

Lemma 3.16. *For $\chi \in V_{hp}$, there holds $g_i = \chi(x_i)$ with $x_i \in G_{hp}$ the set of Gauss-Lobatto points which defines ϕ_i, $1 \leq i \leq \dim V_{hp}$ and g_i as defined in Theorem 3.4.*

Proof. If $\chi \in V_{hp}$, then $\chi(x) = \sum_{j=1}^{\dim V_{hp}} \chi_j \phi_j(x)$. Then, by the biorthogonality of ϕ_i to ψ_j there holds

$$g_i := \frac{1}{D_i} \int_\Omega \sum_{j=1}^{\dim V_{hp}} \chi_j \phi_j(x) \psi_i(x) \, dx = \chi_i \frac{\int_\Omega \phi_i(x) \, dx}{D_i} = \chi_i.$$

On the other hand, since ϕ_j are nodal basis functions there holds $\chi(x_i) = \chi_i\phi_i(x_i) = \chi_i$, which completes the proof. $\qquad\square$

As a consequence of Lemma 3.16 there holds

$$K_{hp} := \{v \in V_{hp} : \ v_i \geq \chi_i\} = \{v \in V_{hp} : \ v(x_i) \geq \chi(x_i) \ \forall \ x_i \in G_{hp}\} \qquad (3.50)$$

if $\chi \in V_{hp}$. This means that for a piecewise polynomial obstacle the weak non-penetration is equivalent to a pointwise non-penetration condition in the Gauss-Lobatto interpolation points.

Lemma 3.17. *If $\chi \in V_{hp}$, $g \in V_{hp}|_\Gamma$ and $E \in \mathcal{E}_h$ are axis-parallel quadrilaterals, then there holds*

$$\inf_{v_h \in K_{hp}} \|u - v_h\|_{1,hp} \leq Ch^{\mu-1}p^{1-\mu} \|u\|_{H^\mu(\Omega)}, \ \mu \geq 1 \qquad (3.51)$$

for $u \in K \cap H^\mu(\Omega)$ with $C > 0$ independent of h and p.

Proof. Let $v_h = \mathcal{I}_{hp}u$ be the interpolation of u in the Gauss-Lobatto points G_{hp}. Since $u \in K$, there holds $v_h(x_i) = u(x_i) \geq \chi(x_i)$. With $\chi \in V_{hp}$, Lemma 3.16 implies $v_h \in K_h$. Furthermore, since $u \in H^\mu(\Omega)$ and v_h is the piecewise polynomial interpolant of a continuous function, this yields in conjunction with $g \in V_{hp}|_\Gamma$, i.e. $(u - v_h)|_\Gamma = 0$, that $\|u - v_h\|_{1,hp} = \|u - v_h\|_{H^1(\Omega)}$. The approximation property follows from [5, Theorem 5.7]. $\qquad\square$

An estimate for the term $\|u_h - v\|_{1,hp}$ remains open. Due to the mesh-dependent norm in conjunction with the weakly enforced Dirichlet condition, Glowinski [31, Theorem I.5.2] cannot be applied to obtain simple convergence. Contrary to DG, it can be applied for the continuous Galerkin case in which the convex cone $K_{hp} \subset H^1(\Omega)$ satisfies non-penetration condition in the Gauss-Lobatto quadrature points [48, 47].

3.5 Numerical Experiments

As a numerical experiment, the example by Bartels and Carstensen [4] is considered, however with additional convection and mass term. The space domain is $\Omega = [-1.5, 1.5]^2$ and the convection coefficient is $\vec{\gamma} = (1, -1)^T$. For the obstacle $\chi \equiv 0$ and volume force $f = -2 + \vec{\gamma} \cdot \nabla u + u$ the exact solution is

$$u(x, t) = \begin{cases} \frac{r^2}{2} - \ln(r) - \frac{1}{2}, & \text{if } r := |x|_2 \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

and is plotted in Figure 3.1 with the green circle indicating the free boundary. Knowing the exact solution, the approximation error of u_{hp} solving (3.16) can be computed in the energy norm for families of meshes. These results and the estimated error η are
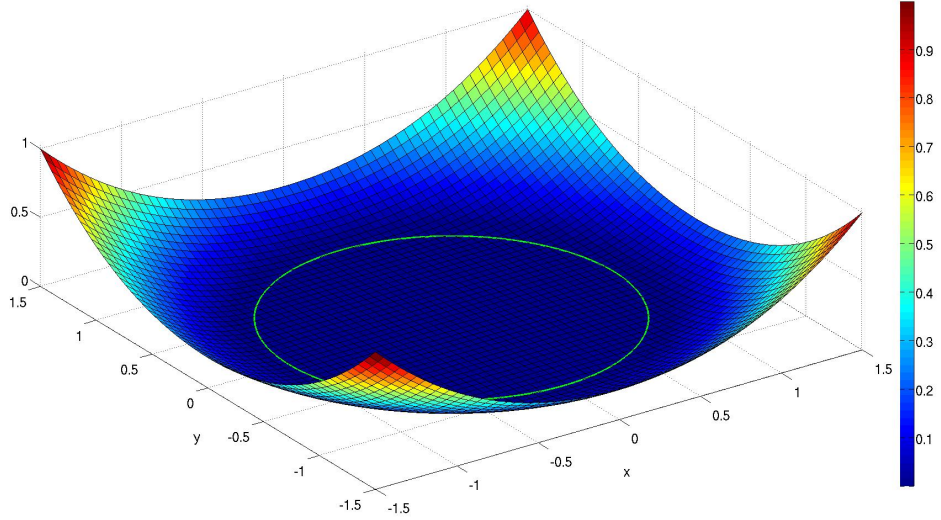
Figure 3.1: Solution with free boundary for the elliptic obstacle problem

plotted in a semilogarithmic scale versus the third root of the degree of freedom (dof). In such scaling, a straight line indicates exponential convergence, i.e.

$$\|u - u_{hp}\|_{1,hp} \leq C e^{-b\sqrt[3]{\mathrm{dof}}} \quad C, b > 0.$$

Since $u \in H^2(\Omega)$, the lowest order uniform h-version has an experimental convergence rate of $\frac{1}{2}$ with respect to the dof, i.e. 1 with respect to h, and the estimated error is almost the same. Consequently, an h-adaptive scheme cannot improve the convergence rate, and only reduces the error constant, which Figure 3.2 clearly displays. For the uniform p-version the error reduces algebraically with respect to the dof, i.e. u is not analytic, whereas the estimated error has a reduced convergence rate. Figure 3.2 indicates that the hp-adaptive method with $\delta = 0.5$ and $\theta = 0.9$ converges exponentially fast and that the estimated error is very close to the true error. Only the kink at 15 to 17 is unsatisfactory, which depends on the shape regularity and may result from a non-optimal marking. By Theorem 3.9, there exists a connection between the saturation constant and the effectivity constant of the error estimator. Figure 3.3 shows such a connection between the experimental effectivity index (EEI) $\eta / \|u - u_{hp}\|_{1,hp}$ and the experimental saturation constant (ESC) $\|u - u_{hp}\|_{1,hp} / \|u - u_{hp+1}\|_{1,hp+1}$, although it is less severe than expected by Theorem 3.9. For the uniform and adaptive h-versions the EEI tends towards one and the ESC towards some value much smaller than 0.1. For the uniform p-version the EEI increases continuously but slowly although the ESC tends towards one rapidly. The reason for the violation of the uniform boundedness of the saturation constant q_S is that the relative FE-space enrichment $\frac{2p+3}{(p+1)^2}$ tends towards zero as $p \to \infty$. This can be avoided by choosing not a $p - (p + 1)$ indicator but a $p - \lceil \sqrt{2}p \rceil$ for which the ESC is around 0.55 and the EEI between 0.9 and 1.1. However, this is extremely CPU-time and memory consuming. For the hp-adaptive method
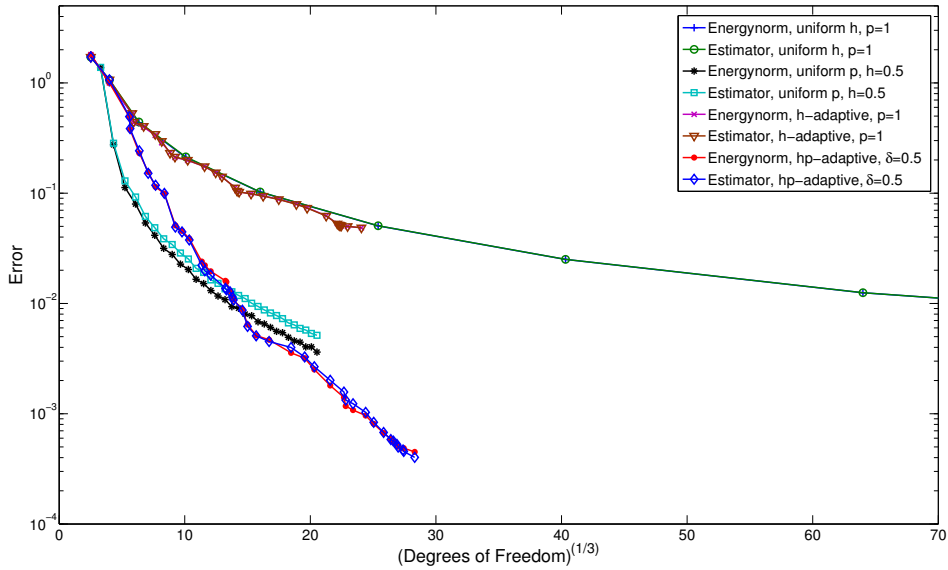
Figure 3.2: Error in energy norm and error indicator for the elliptic obstacle problem

the ESC has an increasing tendency but the EEI is always between 0.8 and 1.2. One advantage of discontinuous Galerkin methods is that irregular meshes impose no (significant) challenges in the implementation of the algorithm, but the irregularity seems to influence the approximation error. Although, shape regularity requirements results in additional refinement within adaptive schemes, Figure 3.4a shows that for this specific experiment, allowing three hanging nodes in an h-adaptive method, yields a better error constant than with only one or even infinity many hanging nodes. Similarly, Figure 3.4b implies that allowing the polynomial degree to differ at most by one between two adjacent elements is superior to a difference of two or infinity in an hp-adaptive method with at most three hanging nodes. Here, superiority is in terms of error constant, convergence rate and the kink discussed above. Since the obstacle is constant, the adaptive algorithm should identify the true contact set and refine the mesh in that area as little as necessary. The tenth and 31st mesh generated by the h-adaptive scheme are displayed in Figure 3.6a-b. These meshes are symmetric and the free boundary is identified. In the non-contact set $\mathcal{N}$ they are almost uniform meshes and in the contact set $\mathcal{C}$ only those refinements necessary for the shape regularity constraint are carried out. Analogous observations can be made for the hp-adaptive meshes, Figure 3.6c-d, with some significant differences. The free boundary is finer resolved and away from it, where the solution is smooth, the polynomial degree is increased in $\mathcal{N}$. The high polynomial degrees within the contact set are a result of the shape regularity on the polynomial degree distribution. If the free boundary is (almost) parallel to one of the element axis and only slightly intersects this specific element (c.f. Figure 3.5), then the solution is smooth in the parallel direction and singular in the perpendicular direction. However, many points in that direction are relatively "far" away from the singularity,
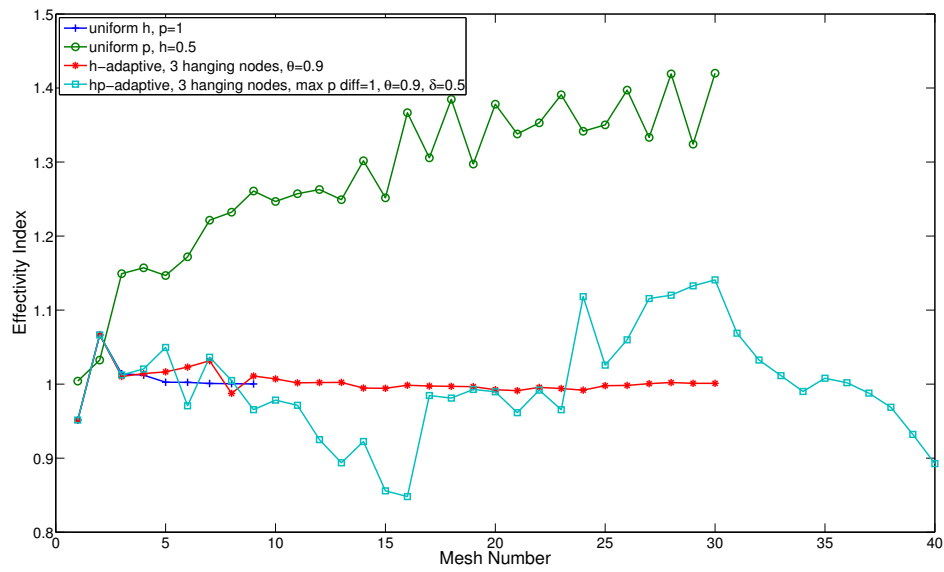
(a) Experimental efficiency index
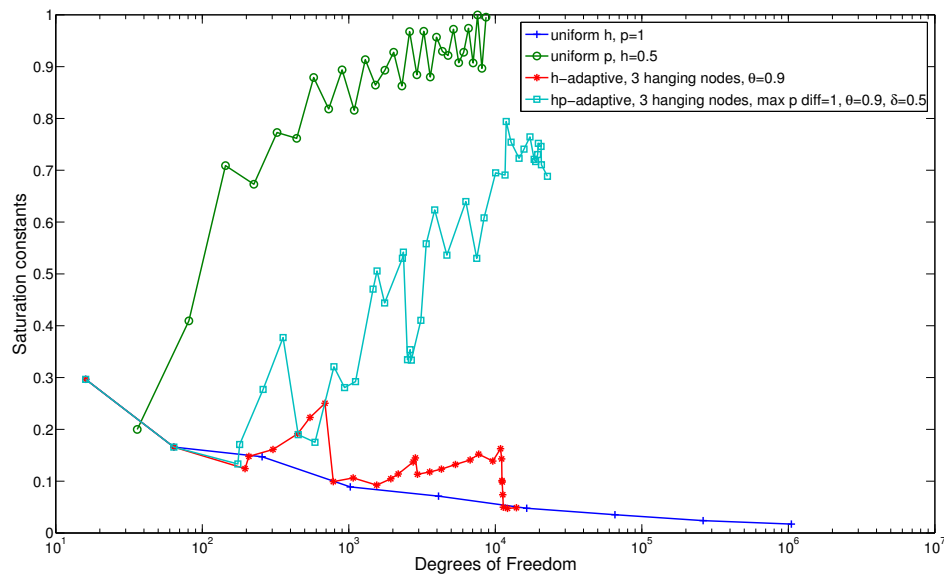
(b) Experimental saturation constant

Figure 3.3: Experimental efficiency index and saturation constant for the elliptic obstacle problem
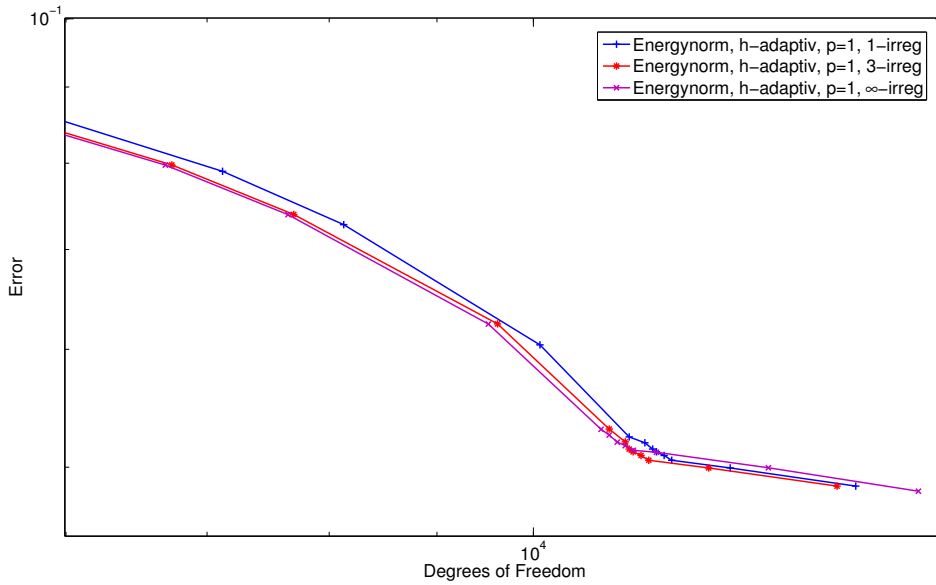
(a) Different maximal number of hanging nodes
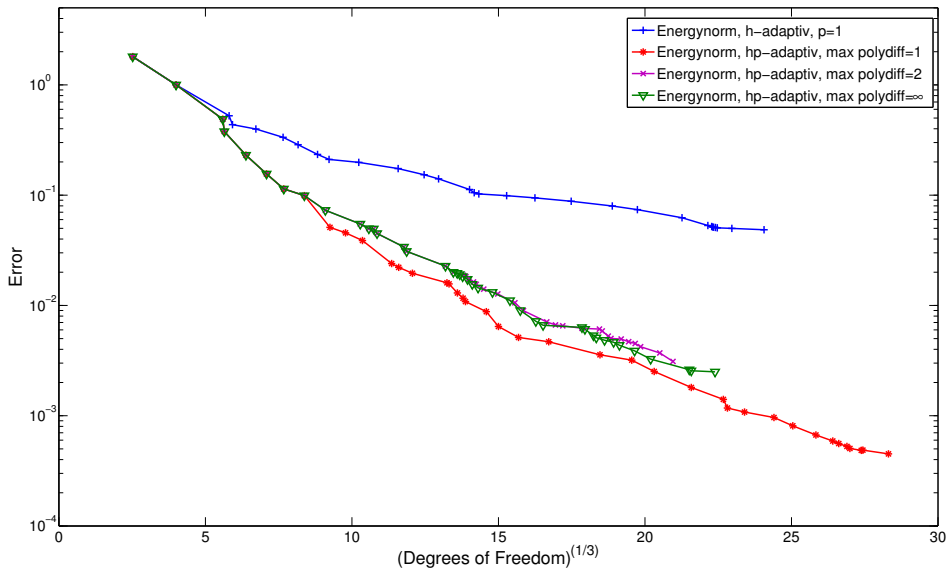
(b) Different maximal polynomial degree difference

Figure 3.4: Influence of the shape regularity on the error and indicator for the elliptic obstacle problem
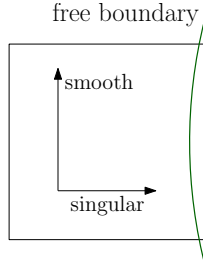
Figure 3.5: Singular direction induced by free boundary

and, therefore, the isotropic refinement with the analyticity estimate described in Section 3.3 yields an increase in p. An often noted, but still unproven observation is that the contact set identified by the discrete solution is an "optimal" approximation of the true contact set on the given mesh. The red stars in Figure 3.7 are the Gauss-Lobatto points (which were used to define the basis functions) in which the discrete solution is in contact.

Hypothesis 3.1. *For the uniform h- and p-version there seems to hold that the node is correctly identified to be a contact node if it lies is in the contact set. Otherwise, if twice the distant to the free boundary is less than the distant to the next Gauss-Lobatto point, it is also identified to be a contact node.*

Such a result would be very important as the correct identification of the free boundary is required in the optimal stopping time for American put options [63]. Finally, the choice of the method's parameter $\epsilon \in \{-1, 0, +1\}$ has no significant influence on the error within the hp-adaptive method as displayed in Figure 3.8a. On the other hand, Figure 3.8b shows that the choice of the penalty parameter σ_e is crucial. If σ_e is sufficiently large, the hp-adaptive method converges exponentially fast with the error in the energy norm depending on σ_e since the norm itself depends on σ_e. If σ_e is chosen too small than the bilinear form is no longer coercive and the method does not converge. Interestingly, the estimated error is still (almost) identical to the error and therefore identifies the non-convergence, although the non-convergence implies that the saturation assumption required for the error estimator is not valid. Crucial for the hp-adaptivity is the decision when to refine h and when to increase p. If δ is close to zero, h refinements will be favored and for δ close to one, p refinements will be favored. This means that the hp-adaptivity deteriorates to an h- or p-adaptive scheme, with only algebraic convergence. The parameter study displayed in Figure 3.9 indicates that the best convergence rate is obtained for $\delta = 0.5$.
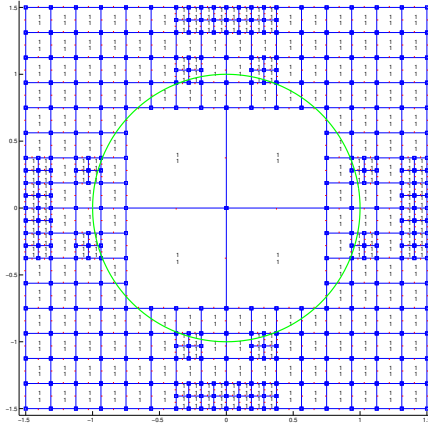
(a) Mesh nr. 10, *h*-adaptive
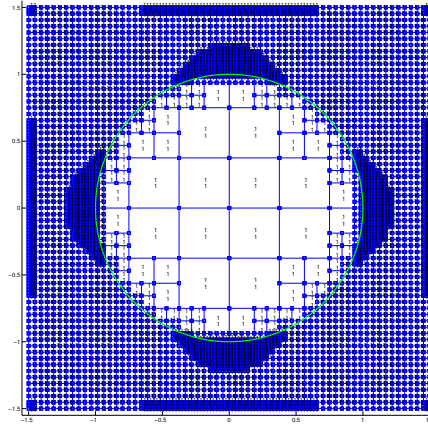
(b) Mesh nr. 31, *h*-adaptive
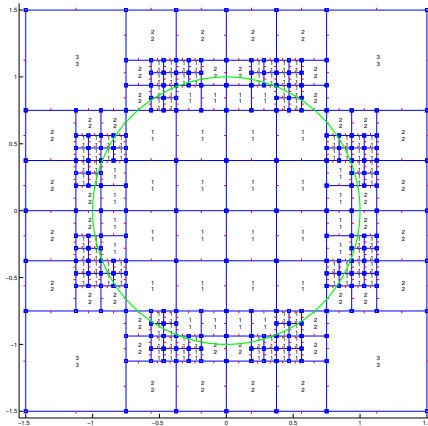
(c) Mesh nr. 10, *hp*-adaptive
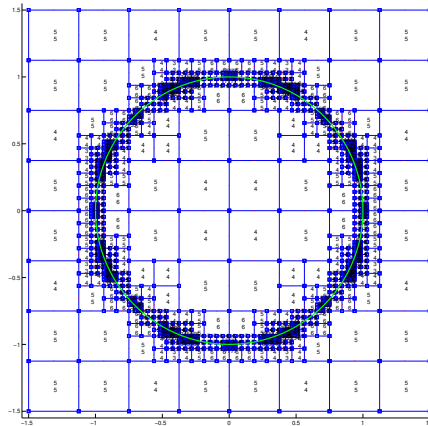
(d) Mesh nr. 41, *hp*-adaptive

Figure 3.6: Different meshes generated by the *h*- and *hp*-adaptive algorithm with a maximum of three hanging nodes and polynomial difference of one for the elliptic obstacle problem
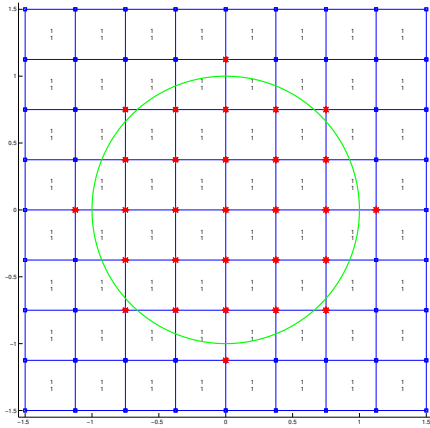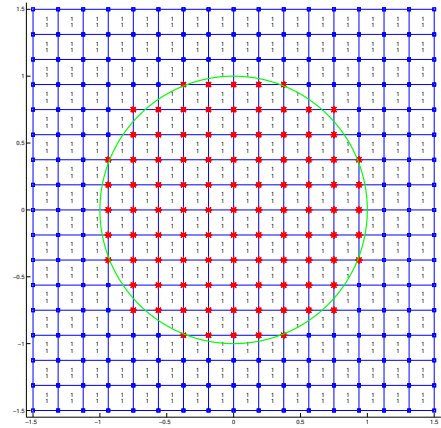
(a) uniform $h = 3/8$, $p = 1$

(b) uniform $h = 3/16$, $p = 1$
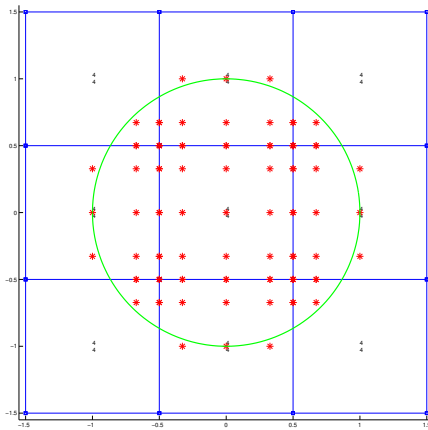
(c) uniform $h = 1/2$, $p = 4$
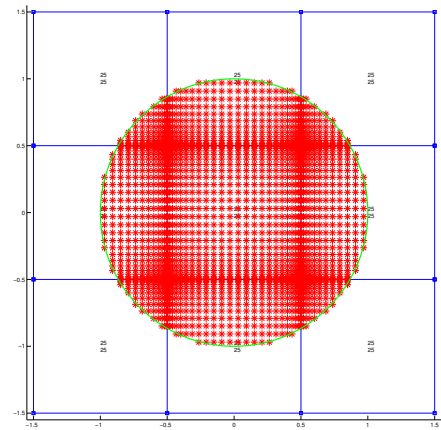
(d) uniform $h = 1/2$, $p = 25$

Figure 3.7: Identified contact set by discrete solution for different h- and p-uniform meshes for the elliptic obstacle problem
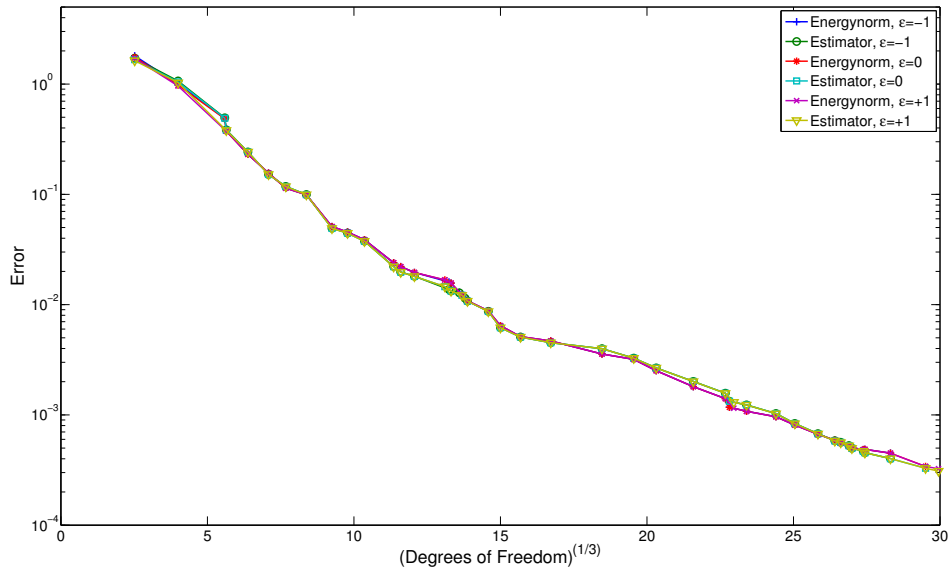
(a) Different interior penalty methods, variation of ϵ
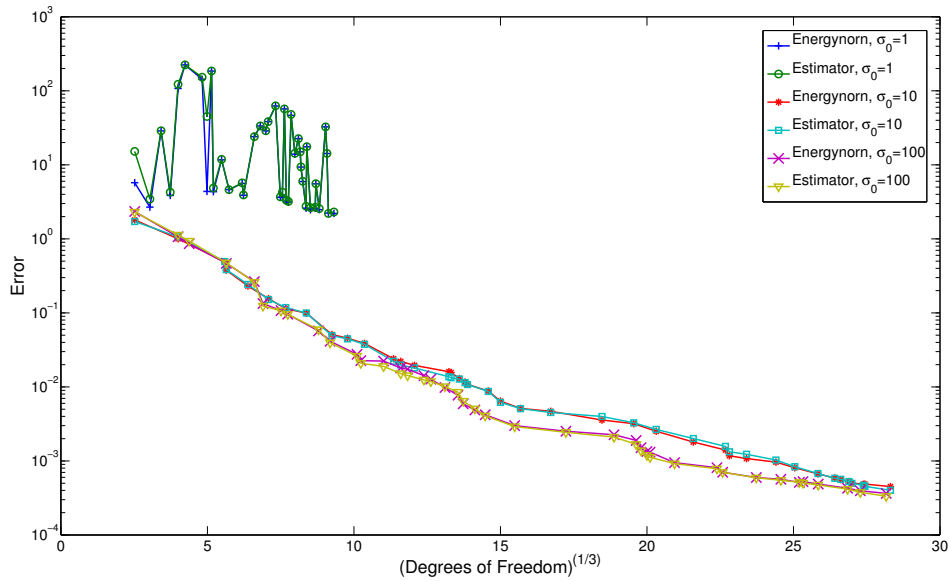
(b) Different penalty factors, variation of σ_e

Figure 3.8: Influence of ϵ and σ_e on error and indicator for hp-adaptivity for the elliptic obstacle problem
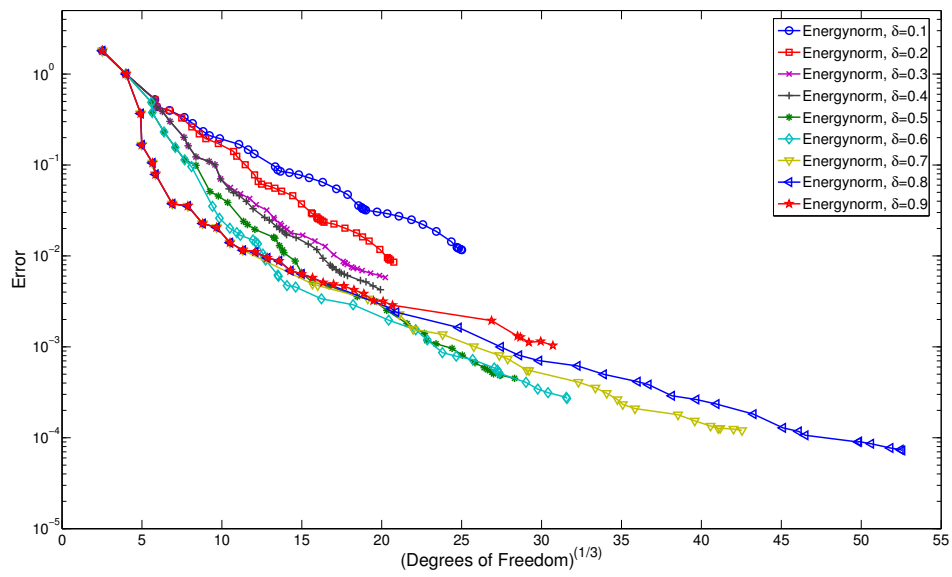
Figure 3.9: Influence of δ on the hp-adaptivity for the elliptic obstacle problem

4 Parabolic Obstacle Problems

Over the last decades, options became an important feature of the financial markets. In 2010 over 3.899 billion option contracts were traded with a total amount of 1.2 trillion US dollars as reported by the Options Industry Council on January the 3rd 2011. This is the eighth year in a row in which the trading volume has increased. Most of the traded options are of American type [79, 46] which allows the holder to exercise his right at any time prior to maturity. Within the Black-Scholes model [6] the "fair" price of an American put basket option is the solution of a parabolic obstacle problem similar to (4.1). It is therefore indispensable to find solution procedures which are both efficient and accurate.

4.1 Weak Formulations for Parabolic Obstacle Problems

Let $I = (0, T)$ be the time interval and $\Omega \subset \mathbb{R}^d$ the open, bounded, polygonal space domain with boundary $\partial\Omega$. For given volume data $f \in L^2(I; H^{-1}(\Omega))$, Dirichlet data $g_D := G_D|_{\partial\Omega} \in L^2(I; H^{\frac{1}{2}}(\partial\Omega))$ with $\frac{\partial G_D}{\partial t} \in L^2(I; H^{-1}(\Omega))$, obstacle $\chi \in L^2(I; H^1(\Omega))$, convection coefficient $\vec{\gamma}$ with $\chi|_{\partial\Omega} = g_D$, $\vec{\gamma} \in \mathbb{R}^d$ for simplicity, and initial data $u_0 \in L^2(\Omega)$. The parabolic obstacle problem is to find a function u such that

$$\dot{u} - \Delta u + \vec{\gamma} \cdot \nabla u + u \geq f \quad \text{in } \Omega \times I \tag{4.1a}$$

$$u = g_D \quad \text{on } \partial\Omega \times I \tag{4.1b}$$

$$u \geq \chi \quad \text{in } \Omega \times I \tag{4.1c}$$

$$u(0) = u_0 \quad \text{in } \Omega. \tag{4.1d}$$

The first time derivative is abbreviated by $\dot{v} := \frac{\partial v}{\partial t}$ and $u(0)$ is an abbreviation for $u(\cdot, 0)$. As in the elliptic case of Chapter 3 there exists a weak formulation based on a variational inequality approach and one based on a mixed method. Contrary to the contact problem in Chapter 5 this problem is not equivalent to a constraint minimization problem since the differential operator is not symmetric. For the mixed method let

$$L^+ := \left\{ \mu \in L^2(I; H^{-1}(\Omega)) : \int_I \langle v, \mu \rangle \, dt \leq 0 \text{ for } v \in L^2(I; H_0^1(\Omega)), \ v \leq 0 \right\} \tag{4.2}$$

be the convex Lagrange multiplier set with $L^2(I; H^{-1}(\Omega))$ the dual space to $L^2(I; H_0^1(\Omega))$ [78, Chapter 23.2-23.3]. Then, the corresponding variational mixed formulation is to

find $(u, \lambda) \in W_2^1(I; H^1(\Omega), L^2(\Omega)) \times L^+$ such that $u|_{\partial\Omega} = g_D$ and

$$B(u, v) - \int_I \langle v, \lambda \rangle \ dt = \int_I \langle f, v \rangle \ dt \qquad \forall \, v \in L^2(I; H_0^1(\Omega)) \qquad (4.3\text{a})$$

$$\int_I \langle u, \mu - \lambda \rangle \ dt \geq \int_I \langle \chi, \mu - \lambda \rangle \ dt \quad \forall \, \mu \in L^+ \qquad (4.3\text{b})$$

$$\langle u(0), v \rangle = \langle u_0, v \rangle \qquad \forall \, v \in L^2(\Omega) \qquad (4.3\text{c})$$

with the bilinear form

$$B(u, v) := \int_I \langle \dot{u} + \vec{\gamma} \cdot \nabla u + u, v \rangle + \langle \nabla u, \nabla v \rangle \ dt. \qquad (4.4)$$

Here $W_2^1(I; H^1(\Omega), L^2(\Omega)) := \{ u \in L^2(I; H^1(\Omega)) : \dot{u} \in L^2(I; H^{-1}(\Omega)) \}$ is a Sobolev space with the notation taken from [78, Chapter 23.6] and $\langle \cdot, \cdot \rangle$ the duality pairing or simply the L^2-inner product for L^2-functions. In particular, the embedding $W_2^1(I; H^1(\Omega), L^2(\Omega)) \subseteq C(\bar{I}, L^2(\Omega))$ is continuous, i.e. $u \in C(\bar{I}, L^2(\Omega))$.

Lemma 4.1. *Any solution of* (4.1) *is a solution of* (4.3). *The converse holds in a distributional sense.*

Proof. Follows from Lemma 4.2 and Theorem 4.1. $\qquad \square$

For an alternative weak formulation without a Lagrange multiplier which is based on a variational inequality let

$$K := \{ v \in L^2(I; H^1(\Omega)) : v \geq \chi \text{ a.e. in } \Omega \times I, \ v|_{\partial\Omega} = g_D \} \qquad (4.5)$$

be the convex cone of admissible functions. Then, the variational inequality formulation is to find $u \in W_2^1(I; H^1(\Omega), L^2(\Omega)) \cap K$ such that

$$B(u, v - u) \geq \int_I \langle f, v - u \rangle \ dt \quad \forall \, v \in K \qquad (4.6\text{a})$$

$$\langle u(0), v \rangle = \langle u_0, v \rangle \qquad \forall \, v \in L^2(\Omega). \qquad (4.6\text{b})$$

Lemma 4.2. *Any solution of* (4.1) *is a solution of* (4.6). *The converse holds in a distributional sense.*

Proof. (4.1) $\Rightarrow$ (4.6) : From (4.1b)-(4.1c) follows $u \in K$ and (4.1d) implies (4.6b). Testing (4.1a) with $v - u \in K$ and integrating by parts in space yields (4.6a).
(4.1) $\Leftarrow$ (4.6) : From $u \in K$ follows (4.1b)-(4.1c) and (4.6b) implies (4.1d). Partial integration of (4.6a) in a distributional sense yields with $v = w + u$, $w(t) \in C_0^\infty(\Omega)$

$$0 \leq \int_I \int_\Omega (\dot{u} - \Delta u + \vec{\gamma} \cdot \nabla u + u - f) \, w \ dx$$

and with $w \geq 0$, i.e. $v \geq u \geq \chi \Rightarrow v \in K$, follows (4.1a). $\qquad \square$

4.1.1 Existence and Uniqueness of a Weak Solution

To proof existence and uniqueness of a weak solution, the following coercivity result for the bilinear form $B(\cdot, \cdot)$ is required.

Lemma 4.3. *There exists a constant $\alpha > 0$ such that for all $v \in W_2^1(I; H_0^1(\Omega), L^2(\Omega))$ there holds*

$$B(v, v) \geq \alpha \|v\|_{L^2(I;H_0^1(\Omega))}^2 + \frac{1}{2}\|v(T)\|_{L^2(\Omega)}^2 - \frac{1}{2}\|v(0)\|_{L^2(\Omega)}^2. \tag{4.7}$$

Proof. Using Green's formula, it is well known that for $v \in H_0^1(\Omega)$

$$\int_\Omega \vec{\gamma}v \cdot \nabla v\ dx = \frac{1}{2}\int_\Omega \vec{\gamma} \cdot \nabla(v^2)\ dx = \frac{1}{2}\int_{\partial\Omega} n \cdot \vec{\gamma}v^2\ ds = 0$$

and therewith

$$\langle \vec{\gamma} \cdot \nabla v + v, v \rangle + \langle \nabla v, \nabla v \rangle \geq \alpha \|v\|_{H_0^1(\Omega)}^2$$

with $\alpha = 1$. Now, partial integration in time of the first term in $B(\cdot, \cdot)$ yields the assertion. $\square$

Lemma 4.4. *The bilinear form $-\int_I \langle v, \mu \rangle\ dt$ is continuous and satisfies the inf-sup condition with unity constant.*

Proof. The norm of the dual space is given by

$$\|\mu\|_{L^2(I;H^{-1}(\Omega))} = \sup_{0 \neq v \in L^2(I;H_0^1(\Omega))} \frac{\int_I \langle v, \mu \rangle\ dt}{\|v\|_{L^2(I;H_0^1(\Omega))}}.$$

Hence, the continuity condition

$$\left| -\int_I \langle v, \mu \rangle\ dt \right| \leq \|v\|_{L^2(I;H_0^1(\Omega))} \|\mu\|_{L^2(I;H^{-1}(\Omega))}$$

and the inf-sup condition

$$\inf_{0 \neq \mu \in L^+} \sup_{0 \neq -v \in L^2(I;H_0^1(\Omega))} \frac{-\int_I \langle -v, \mu \rangle\ dt}{\|-v\|_{L^2(I;H^1(\Omega))}\|\mu\|_{L^2(I;H^{-1}(\Omega))}} = 1$$

are satisfied where the supremum is taken over $-v$ due to the sign condition of L^+. $\square$

Theorem 4.1. *The problems (4.3) and (4.6) are equivalent.*

Proof. Only the equivalence of (4.3a)-(4.3b) to (4.6a) must be shown.
"(4.3) $\Rightarrow$ (4.6)": Choosing $\mu = 0$, $\mu = 2\lambda$ respectively, in (4.3b) yields

$$\int_I \langle u - \chi, \lambda \rangle\ dt = 0 \quad \text{and} \quad \int_I \langle u - \chi, \mu \rangle\ dt \geq 0 \quad \forall \mu \in L^+. \tag{4.8}$$

Assume there were to exist a Lebesgue measurable set $O^- \subseteq \Omega \times I$ in which $u < \chi$. Choosing $\mu = 1_{O^-} \in L^+$ which is one in O^- and zero elsewhere yields

$$\int_I \langle u - \chi, 1_{O^-} \rangle \ dt = \int_{O^-} u - \chi \ dx \ dt < 0$$

which is a contradiction to (4.8). Hence, $u \in W_2^1(I; H^1(\Omega), L^2(\Omega)) \cap K$. Choosing $v = w - u$ in (4.3a) with $w \in K$ and using that

$$\int_I \langle w - u, \lambda \rangle \ dt = \int_I \langle w - \chi, \lambda \rangle \ dt \geq 0 \tag{4.9}$$

by (4.8), yields the variational inequality (4.6a) exploiting $\lambda \in L^+$ and $w \in K$.
"(4.3) $\Leftarrow$ (4.6)": Firstly, define the residual $Res \in L^2(I; H^{-1}(\Omega))$ of (4.6a) by

$$\int_I \langle Res, v \rangle \ dt := B(u, v) - \int_I \langle f, v \rangle \ dt \quad \forall v \in L^2(I; H_0^1(\Omega)). \tag{4.10}$$

Hence, with u solving (4.6a) there holds additionally: Find $u \in W_2^1(I; H^1(\Omega), L^2(\Omega))$ s.t. $u|_{\partial\Omega} = g_D$ and

$$B(u, w) - \int_I \langle Res, w \rangle \ dt = \int_I \langle f, w \rangle \ dt \quad \forall w \in L^2(I; H_0^1(\Omega))$$

and also by (4.6a)

$$\int_I \langle Res, v - u \rangle \ dt \geq 0 \quad \forall v \in K. \tag{4.11}$$

Choosing $v = \chi$ and $v = 2u - \chi$ yields

$$\int_I \langle Res, u - \chi \rangle \ dt \leq 0, \quad \int_I \langle Res, u - \chi \rangle \ dt \geq 0 \quad \Rightarrow \quad \int_I \langle u - \chi, Res \rangle \ dt = 0.$$

Recall that $u \in K$ then for $\mu \in L^+$, $\int_I \langle u - \chi, \mu \rangle \ dt \geq 0$. In conjunction with the above result this yields (4.3b). Choosing $v = w + u$ with $w \geq 0$ in (4.11) yields $Res \in L^+$ which completes the proof. $\qquad\square$

Lemma 4.5. *If u is the unique solution of (4.6) then there exists a $\lambda \in L^+$ such that (u, λ) is the unique solution of (4.3).*

Proof. Let u be the unique solution of (4.6) then by Theorem 4.1 there exists a λ such that (u, λ) is a solution of (4.3). Also, by this theorem u is the unique solution for the mixed formulation. Now, assume there were to exists λ_1 and λ_2 such that both (u, λ_1) and (u, λ_2) were solutions to (4.3). Then, equation (4.3a) yields

$$\int_I \langle v, \lambda_1 - \lambda_2 \rangle \ dt = 0 \quad \forall v \in L^2(I; H_0^1(\Omega)),$$

which using the definition of the dual norm directly yields $\|\lambda_1 - \lambda_2\|_{L^2(I; H^{-1}(\Omega))} = 0$. $\quad\square$

With the equivalence and coercivity result, the following existence and uniqueness result can be proven.

Theorem 4.2. *There exists exactly one solution to the problems* (4.3) *and* (4.6).

Proof. By Theorem 4.1 and Lemma 4.5 only the uniqueness and existence of a solution u to (4.6) must be proven.

Assume $u_1 \in K$ and $u_2 \in K$ were two solutions to (4.6). Then their difference $\delta := u_1 - u_2 \in W_2^1(I; H_0^1(\Omega), L^2(\Omega))$ satisfies

$$\alpha \|\delta\|_{L^2(I;H_0^1(\Omega))}^2 + \frac{1}{2} \|\delta(T)\|_{L^2(\Omega)}^2 - \frac{1}{2} \|\delta(0)\|_{L^2(\Omega)}^2 \leq B(\delta, \delta) \leq 0$$

by adding (4.6a) with $v_1 = u_2$, $v_2 = u_1$, and by employing Lemma 4.3. Since u_1 and u_2 satisfy the same initial condition (4.6b), i.e. $\|\delta(0)\|_{L^2(\Omega)}^2 = 0$, this provides uniqueness. The proof of existence is based on a finite difference approximation in time similar to [39] yet without a Yosida-Moreau approximation of the complementarity condition for the Lagrange multiplier. First, decompose $u = \tilde{u} + u_D$ with $\tilde{u} \in W_2^1(I; H_0^1(\Omega), L^2(\Omega))$, $\tilde{u}(0) = 0$ and a Dirichlet lift $u_D \in W_2^1(I; H^1(\Omega), L^2(\Omega))$, s.t. $u_D|_{\partial\Omega} = g_D$ and

$$\int_I \langle \dot{u}_D + u_D, v \rangle + \langle \nabla u_D, \nabla v \rangle \; dt = 0 \qquad \forall v \in L^2(I; H_0^1(\Omega)) \tag{4.12a}$$

$$\langle u_D(0), v \rangle = \langle u_0, v \rangle \quad \forall v \in L^2(\Omega). \tag{4.12b}$$

The existence of such a u_D is well known, and follows directly from the extension operator mapping from $H^{\frac{1}{2}}(\partial\Omega)$ to $H^1(\Omega)$ and the solvability of the heat equation. Let for a fixed time t

$$\mathcal{K}(t) = \left\{ v \in H_0^1(\Omega) : \; v \geq \chi(t) - u_D(t) \text{ a.e. in } \Omega \right\}$$

be the new time dependent, convex cone of admissible functions consisting of time independent functions, and set for $n \in \mathbb{N}$

$$\Delta t = \frac{T}{n}, \quad t^j = j\Delta t, \quad \tilde{u}^0 = 0, \quad \langle f^j, v \rangle = \left\langle \frac{1}{\Delta t} \int_{t^{j-1}}^{t^j} f(t) - \vec{\gamma} \cdot \nabla u_D(t) \; dt, v \right\rangle. \tag{4.13}$$

Then, due to (4.6a) the finite difference problem is to find $\tilde{u}^j \in \mathcal{K}(t^j)$ $(1 \leq j \leq n)$ such that

$$\left\langle \frac{\tilde{u}^j - \tilde{u}^{j-1}}{\Delta t}, v - \tilde{u}^j \right\rangle + a(\tilde{u}^j, v - \tilde{u}^j) \geq \langle f^j, v - \tilde{u}^j \rangle \quad \forall v \in \mathcal{K}(t^j) \tag{4.14}$$

with the $H_0^1(\Omega)$-coercive and $H^1(\Omega)$-continuous bilinear form

$$a(u, v) := \langle \vec{\gamma} \cdot \nabla u + u, v \rangle + \langle \nabla u, \nabla v \rangle.$$

Stampacchia and Kinderlehrer [45, Theorem 2.1] provide the unique existence of $\tilde{u}^j$ since $\mathcal{K}(t^j)$ is a nonempty, closed, convex subset of $H_0^1(\Omega)$ due to the assumption of $\chi|_{\partial\Omega} = g_D = u_D|_{\partial\Omega}$. Multiplying equation (4.14) with (-1) and adding a zero yields

$$\left\langle \frac{\tilde{u}^j - \tilde{u}^{j-1}}{\Delta t}, \tilde{u}^j - v \right\rangle + a(\tilde{u}^j - v + v, \tilde{u}^j - v) \leq \langle f^j, \tilde{u}^j - v \rangle.$$

Then, the same arguments as in [39, p. 421] immediately imply

$$
\frac{1}{2\Delta t}\left(\left\| \tilde{u}^j - v \right\|_{L^2(\Omega)}^2 - \left\| \tilde{u}^{j-1} - v \right\|_{L^2(\Omega)}^2 + \left\| \tilde{u}^j - \tilde{u}^{j-1} \right\|_{L^2(\Omega)}^2 \right) + \alpha \left\| \tilde{u}^j - v \right\|_{H^1(\Omega)}^2 \tag{4.15}
$$
$$
\leq C_a \left\| v \right\|_{H^1(\Omega)} \left\| \tilde{u}^j - v \right\|_{H^1(\Omega)} + \left\| f^j \right\|_{H^{-1}(\Omega)} \left\| \tilde{u}^j - v \right\|_{H^1(\Omega)}
$$

where α, C_a are the ellipticity and continuity constants of $a(\cdot, \cdot)$. Multiplying this inequality with $2\Delta t$, applying Young's inequality $ab \leq \frac{\alpha}{2}a^2 + \frac{1}{2\alpha}b^2$, and summing from 1 to $k \leq n$ yields

$$
\left\| \tilde{u}^k - v \right\|_{L^2(\Omega)}^2 + \sum_{j=1}^k \left(\left\| \tilde{u}^j - \tilde{u}^{j-1} \right\|_{L^2(\Omega)}^2 + \alpha \Delta t \left\| \tilde{u}^j - v \right\|_{H^1(\Omega)}^2 \right) \leq \left\| \tilde{u}^0 - v \right\|_{L^2(\Omega)}^2
$$
$$
\tag{4.16}
$$
$$
+ \sum_{j=1}^k \frac{\Delta t}{\alpha} \left(\left\| f^j \right\|_{H^{-1}(\Omega)}^2 + C_a^2 \left\| v \right\|_{H^1(\Omega)}^2 \right).
$$

Due to v independent of time, Fubini and Cauchy-Schwarz inequality there holds

$$
\left\| f^j \right\|_{H^{-1}(\Omega)}^2 := \left(\sup_{v \in H_0^1(\Omega)\backslash\{0\}} \frac{\langle f^j, v \rangle}{\left\| v \right\|_{H^1(\Omega)}} \right)^2 = \left(\sup_{v \in H_0^1(\Omega)\backslash\{0\}} \frac{\left\langle (\Delta t)^{-1} \int_{t^{j-1}}^{t^j} \widetilde{f}\, dt, v \right\rangle}{\left\| v \right\|_{H^1(\Omega)}} \right)^2
$$
$$
= \left((\Delta t)^{-1} \int_{t^{j-1}}^{t^j} \sup_{v \in H_0^1(\Omega)\backslash\{0\}} \frac{\langle \widetilde{f}, v \rangle}{\left\| v \right\|_{H^1(\Omega)}}\, dt \right)^2 = (\Delta t)^{-2} \left(\int_{t^{j-1}}^{t^j} 1 \cdot \left\| \widetilde{f} \right\|_{H^{-1}(\Omega)}\, dt \right)^2
$$
$$
\leq (\Delta t)^{-1} \int_{t^{j-1}}^{t^j} \left\| \widetilde{f} \right\|_{H^{-1}(\Omega)}^2\, dt
$$

with $\widetilde{f}(t) := f(t) - \vec{\gamma} \cdot \nabla u_D(t)$. Since $v \in H^1(\Omega)$ is arbitrary and $\sum_{j=1}^k \frac{\Delta t}{\alpha} = \frac{T}{\alpha}\frac{k}{n} \leq \frac{T}{\alpha}$, the right hand side in (4.16) is bounded independently of Δt by

$$
\left\| f - \vec{\gamma} \cdot \nabla u_D \right\|_{L^2(I; H^{-1}(\Omega))}^2 + \frac{TC_a^2}{\alpha} \left\| v \right\|_{H^1(\Omega)}^2 < \infty.
$$

Moreover,

$$
\tilde{u}_{\Delta t}^{(1)} = \tilde{u}^j + \frac{t - t^j}{\Delta t}\left(\tilde{u}^{j+1} - \tilde{u}^j \right) \quad \text{on } \left(t^j, t^{j+1} \right]
$$

is bounded in $W_2^1(I; H_0^1(\Omega), L^2(\Omega))$ independently of Δt. Hence, the Aubin lemma [21, 55] yields the existence of a subsequence which converges to some $\tilde{u}$ weakly in $W_2^1(I; H_0^1(\Omega), L^2(\Omega))$ and strongly in $L^2(I; L^2(\Omega))$. Furthermore, the sequence

$$
\tilde{u}_{\Delta t}^{(2)} = \tilde{u}^{j+1} \quad \text{on } \left(t^j, t^{j+1} \right]
$$

has a subsequences which converges to the same limit $\tilde{u}$ weakly in $L^2(I; H^1(\Omega))$ and strongly in $L^2(I; L^2(\Omega))$ since by (4.16)

$$
\int_I \left\| \tilde{u}_{\Delta t}^{(1)} - \tilde{u}_{\Delta t}^{(2)} \right\|_{L^2(\Omega)}^2\, dt = \frac{\Delta t}{3} \sum_{j=0}^{n-1} \left\| \tilde{u}^j - \tilde{u}^{j+1} \right\|_{L^2(\Omega)}^2 \leq C\frac{\Delta t}{3} \to 0.
$$

Since the functional $v \mapsto \int_I a(v,v) \, dt$ is convex and continuous in $L^2(I; H_0^1(\Omega))$, it is sequentially weakly lower semicontinuous. Then, with (4.14) it follows that the limit $\tilde{u}(t) \in \mathcal{K}(t)$ satisfies

$$\int_I \langle \dot{\tilde{u}}, v - \tilde{u} \rangle + a(\tilde{u}, v - \tilde{u}) \, dt \geq \int_I \langle f(t) - \vec{\gamma} \cdot \nabla u_D, v - \tilde{u} \rangle \, dt \quad \forall \, v(t) \in \mathcal{K}(t) \quad \text{(4.17a)}$$

$$\langle \tilde{u}(0), v \rangle = 0 \qquad \forall \, v \in L^2(\Omega). \quad \text{(4.17b)}$$

Together with the Dirichlet lift u_D, this yields the existence of u solving (4.6) since adding (4.12b) with (4.17b) yields (4.6b), and adding (4.17a) with (4.12a) yields

$$B(\tilde{u} + u_D, v - \tilde{u}) \geq \int_I \langle f, v - \tilde{u} \rangle \, dt \quad \forall \, v(t) \in \mathcal{K}(t).$$

Adding the zero $u_D - u_D$ in the test function and using that $\tilde{u}(t) \in \mathcal{K}(t)$ for almost every $t \in I$ implies that $u = \tilde{u} + u_D \in K$, then this yields (4.6a). $\qquad \square$

4.2 *hp*-TDG/IPDG Discretization for Parabolic Obstacle Problem

Seeking the discrete Lagrange multiplier Λ in a convex set spanned by basis functions which are globally biorthogonal to the basis functions of the discrete primal variable U allows again the componentwise decoupling of the inequality constraints in the mixed method. Hence, the discretization method from the elliptic case of Chapter 3 is extended to the current parabolic case. Let $\mathcal{T}_{hp}^s$ be a subdivision of Ω into non-overlapping rectangulars in 2D, and cubes in 3D, respectively, enhanced by a polynomial degree distribution. Analogously, $\mathcal{T}_{hp}^t$ is a 1D mesh with a polynomial degree distribution for the time interval I. From $\mathcal{T}_{hp}^t$ a decomposition of $\bar{I} = [0, T]$ is obtained such that $\bar{I} = \bigcup_{n=1}^N \bar{I}_n$ with $I_n = (t_{n-1}, t_n)$. For the time discontinuous Galerkin (TDG) method, a notation for the one-sided limits to and jump across the time interval interface is required.

$$v_+^n := \lim_{0 < s \to 0} v(t_n + s), \quad v_-^n := \lim_{0 > s \to 0} v(t_n + s), \quad [v^n] := v_+^n - v_-^n$$

For the spatial part the interior penalty discontinuous Galerkin (IPDG) method is reused with the same notations. For the primal variable, let

$$V_{hp} := \left\{ v \in L^2(\Omega) : \; v|_K \in \mathbb{P}_{p_K}(K) \; \forall \, K \in \mathcal{T}_{hp}^s \right\} = \text{span} \, \{\phi_j\}_{j=1}^{\dim V_{hp}},$$

$$P_{q_n}(I_n) := \left\{ v : I_n \to V_{hp} : \; v(t) = \sum_{i=0}^{q_n} v_i \vartheta_i(t), \; v_i \in V_{hp} \right\}$$

where ϕ_j are affinely transformed Gauss-Lobatto-Lagrange basis functions defined on a reference square/cube which are constructed by a tensor product of 1D functions. The basis functions in time ϑ_j are also affinely transformed Gauss-Lobatto-Lagrange basis functions, yet on the interval I_n. The space V_{hp} may change from one time strip to the next and, therefore, it can take moving singularities into account.

4.2.1 FE Discontinuous Galerkin Discretization for Mixed Formulation

The discrete Lagrange multiplier is sought in

$$
Q_{q_n}(I_n) := \left\{ \begin{array}{l} \mu : I_n \to M_{hp} : \; \mu(t) = \sum_{i=0}^{q_n} \mu_i \zeta_i(t), \; \mu_i \in M_{hp}, \\ \int_{I_n} \langle \mu, v \rangle \, dt \leq 0 \text{ for } v_{i,j} \leq 0, \; v = \sum_{i=1}^{\dim V_{hp}} \sum_{j=0}^{q_n} v_{i,j} \phi_i \vartheta_j \in P_{q_n}(I_n) \end{array} \right\}.
$$

Furthermore, M_{hp} is the dual space of V_{hp} spanned by biorthogonal basis functions ψ_j, i.e.

$$
\int_\Omega \psi_j \phi_i \, dx = \delta_{ij} \int_\Omega \phi_i \, dx, \qquad 1 \leq i, j \leq \dim V_{hp}.
$$

Analogously, ζ_j are time basis functions biorthogonal to ϑ_j. In particular, the Lagrange multiplier Λ inherits the same mesh and polynomial degree distribution from its primal variable U.

Lemma 4.6. *There holds for the integral value of the primal and dual basis functions* $\int_\Omega \psi_i \, dx = \int_\Omega \phi_i \, dx =: D_i^s > 0$ *and* $\int_{I_n} \zeta_j \, dx = \int_{I_n} \vartheta_j \, dx =: D_j^t > 0$.

Proof. Follows from the biorthogonality relationship, a partition of unity and the same arguments as in Lemma 5.11. □

Lemma 4.7. *The primal and dual basis functions span the same set, i.e.*

$$
\text{span} \, \{\phi_j\}_{j=1}^{\dim V_{hp}} = \text{span} \, \{\psi_j\}_{j=1}^{\dim V_{hp}}, \qquad \text{span} \, \{\vartheta_j\}_{j=0}^{q_n} = \text{span} \, \{\zeta_j\}_{j=0}^{q_n}.
$$

Proof. From [52, Equation 2.6] the local basis functions of the one type can be written as a linear combination of the local basis function of the other type. In particular, functions in $P_{q_n}(I_n)$ and $Q_{q_n}(I_n)$ are piecewise polynomials and discontinuous. Since the affinely transformed local basis functions are extended discontinuously by zero to obtain the global basis function, the assertion follows trivially. □

Hence, the discrete mixed DG method is: For $1 \leq n \leq N$, let U_-^{n-1} be known, find $U = U|_{I_n} \in P_{q_n}(I_n)$ and $\Lambda = \Lambda|_{I_n} \in Q_{q_n}(I_n)$ such that

$$
\int_{I_n} \langle \dot{U}, v \rangle + a_\epsilon(U, v) + b_{\vec{\gamma}}(U, v) - \langle v, \Lambda \rangle \, dt + \langle [U^{n-1}], v_+^{n-1} \rangle = \int_{I_n} F_\epsilon(v) \, dt \quad \forall v \in P_{q_n}(I_n)
$$

$$(4.18a)$$

$$
\int_{I_n} \langle U, \mu - \Lambda \rangle \, dt \geq \int_{I_n} \langle \chi, \mu - \Lambda \rangle \, dt \qquad \forall \mu \in Q_{q_n}(I_n) \qquad (4.18b)
$$

with the bilinear, linear forms,

$$a_\epsilon(u,v) := \sum_{E\in\mathcal{E}_h}\int_E \nabla u\nabla v + uv\, dx + \sum_{e\in\Gamma_h\cup\Gamma_D}\int_e -\left\{\frac{\partial u}{\partial\mathbf{n}_e}\right\}[v] + \epsilon\left\{\frac{\partial v}{\partial\mathbf{n}_e}\right\}[u] + \frac{\sigma_e p_e^2}{|e|^\beta}[u][v]\, ds$$

$$(4.19)$$

$$b_{\vec{\gamma}}(u,v) := -\sum_{E\in\mathcal{E}_h}\int_E \vec{\gamma}u\cdot\nabla v\, dx + \sum_{e\in\Gamma_h}\int_e \vec{\gamma}\cdot\mathbf{n}_e u^{\mathrm{up}}[v]\, ds \qquad (4.20)$$

$$F_\epsilon(v) := \int_\Omega fv\, dx + \sum_{e\in\Gamma_D}\int_e \left(\epsilon\frac{\partial v}{\partial\mathbf{n}_e} + \left(\frac{\sigma_e p_e^2}{|e|^\beta} - \vec{\gamma}\cdot\mathbf{n}_e\right)v\right)g_D\, ds, \qquad (4.21)$$

respectively. Here, U_-^0 is an approximation of u_0. The choice of the parameter $\epsilon \in \{-1,0,1\}$ determines which particular IPDG method is used, e.g. for $\epsilon = 1$ and $\sigma_e = 1$ it is called the non-symmetric interior penalty Galerkin (NIPG) method and for $\epsilon = 0$ incomplete interior penalty Galerkin (IIPG), c.f. [67, 66] among others. The penalty parameter σ_e is always non-negative but may vary for different edges [68]. The exponent β is a positive constant depending on the dimension d of Ω such that $\beta(d-1) \geq 1$ and p_e is the maximum of the two polynomial degrees on the edge e. Recall from Chapter 3 the mesh dependent norm

$$\|v\|_{1,hp}^2 := \sum_{E\in\mathcal{E}_h}\int_E(\nabla v)^2 + v^2\, dx + \sum_{e\in\Gamma_h\cup\Gamma_D}\int_e \frac{\sigma_e p_e^2}{|e|^\beta}[v]^2\, ds,\ \ \|v\|_{L^2(I_n;1,hp)}^2 := \int_{I_n}\|v\|_{1,hp}^2\, dt.$$

Lemma 4.8. *If σ_e is sufficiently large, then there exists a constant $\alpha > 0$ such that for all $v \in P_{q_n}(I_n)$*

$$\int_{I_n}\langle\dot{v},v\rangle + a_\epsilon(v,v) + b_{\vec{\gamma}}(v,v)\, dt + \langle v_+^{n-1}, v_+^{n-1}\rangle \geq \alpha\|v\|_{L^2(I_n;1,hp)}^2 + \frac{1}{2}\|v_-^n\|_{L^2(\Omega)}^2 + \frac{1}{2}\|v_+^{n-1}\|_{L^2}^2.$$

$$(4.22)$$

Proof. By [66, p. 38 and p. 99] $a_\epsilon(v,v) \geq \alpha\|v\|_{1,hp}^2$ and $b_{\vec{\gamma}}(v,v) \geq 0$. Partial integration in time yields the assertion. $\qquad\square$

In the following, σ_e is always assumed to be sufficiently large such that Lemma 4.8 can be applied.

Theorem 4.3. *There exists exactly one solution to the discrete problem (4.18).*

Proof. **Uniqueness:** Assume there were to exist two different solutions $(u_1, \lambda_1) \neq (u_2, \lambda_2)$. Then their difference $u_1|_{I_n} - u_2|_{I_n} =: w \in P_{q_n}(I_n)$ and $\lambda_1|_{I_n} - \lambda_2|_{I_n} =: \delta$ satisfy the equation

$$\int_{I_n}\langle\dot{w},v\rangle + a_\epsilon(w,v) + b_{\vec{\gamma}}(w,v) - \langle v,\delta\rangle\, dt + \langle[w^{n-1}], v_+^{n-1}\rangle = 0 \qquad \forall v \in P_{q_n}(I_n).$$

$$(4.23)$$

Choosing $v = w$ and using the coercivity of Lemma 4.8 yields

$$\alpha \left\|w\right\|^2_{L^2(I_n;1,hp)} + \frac{1}{2}\left\|w^n_-\right\|^2_{L^2(\Omega)} + \frac{1}{2}\left\|w^{n-1}_+\right\|^2_{L^2(\Omega)} - \left\langle w^{n-1}_-, w^{n-1}_+ \right\rangle - \int_{I_n} \left\langle w, \delta \right\rangle \, dt \le 0.$$
(4.24)

By choosing $\mu_1 = \lambda_2$ and $\mu_2 = \lambda_1$ in (4.18b) and adding these inequalities, the last term in the above equation is bounded by

$$\int_{I_n} \left\langle u_1 - u_2, \lambda_1 - \lambda_2 \right\rangle \, dt \le 0.$$
(4.25)

Hence, if $w^{n-1}_- = 0$, i.e. u_1 and u_2 have the same initial condition, then $u_1 = u_2$. This trivially holds for $n = 1$ and by induction for all $1 \le n \le N$. Consequently, (4.23) implies with the biorthogonality of the basis functions

$$0 = \int_{I_n} \left\langle v, \delta \right\rangle \, dt = \sum_{i,j} \delta_{i,j} v_{i,j} D^s_i D^t_j$$
(4.26)

with $D^s_i := \int_\Omega \phi_i \, dx$, $D^t_j := \int_{I_n} \vartheta_j \, dt$. From $D^s_i D^t_j > 0$ (c.f. Lemma 4.6) and v arbitrary, i.e. $v_{i,j}$ arbitrary, follows immediately $\delta_{i,j} = 0$, i.e. $\delta = 0 \Rightarrow \lambda_1 = \lambda_2$. This contradicts the assumption of $(u_1, \lambda_1) \ne (u_2, \lambda_2)$ being two different solutions to (4.18).

Existence: Due to Lemma 4.7, it is well known that the problem (4.18b) can be written as the projection equation [51]

$$\lambda = \mathcal{P}_{Q_{q_n}(I_n)}\left(\lambda + r(\chi - u)\right)$$
(4.27)

where $\mathcal{P}_{Q_{q_n}(I_n)}$ is the L^2-projection operator mapping onto $Q_{q_n}(I_n)$ and $r > 0$ is an arbitrary constant. For any given λ, the remaining problem of determining $u(\lambda)$ reduces to solving a system of linear equations. By Lemma 4.8, the real part of the system matrix's eigenvalues are positive and thus provides a corresponding unique $u(\lambda)$. Hence, the solution sequence of an Uzawa iteration is fully described by the fixed point iteration $\lambda^{(k+1)} = T\lambda^{(k)}$ using the mapping

$$T: \; Q_{q_n}(I_n) \to Q_{q_n}(I_n)$$
$$\lambda \mapsto \mathcal{P}_{Q_{q_n}(I_n)}\left(\lambda + r(\chi - u)\right).$$

If T is a contraction, then by the Banach fixed point theorem there exist a λ which satisfies (4.27) and, hence, also a corresponding u solving (4.18). For the ease of presentation denote $\delta\lambda = \lambda_1 - \lambda_2$, $\delta u = u_1 - u_2$ and $\|\cdot\| = \|\cdot\|_{L^2(I_n;L^2(\Omega))}$.

$$\begin{aligned}
\left\|T\lambda_1 - T\lambda_2\right\|^2 &= \left\|\mathcal{P}_{Q_{q_n}(I_n)}\left(\lambda_1 + r(\chi - u_1)\right) - \mathcal{P}_{Q_{q_n}(I_n)}\left(\lambda_2 + r(\chi - u_2)\right)\right\|^2 \\
&\le \left\|\delta\lambda - r\delta u\right\|^2 \\
&= \left\|\delta\lambda\right\|^2 - 2r\int_{I_n} \left\langle \delta\lambda, \delta u \right\rangle \, dt + r^2\left\|\delta u\right\|^2 \\
&= \left\|\delta\lambda\right\|^2 - 2r\left[\int_{I_n}\left\langle \dot{\delta u}, \delta u \right\rangle + a_\epsilon(\delta u, \delta u) + b_{\vec{\gamma}}(\delta u, \delta u)dt + \left\langle \delta u^{n-1}_+, \delta u^{n-1}_+ \right\rangle\right] + r^2\left\|\delta u\right\|^2 \\
&\le \left\|\delta\lambda\right\|^2 - 2\alpha r\left\|\delta u\right\|^2 + r^2\left\|\delta u\right\|^2 \\
&= \left\|\delta\lambda\right\|^2\left(1 - 2\alpha r\gamma^2 + r^2\gamma^2\right)
\end{aligned}$$

with $\gamma = \frac{\|\delta u\|_{L^2(I_n;L^2(\Omega))}}{\|\delta \lambda\|_{L^2(I_n;L^2(\Omega))}}$. The second line is the standard projection result, the fourth line results form (4.23) and the same initial condition, and the fifth line from Lemma 4.8. Hence, for $0 < r < 2\alpha$, T is a strict contraction which completes the proof. $\qquad\square$

The use of biorthogonal basis functions allows the componentwise decoupling of the above variational inequality constraint (4.18b) which is a key element of this approach and allows the application of a semi-smooth Newton solver to an equivalent discrete problem.

Theorem 4.4. *The condition* (4.18b) *is equivalent to the system*

$$u_{i,j} \geq g_{i,j} := \frac{1}{D_i^s D_j^t} \int_{I_n} \int_\Omega \chi \psi_i(x) \zeta_j(t) \, dx \, dt \tag{4.28a}$$

$$\lambda_{i,j} \geq 0 \tag{4.28b}$$

$$\lambda_{i,j} \left(u_{i,j} - g_{i,j} \right) = 0 \tag{4.28c}$$

for $1 \leq i \leq \dim V_{hp}$ and $0 \leq j \leq q_n$. Here, $u_{i,j}$ and $\lambda_{i,j}$ are the expansion coefficients of U and Λ, respectively.

Proof. This proof follows the idea of the proof of [37, Lemma 2.6] generalized to a higher order approach and to higher dimensional problems with tensor product structure. Every function $v \in P_{q_n}(I_n)$ can be written in its linear combination

$$v = \sum_{i,j} v_{i,j} \phi_i(x) \vartheta_j(t).$$

Next, write μ and $\lambda \in Q_{q_n}(I_n)$ as linear combinations of $\psi_i(x)\zeta_j(t)$, i.e.

$$\mu = \sum_{i,j} \mu_{i,j} \psi_i(x)\zeta_j(t), \quad \lambda = \sum_{i,j} \lambda_{i,j} \psi_i(x)\zeta_j(t).$$

Due to the biorthogonality of the employed basis functions, there holds for all $v \in P_{q_n}(I_n)$ with $v_{i,j} \leq 0$

$$\int_{I_n} \int_\Omega \mu v \, dx \, dt = \sum_{i,j} \mu_{i,j} v_{i,j} D_i^s D_j^t \leq 0 \tag{4.29}$$

since $\mu \in Q_{q_n}(I_n)$. With $v_{i,j} \leq 0$ arbitrary and D_i^s, D_j^t positive by Lemma 4.6, equation (4.29) yields

$$\mu_{i,j} v_{i,j} \leq 0, \quad v_{i,j} \leq 0 \implies \mu_{i,j} \geq 0 \quad \forall i,j.$$

Hence, $\lambda \in Q_{q_n}(I_n)$ implies (4.28b). Inserting the linear combinations of μ and λ into (4.18b) yields by the biorthogonality and μ arbitrary after dividing by the positive factor $D_i^s D_j^t$:

$$\text{Find } \lambda_{i,j} \geq 0: \quad u_{i,j} \left(\mu_{i,j} - \lambda_{i,j} \right) \geq g_{i,j} \left(\mu_{i,j} - \lambda_{i,j} \right) \quad \forall \mu_{i,j} \geq 0 \quad \forall i,j. \tag{4.30}$$

Choosing $\mu_{i,j} = \lambda_{i,j} + \eta_{i,j} > 0$ with $\eta_{i,j} > 0$ in (4.30) yields (4.28a). Equation (4.28c) is obtained by choosing $\mu_{i,j} = 0$ and $\mu_{i,j} = 2\lambda_{i,j}$.

For the opposite direction, multiplying (4.28a) with $\mu_{i,j} \geq 0$ and adding the zero (4.28c) to the right hand side yields (4.30). Summing over all i, j and exploiting the biorthogonality yields (4.18b). □

4.2.2 FE Discontinuous Galerkin Discretization of VI Formulation

As on the continuous level, there exists a discrete variational formulation which is equivalent to the discrete mixed formulation. This equivalence is exploited to construct an a posteriori error estimator for the mixed method via the variational inequality approach. Instead of a discrete weak non-penetration condition the discrete solution now satisfies a discrete strong non-penetration condition. More precisely,

$$K_{q_n}(I_n) := \{v \in P_{q_n}(I_n) : v_{ij} \geq \chi_{ij}\} \tag{4.31}$$

is the convex cone of admissible functions where χ_{ij} are one sided box constraints on the solution coefficients. Then, the discrete variational inequality problem is:

For $1 \leq n \leq N$, let U_-^{n-1} be known, find $U = U|_{I_n} \in K_{q_n}(I_n)$ such that

$$\int_{I_n} \langle \dot{U}, v - U \rangle + a_\epsilon(U, v - U) + b_{\bar{\gamma}}(U, v - U) \, dt +$$
$$+ \left\langle [U^{n-1}], (v - U)_+^{n-1} \right\rangle \geq \int_{I_n} F_\epsilon(v - U) \, dt \quad \forall \, v \in K_{q_n}(I_n) \tag{4.32}$$

where U_-^0 is an approximation of u_0.

Theorem 4.5. *There exists exactly one solution to the discrete variational inequality formulation* (4.32).

Proof. It is well known [25, 33] that solving (4.32) is equivalent to solving the system of projection equations

$$x = \mathcal{P}_\Theta \left(x - r(Ax + q) \right) =: Tx, \qquad r > 0 \tag{4.33}$$

with x the coefficient vector to U and $\mathcal{P}_\Theta$ the closest point projection onto Θ which is the feasible set of coefficient vectors corresponding to $K_{q_n}(I_n)$. Here, A is the system matrix and $-q$ the known right hand side of (4.32) consisting of the volume term and the known contribution of the time jump. Denote $\delta_x = x_1 - x_2$, $0 < \bar{\gamma} := \lambda_{\max}(A^T A) < \infty$, $\bar{\alpha} := \alpha \lambda_{\min}(M) > 0$ with M the global, symmetric space-time mass matrix, then

$$\|Tx_1 - Tx_2\|_2^2 = \|\mathcal{P}_\Theta (x_1 - r(Ax_1 + q)) - \mathcal{P}_\Theta (x_2 - r(Ax_2 + q))\|_2^2$$
$$\leq \|\delta_x - rA\delta_x\|_2^2$$
$$= \|\delta_x\|_2^2 - r\delta_x^T A\delta_x - r\delta_x^T A^T \delta_x + r^2 \delta_x^T A^T A\delta_x$$
$$= \|\delta_x\|_2^2 - 2r\delta_x^T A\delta_x + r^2 \delta_x^T A^T A\delta_x$$
$$\leq \|\delta_x\|_2^2 \left(1 - 2r\bar{\alpha} + r^2 \bar{\gamma}\right).$$

A is positive definite and finite dimensional. Therefore, the Rayleigh quotient implies $\delta_x^T A^T A \delta_x \leq \bar{\gamma} \delta_x^T \delta_x$. Unfortunately, the Rayleigh quotient cannot directly be applied to the middle term since A is not symmetric. However, abbreviating the whole bilinear form in Lemma 4.8 by $a(\cdot, \cdot)$ there holds

$$\delta_x^T A \delta_x = a(u_1 - u_2, u_1 - u_2) \geq \alpha \|u_1 - u_2\|^2_{L^2(I_n; L^2(\Omega))} = \alpha \delta_x^T M \delta_x \geq \alpha \lambda_{\min}(M) \delta_x^T \delta_x.$$

Hence, for $0 < r < 2\frac{\bar{\alpha}}{\bar{\gamma}^2}$ the function T is a strict contraction and Banach's fixed point theorem yields the unique existence of a solution to (4.33) and therewith to (4.32). □

Theorem 4.6. *The discrete problems* (4.18) *and* (4.32) *are equivalent if* $\chi_{i,j} = g_{i,j}$ *of Theorem 4.4.*

Proof. Let (u, λ) solve (4.18). Then Theorem 4.4 implies $u_{i,j} \geq g_{i,j}$, i.e. $u \in K_{q_n}(I_n)$. Furthermore, the biorthogonality and Theorem 4.4 yield for all $w \in K_{q_n}(I_n)$

$$-\int_{I_n} \langle w - u, \lambda \rangle \, dt = -\sum_{i,j} (w_{i,j} - u_{i,j}) \lambda_{i,j} D_i^s D_j^t = -\sum_{i,j} (w_{i,j} - g_{i,j}) \lambda_{i,j} D_i^s D_j^t \leq 0.$$

Hence, choosing $v = w - u$ with $w \in K_{q_n}(I_n)$ in (4.18) yields (4.32).
For the opposite direction let $u \in K_{q_n}(I_n)$ solve (4.32), i.e. $u_{i,j} \geq g_{i,j}$. Define the Lagrange multiplier by $\lambda = \sum_{i,j} \lambda_{i,j} \psi_i(x) \zeta_j(t)$ with

$$\lambda_{i,j} = \frac{\int_{I_n} \langle \dot{u}, \phi_i \vartheta_j \rangle + a_\epsilon(u, \phi_i \vartheta_j) + b_{\bar{\gamma}}(u, \phi_i \vartheta_j) - F_\epsilon(\phi_i \vartheta_j) \, dt + \langle [u^{n-1}], (\phi_i \vartheta_j)_+^{n-1} \rangle}{D_i^s D_j^t},$$
(4.34)

i.e. $\int_{I_n} \langle \lambda, v \rangle \, dt$ is the residual of the discrete variational inequality for all $v \in P_{q_n}(I_n)$. Choosing $v = u + \phi_i \vartheta_j$, i.e. $v \in K_{q_n}(I_n)$, in (4.32) yields

$$0 \leq \int_{I_n} \langle \dot{u}, \phi_i \vartheta_j \rangle + a_\epsilon(u, \phi_i \vartheta_j) + b_{\bar{\gamma}}(u, \phi_i \vartheta_j) - F_\epsilon(\phi_i \vartheta_j) \, dt + \langle [u^{n-1}], (\phi_i \vartheta_j)_+^{n-1} \rangle$$
$$= \lambda_{i,j} D_i^s D_j^t \Rightarrow \lambda_{i,j} \geq 0.$$

Finally, choose $v \in P_{q_n}(I_n)$ such that $v_{i,j} = g_{i,j}$ and $v_{i,j} = 2u_{i,j} - g_{i,j}$ in (4.32) to obtain

$$0 = \sum_{i,j} \left[\int_{I_n} \langle \dot{u}, \phi_i \vartheta_j \rangle + a_\epsilon(u, \phi_i \vartheta_j) + b_{\bar{\gamma}}(u, \phi_i \vartheta_j) - F_\epsilon(\phi_i \vartheta_j) dt + \langle [u^{n-1}], (\phi_i \vartheta_j)_+^{n-1} \rangle \right] (u_{i,j} - g_{i,j})$$
$$= \sum_{i,j} \lambda_{i,j} (u_{i,j} - g_{i,j}) D_i^s D_j^t = \int_{I_n} \langle u - \chi, \lambda \rangle \, dt.$$

The assertion follows by Theorem 4.4. □

Remark 4.1. *The problem* (4.32) *can be solved with Han's self-adaptive projection method, in which the projection constant is variable and self-adapting [33].*

Proof. Note that there exists a solution to (4.33) and its global system matrix is positive definite. Then the convergence is given by [33, Theorem 2.4]. □

Remark 4.2. *Computations for the numerical experiments in Section 4.5 have shown that the projected SOR method and the active set method of [50] fail to solve (4.32) once the polynomial degree in time is at least one.*

4.2.3 An Algorithm for Solving the Discrete Mixed Formulation

Remark 4.2 and the numerical experiments of Section 4.5 indicate that the variational formulation approach is impracticable due to the lack of appropriate iterative solvers. The mixed method however can be solved very efficiently.

Lemma 4.9. *The discrete problem (4.18) is equivalent to solving*

$$0 \overset{!}{=} F(\vec{u}, \vec{\lambda}) := \begin{pmatrix} A\vec{u} - D\vec{\lambda} - \vec{f} \\ \varphi_\eta(\vec{u}, \vec{\lambda}) \end{pmatrix} \qquad (4.35)$$

where $A\vec{u} - D\vec{\lambda} - \vec{f} = 0$ is the matrix representation of (4.18a) and $\varphi_\eta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is the vector-valued penalized Fischer-Burmeister non-linear complementarity function (NCF) defined by

$$\varphi_\eta(\vec{u}, \vec{\lambda}) = \eta \left(\vec{\lambda} + (\vec{u} - \vec{g}) - \sqrt{\vec{\lambda}^2 + (\vec{u} - \vec{g})^2} \right) + (1 - \eta) \max \left\{ 0, \vec{\lambda} \right\} \max \left\{ 0, \vec{u} - \vec{g} \right\}$$

with $\eta \in (0, 1]$ and a componentwise understood right hand side.

Proof. By Lemma 2.4, φ_η is a NCF, i.e.

$$\varphi_\eta(\vec{u}, \vec{\lambda}) = 0 \iff (u - g)_i \geq 0, \ \lambda_i \geq 0, \ \lambda_i \cdot (u - g)_i = 0 \quad (i = 1, \dots, n).$$

Hence, $\varphi_\eta(\vec{u}, \vec{\lambda}) = 0$ is equivalent to (6.10) which in turn is equivalent to (4.18b) by Theorem 4.4. $\qquad\square$

Lemma 4.10. *The matrix A is (in general) non-symmetric but sparse and all eigenvalues have positive real part if σ_e is sufficiently large. The matrix D is positive definite and diagonal. The function φ_μ is strongly semi-smooth and Lipschitzian.*

Proof. By Lemma 4.8 all eigenvalues of A have positive real part. The positive definiteness and diagonal property of D follows directly from the biorthogonality and the use of Gauss-Lobatto-Lagrange basis functions (c.f. Lemma 4.6). By Lemma 2.4, φ_η is strongly semi-smooth everywhere and with the definition of strongly semi-smoothness also Lipschitzian. $\qquad\square$

Remark 4.3. *The choice of the NCF significantly influences the properties of the SSN method to solve (4.35) iteratively [40]. In a 1D case with standard Galerkin in space, the use of $\varphi(u, \lambda) = \lambda - \max \{0, \lambda - c(u - g)\}$ with $c > 0$ as studied in [49, 38] realized as a primal-dual active set strategy was not free of cycling for higher order time polynomials. Consequently the SSN method failed to solve the above problem.*

The non-linear problem (4.35) can be solved using the following globalized SSN algorithm a realization of Algorithm 2.1. For the globalization, first note that the non-negative merit function

$$\Psi(u, \lambda) := \frac{1}{2}F(u, \lambda)^2 \tag{4.36}$$

is continuously differentiable and that solving (4.35) is equivalent to finding the minimizers of Ψ. Global convergence is ensured by a line search to minimize the smooth merit function Ψ and, additionally, if the Newton search direction does not satisfy the descent condition (4.38), the algorithm switches to one globally converging gradient descent step. This globalization strategy has been introduced and used in [22, 40].

Algorithm 4.1. *(Semi-smooth Newton algorithm for parabolic obstacle)*

---

1. *Choose initial solution u^0, $\lambda^0 \in \mathbb{R}^n$, $\rho > 0$, $\beta \in (0, 1)$, $\sigma \in (0, \frac{1}{2})$, $p > 2$, tol > 0*

2. *For $k = 0, 1, 2, \ldots$ do*

   a) *If $\left\| \nabla \Psi(u^k, \lambda^k) \right\| < tol$ or $\left\| \Psi(u^k, \lambda^k) \right\| < tol$ then stop.*

   b) *Compute subdifferential $H_k \in \partial F(u^k, \lambda^k)$ and find $d^k = \left(d_u^k, d_\lambda^k \right) \in \mathbb{R}^{2n}$ s.t.*

   $$H_k d^k = -F(u^k, \lambda^k). \tag{4.37}$$

   If (4.37) not solvable, or if the descent condition

   $$\nabla \Psi(u^k, \lambda^k) d^k \leq -\rho \|d^k\|^p \tag{4.38}$$

   is not satisfied, set $d^k := -\nabla \Psi(u^k, \lambda^k)$.

   c) *Compute search length $t_k := \max \left\{ \beta^l : \ l = 0, 1, 2, \ldots \right\}$ s.t.*

   $$\Psi(u^k + t_k d_u^k, \lambda^k + t_k d_\lambda^k) \leq \Psi(u^k, \lambda^k) + \sigma t_k \nabla \Psi(u^k, \lambda^k) d^k.$$

   d) *Update the solution vectors and goto step 2.*

   $$u^{k+1} = u^k + t_k d_u^k, \quad \lambda^{k+1} = \lambda^k + t_k d_\lambda^k$$

For the implementation, the following subdifferential

$$H_k = \begin{pmatrix} A & -D \\ \frac{\partial \varphi_\eta(u^k, \lambda^k)}{\partial u} & \frac{\partial \varphi_\eta(u^k, \lambda^k)}{\partial \lambda} \end{pmatrix}$$

with

$$\frac{\partial \varphi_\eta(u, \lambda)}{\partial u} = \begin{cases} \eta & \text{, if } \lambda = u - g = 0 \\ \eta \left(1 - \frac{u-g}{\sqrt{\lambda^2 + (u-g)^2}} \right) + (1-\eta)\lambda & \text{, if } \lambda > 0 \text{ and } u > g \\ \eta \left(1 - \frac{u-g}{\sqrt{\lambda^2 + (u-g)^2}} \right) & \text{, otherwise} \end{cases}$$

$$\frac{\partial \varphi_\eta(u, \lambda)}{\partial \lambda} = \begin{cases} \eta & \text{, if } \lambda = u - g = 0 \\ \eta \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + (u-g)^2}} \right) + (1-\eta)(u-g) & \text{, if } \lambda > 0 \text{ and } u > g \\ \eta \left(1 - \frac{\lambda}{\sqrt{\lambda^2 + (u-g)^2}} \right) & \text{, otherwise} \end{cases}$$

has been chosen. Other subdifferentials can be chosen but they only differ in the points where φ_η is not classically differentiable. Using this subdifferential, the computation of $\nabla \Psi(u^k, \lambda^k) = H_k^T F(u^k, \lambda^k)$ is straightforward [22]. The most expensive part in the SSN algorithm is the computation of the direction d^k. Note that $\frac{\partial \varphi}{\partial u}$ and $\frac{\partial \varphi}{\partial \lambda}$ are semi-positive definite diagonal matrices and that D is a positive definite diagonal matrix. Hence, the computation of the even sparse Schur complement $S_D = \frac{\partial \varphi}{\partial u} + \frac{\partial \varphi}{\partial \lambda} D^{-1} A$ of H_k is very cheap and halves the dimension of the problem.

$$(4.37) \Leftrightarrow \begin{pmatrix} d_u \\ d_\lambda \end{pmatrix} = \begin{pmatrix} A & -D \\ \frac{\partial \varphi}{\partial u} & \frac{\partial \varphi}{\partial \lambda} \end{pmatrix}^{-1} \begin{pmatrix} -F_u \\ -F_\lambda \end{pmatrix} \Leftrightarrow \begin{array}{l} d_u = -(S_D)^{-1}(F_\lambda + \frac{\partial \varphi}{\partial \lambda} D^{-1} F_u) \\ d_\lambda = D^{-1}(A d_u + F_u) \end{array}$$

A further heuristic possibility to reduce the computational costs is to use a low order extrapolation strategy of the solution of the previous time interval as an initial solution. Assuming the solution u_{hp} behaves benign in time in the sense that its jump at the time interval interface and its higher order time derivatives are small. Then, the solution of the previous time interval is interpolated in time using a low order polynomial interpolation, e.g. $p = 1$, and then extrapolated to the current time interval. The initial solution u^0 is the interpolation of the extrapolated function in $P_{q_n}(I_n)$ as illustrated in Figure 4.1. Since the discrete λ_{hp} does not need to be continuous at the time interval interface, its jumps can be arbitrarily large. Therefore $\lambda^0 = D^{-1}(Au^0 - f)$ is chosen, i.e. the first equation in (4.35) is satisfied. For numerical results see Figure 4.9 in Section 4.5.

Theorem 4.7. *The reduced semi-smooth Newton algorithm*

$$\left(u^{k+1}, \lambda^{k+1} \right)^T = \left(u^k, \lambda^k \right)^T - H_k^{-1} F(u^k, \lambda^k) \tag{4.39}$$

where H_k is a Clarke subdifferential of F at $\left(u^k, \lambda^k \right)^T$ converges locally Q-quadratic.

Proof. The assertion follows from Theorem 2.1 if F is Lipschitzian, strongly semi-smooth, CD-regular and a solution for $F(u, \lambda) = 0$ exists. The existence of a solution follows directly from Lemma 4.9 in conjunction with Theorem 4.3. The first part of
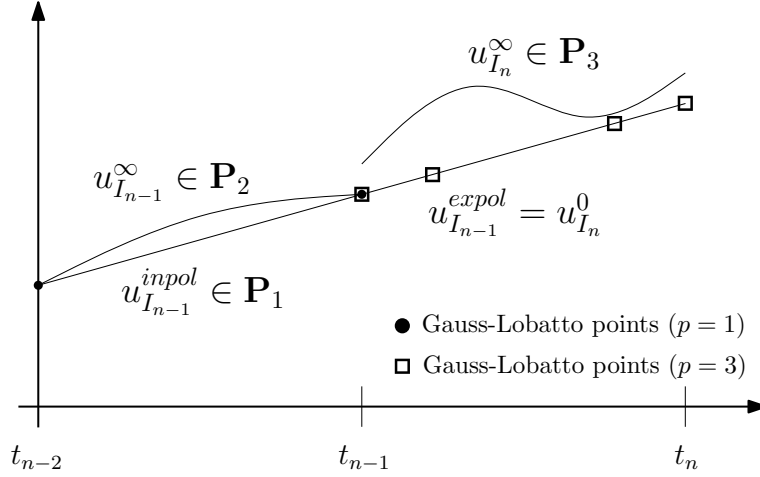
Figure 4.1: An improved initial solution for the SSN method from extrapolation

F is smooth and the second part is strongly semi-smooth by Lemma 4.10. Therewith, simple algebra yields that F is strongly semi-smooth and Lipschitzian. Using the overestimation

$$\partial_C F \subseteq \begin{pmatrix} A & -D \\ D_a & D_b \end{pmatrix} =: E \tag{4.40}$$

of Lemma 2.3 where D_a and D_b are semi-positive definite diagonal matrices such that $D_a + D_b$ is positive definite, F is CD-regular if all realizations of E are invertible. Recall that all eigenvalues of A have positive real part and that D is positive definite by Lemma 4.10. Let $S = D_a + D_b D^{-1} A$ be the Schur complement of E and assume there exists a vector $0 \neq q \in \mathbb{R}^n$ such that $(D_a + D_b D^{-1} A)q = 0$, i.e. S has a zero eigenvalue. Hence, $D_a q = -D_b D^{-1} A q$ which can be written componentwise as $(D_a)_{ii} q_i = -\frac{(D_b)_{ii}}{(D)_{ii}}(Aq)_i$ and simplified to

$$q_i = -\frac{(D_b)_{ii}}{(D_a)_{ii}(D)_{ii}}(Aq)_i \quad i \in I_+ := \{i \in \{1, \dots, n\} : (D_a)_{ii} > 0\} \tag{4.41}$$

$$0 = (Aq)_i \quad\quad i \in I_+^C := \{1, \dots, n\} \setminus I_+ = \{i \in \{1, \dots, n\} : (D_a)_{ii} = 0\}. \tag{4.42}$$

Multiplying the first equation with q_i yields $0 \leq q_i^2 = -\frac{(D_b)_{ii}}{(D_a)_{ii}(D)_{ii}} q_i (Aq)_i \Rightarrow q_i (Aq)_i \leq 0$. The second equation directly yields $q_i(Aq)_i = 0$. Summing over all i yields

$$\sum_{i=1}^{n} q_i (Aq)_i = q^T A q \leq 0$$

which is a contradiction to A having only eigenvalues with positive real part. Hence, S, and therewith E, are invertible which completes the proof. □

Remark 4.4. *Using the same arguments as in the proof of existence in Theorem 4.3, it can be shown that the standard Uzawa algorithm with the componentwise projection*

$$\lambda_{i,j} = \max\{0, \lambda_{i,j} + r(u_{i,j} - g_{i,j})\} \quad for \ 1 \leq i \leq \dim V_{hp}, \ 0 \leq j \leq q_n$$

converges if the projection constant $r > 0$ is sufficiently small.

4.3 A Posteriori Error Estimate and hp-Adaptivity

In this section a hierarchical error estimator for the variational inequality formulation and by Theorem 4.6 also for the mixed method is presented. It is based on the ideas presented in [26] and in Section 3.3.

Theorem 4.8. *Let $u \in W_2^1(I; H^1(\Omega), L^2(\Omega)) \cap K$ solve (4.6) and $u_p \in \sum_{n=1}^{N} K_{q_n}(I_n) =: K_{hp,kq}$, $u_{p+1} \in K_{hp+1,k,q+1}$ solve (4.32). If the saturation assumption holds, i.e.*

$$\|u - u_{p+1}\|_{L^2(I;1,hp+1)} \leq q_S \|u - u_p\|_{L^2(I;1,hp)} \tag{4.43}$$

with q_S uniformly in $(0, 1)$, then

$$\frac{1}{1 + q_S}\eta \leq \|u - u_p\|_{L^2(I;1,hp)} \leq \frac{1}{1 - q_S}\eta \tag{4.44}$$

with the global error indicator

$$\eta := \|u_{p+1} - u_p\|_{L^2(I;1,hp)}. \tag{4.45}$$

Proof. As in the proof of Theorem 3.9, the triangle inequality and an estimation of the mesh dependent norm yields the assertion. □

Since the $L^2(I; 1, hp)$-norm is local, the error estimator η can be written as a sum of local error indicators.

Lemma 4.11. *The error indicator η satisfies*

$$\eta^2 = \sum_{E \in \mathcal{T}_{hp}^s} \sum_{I_n \in \mathcal{T}_{kq}^t} \eta^2(E, I_n) \tag{4.46}$$

with the local error indicators

$$\begin{aligned}
\eta^2(E, I_n) = &\int_{I_n} \int_E \left(\nabla(u_{p+1} - u_p)\right)^2 + (u_{p+1} - u_p)^2 \ dx \ dt \\
&+ \sum_{e \in \partial E \cap \Gamma_h} \int_{I_n} \int_e \frac{\sigma_e p_e^2}{2|e|^\beta} [u_{p+1} - u_p]^2 \ ds \ dt \\
&+ \sum_{e \in \partial E \cap \Gamma_D} \int_{I_n} \int_e \frac{\sigma_e p_e^2}{|e|^\beta} (u_{p+1} - u_p)^2 \ ds \ dt.
\end{aligned} \tag{4.47}$$

Proof. Explicitly writing down the norm and separating the jump contributions equally over the adjacent elements. □

Remark 4.5. *The finite element space enrichment does not need to be a p-refinement, but can also be a h-refinement or a combination of both, c.f. Section 3.3.*

Remark 4.6.   *1. The additional computation of u_{p+1} is very expensive but the CPU time can be reduced by using interpolation matrices as in Remark 3.3 for the elliptic case since the FE spaces without the constraints are nested and the constraints are enforced by the iterative solver itself.*

2. It should be analyzed if the computation of u_p can be replaced by a simple projection of u_{p+1} into $K_{hp,kq}$ as in [26] for linear BEM.

3. The indicator η estimates the total space-time error and, thus, allows no refinement as the algorithm goes through time. This is not surprising as the TDG method is a projection method and only the specific choice of time basis function with support in exactly one time strip allows the total problem to be solved as if it were a time stepping method like finite difference methods.

Remark 4.7. *In case of continuous Galerkin with a finite difference approximation in time, a residual based error indicated has been derived in [63]. The proof relies on the Galerkin functional and the $H^1(\Omega)$-conformity of u and seems not to be extendable to IPDG in a straightforward manner.*

It remains to decide which space-time element $E \times I_n$ and how it should be refined. For the numerical experiments in Section 4.5 the following solve-mark-refine algorithm is used.

Algorithm 4.2. *(Solve-mark-refine algorithm for parabolic obstacle)*

1. *Choose initial discretizations $\mathcal{T}_{hp}^s$, $\mathcal{T}_{kq}^t$ and $\theta \in (0,1)$, $tol \geq 0$*

2. *For $k = 0, 1, 2, \dots$ do*

   a) *Solve the variational inequality formulation (4.6) on the current mesh using the mixed method (4.18) with SSN.*

   b) *Compute local error indicators $\eta^2(E, I_n)$.*

   c) *If $\left(\sum_{E,I_n} \eta^2(E, I_n) \right)^{\frac{1}{2}} \leq tol$ then stop.*

   d) *Mark all elements in $M := \left\{ E \times I_n : \eta^2(E, I_n) \geq \theta \cdot \max_{\tilde{E}, \tilde{I}_n} \eta^2(\tilde{E}, \tilde{I}_n) \right\}$ for refinement.*

      i. *For all $E \times I_n \in M$ estimate the local analyticity of $\bar{u}_h(x) = \frac{1}{|I_n|} \int_{I_n} u_h(x,t) dt$.*

      ii. *For all $E \times I_n \in M$ let $S_{I_n} := \sum_{E \times I_n \in M} E$ with I_n fixed and estimate the local analyticity of $\tilde{u}_h(t) = \frac{1}{|S_{I_n}|} \int_{S_{I_n}} u_h(x,t) \, dx$.*

   e) *Refine $\mathcal{T}_{hp}^s$, $\mathcal{T}_{kq}^t$ using M and the analyticity estimates as in [36].*

The parameter θ steers the amount of marked elements. For θ close to zero, the sequence of refined meshes will be close to uniformed meshes with similar experimental

convergence rates. For θ close to one, only very few elements are refined resulting in a highly localized refinement, yet many refinement cycles are required to reduce the total error. The tensor structure of the space-time discretization implies, that if an element $E \times I_n$ is refined in time, then the whole time strip is refined, i.e. all elements in that particular time strip are refined in time. Many of these elements are refined unnecessarily with respect to the marking set M. Therefore, the decision weather a h- or p-refinement should be carried out should only depend on the marked elements in that time strip. Furthermore, the space refinement should be independent of the time regularity. Hence, step 2.d) in Algorithm 4.2. The least square approximation within the analyticity estimate explained in Section 3.3 requires at least a polynomial degree of one to be solvable. Therefore, if the time polynomial is zero, a p-refinement will be carried out such that in the case that this element is marked for refinement again, the analyticity estimate can now be carried out.

4.4 First Results on an a Priori Error Estimate

In this section an a priori error estimate without convergence rate is derived for the variational inequality formulation (4.32) and by Theorem 4.6 also for the primal variable of the mixed method (4.18). For simplicity, it is assumed throughout this section that $\vec{\gamma} = \vec{0}$ and $\beta = 1$. Therewith, the same extension operators and (bi)-linear forms as in Section 3.4 can be used. Let $K_{hp,kq} := \sum_{n=1}^{N} K_{q_n}(I_n)$ be the total discrete convex cone of admissible functions over the entire time interval and let

$$B_{DG}(u,v) := \sum_{n=1}^{N} \int_{I_n} \langle \dot{u}, v \rangle + \widetilde{a}_\epsilon(u,v) \, dt, \qquad F_{DG}(v) := \sum_{n=1}^{N} \int_{I_n} \widetilde{F}_\epsilon(v) \, dt.$$

Then, the discrete problem (4.32) is equivalent to finding $U \in K_{hp,kq}$ such that

$$B_{DG}(U, V-U) + \sum_{n=2}^{N} \left\langle [U^{n-1}], (V-U)_+^{n-1} \right\rangle + \left\langle U_+^0, (V-U)_+^0 \right\rangle \geq F_{DG}(V-U) + \tag{4.48}$$

$$+ \left\langle u_0, (V-U)_+^0 \right\rangle \qquad \forall \, V \in K_{hp,kq}.$$

If $u \in W_2^1(I; H^1(\Omega), L^2(\Omega)) \cap K$ solves (4.6), than it also satisfies

$$B_{DG}(u, v-u) \geq F_{DG}(v-u) \quad \forall \, v \in K. \tag{4.49}$$

Lemma 4.12. *There exists a constant $\alpha > 0$ such that for all $v \in W_2^1(I; H^1(\Omega), L^2(\Omega)) \cup \sum_{n=1}^{N} P_{q_n}(I_n)$ there holds*

$$B_{DG}(v,v) + \sum_{n=2}^{N} \left\langle [v^{n-1}], v_+^{n-1} \right\rangle + \left\langle v_+^0, v_+^0 \right\rangle \geq \alpha \|v\|_{L^2(I;1,hp)}^2 + \frac{1}{2} \|v_+^0\|_{L^2(\Omega)}^2$$

$$+ \frac{1}{2} \sum_{n=1}^{N-1} \|[v^n]\|_{L^2(\Omega)}^2 + \frac{1}{2} \|v_-^N\|_{L^2(\Omega)}^2$$

Proof. Follows directly from [69, Lemma 2.7], Lemma 4.8 and Lemma 3.15. $\qquad\square$

Theorem 4.9. *Let* $u \in W_2^1(I; H^1(\Omega), L^2(\Omega)) \cap K$ *solve* (4.6) *and* $U \in K_{hp,kq}$ *solve* (4.32), *then there holds for arbitrary* $v \in K$ *and* $V \in K_{hp,kq} \cap C(I, L^2(\Omega))$

$$\alpha \|u - U\|_{L^2(I;1,hp)}^2 + \|(u - U)_+^0\|_{L^2(\Omega)}^2 + \sum_{n=1}^{N-1} \|[(u - U)^n]\|_{L^2(\Omega)}^2 + \frac{1}{2} \|(u - U)_-^N\|_{L^2(\Omega)}^2$$

$$\leq 2 \|(u - V)_-^N\|_{L^2(\Omega)}^2 + \frac{1}{\alpha} \left(\|\dot{u} - \dot{V}\|_{L^2(I;L^2(\Omega))} + C_C \|u - V\|_{L^2(I;1,hp)} \right)$$

$$+ 2 \left(C_{F_{DG}} + \|\dot{u}\|_{L^2(I;L^2(\Omega))} + C_C \|u\|_{L^2(I;1,hp)} \right) \left(\|u - V\|_{L^2(I;1,hp)} + \|U - v\|_{L^2(I;1,hp)} \right)$$

if $\vec{\gamma} = \vec{0}$, $\beta = 1$ *and* $\dot{u}, f \in L^2(I; L^2(\Omega))$ *with* C_C *defined in Lemma 3.14,* α *defined in Lemma 4.12 and* $C_{F_{DG}}$ *the continuity constant of* $F_{DG}(\cdot)$.

Proof. Adding (4.49) and (4.48) yields

$$B_{DG}(u, u) + B_{DG}(U, U) + \sum_{n=2}^{N} \left\langle [U^{n-1}], U_+^{n-1} \right\rangle + \left\langle U_+^0, U_+^0 \right\rangle \leq F_{DG}(u - V) + F_{DG}(U - v)$$

$$+ B_{DG}(u, v) + B_{DG}(U, V) + \left\langle u_0, (U - V)_+^0 \right\rangle + \sum_{n=2}^{N} \left\langle [U^{n-1}], V_+^{n-1} \right\rangle + \left\langle U_+^0, V_+^0 \right\rangle.$$

Next, subtract $B_{DG}(u, U) + B_{DG}(U, u) + \sum_{n=2}^{N} \left\langle [U^{n-1}], u_+^{n-1} \right\rangle + \left\langle U_+^0, u_+^0 \right\rangle$ to obtain

$$B_{DG}(u - U, u - U) + \sum_{n=2}^{N} \left\langle [-U^{n-1}], (u - U)_+^{n-1} \right\rangle + \left\langle -U_+^0, (u - U)_+^0 \right\rangle \leq F_{DG}(u - V)$$

$$+ F_{DG}(U - v) + \left\langle u_0, (U - V)_+^0 \right\rangle + \sum_{n=2}^{N} \left\langle [U^{n-1}], (V - u)_+^{n-1} \right\rangle$$

$$+ \left\langle U_+^0, (V - u)_+^0 \right\rangle - B_{DG}(u, U - v) - B_{DG}(u, u - V) + B_{DG}(u - U, u - V).$$

Since $[u^n] = 0$ and $u_0 = u_+^0$ in $L^2(\Omega)$ there holds

$$B_{DG}(u - U, u - U) + \sum_{n=2}^{N} \left\langle [(u - U)^{n-1}], (u - U)_+^{n-1} \right\rangle + \left\langle (u - U)_+^0, (u - U)_+^0 \right\rangle$$

$$\leq F_{DG}(U - v) - B_{DG}(u, U - v) + F_{DG}(u - V) - B_{DG}(u, u - V)$$

$$+ B_{DG}(u - U, u - V) + \sum_{n=2}^{N} \left\langle [(u - U)^{n-1}], (u - V)_+^{n-1} \right\rangle$$

$$+ \left\langle (u - U)_+^0, (u - V)_+^0 \right\rangle.$$

For $\dot{u}, f \in L^2(I; L^2(\Omega))$, [69, Lemma 2.7], Lemma 4.12 and the continuity result of

Lemma 3.14 yield

$$\alpha \|u - U\|^2_{L^2(I;1,hp)} + \frac{1}{2} \left\|(u-U)^0_+\right\|^2_{L^2(\Omega)} + \frac{1}{2} \sum_{n=1}^{N-1} \|[(u-U)^n]\|^2_{L^2(\Omega)} + \frac{1}{2} \left\|(u-U)^N_-\right\|^2_{L^2(\Omega)}$$

$$\leq \left(C_{F_{DG}} + \|\dot{u}\|_{L^2(I;L^2(\Omega))} + C_C \|u\|_{L^2(I;1,hp)}\right) \left(\|u-V\|_{L^2(I;1,hp)} + \|U-v\|_{L^2(I;1,hp)}\right)$$

$$+ \sum_{n=1}^{N} \int_{I_n} \left\langle -u+U, \dot{u}-\dot{V}\right\rangle + \widetilde{a}_\epsilon(u-U, u-V) \, dt$$

$$- \sum_{n=1}^{N-1} \left\langle (u-U)^n_-, [(u-V)^n]\right\rangle + \left\langle (u-U)^N_-, (u-V)^N_-\right\rangle.$$

Choosing $V \in C(I; L^2(\Omega))$ and using the Cauchy-Schwarz inequality yields

$$\alpha \|u - U\|^2_{L^2(I;1,hp)} + \frac{1}{2} \left\|(u-U)^0_+\right\|^2_{L^2(\Omega)} + \frac{1}{2} \sum_{n=1}^{N-1} \|[(u-U)^n]\|^2_{L^2(\Omega)} + \frac{1}{2} \left\|(u-U)^N_-\right\|^2_{L^2(\Omega)}$$

$$\leq \left(C_{F_{DG}} + \|\dot{u}\|_{L^2(I;L^2(\Omega))} + C_C \|u\|_{L^2(I;1,hp)}\right) \left(\|u-V\|_{L^2(I;1,hp)} + \|U-v\|_{L^2(I;1,hp)}\right)$$

$$+ \|u - U\|_{L^2(I;1,hp)} \left(\left\|\dot{u}-\dot{V}\right\|_{L^2(I;L^2(\Omega))} + C_C \|u-V\|_{L^2(I;1,hp)}\right)$$

$$+ \left\|(u-U)^N_-\right\|_{L^2(\Omega)} \left\|(u-V)^N_-\right\|_{L^2(\Omega)}.$$

Applying Young's inequality twice yields

$$\alpha \|u - U\|^2_{L^2(I;1,hp)} + \left\|(u-U)^0_+\right\|^2_{L^2(\Omega)} + \sum_{n=1}^{N-1} \|[(u-U)^n]\|^2_{L^2(\Omega)} + \frac{1}{2} \left\|(u-U)^N_-\right\|^2_{L^2(\Omega)}$$

$$\leq 2 \left\|(u-V)^N_-\right\|^2_{L^2(\Omega)} + \frac{1}{\alpha} \left(\left\|\dot{u}-\dot{V}\right\|_{L^2(I;L^2(\Omega))} + C_C \|u-V\|_{L^2(I;1,hp)}\right)$$

$$+ 2 \left(C_{F_{DG}} + \|\dot{u}\|_{L^2(I;L^2(\Omega))} + C_C \|u\|_{L^2(I;1,hp)}\right) \left(\|u-V\|_{L^2(I;1,hp)} + \|U-v\|_{L^2(I;1,hp)}\right)$$

which completes the proof. $\qquad\square$

4.5 Numerical Experiments

As a numerical experiment the example by Moon et al. [63, Section 5.6] of a 2d oscillating moving circle is reconsidered but with additional convection and mass term. The space domain is $\Omega = [-1, 1]^2$, the time interval is $I = (0, 0.8)$ and the (non)-contact sets are

$$\mathcal{N} = \{|x - c(t)|_2 > r_0(t)\} \text{ and } \mathcal{C} = \{|x - c(t)|_2 \leq r_0(t)\}$$

with

$$r_0(t) = \frac{1}{3} + 0.3\sin(16\pi t) \text{ and } c(t) = \frac{1}{3} \left(\cos(4\pi t), \sin(4\pi t)\right)^T.$$

(a) Solution with free boundary for $t = 0.15$

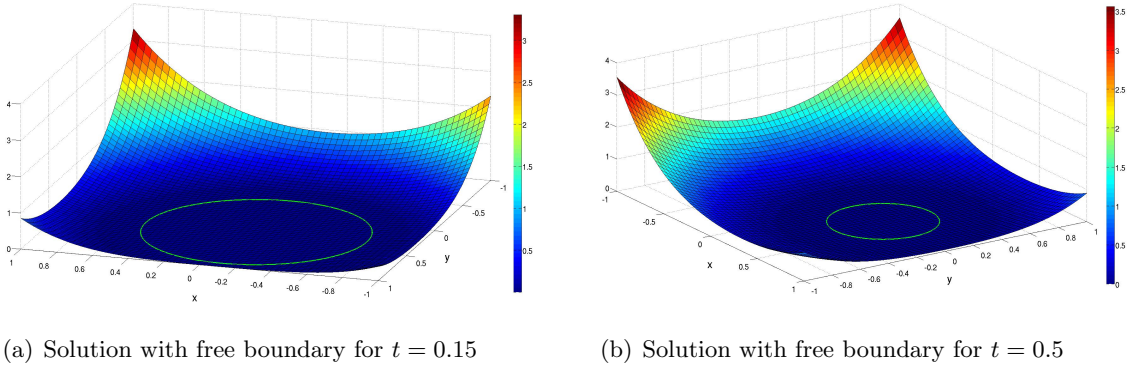(b) Solution with free boundary for $t = 0.5$

Figure 4.2: Solution with free boundary for parabolic obstacle problem

For the obstacle $\chi \equiv 0$, convection coefficient $\vec{\gamma} = (1, 1)^T$ and volume force

$$f(x, t) = \begin{cases} -2 \left(|x - c(t)|_2^2 - r_0^2(t) \right) \left[(x - c(t)) \cdot \dot{c}(t) + r_0(t)\dot{r}_0(t) \right] + 4 \left(r_0^2(t) - 2|x - c(t)|_2^2 \right) \\ \quad +2 \left(|x - c(t)|_2^2 - r_0^2(t) \right) \vec{\gamma} \cdot (x - c(t)) + \frac{1}{2} \left(|x - c(t)|_2^2 - r_0^2(t) \right)^2, \quad \text{in } \mathcal{N} \\ -4r_0(t) \left(1 - |x - c(t)|_2^2 - r_0^2(t) \right), \quad \text{in } \mathcal{C} \end{cases}$$

the exact solution is

$$u(x, t) = \begin{cases} \frac{1}{2} \left(|x - c(t)|_2^2 - r_0^2(t) \right)^2, & \text{in } \mathcal{N} \\ 0, & \text{in } \mathcal{C}. \end{cases}$$

For the time points $t = 0.15$ and $t = 0.5$, the solution is plotted in Figure 4.2 where the green circle indicates the free boundary. In fact, the contact set is a counter clockwise moving spiral in the three dimensional space-time domain $\Omega \times I$ with oscillating radius $r_0(t)$ as visualized in Figure 4.3.

For sequences of differently refined meshes, the $L^2(I; 1, hp)$-error and the error indicator η are plotted in a semi-logarithmic scale versus the forth root of the degrees of freedom in Figure 4.4. The lowest order uniform h-version shows its algebraic convergence rate of 0.175 wrt. the dof and the estimated error is almost the same. Using an h-adaptive refinement strategy the convergence rate is improved with a slight reduction in the effectivity index of the error estimator compared to the uniform h-version. The two hp-adaptive strategies only differ in their initial mesh, where the second, finer mesh is a uniform refinement of the first, coarser mesh. Both show exponential convergence at the beginning which turns into an algebraic one. The reason for this is the space-time discretization based on a tensor product combined with the contact set $\mathcal{C}$ as displayed in Figure 4.3a. For this set, in each time strip there exists a fixed space point x which goes from non-contact to contact or vise versa. This implies that the solution at x is singular in time and therefore the space-time cube containing x at that time should be refined in time. By the tensor structure of the space-time discretizations all space-time-cubes in that particular time strip are refined in time. Since for every time strip such a point x exists, every time strip is refined leading to a uniform-like refinement in time
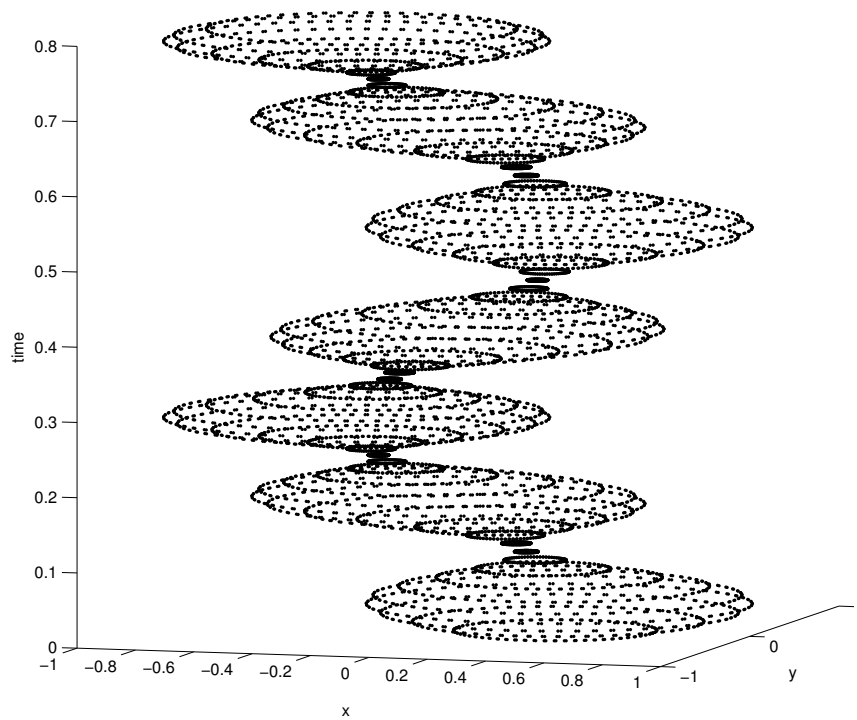
(a) Free boundary for parabolic obstacle problem
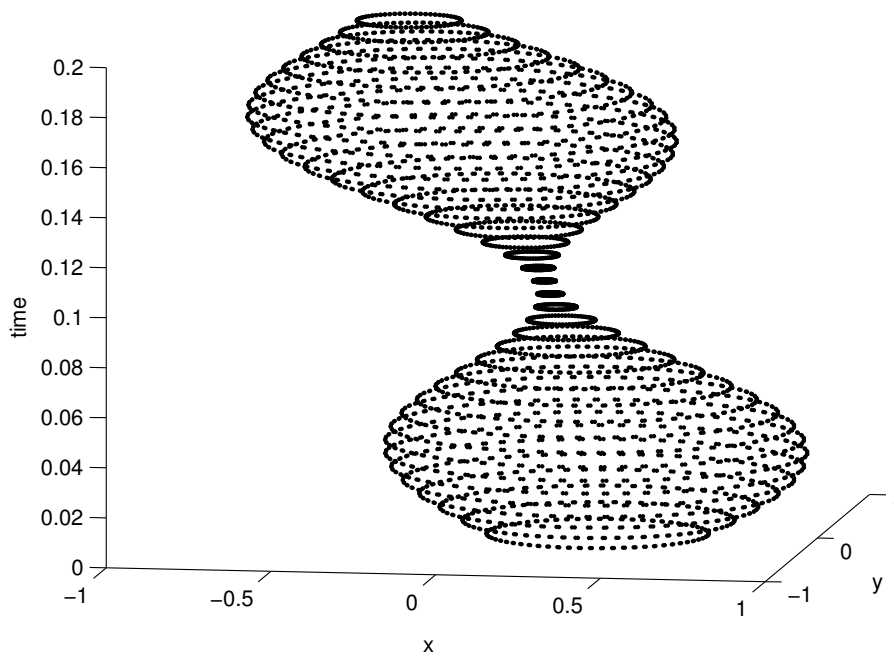
(b) Free boundary for parabolic obstacle problem, zoomed in on $t \in (0, 0.2)$

Figure 4.3: Free boundary for parabolic obstacle problem
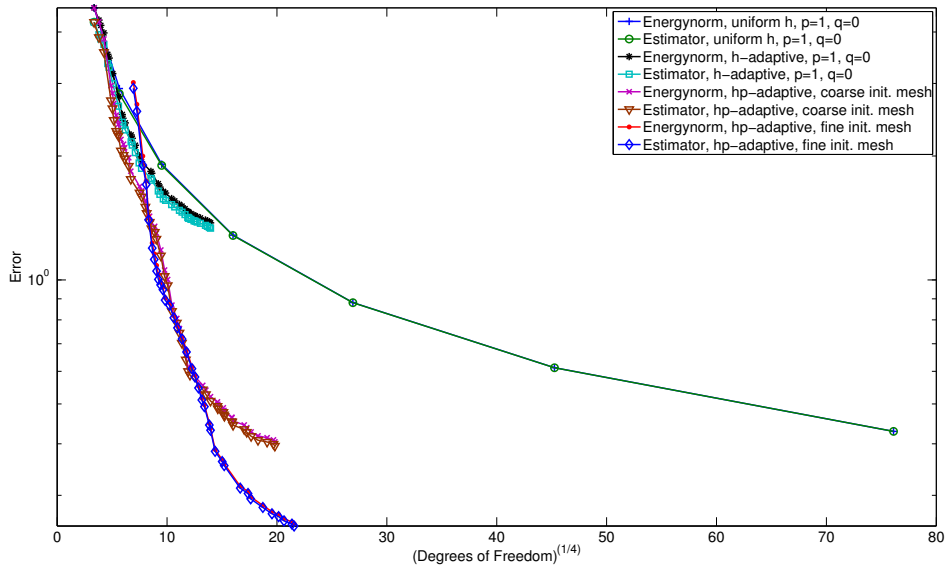
Figure 4.4: Error in energy norm and error indicator for parabolic obstacle problem

and therefore only to algebraic convergence. For the space discretization error similar observations as in the elliptic case of Section 3.5 can be made. Hence, a discretization based on tensor products seems to be inappropriate if the contact set itself is not based on a tensor product and exponential convergence should be achieved.

The above observation also explains the behavior of the experimental effectivity index (EEI) displayed in Figure 4.5a. For the uniform h and h-adaptive method the EEI quickly tends towards a constant very close to one as has also been seen in Section 3.5. At the point where the hp-adaptive methods turn into an algebraic converging approach their EEI also tends towards a constant between 0.97 and 0.99. All in all, the EEI is always very good and lies between 0.91 and 1.

Similarly, Figure 4.5b shows that the experimental saturation constants tends towards some constant smaller than 0.2. Varying the penalty factor σ_e or the method's parameter ϵ, similar observations as in the elliptic case can be made and are omitted here for brevity. Changing the regularity estimate parameter δ, only influences at which point the hp-method turns to algebraic convergence.

Remark 4.8. *Most of the theoretical results of this chapter can be carried over to the case of the H^1-conforming continuous Galerkin approach.*

In the following, a continuous Galerkin approach for pricing the $1d$ American put option
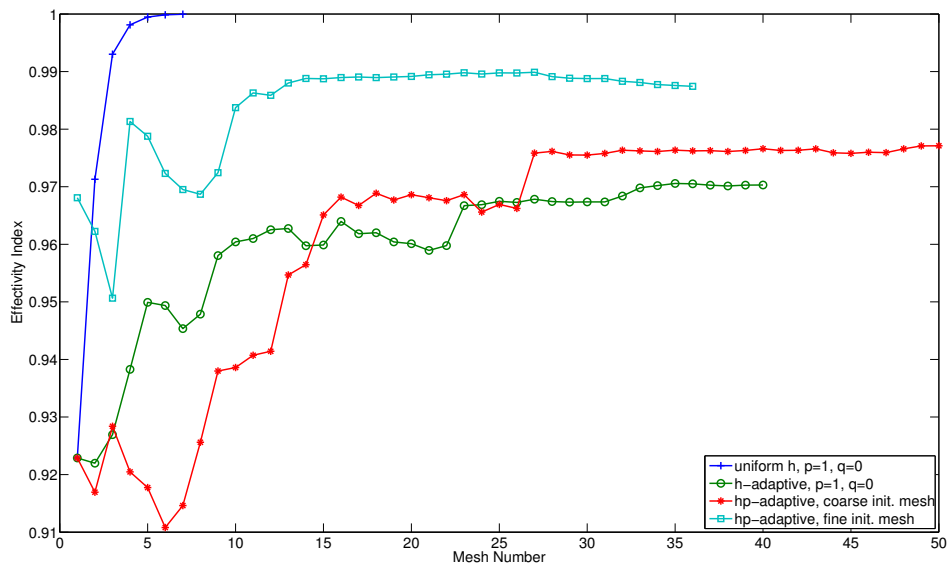
(a) Experimental efficiency index
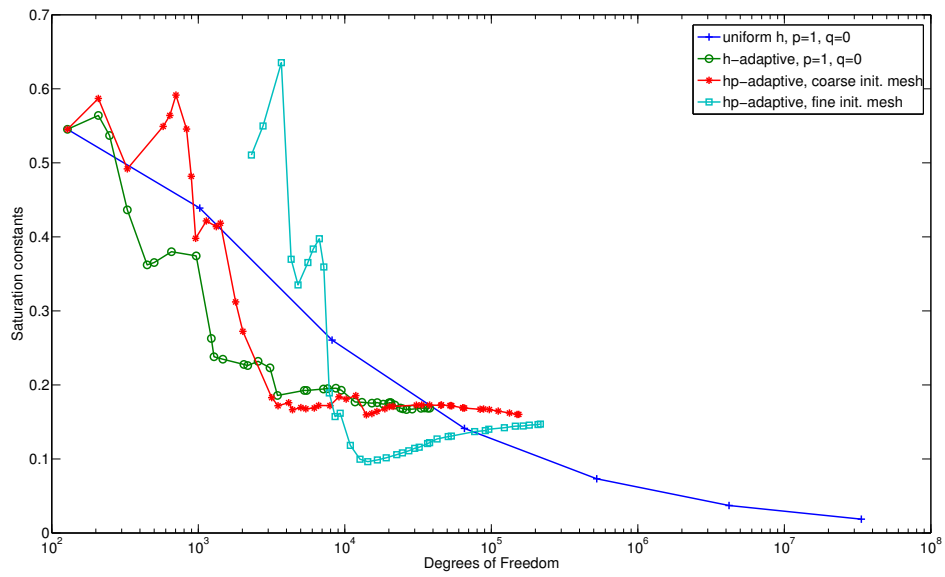
(b) Experimental saturation constant

Figure 4.5: Experimental efficiency index and saturation constant for parabolic obstacle problem

is considered, i.e. a "fair" price $P(S, \tau)$ is sought such that

$$\frac{\partial P}{\partial \tau} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 P}{\partial S^2} + rS \frac{\partial P}{\partial S} - rP \leq 0 \qquad \text{in } [0, \infty) \times [0, T] \qquad (4.50a)$$

$$P(0, \tau) = E \qquad \text{in } [0, T] \qquad (4.50b)$$

$$P(S, \tau) = 0 \text{ as } S \to \infty \qquad \text{in } [0, T] \qquad (4.50c)$$

$$P(S, \tau) \geq \max\{0, E - S\} \qquad \text{in } [0, \infty) \times [0, T] \qquad (4.50d)$$

$$P(S, T) = \max\{0, E - S\} \qquad \text{in } [0, \infty) \qquad (4.50e)$$

where S is the price of the underlying, $\sigma = 0.3604$ is the volatility, $r = 0.01845$ the risk free interest rate, $T = 0.5$ the expiry date, and $E = 125$ the strike price. Typically a change of variables is performed to better deal with the unbounded domain and to obtain an initial boundary value problem. Let $t = T - \tau$ be the time to maturity, $x = \ln S$, $u(x, t) = P(e^x, T - t)$ and truncate the arising domain $(-\infty, +\infty)$ by $\pm R = \pm 20$. Hence, (4.50) becomes

$$\frac{\partial u}{\partial t} - \frac{\sigma^2}{2} \frac{\partial^2 u}{\partial x^2} + (\frac{\sigma^2}{2} - r) \frac{\partial u}{\partial x} + ru \geq 0 \quad \text{in } \Omega \times I := (-R, R) \times (0, T] \qquad (4.51a)$$

$$u(\pm R, t) = \chi(\pm R) \qquad \text{in } (0, T] \qquad (4.51b)$$

$$u(x, t) \geq \chi(x) \qquad \text{in } (-R, R) \times (0, T] \qquad (4.51c)$$

$$u(x, 0) = \chi(x) \qquad \text{in } (-R, R) \qquad (4.51d)$$

with $\chi(x) = \max\{0, E - e^x\}$ the profit from exercising the option and with the truncation error decreasing exponentially fast with R [63]. Since no analytic solution is available, convergence is measured in

$$e(u_{hp}) := \left(\|u\|^2_{L^2(I; H^1(\Omega))} - \|u_{hp}\|^2_{L^2(I; H^1(\Omega))} \right)^{1/2}. \qquad (4.52)$$

where the value $\|u\|_{L^2(I; H^1(\Omega))} \approx 430.7537067$ has been obtained from extrapolation of a family of lowest order h-version approximations of the mixed method. The expression (4.52) is not a norm since the Galerkin orthogonality does not hold. However, if $u_{hp} \to u$ then certainly $e(u_{hp}) \to 0$. In Figure 4.6, the h-version shows for more than 30,000 degrees of freedom a promising stable algebraic convergence behavior of 0.246 wrt. dof. Contrary, the two hp-versions show the aspired exponential error reduction with almost identical course. The courses must not be identical since the obstacle condition is enforced in different, not equivalent ways. For the mixed method it is weakly enforced and for the variational inequality by a pointwise interpolation condition. If it was weakly enforced than the approximation error would be identical up to rounding errors and would depend on the different stopping criterions of the iterative solvers. Figure 4.7 plots the invested CPU time resources versus the pseudo-error to underline the efficiency of the two hp-methods and of the hp-mixed method in particular. Due to the exponential error reduction of the hp-versions, there exists a break-even point with the uniform h-version of the mixed method. The slow convergence of the projection solver for the variational inequality can be seen clearly and results in a less steep course of its corresponding curve compared to the mixed hp-method. In particular, the calculation
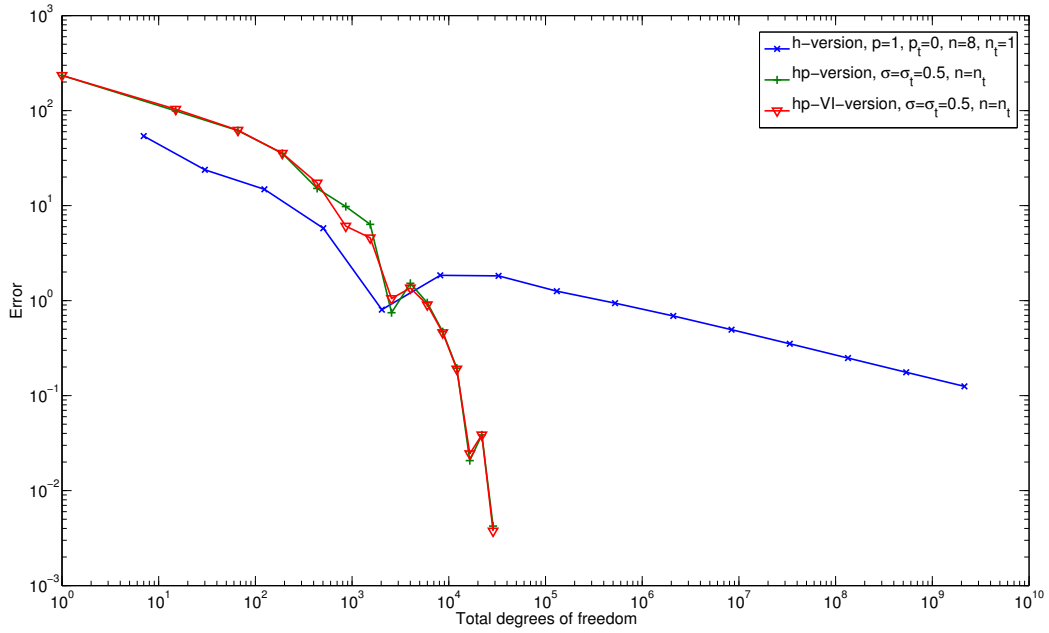
Figure 4.6: Error $e(u_{hp}) := \left(\|u\|^2_{L^2(I;H^1)} - \|u_{hp}\|^2_{L^2(I;H^1)} \right)^{1/2}$ for h- and hp-methods, grading towards $x = \ln E$ and $t = 0$

of the finest resolved solution of the mixed method takes only six seconds whereas that of the variational inequality methods takes about three hours.

Theorem 4.7 provides an optimal theoretical convergence behavior of the SSN method and, together with the additionally proposed heuristic strategy, suggests efficiency for practical applications. For the first of three experiments the tolerance is set to 10^{-14} to investigate the Q-quadratic convergence of the reduced SSN method (4.39) of Theorem 4.7. Let the matrices A and D and the obstacle vector $g_{i,j}$ correspond to the twelfth time step of a hp-discretization using twelve levels in both space and time. The right hand side is modified such that $u = (1, 0, \dots, 1, 0)^T + g \in \mathbb{R}^{1080}$ and $\lambda = (0, 1, \dots, 0, 1)^T \in \mathbb{R}^{1080}$ is the exact solution. The Newton iteration starts with the initial solution $u^0 = \lambda^0 = \vec{1}$. The error $e^k := \left(\|u - u^k\|^2_2 + \|\lambda - \lambda^k\|^2_2 \right)^{\frac{1}{2}}$ measured in the Euclidean norm and the error ratio $r^k := \frac{e^{k+1}}{(e^k)^2}$ are displayed in Table 4.1. The ratio r^k is clearly bounded with only small oscillations for the later iterations which indicates Q-quadratic convergence. By Remark 4.4 the Uzawa algorithm converges as well but with more than 100 iterations making it much slower than the SSN method. The number of iterations can be reduced if the mesh and problem dependent projection parameter is adapted. However, this is a hideous and above all impracticable task.

For the next two numerical experiments, the tolerance is reduced to 10^{-10} and the mesh is graded towards the strike price $\ln E$ and towards the endpoints $\pm R$. Figure 4.8 shows the influence of the penalty parameter in the penalized Fischer-Burmeister NCF on the
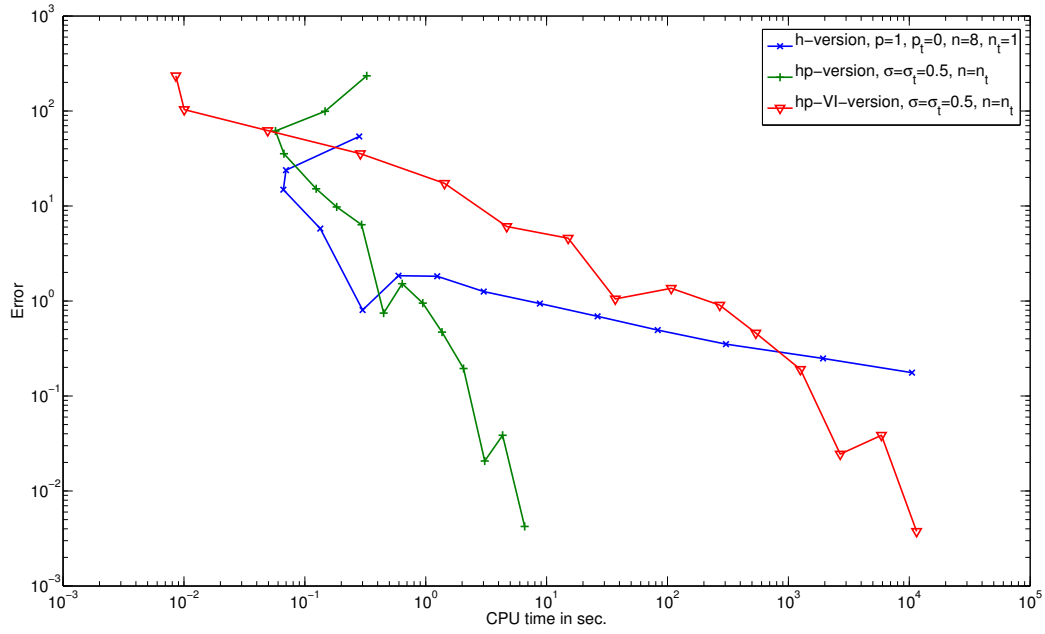
Figure 4.7: Pseudo-error vs. CPU time in seconds for the *h*-, *hp*-mixed version and the *hp*-variational inequality version

Iter	e^k	r^k
0	$2.888 \cdot 10^3$	$6.237 \cdot 10^{-6}$
1	$5.204 \cdot 10^1$	$6.446 \cdot 10^{-3}$
2	$1.746 \cdot 10^1$	$3.028 \cdot 10^{-2}$
3	$9.235 \cdot 10^0$	$1.634 \cdot 10^{-1}$
4	$1.393 \cdot 10^1$	$3.888 \cdot 10^{-2}$
5	$7.552 \cdot 10^0$	$9.938 \cdot 10^{-2}$
6	$5.669 \cdot 10^0$	$1.488 \cdot 10^{-1}$
7	$4.782 \cdot 10^0$	$3.537 \cdot 10^{-2}$
8	$8.090 \cdot 10^{-1}$	$2.070 \cdot 10^{-3}$
9	$1.355 \cdot 10^{-3}$	$1.119 \cdot 10^{-3}$
10	$2.056 \cdot 10^{-9}$	

Table 4.1: Locally Q-quadratic convergence of the reduced SSN method
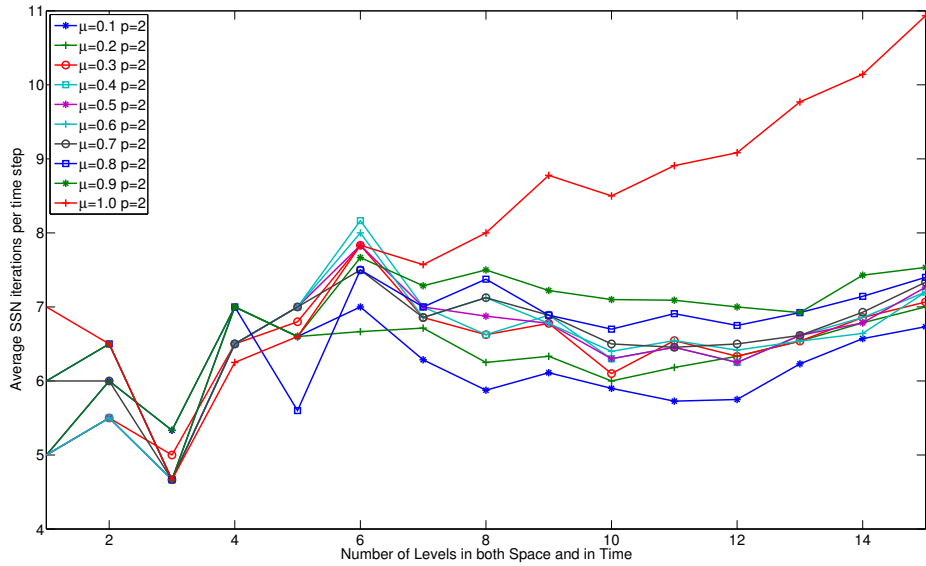
Figure 4.8: Average numbers of SSN iterations per time step vs. number of levels in both space and in time for different penalization parameters μ exploiting a quadratic polynomial extrapolation for the initial solution in the mixed hp-version

average number of SSN iterations per time step for the mixed hp-version. Choosing $\mu = 1$, i.e. no penalization, leads to a steady increase of the SSN iterations with the number of levels. However, even a small penalization, e.g. $\mu = 0.9$, can eliminate this effect completely such that there is almost no interdependence between the number of levels and carried out iterations. Further changes in μ lead to only marginal changes in the numbers of iterations and, thus, unburdens the user from fine tuning this parameter.

Figure 4.9 shows the effect of the choice of the initial solution on the number of semismooth Newton iterations in the case of the hp-mixed version. The same number of levels is used in space and time and is plotted versus the average numbers of iterations per time step. Choosing $u^0 = \lambda^0 = 0$ is the standard naive initial solution and works as a reference case to investigate the effect of different initial solutions which are obtained as described in Section 4.2.3. Good results are achieved if the extrapolation polynomial degree is chosen to be one or two which leads to an average reduction of two SSN iterations per time step compared to the naive initial solution. Moreover, this method requires only seven iterations on average per time step which is very few. Further, the changes in SSN iterations with the number of levels is benign and not systematically. Additional computational experiments have shown that using a large time polynomial degree for the extrapolation, e.g. 8, leads to a poor initial solution due to the large higher order time derivatives in the new, current time interval of the extrapolent and, therewith, to a dramatic increase in the number of iterations for higher levels.
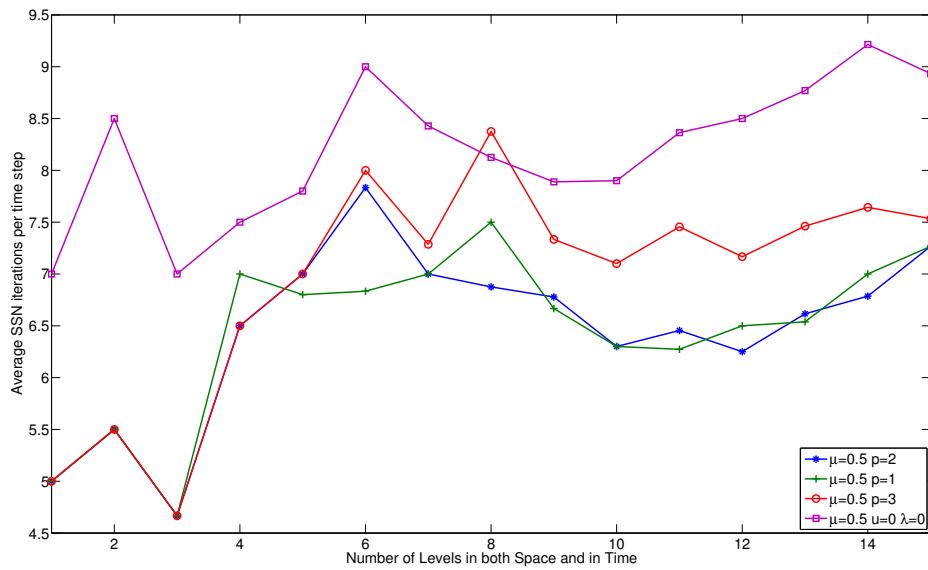
Figure 4.9: Average numbers of SSN iterations per time step vs. number of levels in both space and in time for different polynomial degrees for extrapolation for the initial solution and the naive zero initial solution for penalization parameter $\mu = 0.5$ in the mixed hp-version

5 Elliptic Stochastic Contact Problems

In this chapter an exterior, elliptic, stochastic contact model problem is analyzed. The problem may then serve as a benchmark problem for more realistic but more complex formulations. After transforming the stochastic problem into a deterministic but high dimensional formulation, the weak mixed formulation is similar to the one from the parabolic obstacle problem in Chapter 4. The discretization is optimized for the construction of a fast iterative solver for the resulting non-linear problem. Furthermore, an *hp*-adaptive mesh refinement may lead to improved convergence rates compared to uniform *h*- and *p*-versions.

5.1 Boundary Weak Formulations for Exterior Elliptic Stochastic Contact Problems

Let $R \subset \mathbb{R}^2$ be a bounded polygonal domain with boundary Γ and $cap(\Gamma) < 1$, decomposable into the disjoint parts $\overline{\Gamma} = \overline{\Gamma}_D \cup \overline{\Gamma}_\Sigma$ with $\overline{\Gamma}_\Sigma := \overline{\Gamma}_N \cup \overline{\Gamma}_C$ and exterior normal n. Furthermore, let $\overline{\Gamma}_D \cap \overline{\Gamma}_C = \emptyset$. Here, $meas(\Gamma_D) > 0$ is assumed such that the results and formulations can be carried over directly to the interior problem by replacing the Steklov-Poincaré operator with the interior one. However, the absence of a Dirichlet boundary part would cause no further difficulties in the exterior problem. Further, let $(\Omega, \mathcal{F}, P)$ be a probability space as in Definition 2.3 with Ω the set of outcomes, $\mathcal{F} \subset 2^\Omega$ the σ-algebra of events, and $P : \mathcal{F} \to [0,1]$ a probability measure. For given Neumann force t and obstacle χ, both stochastic functions, a third stochastic function $u : \mathbb{R}^2 \setminus \overline{R} \times \Omega \to \mathbb{R}$ is sought such that P-a.e. in Ω

$$-\Delta u(x, \omega) = 0 \qquad\qquad \text{in } \mathbb{R}^2 \setminus \overline{R} \qquad (5.1\text{a})$$

$$u(x, \omega) = 0 \qquad\qquad \text{on } \Gamma_D \qquad (5.1\text{b})$$

$$\frac{\partial u}{\partial n}(x, \omega) = t(x, \omega) \qquad\qquad \text{on } \Gamma_N \qquad (5.1\text{c})$$

$$u \geq \chi(x, \omega), \ \frac{\partial u}{\partial n} \leq 0, \ \frac{\partial u}{\partial n}(u - \chi) = 0 \qquad \text{on } \Gamma_C \qquad (5.1\text{d})$$

$$u(x) = a \cdot \log(x) + b + o(1) \quad \text{as } \|x\| \to \infty \qquad (5.1\text{e})$$

holds with a, b real constants. Contrary to an interior problem $\frac{\partial u}{\partial n}|_{\Gamma_C} \leq 0$ and not greater or equal to zero due to the orientation of the normal n. The uncertainty in (5.1) lies within the magnitude of the Neumann force and the shape of the obstacle. In the following, two different approaches are considered. One is based on a primal

variational inequality and the other is based on a mixed method in which the negative of the unknown normal derivative on Γ_C is represented by a Lagrange multiplier λ. Due to the Signorini conditions (5.1d) λ satisfies

$$\lambda(x, \omega) = \begin{cases} 0 & \text{if } (x, \omega) \in \mathcal{N} \\ \geq 0 & \text{if } (x, \omega) \in \mathcal{C} \end{cases} \tag{5.2}$$

where $\mathcal{C} := \{(x, \omega) \in \Gamma_C \times \Omega : u = \chi\}$ is the contact set and $\mathcal{N}$ its complement in $\Gamma_C \times \Omega$. The main difficulty in the numerical analysis is caused by the Signorini condition. Since it is only applied to one part of the space boundary, it seems natural to use a boundary integral variational formulation in space and a standard weak formulation in the stochastic component. For the spatial part of the solution it is sufficient to complete its Cauchy data which is achieved by employing the Steklov-Poincaré operator S as a Dirichlet-to-Neumann map. For the mixed method let

$$L^+ := \left\{ \mu \in L_P^2(\Omega; H^{\frac{1}{2}}(\Gamma_C))' : E\left[\langle \mu, v \rangle_{\Gamma_C}\right] \leq 0 \ \forall 0 \geq v \in L_P^2(\Omega; H^{\frac{1}{2}}(\Gamma_C)) \right\} \tag{5.3}$$

be the convex Lagrange multiplier set with $L_P^2(\Omega; H^{\frac{1}{2}}(\Gamma_C))' = L_P^2(\Omega; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))$ the dual space to $L_P^2(\Omega; H^{\frac{1}{2}}(\Gamma_C))$ and $E\left[\langle \mu, v \rangle_{\Gamma_C}\right]$ the expected value of the duality pairing. The corresponding variational mixed formulation within a Galerkin setting is:
Find $(u, \lambda) \in L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)) \times L^+$ such that

$$A(u, v) - E\left[\langle v, \lambda \rangle_{\Gamma_C}\right] = L(v) \qquad \forall v \in L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)) \tag{5.4a}$$

$$E\left[\langle u, \mu - \lambda \rangle_{\Gamma_C}\right] \geq E\left[\langle \chi, \mu - \lambda \rangle_{\Gamma_C}\right] \quad \forall \mu \in L^+ \tag{5.4b}$$

with the bilinear, linear forms

$$A(u, v) := E\left[\langle Su, v \rangle_{\Gamma_\Sigma}\right], \qquad L(v) := -E\left[\langle t, v \rangle_{\Gamma_N}\right]. \tag{5.5}$$

Lemma 5.1. *Any solution of* (5.1) *is a solution of* (5.4). *The converse holds in a distributional sense.*

Proof. From the Calderon projector it is well known that for u and $v \in \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)$

$$\langle Su, v \rangle = \langle Su, v \rangle_{\Gamma_\Sigma} = -\left\langle \frac{\partial u}{\partial n}, v \right\rangle = \langle -t, v \rangle_{\Gamma_N} + \langle \lambda, v \rangle_{\Gamma_C}. \tag{5.6}$$

Taking the expected value of (5.6) yields (5.4a). Equation (5.1d)$_2$ directly yields $\lambda \in L^+$ and testing (5.1d)$_1$ with $\mu \in L^+$ yields

$$\int_\Omega \int_{\Gamma_C} (u - \chi)\mu \, ds \, dP(\omega) \geq 0$$

which in turn yields (5.4b) after subtracting the zero (5.1d)$_3$. The Dirichlet condition is strongly satisfied in $\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)$.
For the opposite direction choose $v \in C_0^\infty \cap L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$ in (5.4a), then the representation formula (2.2) yields with the Cauchy data $(u, V^{-1}(K - \frac{1}{2})u)$ equation (5.1a)

and (5.1e). Now choosing $v \in C^\infty \cap L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$ such that additionally $v|_{\Gamma_C} = 0$. Then the Dirichlet-to-Neumann mapping (5.6) yields (5.1c). The constraint (5.1b) is satisfied immediately. Next, choosing $\mu = 0$, $\mu = 2\lambda$ in (5.4b) yields

$$\int_\Omega \int_{\Gamma_C} (u - \chi)\lambda \, ds \, dP(\omega) = 0 \quad \text{and} \quad \int_\Omega \int_{\Gamma_C} (u - \chi)\mu \, ds \, dP(\omega) \geq 0 \quad \forall \mu \in L^+.$$

$$(5.7)$$

For $\mu \in C_0^\infty(\Gamma_C) \cap L^+$ this yields $(5.1d)_1$. Analogously, $\lambda \in L^+$ yields $(5.1d)_2$. These together with (5.7) yield equation $(5.1d)_3$. $\qquad\square$

A drawback of mixed methods is that an additional, later to be approximated, unknown is introduced. This can be avoided by dealing with variational inequality formulations in which the non-penetration condition is incorporated in the convex set of the ansatz and test functions. Let

$$K := \left\{ v \in L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)) : u \geq \chi \ (P)\text{-a.e. on } \Gamma_C \times \Omega \right\} \qquad (5.8)$$

be this convex set, then the variational inequality formulation is:

$$\text{Find } u \in K: \quad A(u, v - u) \geq L(v - u) \quad \forall v \in K. \qquad (5.9)$$

Since the bilinear form $A(\cdot, \cdot)$ is symmetric and coercive (c.f. Lemma 5.3) (5.9) is equivalent to the minimization problem

$$\text{Find } u \in K: \quad J(u) \leq J(v) \quad \forall v \in K \qquad (5.10)$$

with energy functional

$$J(v) := \frac{1}{2} A(v, v) - L(v). \qquad (5.11)$$

Lemma 5.2. *Any solution of (5.1) is a solution of (5.9) and (5.10). The converse holds in a distributional sense.*

Proof. Follows immediately from the equivalence Theorem 5.2 and Lemma 5.1. $\qquad\square$

5.1.1 Existence and Uniqueness of a Weak Solution

The proof of existence and uniqueness is an application of the Lions-Stampacchia theorem. Hence, only the continuity and coercivity of the bilinear form and the equivalence of the different weak formulations must be shown.

Lemma 5.3. *The bilinear form $A(\cdot, \cdot)$ is symmetric, continuous and $L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$-coercive, i.e. there exists constants $C_A > 0$ and $\alpha > 0$ such that there holds for all $u, v \in L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$*

$$A(u, v) \leq C_A \|u\|_{L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} \|v\|_{L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}, \qquad (5.12)$$

$$A(v, v) \geq \alpha \|v\|_{L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}^2. \qquad (5.13)$$

5 Elliptic Stochastic Contact Problems

Proof. Lemma 2.2 yields the continuity and $\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)$-coercivity of S. The continuity assertion follows with the Cauchy-Schwarz inequality and the coercivity by the definition of the norm. The symmetry follows directly from the symmetry of S. $\qquad\square$

In order to proof the inf-sup-condition the abstract result from [15, Theorem 3.2.1] is first generalized to the $L^2_P(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$-norm.

Lemma 5.4. *Let Γ be a closed Lipschitz curve with two open and connected but disjoint subsets $\gamma_0 \subset \Gamma$, $\gamma_1 \subset \Gamma$, $\overline{\gamma}_0 \cap \overline{\gamma}_1 = \emptyset$ and $\gamma_0^* := \Gamma \setminus \overline{\gamma}_0$, $\gamma_{01}^* := \Gamma \setminus (\overline{\gamma}_0 \cup \overline{\gamma}_1)$. Further let*

$$X_{\gamma_0,\gamma_1} := \left\{ \chi \geq 0 : ||\chi||_{L^\infty(\Omega;L^\infty(\Gamma))} = 1,\ \chi' \in L^\infty(\Omega; L^\infty(\Gamma)),\ \chi|_{\gamma_0} \equiv 0 \right.$$
$$\left. and\ \chi|_{\gamma_1} \equiv 1\ for\ P\text{-}a.e.\ \omega \in \Omega \right\}.$$

Then for arbitrary $v \in L^2_P(\Omega; H^{\frac{1}{2}}(\Gamma))$ and for arbitrary $\chi \in X_{\gamma_0,\gamma_1}$ there holds

$$\chi v \in L^2_P(\Omega; H^{\frac{1}{2}}(\Gamma)), \qquad \chi v|_{\gamma_0^*} \in L^2_P(\Omega; \tilde{H}^{\frac{1}{2}}(\gamma_0^*))$$

and

$$||\chi v||_{L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))} := ||\chi v||_{L^2_P(\Omega;H^{\frac{1}{2}}(\Gamma))} \leq C_{\chi'}||v||_{L^2_P(\Omega;H^{\frac{1}{2}}(\gamma_0^*))} \leq C_{\chi'}||v||_{L^2_P(\Omega;H^{\frac{1}{2}}(\Gamma))}. \tag{5.14}$$

where $C_{\chi'} = 2^{\frac{1}{4}} \left\| \left(1 + ||\chi'||^2_{L_\infty(\gamma_{01}^*)} \right)^{\frac{1}{2}} \right\|^{\frac{1}{2}}_{L^\infty(\Omega)}$.

Proof. Obviously there holds $||\chi v||_{L_2(\Gamma)} \leq ||v||_{L_2(\gamma_0^*)}$ for P-a.e. $\omega \in \Omega$. Further, for $v \in H^1(\Gamma)$ holds

$$||\chi v||_{H^1(\Gamma)} := \left(\int_\Gamma (\chi' v + \chi v')^2\, ds + ||\chi v||^2_{L_2(\Gamma)} \right)^{1/2}$$
$$\leq \left(2 \int_\Gamma (\chi' v)^2 + (\chi v')^2\, ds + ||\chi v||^2_{L_2(\Gamma)} \right)^{1/2}$$
$$\leq \sqrt{2} \left(||\chi'||^2_{L_\infty(\gamma_{01}^*)} ||v||^2_{L_2(\gamma_{01}^*)} + ||v'||^2_{L_2(\gamma_0^*)} + ||v||^2_{L_2(\gamma_0^*)} \right)^{1/2}$$
$$\leq \sqrt{2} \left(1 + ||\chi'||^2_{L_\infty(\gamma_{01}^*)} \right)^{1/2} ||v||_{H^1(\gamma_0^*)}.$$

Then, the first inequality in the assertion follows from the real interpolation between L_2 and H^1 as well as the definition of the $L^2_P(\Omega; \cdot)$-norm. The second inequality follows trivially by definition of the Sobolev spaces on open curves. $\qquad\square$

Lemma 5.5. *Let the assumptions of Lemma 5.4 hold. Then for all $\phi \in L^2_P(\Omega; H^{\frac{1}{2}}(\gamma_1))$ there exists an extension $f_\phi \in L^2_P(\Omega; \tilde{H}^{\frac{1}{2}}(\gamma_0^*))$ of ϕ onto γ_0^*, such that $f_\phi|_{\gamma_1} = \phi$ for P-a.e. $\omega \in \Omega$ and*

$$\exists\, \alpha > 0 : \quad ||\phi||_{L^2_P(\Omega;H^{\frac{1}{2}}(\gamma_1))} \geq \alpha ||f_\phi||_{L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))}, \tag{5.15}$$

where the constant $\alpha > 0$ is independent of ϕ.

Proof. Using the definition of the $H^{\frac{1}{2}}$-norm on open curves and Lemma 5.4 yields for arbitrary but fixed $\chi \in X_{\gamma_0,\gamma_1}$ that there holds

$$\|\phi\|_{L^2_P(\Omega;H^{\frac{1}{2}}(\gamma_1))} \geq \inf_{v \in L^2_P(\Omega;H^{\frac{1}{2}}(\Gamma))} \left\{ \|v\|_{L^2_P(\Omega;H^{\frac{1}{2}}(\Gamma))} : v|_{\gamma_1} = \phi \right\}$$

$$\geq C_{\chi'}^{-1} \inf_{v \in L^2_P(\Omega;H^{\frac{1}{2}}(\Gamma))} \left\{ \|\chi v\|_{L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))} : \chi v|_{\gamma_1} = \phi \right\}$$

$$\geq C_{\chi'}^{-1} \inf_{f \in L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))} \left\{ \|f\|_{L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))} : f|_{\gamma_1} = \phi \right\}.$$

The last inequality holds due to inclusion

$$\left\{ \chi v|_{\gamma_0^*} : v \in H^{\frac{1}{2}}(\Gamma) \right\} \subset \tilde{H}^{\frac{1}{2}}(\gamma_0^*) \qquad \forall \chi \in X_{\gamma_0,\gamma_1} \text{ and } P\text{-a.e. } \omega \in \Omega.$$

Further, there exists $f_\phi \in \left\{ w \in L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*)) : w|_{\gamma_1} = \phi \right\}$ such that

$$\|f_\phi\|_{L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))} \leq 2 \inf_{f \in L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))} \left\{ \|f\|_{L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))} : f|_{\gamma_1} = \phi \right\}.$$

and therefore

$$\|\phi\|_{L^2_P(\Omega;H^{\frac{1}{2}}(\gamma_1))} \geq (2C_{\chi'})^{-1}\|f_\phi\|_{L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))},$$

The largest possible constant α in the above estimate is $\alpha := \left(2 \inf_{\chi \in X_{\gamma_0,\gamma_1}} C_{\chi'} \right)^{-1} > 0.$

$\square$

Theorem 5.1. *Let the assumptions of Lemma 5.4 hold. Then there holds the following inf-sup condition:*

$$\exists\, \alpha > 0 : \quad \sup_{v \in L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))\setminus\{0\}} \frac{\int_\Omega \langle \mu, v \rangle_{\gamma_1}\, dP(\omega)}{\|v\|_{L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))}} \geq \alpha\|\mu\|_{L^2_P(\Omega;\tilde{H}^{-\frac{1}{2}}(\gamma_1))} \ \forall \mu \in L^2_P(\Omega;\tilde{H}^{-\frac{1}{2}}(\gamma_1)).$$

$$(5.16)$$

Moreover, the constant $\alpha > 0$ is independent of μ and v.

Proof. The proof is an application of the previous Lemma 5.5.

$$\|\mu\|_{L^2_P(\Omega;\tilde{H}^{-\frac{1}{2}}(\gamma_1))} = \sup_{\phi \in L^2_P(\Omega;H^{\frac{1}{2}}(\gamma_1))\setminus\{0\}} \frac{\int_\Omega \langle \mu, \phi \rangle_{\gamma_1}\, dP(\omega)}{\|\phi\|_{L^2_P(\Omega;H^{\frac{1}{2}}(\gamma_1))}}$$

$$\leq \alpha^{-1} \sup_{\phi \in L^2_P(\Omega;H^{\frac{1}{2}}(\gamma_1))\setminus\{0\}} \frac{\int_\Omega \langle \mu, f_\phi \rangle_{\gamma_1}\, dP(\omega)}{\|f_\phi\|_{L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))}}$$

$$\leq \alpha^{-1} \sup_{v \in L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))\setminus\{0\}} \frac{\int_\Omega \langle \mu, v \rangle_{\gamma_1}\, dP(\omega)}{\|v\|_{L^2_P(\Omega;\tilde{H}^{\frac{1}{2}}(\gamma_0^*))}}$$

where the constant α is $\left(2 \inf_{\chi \in X_{\gamma_0, \gamma_1}} C_{\chi'}\right)^{-1}$. Since for an arbitrary $\phi \in L_P^2(\Omega; H^{\frac{1}{2}}(\gamma_1))$ there holds $f_\phi \in L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\gamma_0^*))$ by construction. Therefore, taking the supremum not only over $\phi \in L_P^2(\Omega; H^{\frac{1}{2}}(\gamma_1))$ and working with $f_\phi \in L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\gamma_0^*))$ but taking the supremum over the the entire space $L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\gamma_0^*)) \setminus \{0\}$ yields the last inequality. $\qquad \square$

Lemma 5.6. *The bilinear form* $E\left[\langle v, \mu \rangle_{\Gamma_C}\right]$ *is continuous and satisfies the inf-sup condition, i.e.* $\exists \, \alpha > 0$:

$$\sup_{v \in L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)) \setminus \{0\}} \frac{E\left[\langle v, \mu \rangle_{\Gamma_C}\right]}{\|v\|_{L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}} \geq \alpha \|\mu\|_{L_P^2(\Omega; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} \quad \forall \mu \in L_P^2(\Omega; \tilde{H}^{-\frac{1}{2}}(\Gamma_C)).$$

$$(5.17)$$

Proof. The continuity condition

$$\left| E\left[\langle v, \mu \rangle_{\Gamma_C}\right] \right| \leq \|v\|_{L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} \|\mu\|_{L_P^2(\Omega; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))}$$

holds trivially by the definition of the dual norm and the embedding $H^{\frac{1}{2}}(\Gamma_C) \subset \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)$, since $\overline{\Gamma}_D \cap \overline{\Gamma}_C = \emptyset$. The inf-sup condition is an application of Theorem 5.1 with $\gamma_0 = \Gamma_D$, $\gamma_1 = \Gamma_C$, $\gamma_{01}^* = \Gamma_N$ and $\gamma_0^* = \Gamma_\Sigma$. $\qquad \square$

Lemma 5.7. *The energy functional J is coercive, i.e.* $J(v) \to \infty$ *as* $\|v\|_{L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} \to \infty$.

Proof. Follows immediately from the coercivity of the bilinear form $A(\cdot, \cdot)$, the Cauchy-Schwarz inequality and the trace theorem. $\qquad \square$

With the coercivity and the continuous inf-sup property, the following equivalence theorem can be proven.

Theorem 5.2. *The problems* (5.4), (5.9) *and* (5.10) *are equivalent.*

Proof. (5.9) $\Rightarrow$ (5.10): The linearity and coercivity yield

$$0 \leq A(u, v - u) - L(v - u) = J(v) - J(u) - \frac{1}{2} A(v - u, v - u) \leq J(v) - J(u).$$

(5.9) $\Leftarrow$ (5.10): Since K is convex there holds with $\lambda \in (0, 1)$ and the symmetry of A

$$0 \leq J(u + \lambda(v - u)) - J(u) = \lambda A(u, v - u) + \frac{\lambda^2}{2} A(v - u, v - u) - \lambda L(v - u).$$

Dividing by λ and taking the limit $\lambda \to 0^+$ yields (5.9).
(5.4) $\Rightarrow$ (5.9): Assume there exists a measurable set $O \subseteq \Gamma_C \times \Omega$, i.e.

$$\int_\Omega \int_{\Gamma_C} 1_O \, ds \, dP(\omega) > 0$$

where 1_O is one in O and zero elsewhere, in which $u < \chi$. Hence, choosing $\mu = 1_O \in L^+$ yields

$$E\left[\langle u - \chi, 1_O\rangle_{\Gamma_C}\right] < 0$$

which contradicts (5.7). Consequently $u \in K$. Now, choosing $v = w - u$ in (5.4a) with $w \in K$ and using

$$E\left[\langle w - u, \lambda\rangle_{\Gamma_C}\right] = E\left[\langle w - \chi, \lambda\rangle_{\Gamma_C}\right] \geq 0$$

yields (5.9).

(5.4) $\Leftarrow$ (5.9): Let $Res \in L_P^2(\Omega; H^{-\frac{1}{2}}(\Gamma_\Sigma))$ be the residual of (5.9) which is defined by

$$E\left[\langle Res, v\rangle_{\Gamma_\Sigma}\right] := A(u, v) - L(v) \quad \forall v \in L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)).$$

Obviously, there holds $E\left[\langle Res, v - u\rangle_{\Gamma_\Sigma}\right] \geq 0$ for all $v \in K$. Choosing $v \in K$ arbitrary but with $v|_{\Gamma_C} = u|_{\Gamma_C}$ yields

$$E\left[\langle Res, v\rangle_{\Gamma_N}\right] = 0 \quad \forall v \in L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_N)).$$

Hence,

$$\|Res\|_{L_P^2(\Omega; H^{-\frac{1}{2}}(\Gamma_N))} = 0.$$

Consequently, it remains to show that $Res|_{\Gamma_C} \in L^+$ and satisfies (5.4b). Choosing $v|_{\Gamma_C} = \chi$ and $v|_{\Gamma_C} = 2u|_{\Gamma_C} - \chi$ yields

$$E\left[\langle Res, u - \chi\rangle_{\Gamma_C}\right] \leq 0, \quad E\left[\langle Res, u - \chi\rangle_{\Gamma_C}\right] \geq 0 \quad \Rightarrow \quad E\left[\langle u - \chi, Res\rangle_{\Gamma_C}\right] = 0.$$

This result together with $E\left[\langle u - \chi, \mu\rangle_{\Gamma_C}\right] \geq 0$ for all $\mu \in L^+$ yields (5.4b). Finally, choosing $v|_{\Gamma_C} = w + u|_{\Gamma_C}$ with $w \geq 0$ yields $Res|_{\Gamma_C} \in L^+$ and setting $\lambda = Res|_{\Gamma_C}$ completes the proof. $\qquad\square$

Lemma 5.8. *Let $u \in K$ be the unique solution of (5.9) then there exists a $\lambda \in L^+$ such that (u, λ) is the unique solution of (5.4).*

Proof. Theorem 5.2 guaranties the existence of λ and carries the uniqueness of u over from (5.9) to (5.4). Assume (u, λ_1) and (u, λ_2) both solve (5.4). Then (5.4a) yields

$$E\left[\langle v, \lambda_1 - \lambda_2\rangle_{\Gamma_C}\right] = 0 \quad \forall v \in L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$$

and the continuous inf-sup condition of Lemma 5.6 implies $\|\lambda_1 - \lambda_2\|_{L_P^2(\Omega; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} = 0$.
$\qquad\square$

This equivalence of the different weak formulations simplifies the proof of existence and uniqueness.

Theorem 5.3. *There exists exactly one solution to the problems* (5.4), (5.9) *and* (5.10).

Proof. By Theorem 5.2 and Lemma 5.8 it is sufficient to proof existence and uniqueness for (5.9). Since $K \subset L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$ is a non-empty, closed, convex subset of a Hilbert space and A is a continuous, $L_P^2(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$-coercive bilinear form and $L \in L_P^2(\Omega; H^{-\frac{1}{2}}(\Gamma_\Sigma))$, existence and uniqueness of u follows from the Stampacchia theorem in [45, Theorem 2.1]. □

5.1.2 Equivalent Deterministic Formulation

The previous sections deal with the abstract probability space $(\Omega, \mathcal{F}, P)$. For the use of finite elements it is crucial to express, or at least to approximate, the random fields by a finite number of mutually independent random variables [1, 41]. The truncated Karhunen-Loève expansion

$$t(x, \omega) \approx E[t](x) + \sum_{n=1}^{N} \sqrt{\hat{\lambda}_n} \hat{b}_n(x) Y_n(\omega) \tag{5.18}$$

is a widely used tool to find such a set of random variables. In [41] Keese explains this with reference to [73] by the fact that the Karhunen-Loève expansion is an optimal linear approximation of t in the sense that the $L^2((\mathbb{R}^2 \setminus \bar{R}) \times \Omega)$-error would be larger if $\sqrt{\hat{\lambda}_n} Y_n(\omega)$ were chosen differently.

Assumption 5.1 (Finite dimensional noise, Assumption 2.1 in [1])**.** *The functions* $t : \Gamma_N \times \Omega \to \mathbb{R}$ *and* $\chi : \Gamma_C \times \Omega \to \mathbb{R}$ *are of finite Karhunen-Loève expansions, i.e.*

$$t(x,\omega) = E[t](x) + \sum_{n=1}^{N} \sqrt{\hat{\lambda}_n} \hat{b}_n(x) Y_n(\omega) \ \text{and} \ \chi(x,\omega) = E[\chi](x) + \sum_{n=1}^{N} \sqrt{\tilde{\lambda}_n} \tilde{b}_n(x) Y_n(\omega),$$

where $\{Y_n\}_{n=1}^{N}$ *are real random variables with zero mean and unit variance, are uncorrelated, and have images,* $\Theta_n \equiv Y_n(\Omega)$, *that are bounded intervals in* $\mathbb{R}$ *for* $1 \leq n \leq N$. *Moreover, it is assumed that each* Y_n *has a density function* $\rho_n : \Theta_n \to \mathbb{R}_{>0}$.

For the ease of presentation denote the tensor product of the images by $\Theta \equiv \prod_{n=1}^{N} \Theta_n \subset \mathbb{R}^N$ and the joint density function by $\rho(y)$ for all $y \in \Theta$. Assumption 5.1 is most crucial since it allows to describe u and therefore also λ by a finite number of random variables. More precisely, the Doob-Dynkin lemma yields $u(x, \omega) = u(x, Y_1(\omega), \ldots, Y_N(\omega))$ and $\lambda(x, \omega) = \lambda(x, Y_1(\omega), \ldots, Y_N(\omega))$, c.f. [1]. Using the transformation $y = Y(\omega)$ and $E[Y] = \int_\Theta y \rho(y) \, dy$ the formulations (5.4), (5.9), and (5.10) are equivalent to the deterministic formulations: Let

$$\tilde{L}^+ := \left\{ \mu \in L_\rho^2(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C)) : \int_\Theta \langle \mu, v \rangle_{\Gamma_C} \rho \, dy \leq 0 \ \ \forall 0 \geq v \in L_\rho^2(\Theta; H^{\frac{1}{2}}(\Gamma_C)) \right\} \tag{5.19}$$

and

$$\tilde{K} := \left\{ v \in L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)) : u \geq \chi \ (\rho)\text{-a.e. on } \Gamma_C \times \Omega \right\}. \tag{5.20}$$

Then, these may be formulated as a mixed method: Find $(u, \lambda) \in L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)) \times \tilde{L}^+$ such that

$$\tilde{A}(u, v) - \int_\Theta \langle v, \lambda \rangle_{\Gamma_C} \rho \, dy = \tilde{L}(v) \qquad \forall \, v \in L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)) \tag{5.21a}$$

$$\int_\Theta \rho(y) \langle u, \mu - \lambda \rangle_{\Gamma_C} \, dy \geq \int_\Theta \rho(y) \langle \chi, \mu - \lambda \rangle_{\Gamma_C} \, dy \quad \forall \, \mu \in \tilde{L}^+ \tag{5.21b}$$

as a variational inequality formulation:

$$\text{Find } u \in \tilde{K} : \quad \tilde{A}(u, v - u) \geq \tilde{L}(v - u) \quad \forall \, v \in \tilde{K} \tag{5.22}$$

or as a minimization problem:

$$\text{Find } u \in \tilde{K}: \quad \tilde{J}(u) \leq \tilde{J}(v) \quad \forall \, v \in \tilde{K} \tag{5.23}$$

with the bilinear form, linear form, and energy functional

$$\tilde{A}(u, v) := \int_\Theta \langle Su, v \rangle_{\Gamma_\Sigma} \rho \, dy, \quad \tilde{L}(v) := - \int_\Theta \langle t, v \rangle_{\Gamma_N} \rho \, dy, \quad \tilde{J}(v) := \frac{1}{2} \tilde{A}(v, v) - \tilde{L}(v), \tag{5.24}$$

respectively.

5.2 *hp*-**FEM/BEM Discretizations for Elliptic Stochastic Contact Problem**

The multi-dimensional deterministic problems of the previous section can be solved approximatively by an *hp*-FEM/BEM discretization technique. Therewith two problems arise. The first one is the discretization of the bilinear form $\tilde{A}(\cdot, \cdot)$, more precisely of the Steklov-Poincaré operator S, and the second one is the discretization of the dual variable λ or non-penetration condition in $\tilde{K}$. The discretization strategies for both the mixed and variational inequality problem are $L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$-conforming in the primal variable u but non-conforming in $\tilde{K}$ and $\tilde{L}^+$.

Let $\mathcal{T}_{hp}$ be a subdivision of Γ_Σ with a corresponding polynomial degree distribution. Analogously, let $\mathcal{T}_{kq}^{stoch}$ be a mesh of Θ with elements constructed by tensor products of 1d elements. With these meshes the following discrete function spaces

$$V_{hp} := \left\{ v \in \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma) : \ v|_I \in \mathbb{P}_{p_I}(I) \ \forall \, I \in \mathcal{T}_{hp} \right\} = \text{span} \left\{ \phi_j \right\}_{j=1}^{\dim V_{hp}},$$

$$V_{hp}^D := \left\{ v \in H^{-\frac{1}{2}}(\Gamma_\Sigma) : \ v|_I \in \mathbb{P}_{p_I - 1}(I) \ \forall \, I \in \mathcal{T}_{hp} \right\} = \text{span} \left\{ \phi_j^D \right\}_{j=1}^{\dim V_{hp}^D},$$

$$W_{kq} := \left\{ v \in L_\rho^2(\Omega) : \ v|_Q \in \mathbb{P}_{p_Q}(Q) \ \forall \, Q \in \mathcal{T}_{kq}^{stoch} \right\} = \text{span} \left\{ \vartheta_j \right\}_{j=1}^{\dim W_{kq}}$$

are well defined. The basis functions ϕ_j^D are monomials which allows an easy and stable analytic evaluation of the involved potentials [57, 58]. Whereas the basis functions ϕ_j are 1-D affinely transformed Gauss-Lobatto-Lagrange basis functions which are also continuous. Since Q is a N-dimensional cube, the basis functions ϑ_j are simply N-D affinely transformed Gauss-Lobatto-Lagrange basis functions. Let G_{hp} be the global set of affinely transformed Gauss-Lobatto points which are used to construct ϕ_j restricted to Γ_C and G_{kq}^{stoch} the analogous set in the stochastic domain. For the ease of presentation and implementation it is assumed that the basis functions are sorted, such that the ϕ_j, ϕ_j^D with support in Γ_C are the first ones, starting with index one. At this point the choice of basis function is still arbitrary but it will play a significant role for the construction of the discrete sets $K_{hp,kq}$ and $L_{hp,kq}^+$. In the discrete formulation the bilinear form $\tilde{A}(\cdot,\cdot)$ is replaced by the discrete bilinear form

$$A_h(u_h, v_h) := \int_\Theta \langle S_h u_h, v_h \rangle_{\Gamma_\Sigma} \, \rho \, dy = \int_\Theta \left\langle W u_h + (K' - \frac{1}{2})\Psi_h, v_h \right\rangle_{\Gamma_\Sigma} \rho \, dy \quad (5.25)$$

where $\Psi_h(y) \in V_{hp}^D$ solves for ρ-a.e. $y \in \Theta$ the auxiliary problem

$$\langle V\Psi_h, v_h \rangle_{\Gamma_\Sigma} = \left\langle (K - \frac{1}{2})u_h, v_h \right\rangle_{\Gamma_\Sigma} \quad \forall\, v \in V_{hp}^D. \quad (5.26)$$

The equation (5.26) approximates the action of V^{-1} on $(K - \frac{1}{2})u_h$.

Lemma 5.9 ([11] and [72], Lemma 12.4)**.** *The symmetric discrete Poincaré-Steklov operator $S_h : H^{\frac{1}{2}}(\Gamma) \to H^{-\frac{1}{2}}(\Gamma)$ is continuous, i.e.*

$$\exists\, C_{S_h} > 0 \quad \forall\, v \in H^{\frac{1}{2}}(\Gamma): \quad \|S_h v\|_{H^{-\frac{1}{2}}(\Gamma)} \leq C_{S_h} \|v\|_{H^{\frac{1}{2}}(\Gamma)}$$

elliptic, i.e.

$$\exists\, \alpha_{S_h} > 0 \quad \forall\, v \in H^{\frac{1}{2}}(\Gamma): \quad \langle S_h v, v \rangle_\Gamma \geq \alpha_{S_h} \|v\|_{H^{\frac{1}{2}}(\Gamma)}^2$$

$$\exists\, \alpha_{S_h} > 0 \quad \forall\, v \in \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma): \quad \langle S_h v, v \rangle_\Gamma \geq \alpha_{S_h} \|v\|_{H^{\frac{1}{2}}(\Gamma)}^2$$

and satisfies the error estimate

$$\|(S - S_h)v\|_{H^{-\frac{1}{2}}(\Gamma)} \leq c \inf_{\Psi \in V_{hp}^D} \left\| V^{-1}(K - \frac{1}{2})v - \Psi \right\|_{H^{-\frac{1}{2}}(\Gamma)}.$$

Lemma 5.10. *The discrete bilinear form $A_h(u_h, v_h)$ is continuous and $L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$-coercive.*

Proof. Analogously to the proof of Lemma 5.3 but now with Lemma 5.9. $\qquad\square$

5.2.1 A Discrete Mixed Formulation

The discretization of the variational equation (5.21a) is straight forward. However, the discretization of the variational inequality constraint (5.21b) is not. The key challenge in mixed methods is the construction of the discrete Lagrange multiplier space. Wohlmuth et al. studied the use of dual basis functions for low order methods for such purposes in [77, 37] among others and the construction of higher order dual basis functions in [52]. Let

$$M_{hp} := \operatorname{span}\{\psi_j\}_{j=1}^{\dim V_{hp}|_{\Gamma_C}},$$

$$T_{kq} := \operatorname{span}\{\zeta_j\}_{j=1}^{\dim W_{kq}},$$

$$L^+_{hp,kq} := \left\{ \begin{array}{c} \mu \in M_{hp} \otimes T_{kq} : \int_\Theta \langle \mu, v \rangle_{\Gamma_C} \, \rho \, dy \le 0 \text{ for all} \\ v = \sum_{i=1}^{\dim M_{hp}} \sum_{j=1}^{\dim W_{kq}} v_{i,j} \phi_i \vartheta_j \in V_{hp}|_{\Gamma_C} \otimes W_{kq} \text{ with } v_{i,j} \le 0 \end{array} \right\}.$$

The discrete dual spaces M_{hp}, T_{kq} are spanned by biorthogonal basis functions, i.e.

$$\int_{\Gamma_C} \psi_i \phi_j \, ds = \delta_{ij} \int_{\Gamma_C} \phi_j \, ds, \qquad \int_\Theta \zeta_i \vartheta_j \rho \, dy = \delta_{ij} \int_\Theta \vartheta_j \rho \, dy, \qquad (5.27)$$

respectively. These biorthogonal basis function work particularly well with Gauss-Lobatto-Lagrange basis functions. Especially, the same mesh and polynomial degree distribution as for the discrete primal variable u_{hp} is used. As in the construction of the forthcoming discrete cone $K_{hp,kq}$ the weak non-negativity constraint of λ is relaxed to a discrete one. In particular, this method is non-conforming $L^+_{hp,kq} \not\subseteq L^+$ in the dual variable.

The following lemma which establishes a relationship of the integral value for the primal basis function with its dual is essential for reducing the discrete constraints to a simple complementarity problem for the expansion coefficients.

Lemma 5.11. *There holds for the integral value of the primal and dual basis functions* $\int_{\Gamma_C} \psi_i \, ds = \int_{\Gamma_C} \phi_i \, ds =: D^{space}_i > 0$ $(1 \le i \le \dim M_{hp})$ *and* $\int_\Theta \zeta_j \rho \, dy = \int_\Theta \vartheta_j \rho \, dy =: D^{stoch}_j > 0.$

Proof. First note that $\sum_{j=1}^{\dim V_{hp}} \phi_j|_{\Gamma_C} \equiv 1|_{\Gamma_C}$ is a partition of unity. Then by biorthogonality

$$\int_{\Gamma_C} \psi_i \, ds = \int_{\Gamma_C} \psi_i \sum_{j=1}^{\dim V_{hp}} \phi_j \, ds = \int_{\Gamma_C} \psi_i \phi_i \, ds = \int_{\Gamma_C} \phi_i \, ds.$$

Since $\phi_i|_I$ with $\mathcal{T}_{hp} \ni I \subseteq \operatorname{supp} \phi_i$ is a polynomial of given degree, denoted by p, the integral $\int_I \phi_i \, ds$ can be evaluated exactly with a Gauss-Lobatto quadrature with $p+1$ nodes and positive weights. Since ϕ_i is a Gauss-Lobatto-Lagrange basis function, it is zero in every quadrature node but one in which it takes the value one. Summing over all elements I yields the assertion. The second assertion follows analogously but also exploits the tensor product structure of the elements Q. $\square$

5 Elliptic Stochastic Contact Problems

The discrete mixed formulation is: Find $u_h \in V_{hp} \otimes W_{kq}$ and $\lambda_h \in L^+_{hp,kq}$:

$$A_h(u_h, v_h) - \int_\Theta \langle v_h, \lambda_h \rangle_{\Gamma_C} \, \rho \, dy = \tilde{L}(v_h) \qquad \forall \, v_h \in V_{hp} \otimes W_{kq} \quad \text{(5.28a)}$$

$$\int_\Theta \langle u_h, \mu_h - \lambda_h \rangle_{\Gamma_C} \, \rho \, dy \geq \int_\Theta \langle \chi, \mu_h - \lambda_h \rangle_{\Gamma_C} \, \rho \, dy \quad \forall \, \mu_h \in L^+_{hp,kq} \qquad \text{(5.28b)}$$

Due to the use of biorthogonal basis functions the system matrices of the discrete variational inequalities in (5.28b) and in $L^+_{hp,kq}$ are diagonal matrices which implies a compontentwise decoupling.

Theorem 5.4. *The condition* (5.28b) *is equivalent to the system*

$$U_{i,j} \geq g_{i,j} := \frac{1}{D_i^{space} D_j^{stoch}} \int_\Theta \rho \int_{\Gamma_C} \chi(x, \omega) \psi_i(x) \zeta_j(\omega) \, ds \, dy \qquad \text{(5.29a)}$$

$$\Lambda_{i,j} \geq 0 \qquad \text{(5.29b)}$$

$$\Lambda_{i,j} \, (U_{i,j} - g_{i,j}) = 0 \qquad \text{(5.29c)}$$

for $1 \leq i \leq \dim M_{hp}$ *and* $1 \leq j \leq \dim T_{kq}$. *Here* $U_{i,j}$, $\Lambda_{i,j}$ *are the expansion coefficients of* u_h, λ_h *respectively, and* $D_i^{space} := \int_{\Gamma_C} \phi_i \, ds$, $D_j^{stoch} := \int_\Theta \rho \vartheta_j \, dy$.

Proof. For the ease of presentation if not otherwise mentioned i ranges from 1 to $\dim M_{hp}$ and j from 1 to $\dim T_{kq}$. For every function $v \in V_{hp} \otimes W_{kq}$ there exists a unique $\{v_{i,j}\}_{i=1,\dots,\dim V_{hp}}^{j=1,\dots,\dim W_{kq}}$ such that

$$v = \sum_{i=1}^{\dim V_{hp}} \sum_{j=1}^{\dim W_{kq}} v_{i,j} \phi_i(x) \vartheta_j(y).$$

Also μ and $\lambda \in L^+_{hp,kq}$ can be written as a linear combination of the dual basis functions.

$$\mu = \sum_{i,j} \mu_{i,j} \psi_i(x) \zeta_j(y), \quad \lambda = \sum_{i,j} \lambda_{i,j} \psi_i(x) \zeta_j(y)$$

Due to the biorthogonality of the employed basis functions, there holds for all $v \in V_{hp}|_{\Gamma_C} \otimes W_{kq}$ with $v_{i,j} \leq 0$

$$\int_\Theta \rho \, \langle \mu, v \rangle_{\Gamma_C} \, dy = \sum_{i=1}^{\dim M_{hp}} \sum_{j=1}^{\dim T_{hp}} \mu_{i,j} v_{i,j} D_i^{space} D_j^{stoch} \leq 0 \qquad \text{(5.30)}$$

if $\mu \in L^+_{hp,kq}$. With $v_{i,j} \leq 0$ arbitrary and D_i^{space}, D_j^{stoch} positive by Lemma 5.11, the equation (5.30) yields $\mu_{i,j} v_{i,j} \leq 0$ and together with $v_{i,j} \leq 0$ this implies

$$\mu_{i,j} \geq 0 \quad \forall \, i, j.$$

Hence, $\lambda \in L^+_{hp,kq}$ implies (5.29b). On the other hand, equation (5.28b) yields with $\mu \in L^+_{hp,kq}$ arbitrary after exploiting the biorthogonality and dividing by the positive factor $D_i^{space} D_j^{stoch}$:

$$\text{Find } \lambda_{i,j} \geq 0 : \quad u_{i,j} \, (\mu_{i,j} - \lambda_{i,j}) \geq g_{i,j} \, (\mu_{i,j} - \lambda_{i,j}) \quad \forall \mu_{i,j} \geq 0 \quad \forall i, j. \qquad \text{(5.31)}$$

Choosing $\mu_{i,j} = \lambda_{i,j} + \eta_{i,j} > 0$ with $\eta_{i,j} > 0$ in (5.31) yields (5.29a). Choosing $\mu_{i,j} = 0$ and $\mu_{i,j} = 2\lambda_{i,j}$ the above equation yields (5.29c).

For the opposite direction, simply multiply (5.29a) with $\mu_{i,j} \geq 0$ and add the zero (5.29c) to obtain (5.31). Summing over all i, j and exploiting the biorthogonality yields (5.28b). $\qquad\square$

Theorem 5.5. *There exists exactly one solution to the discrete mixed formulation* (5.28).

Proof. **Uniqueness:** Assume (u_1, λ_1) and (u_2, λ_2) solve (5.28), then their difference satisfies

$$A_h(u_1 - u_2, v) - \int_\Theta \langle v, \lambda_1 - \lambda_2 \rangle_{\Gamma_C} \rho \, dy = 0 \quad \forall\, v \in V_{hp} \otimes W_{kq}. \qquad (5.32)$$

From (5.28b) follows with $\mu_1 = \lambda_2$ and $\mu_2 = \lambda_1$

$$\int_\Theta \langle u_1 - u_2, \lambda_1 - \lambda_2 \rangle_{\Gamma_C} \rho \, dy \leq 0.$$

Hence, choosing $v = u_1 - u_2$ in (5.32) and using the coercivity of Lemma 5.10 yields

$$\alpha \left\| u_1 - u_2 \right\|^2_{L^2_P(\Omega; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} \leq A_h(u_1 - u_2, u_1 - u_2) - \int_\Theta \langle u_1 - u_2, \lambda_1 - \lambda_2 \rangle_{\Gamma_C} \rho \, dy = 0.$$

Thus $u_1 = u_2$ and (5.32) reduces to

$$0 = \int_\Theta \rho \int_{\Gamma_C} v(\lambda_1 - \lambda_2) \, ds \, dy = \sum_{i=1}^{\dim M_{hp}} \sum_{j=1}^{\dim T_{kq}} v_{i,j} (\lambda_1 - \lambda_2)_{i,j} D_i^{space} D_j^{stoch} \quad \forall\, v \in V_{hp} \otimes W_{kq}$$

by the ordering of the basis function, i.e. for $\operatorname{supp} \phi_i \cap \Gamma_C = \emptyset \Rightarrow i > \dim M_{hp}$ and by exploiting the biorthogonality of the basis functions. Since v and therewith $v_{i,j}$ is arbitrary, it follows with $D_i^{space} D_j^{stoch} > 0$ that $(\lambda_1 - \lambda_2)_{i,j} = 0$ and hence $\lambda_1 = \lambda_2$.

Existence: It is well known that (5.29) is equivalent to the projection equation

$$\Lambda_{i,j} = \max\left\{0, \Lambda_{i,j} + r(g_{i,j} - U_{i,j})\right\} =: T\Lambda_{i,j} \quad 1 \leq i \leq \dim M_{hp}, 1 \leq j \leq \dim T_{kq} \tag{5.33}$$

where $r > 0$ is an arbitrary constant. Since for any given $\lambda_h \in L^+_{hp,kq}$ problem (5.28) reduces to a linear, finite dimensional problem, the above proven uniqueness also implies the unique existence of a discrete $u_h(\lambda_h)$. Hence, it is sufficient to show that T is a strict contraction to proof existence of (u_h, λ_h) solving (5.28). For the ease of presentation denote $\delta\Lambda = \Lambda_1 - \Lambda_2$ and $\delta U = U_1 - U_2$ where $\delta U|_{\Gamma_C}$ is a restriction to the indices

associated with the contact boundary, i.e. $1 \leq i \leq \dim M_{hp}, 1 \leq j \leq \dim T_{kq}$.

$$
\begin{aligned}
\|T\Lambda_1 - T\Lambda_2\|_2^2 &\leq \|\delta\Lambda - r\delta U|_{\Gamma_C}\|_2^2 \\
&= \|\delta\Lambda\|_2^2 - 2r\delta\Lambda^T \delta U|_{\Gamma_C} + r^2 \|\delta U|_{\Gamma_C}\|_2^2 \\
&= \|\delta\Lambda\|_2^2 - 2r \sum_i \sum_j \frac{D_i^{space} D_j^{stoch}}{D_i^{space} D_j^{stoch}} \delta\Lambda_{i,j} \delta U_{i,j} + r^2 \|\delta U|_{\Gamma_C}\|_2^2 \\
&\leq \|\delta\Lambda\|_2^2 - 2r \min_{i,j} \left\{ D_i^{space} D_j^{stoch} \right\} \delta U^T D\delta\Lambda + r^2 \|\delta U\|_2^2 \\
&= \|\delta\Lambda\|_2^2 - 2r \min_{i,j} \left\{ D_i^{space} D_j^{stoch} \right\} \delta U^T A_h \delta U + r^2 \|\delta U\|_2^2 \\
&\leq \|\delta\Lambda\|_2^2 - 2r\beta \|\delta U\|_2^2 + r^2 \|\delta U\|_2^2 \\
&= \|\delta\Lambda\|_2^2 \left(1 - 2\beta r\gamma^2 + r^2\gamma^2 \right)
\end{aligned}
$$

with $\gamma = \|\delta U\|_2 / \|\delta\Lambda\|_2$, $\beta = \min_{i,j} \left\{ D_i^{space} D_j^{stoch} \right\} \cdot$ minimal eigenvalue$(A_h) > 0$ and A_h, D the matrix representation of the operators in (5.28a) . Hence, for $0 < r < 2\beta$, T is a strict contraction. $\qquad\square$

5.2.2 A Discrete Variational Inequality Approach

Within the continuous formulation of the variational inequality problem, the non-penetration condition is almost everywhere strongly enforced. For higher order methods or general obstacle, this condition can not easily be described by a finite number of conditions. Therefore, the continuous non-penetration condition in the convex cone $\tilde{K}$ is replaced by a discrete one. Let

$$
K_{hp,kq} := \{ v \in V_{hp} \otimes W_{kq} : \ v_{i,j} \geq g_{i,j} \} \tag{5.34}
$$

be the discrete convex cone, which in general is not a subset of $\tilde{K} \not\supseteq K_{hp,kq}$. If $\chi \in C(\overline{\Gamma}_C \times \overline{\Theta})$, which requires more regularity than for the well posedness of (5.22) needed, then often $v_{i,j} = v(x_{ij}) \geq g_{i,j} := \chi(x_{ij})$ is chosen where x_{ij} are Gauss-Lobatto quadrature points. In the forthcoming $v_{i,j}$ are the coefficients of the discrete function $v \in V_{hp} \otimes W_{kq}$ and $g_{i,j}$ as defined in (5.29a). The biorthogonality of the basis functions immediately yields that the discrete cone $K_{hp,kq}$ is equivalent to

$$
K_{hp,kq} = \left\{ v \in V_{hp} \otimes W_{kq} : \ \int_\Theta \rho \int_{\Gamma_C} (v - \chi)\mu \, ds \, dy \geq 0 \ \ \forall \mu \in L_{hp,kq}^+ \right\}. \tag{5.35}
$$

Therewith, the discrete variational inequality formulation is:

$$
\text{Find } u_h \in K_{hp,kq}: \quad A_h(u_h, v_h - u_h) \geq \tilde{L}(v_h - u_h) \quad \forall v_h \in K_{hp,kq}. \tag{5.36}
$$

Theorem 5.6. *Let $g_{i,j}$ be as defined in (5.29a), then the discrete mixed formulation (5.28) is equivalent to the discrete variational inequality (5.36).*

Proof. For the ease of presentation if not otherwise mentioned i ranges from 1 to $\dim M_{hp} < \dim V_{hp}$ and j from 1 to $\dim T_{kq} = \dim W_{kq}$.

(5.28) $\Rightarrow$ (5.36) : Let (u, λ) be the solution of (5.28), then by Theorem 5.4 $u_{i,j} \geq g_{i,j}$ for all $1 \leq i \leq \dim M_{hp}$ and $1 \leq j \leq \dim T_{kq}$ which implies $u \in K_{hp,kq}$. By the biorthogonality of the basis functions there holds for $w \in K_{hp,kq}$

$$- \int_\Theta \langle w - u, \lambda \rangle_{\Gamma_C} \rho \, dy = - \sum_{i,j} D_i^{space} D_j^{stoch} (w_{i,j} - u_{i,j}) \lambda_{i,j} \leq 0$$

where the last inequality follows from Theorem 5.4, (5.29c), $w \in K_{hp,kq}$ and Lemma 5.11. Consequently, choosing $v = w - u$ with $w \in K_{hp,kq}$ in (5.28) yields (5.36).

(5.28) $\Leftarrow$ (5.36) : Let $u \in K_{hp,kq}$ be the solution of (5.36), then $u_{i,j} \geq g_{i,j}$ for all $1 \leq i \leq \dim M_{hp}$ and $1 \leq j \leq \dim T_{kq}$. Next, define the Lagrange multiplier by

$$\lambda = \sum_{i,j} \lambda_{i,j} \psi_i(x) \zeta_j(y) \in M_{hp} \otimes T_{kq} \text{ with } \lambda_{i,j} = \frac{A_h(u, \phi_i \vartheta_j) - \tilde{L}(\phi_i \vartheta_j)}{D_i^{space} D_j^{stoch}}.$$

Then, λ is the residual of (5.36), i.e.

$$A_h(u, v) - \int_\Theta \langle v, \lambda \rangle_{\Gamma_C} \rho \, dy - \tilde{L}(v) = 0 \quad \forall v \in V_{hp} \otimes W_{kq}. \tag{5.37}$$

First assume $v = \sum_{i=1}^{\dim M_{hp}} \sum_{j=1}^{\dim T_{kq}} v_{i,j} \phi_i \vartheta_j$, i.e. the coefficients $v_{i,j}$ with $i > \dim M_{hp}$ to the basis functions with $\operatorname{supp} \phi_i \vartheta_j \cap \Gamma_C \times \Theta = \emptyset$ are zero. Then by the biorthogonality and the definition of λ

$$- \int_\Theta \langle v, \lambda \rangle_{\Gamma_C} \rho \, dy = - \sum_{i,j} \lambda_{i,j} v_{i,j} D_i^{space} D_j^{stoch} = - \sum_{i,j} \left[A_h(u, \phi_i \vartheta_j) - \tilde{L}(\phi_i \vartheta_j) \right] v_{i,j}$$

$$= -A_h(u, v) + \tilde{L}(v)$$

Now assume the opposite case, i.e. $v_{i,j} = 0$ for all $1 \leq i \leq \dim M_{hp}$ and $1 \leq j \leq \dim T_{kq}$, then the residual of the discrete variational inequality is zero. Let

$$w = \sum_{i=1}^{\dim M_{hp}} \sum_{j=1}^{\dim T_{kq}} u_{i,j} \phi_i \vartheta_j + \sum_{i=\dim M_{hp}+1}^{\dim V_{hp}} \sum_{j=1}^{\dim T_{kq}} u_{i,j} \pm v_{i,j} \phi_i \vartheta_j.$$

If $v := w - u = \pm \sum_{i=\dim M_{hp}+1}^{\dim V_{hp}} \sum_{j=1}^{\dim T_{kq}} v_{i,j} \phi_i \vartheta_j$ is chosen in (5.36), this equation yields

$$A_h(u, v) - \tilde{L}(v) = 0.$$

Since all involved operators are linear these two cases yield (5.37). Choosing $v = u + \phi_i \vartheta_j$, i.e. $v \in K_{hp,kq}$, in (5.36) yields

$$0 \leq A_h(u, \phi_i \vartheta_j) - \tilde{L}(\phi_i \vartheta_j) = \lambda_{i,j} D_i^{space} D_j^{stoch} \Rightarrow \lambda_{i,j} \geq 0.$$

Finally, choose $v = \sum_{i=1}^{\dim M_{hp}} \sum_{j=1}^{\dim T_{kq}} g_{i,j} \phi_i \vartheta_j + \sum_{i=\dim M_{hp}+1}^{\dim V_{hp}} \sum_{j=1}^{\dim T_{kq}} u_{i,j} \phi_i \vartheta_j$ and
$v = \sum_{i=1}^{\dim M_{hp}} \sum_{j=1}^{\dim T_{kq}} (2u_{i,j} - g_{i,j}) \phi_i \vartheta_j + \sum_{i=\dim M_{hp}+1}^{\dim V_{hp}} \sum_{j=1}^{\dim T_{kq}} u_{i,j} \phi_i \vartheta_j$ in (5.36) to obtain

$$
\begin{aligned}
0 &= \sum_{i=1}^{\dim M_{hp}} \sum_{j=1}^{\dim T_{kq}} \left[A_h(u, \phi_i \vartheta_j) - \tilde{L}(\phi_i \vartheta_j) \right] (u_{i,j} - g_{i,j}) \\
&= \sum_{i=1}^{\dim M_{hp}} \sum_{j=1}^{\dim T_{kq}} \lambda_{i,j} (u_{i,j} - g_{i,j}) D_i^{space} D_j^{stoch} = \int_\Theta \langle u - \chi, \lambda \rangle_{\Gamma_C} \, \rho \, dy.
\end{aligned}
$$

The assertion follows by Theorem 5.4. $\qquad\square$

Theorem 5.7. *There exists exactly one solution to* (5.36).

Proof. For general $g_{i,j}$, again by the Stampacchia theorem [45, Theorem 2.1]. Alternatively, for $g_{i,j}$ as defined in (5.29a) by Theorem 5.6 and Theorem 5.5. $\qquad\square$

Since the discrete bilinear form $A_h(\cdot, \cdot)$ is symmetric and coercive and the discrete non-penetration condition is a simple box constraint on the coefficients, the problem (5.36) can be solved very efficiently by the primal-dual active set strategy of [50].

5.2.3 An Algorithm for Solving the Discrete Mixed Formulation

While for the variational formulation an efficient iterative solver exists, it still must be constructed for the mixed method. For every (i, j) the problem (5.29) is of the same type as (2.4) and can be dealt with by the ideas presented in Section 2.2.

Lemma 5.12. *The discrete problem* (5.28) *is equivalent to finding the root of*

$$
0 \stackrel{!}{=} F(\vec{u}_h, \vec{\lambda}_h) := \begin{pmatrix} A_h \vec{u}_h - D\vec{\lambda}_h - f \\ \varphi_\eta(\vec{u}_h, \vec{\lambda}_h) \end{pmatrix} \tag{5.38}
$$

where $A_h \vec{u}_h - D\vec{\lambda}_h - f = 0$ is the matrix representation of (5.28a) *and $\varphi_\eta : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ ($n = d \cdot \tilde{d}$, $m = \hat{d} \cdot \tilde{d}$, $d = \dim M_{hp}$, $\hat{d} = \dim V_{hp}$ and $\tilde{d} = \dim T_{kq} = \dim W_{kq}$) is the vector-valued penalized Fischer-Burmeister non-linear complementarity function (NCF), c.f.* (2.6). *The entry at position $(j-1)d + i$ ($1 \le j \le \tilde{d}$, $1 \le i \le d$) is defined by*

$$
\varphi_\eta(v, \mu) = \eta \left(\mu + v - \sqrt{\mu^2 + v^2} \right) + (1 - \eta) \max\{0, \mu\} \max\{0, v\}
$$

with $v = \vec{u}_h((j-1)\hat{d} + i) - \vec{g}((j-1)d + i)$, $\mu = \vec{\lambda}_h((j-1)d + i)$ and $\eta \in (0, 1]$.

Proof. Since φ_η is a NCF by Lemma 2.4, there holds

$$
\varphi_\eta(v, \mu) = 0 \iff v_i \ge 0, \ \mu_i \ge 0, \ \mu_i \cdot v_i = 0 \quad (i = 1, \dots, n).
$$

Hence, $\varphi_\eta(\vec{u}_h, \vec{\lambda}_h) = 0$ is equivalent to (5.29) which in turn is equivalent to (5.28b) by Theorem 5.4. $\qquad\square$

Lemma 5.13. *The matrix A_h is symmetric positive definite. The non-zero entries in $D = Diag(D^{stoch}) \otimes Diag([D^{space}; 0])$ are positive. The function φ_η is strongly semi-smooth and Lipschitzian.*

Proof. By Lemma 5.10 all eigenvalues of A_h are positive. The non-zero entries in D are of the kind $D_i^{space} D_j^{stoch} > 0$ by Lemma 5.11. The structure of D follows directly from the tensor structure of the space-stochastic domain and the numbering of the dofs. By Lemma 2.4, φ_η is strongly semi-smooth and Lipschitzian. $\qquad\square$

The non-linear problem (5.38) with the continuously differentiable, non-negative merit function

$$\Psi(u, \lambda) := \frac{1}{2} F(u, \lambda)^2 \qquad (5.39)$$

can be solved by the SSN Algorithm 5.1 being a realization of the Algorithm 2.1.

Algorithm 5.1. *(Semi-smooth Newton algorithm for elliptic stochastic contact)*

1. *Choose initial solution $u^0 \in \mathbb{R}^m$, $\lambda^0 \in \mathbb{R}^n$, $\tilde{\rho} > 0$, $\beta \in (0, 1)$, $\sigma \in (0, \frac{1}{2})$, $p > 2$*

2. *For $k = 0, 1, 2, \ldots$ do*

   a) *If $\left\| \nabla \Psi(u^k, \lambda^k) \right\| < tol$ or $\left\| \Psi(u^k, \lambda^k) \right\| < tol$ then stop.*

   b) *Compute subdifferential $H_k \in \partial F(u^k, \lambda^k)$ and find $d^k = (d_u^k, d_\lambda^k) \in \mathbb{R}^{m+n}$ s.t.*

   $$H_k d^k = -F(u^k, \lambda^k). \qquad (5.40)$$

   If (5.40) not solvable or if the descent condition

   $$\nabla \Psi(u^k, \lambda^k) d^k \leq -\tilde{\rho} ||d^k||^p \qquad (5.41)$$

   is not satisfied, set $d^k := -\nabla \Psi(u^k, \lambda^k)$.

   c) *Compute search length $t_k := \max \left\{ \beta^l : l = 0, 1, 2, \ldots \right\}$ s.t.*

   $$\Psi(u^k + t_k d_u^k, \lambda^k + t_k d_\lambda^k) \leq \Psi(u^k, \lambda^k) + \sigma t_k \nabla \Psi(u^k, \lambda^k) d^k.$$

   d) *Update the solution vectors and goto step 2.*

   $$u^{k+1} = u^k + t_k d_u^k, \quad \lambda^{k+1} = \lambda^k + t_k d_\lambda^k$$

For the implementation it is sufficient to choose one subdifferential, e.g.

$$H_k = \begin{pmatrix} A & -D \\ \frac{\partial \varphi_\eta(u^k, \lambda^k)}{\partial u} & \frac{\partial \varphi_\eta(u^k, \lambda^k)}{\partial \lambda} \end{pmatrix}$$

with

$$\frac{\partial \varphi_\eta(u, \lambda)}{\partial u} = \begin{cases} \eta & \text{, if } \lambda = u - g = 0 \\ \eta \left(1 - \frac{u-g}{\sqrt{\lambda^2+(u-g)^2}} \right) + (1-\eta)\lambda & \text{, if } \lambda > 0 \text{ and } u > g \\ \eta \left(1 - \frac{u-g}{\sqrt{\lambda^2+(u-g)^2}} \right) & \text{, otherwise} \end{cases}$$

for the coefficients of u if they are associated with the contact boundary $\Gamma_C \times \Omega$ and simply zero for all other coefficients and

$$\frac{\partial \varphi_\eta(u, \lambda)}{\partial \lambda} = \begin{cases} \eta & \text{, if } \lambda = u - g = 0 \\ \eta \left(1 - \frac{\lambda}{\sqrt{\lambda^2+(u-g)^2}} \right) + (1-\eta)(u-g) & \text{, if } \lambda > 0 \text{ and } u > g \\ \eta \left(1 - \frac{\lambda}{\sqrt{\lambda^2+(u-g)^2}} \right) & \text{, otherwise.} \end{cases}$$

Other subdifferentials can be chosen but they only differ in the points where φ_η is not classically differentiable. Using this subdifferential the computation of $\nabla \Psi(u^k, \lambda^k) = H_k^T F(u^k, \lambda^k)$ is straightforward [22].

Theorem 5.8. *The reduced semi-smooth Newton algorithm* (5.42) *converges locally Q-quadratic.*

$$\left(u^{k+1}, \lambda^{k+1} \right)^T = \left(u^k, \lambda^k \right)^T - H_k^{-1} F(u^k, \lambda^k) \tag{5.42}$$

with H_k a Clarke subdifferential of F at $\left(u^k, \lambda^k \right)^T$.

Proof. The assertion follows from Theorem 2.1 if F is Lipschitzian, strongly semi-smooth, CD-regular and a solution for $F(u, \lambda) = 0$ exists. The existence of a solution follows directly from Lemma 5.12 in conjunction with Theorem 5.5. The first part of F is linear and the second part is strongly semi-smooth by Lemma 5.13. Therewith, simple algebra yields that F is strongly semi-smooth and also Lipschitzian. To show the CD-regularity, first rearrange the degrees of freedom such that the ones associated with $\Gamma_C \times \Theta$ are the first and grouped together to the set $\mathcal{C}$, the remaining dofs are grouped to $\mathcal{N}$, i.e.

$$A := \begin{pmatrix} A_\mathcal{C} & A_{\mathcal{CN}} \\ A_{\mathcal{NC}} & A_\mathcal{N} \end{pmatrix}, \qquad D := \begin{pmatrix} D_\mathcal{C} \\ 0 \end{pmatrix}.$$

Since F is Lipschitzian everywhere, the set of Clark subdifferentials can be overestimated by Lemma 2.3. Together with the rearranging of the dofs there holds

$$\partial_C F \subseteq \begin{pmatrix} A_\mathcal{C} & A_{\mathcal{CN}} & -D_\mathcal{C} \\ A_{\mathcal{NC}} & A_\mathcal{N} & 0 \\ D_a & 0 & D_b \end{pmatrix} =: E \tag{5.43}$$

where D_a and D_b are semi-positive definite diagonal matrices such that $D_a + D_b$ is positive definite. If all realizations of E are invertible, then F must be CD-regular. Since A is invertible, the Schur complement

$$S = D_b + (D_a, 0)A^{-1}D$$

of E can be computed. Assume that S has a zero eigenvalue, i.e. there exists a vector $0 \neq q \in \mathbb{R}^n$ such that $Sq = 0$. Hence, $D_b q = -(D_a, 0)A^{-1}Dq$ which can be written componentwise as $(D_b)_{ii}q_i = -(D_a)_{ii}(A^{-1}Dq)_i$ and can be simplified to

$$q_i = -\frac{(D_a)_{ii}}{(D_b)_{ii}}(A^{-1}Dq)_i \quad i \in I_+ := \{i \in \{1, \ldots, n\} : (D_b)_{ii} > 0\} \tag{5.44}$$

$$0 = (A^{-1}Dq)_i \qquad i \in I_+^C := \{1, \ldots, n\} \setminus I_+ = \{i \in \{1, \ldots, n\} : (D_b)_{ii} = 0\}. \tag{5.45}$$

Multiplying (5.44) with q_i yields $0 \leq q_i^2 = -\frac{(D_a)_{ii}}{(D_b)_{ii}}q_i(A^{-1}Dq)_i$ and with $D_{ii} > 0$ from Lemma 5.13 it becomes $D_{ii}q_i(A^{-1}Dq)_i \leq 0$. Equation (5.45) directly yields $D_{ii}q_i(A^{-1}Dq)_i = 0$. Summing over all i yields

$$\sum_{i=1}^{n} D_{ii}q_i(A^{-1}Dq)_i = \sum_{i=1}^{m} ((D_{\mathcal{C}}, 0)q)_i (A^{-1}Dq)_i = (Dq)^T A^{-1}(Dq) \leq 0.$$

Since $D_{\mathcal{C}}$ is a positive definite diagonal matrix, this contradicts A, and therefore, A^{-1} only having positive eigenvalues. Hence, S and therewith E are invertible which completes the proof. $\qquad \square$

In practical applications the dimension N of the stochastic domain Θ may be very large, e.g. $4 \leq N \leq 10$. Hence, the size of the algebraic system (5.40) grows rapidly and solving it exactly becomes impracticable. Therefore, the SSN has been replaced by an inexact SSN method in which (5.40) is solved only inexactly and iteratively such that

$$\left\| H_k d^k + F(u^k, \lambda^k) \right\| \leq \min \left\{ \frac{10^{-3}}{k}, \left\| F(u^k, \lambda^k) \right\| \right\} \left\| F(u^k, \lambda^k) \right\|. \tag{5.46}$$

Its heuristic idea is to find only an approximation of the Newton's search direction, with the approximation error decreasing sufficiently fast as $x^k \to x^\star$. This is applicable since in neither case the search direction is globally correct and must always be corrected in the next Newton step. Such inexact SSN methods still converge Q-quadratic, c.f. [40] and the references therein. Elementary to any modern iterative solvers, e.g. GMRES, is a fast matrix-vector multiplication and an efficient preconditioner. Since the discretization is based on the tensor product structure of the stochastic domain in itself and to the space domain, the global system matrix can be written as $A = M_{st} \otimes S_h$ with $M_{st} = \bigotimes_{n=1}^{N} M_{st,n}$. Here S_h is the matrix representation of the discrete Steklov-Poincaré operator and $M_{st,n}$ is the mass matrix to the 1d-domain Θ_n. This representation has two advantages. Firstly, the system matrix does not need to be assembled which significantly reduces the memory storage requirement since S_h is a dense BEM

matrix and M_{st} is a FEM mass matrix of a high dimensional domain. Secondly, the matrix-vector multiplication can exploit the tensor structure, which is faster than to assemble the system matrix and carry out a sparse matrix-vector multiplication. The exact inverse of H is due to its 2×2-Block structure

$$H^{-1} = \begin{pmatrix} A & -D \\ D_a & D_b \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} - A^{-1}DS_A^{-1}D_aA^{-1} & A^{-1}DS_A^{-1} \\ -S_A^{-1}D_aA^{-1} & S_A^{-1} \end{pmatrix}$$

with the Schur complement $S_A = D_b + D_a A^{-1}D$ and $A^{-1} = (M_{st} \otimes S_h)^{-1} = M_{st}^{-1} \otimes S_h^{-1}$ with $M_{st}^{-1} = \bigotimes_{n=1}^{N} M_{st,n}^{-1}$. The matrices D_a and D_b change in every Newton step and therefore H^{-1} changes as well. Furthermore D_b is in general not invertible and the influences of D_a and D are not negligible. Consequently, the computation or at least approximation of both S_A and S_A^{-1} is very cost intensive. The construction of efficient preconditioners for differential equations with algebraic constraints resulted from subdifferentials is still an open and important question. In the experiments S_A is "approximated" by the identity. For the lowest order uniform h-version with 2048 dof in the primal variable and 528 dof in the dual variable the preconditioning has two important effects. First, the condition number is reduced from 2000 to 50 in the eights step as displayed in Figure 5.1 which leads to fewer GMRES iterations per SSN step. Second, the preconditioned cases requires two SSN steps less than the non preconditioned case. Table 5.1 shows that the GMRES iterations steadily increases with the number of SSN iterations. This is a result from the reduction of the tolerance (5.46), to which the system of linear equations is solved, in every inexact SSN step. For the preconditioned case the number of GMRES iterations is very small due the small condition number. For the non preconditioned case, they explode as the condition number tends to 2700. In the tenth step, the GMRES is even restarted after 50 iterations. A further consequence of the preconditioning is that the system of linear equations is solved much more accurate than the tolerance requires without additional GMRES iterations. Therefore, the inexact search direction is much closer to the exact search direction than in the non preconditioned case. Consequently, the effect of additional SSN iterations in the inexact variant with preconditioning is much smaller than without preconditioning.

5.3 A Posteriori Error Estimates

The SSN method allows to solve the discrete problem very efficiently for a given discretization. But at least as important is the construction of well adapted meshes to reduce the discretization error at optimal rate. Since the solution typically exhibits reduced regularity across the a priori unknown free boundary, local a posteriori error estimators which steer an adaptive mesh refinement are inevitable to obtain the optimal convergence rates.

In this section a residual type a posteriori error estimator for the mixed formulation (5.28) is derived and by Theorem 5.6 also for the variational inequality (5.36). The ideas are based on the work of Braess [8] to split discretization and consistency error
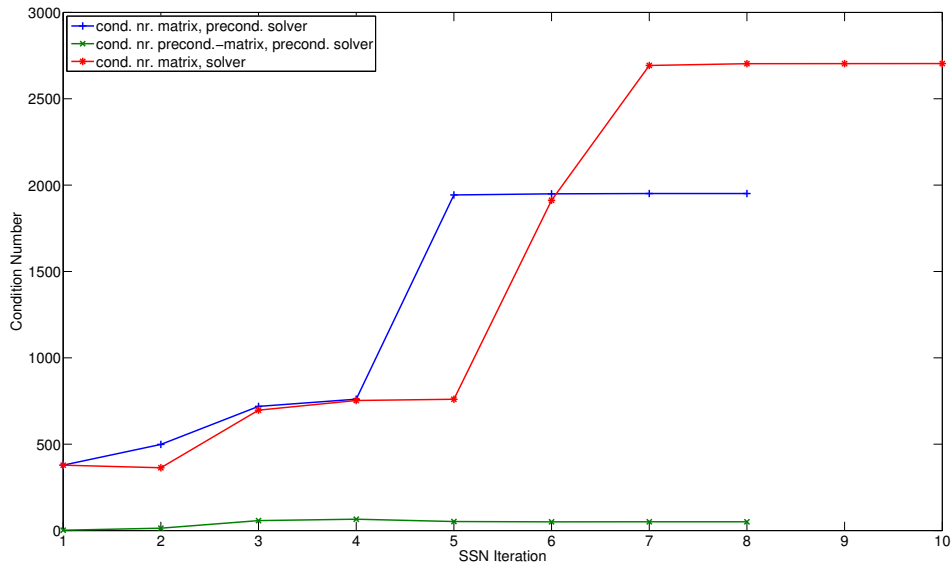
Figure 5.1: Condition number of system matrix per SSN step, $\mathrm{dof}(u) = 2048$, $\mathrm{dof}(\lambda) = 528$

GMRES Iter	
Preconditioned	Not Preconditioned
2	1
5	5
7	10
5	10
8	14
7	29
7	10
10	44
	46
	50+38

Table 5.1: GMRES iterations per SSN step, $\mathrm{dof}(u) = 2048$, $\mathrm{dof}(\lambda) = 528$

and on Carstensen [10] for the discretization error of a Neumann problem. However, here a general obstacle χ is considered. First, an auxiliary problem with given Lagrange multiplier λ_h is introduced. That is, find $z \in L^2_\rho(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$ such that

$$\tilde{A}(z, v) = \tilde{L}(v) + \int_\Theta \langle v, \lambda_h \rangle_{\Gamma_C} \rho \, dy \quad \forall \, v \in L^2_\rho(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)). \tag{5.47}$$

The following lemma separates the error introduced by the discretization of a variational formulation and by the presence of contact. It works for methods which are continuous and coercive in the same norm and conforming in the primal variable.

Lemma 5.14. *Let (u, λ), (u_h, λ_h) and z be the solution of (5.21), (5.28), (5.47) respectively. Then there exists constants ϵ_1, ϵ_2, C_A, $\alpha > 0$ independent of h, p, k and q such that*

$$-\epsilon_2 \|\lambda_h - \lambda\|^2_{L^2_\rho(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} + (\alpha - \epsilon_1) \|u - u_h\|^2_{L^2_\rho(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} \leq \frac{1}{4\epsilon_2} \left\| (u_h - \chi)^- \right\|^2_{L^2_\rho(\Theta; H^{\frac{1}{2}}(\Gamma_C))}$$

$$+ \frac{1}{4\epsilon_1} \left(C_{\tilde{A}} \|u_h - z\|_{L^2_\rho(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} + \left\| \lambda_h - \lambda_h^+ \right\|_{L^2_\rho(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} \right)^2 + \int_\Theta \langle \lambda_h^+, (u_h - \chi)^+ \rangle_{\Gamma_C} \rho \, dy \tag{5.48}$$

with $v^+ := \max\{v, 0\}$ and $v^- := \min\{v, 0\}$, i.e. $v = v^+ + v^-$.

Proof. First note that for $u_h \in V_{hp} \otimes W_{kq} \subset L^2_\rho(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$

$$\tilde{A}(z - u, u_h - u) = \int_\Theta \langle \lambda_h - \lambda, u_h - u \rangle_{\Gamma_C} \rho \, dy$$

$$= \int_\Theta \langle \lambda_h^+ - \lambda, u_h - u \rangle_{\Gamma_C} \rho \, dy + \int_\Theta \langle \lambda_h^-, u_h - u \rangle_{\Gamma_C} \rho \, dy.$$

The first of the two terms can be estimated by

$$\int_\Theta \langle \lambda_h^+ - \lambda, u_h - \chi + \chi - u \rangle_{\Gamma_C} \rho \, dy = \int_\Theta \langle \lambda_h^+ - \lambda, u_h - \chi \rangle_{\Gamma_C} \rho \, dy + \underbrace{\int_\Theta \langle \lambda, u - \chi \rangle_{\Gamma_C} \rho \, dy}_{=0 \ (5.21\text{b})}$$

$$- \underbrace{\int_\Theta \langle \lambda_h^+, u - \chi \rangle_{\Gamma_C} \rho \, dy}_{\leq 0 \text{ by construction}}$$

$$\leq \int_\Theta \langle \lambda_h^+ - \lambda, u_h - \chi \rangle_{\Gamma_C} \rho \, dy.$$

Using the linearity of the duality pairing and the definition of v^+, v^- and $v = v^+ + v^-$ yields

$$\int_\Theta \langle \lambda_h^+ - \lambda, u_h - \chi \rangle_{\Gamma_C} \rho \, dy = \int_\Theta \langle \lambda_h^+, u_h - \chi \rangle_{\Gamma_C} + \langle -\lambda, (u_h - \chi)^+ + (u_h - \chi)^- \rangle_{\Gamma_C} \rho \, dy$$

$$\leq \int_\Theta \langle \lambda_h^+, u_h - \chi \rangle_{\Gamma_C} \rho \, dy + \int_\Theta \langle \lambda_h - \lambda - \lambda_h^+ - \lambda_h^-, (u_h - \chi)^- \rangle_{\Gamma_C} \rho \, dy$$

$$= \int_\Theta \langle \lambda_h^+, (u_h - \chi)^+ \rangle_{\Gamma_C} \rho \, dy + \int_\Theta \langle \lambda_h - \lambda, (u_h - \chi)^- \rangle_{\Gamma_C} \rho \, dy - \int_\Theta \langle \lambda_h^-, (u_h - \chi)^- \rangle_{\Gamma_C} \rho \, dy$$

$$\leq \int_\Theta \langle \lambda_h^+, (u_h - \chi)^+ \rangle_{\Gamma_C} \rho \, dy + \int_\Theta \langle \lambda_h - \lambda, (u_h - \chi)^- \rangle_{\Gamma_C} \rho \, dy$$

and, hence, with Cauchy-Schwarz

$$\tilde{A}(z-u, u_h-u) \le \int_\Theta \langle \lambda_h^+, (u_h - \chi)^+ \rangle_{\Gamma_C} \rho \, dy + \int_\Theta \langle \lambda_h - \lambda, (u_h - \chi)^- \rangle_{\Gamma_C} \rho \, dy$$

$$+ \int_\Theta \langle \lambda_h^-, u_h - u \rangle_{\Gamma_C} \rho \, dy$$

$$\le \int_\Theta \langle \lambda_h^+, (u_h-\chi)^+ \rangle_{\Gamma_C} \rho \, dy + \left\| \lambda_h^- \right\|_{L_\rho^2(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} \left\| u_h - u \right\|_{L_\rho^2(\Theta; H^{\frac{1}{2}}(\Gamma_\Sigma))}$$

$$+ \left\| \lambda_h - \lambda \right\|_{L_\rho^2(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} \left\| (u_h - \chi)^- \right\|_{L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_C))}.$$

Consequently, using the ellipticity and continuity of the bilinear form $\tilde{A}$

$$\alpha \left\| u_h - u \right\|_{L_\rho^2(\Theta; H^{\frac{1}{2}}(\Gamma_\Sigma))}^2 \le \tilde{A}(u_h - u, u_h - u) = \tilde{A}(u_h - z, u_h - u) + \tilde{A}(z - u, u_h - u)$$

$$\le C_{\tilde{A}} \left\| u_h - z \right\|_{L_\rho^2(\Theta; H^{\frac{1}{2}}(\Gamma_\Sigma))} \left\| u_h - u \right\|_{L_\rho^2(\Theta; H^{\frac{1}{2}}(\Gamma_\Sigma))} + \tilde{A}(z - u, u_h - u)$$

$$\le \left[C_{\tilde{A}} \left\| u_h - z \right\|_{L_\rho^2(\Theta; H^{\frac{1}{2}}(\Gamma_\Sigma))} + \left\| \lambda_h^- \right\|_{L_\rho^2(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} \right] \left\| u_h - u \right\|_{L_\rho^2(\Theta; H^{\frac{1}{2}}(\Gamma_\Sigma))}$$

$$+ \int_\Theta \langle \lambda_h^+, (u_h - \chi)^+ \rangle_{\Gamma_C} \rho \, dy$$

$$+ \left\| \lambda_h - \lambda \right\|_{L_\rho^2(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} \left\| (u_h - \chi)^- \right\|_{L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_C))}.$$

Finally, employing Young's inequality twice yields the assertion. $\qquad \square$

Since the factor in front of the Lagrange multiplier error is negative, the left hand side is not a norm. This is improved by the following lemma. To obtain an error estimate in which the Lagrange multiplier error is also estimated let $w \in L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$ be the solution of

$$\tilde{A}(w, v) = \int_\Theta \langle \mu, v \rangle_{\Gamma_C} \rho \, dy \quad \forall \, v \in L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)). \tag{5.49}$$

From the continuous inf-sup condition (Lemma 5.6) and the continuity of $\tilde{A}$ follows

$$\beta \left\| \mu \right\|_{L_\rho^2(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} \le \sup_{0 \ne v \in L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} \frac{\int_\Theta \langle \mu, v \rangle_{\Gamma_C} \rho \, dy}{\left\| v \right\|_{L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}}$$

$$= \sup_{0 \ne v \in L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} \frac{\tilde{A}(w, v)}{\left\| v \right\|_{L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}} \le C_A \left\| w \right\|_{L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}. \tag{5.50}$$

Lemma 5.15. *Under the same conditions and notations as in Lemma 5.14 there holds for the total error*

$$(C_2 - \epsilon_2) \left\| \lambda_h - \lambda \right\|_{L_\rho^2(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))}^2 + C_1 \left\| u - u_h \right\|_{L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}^2 \le \frac{2 C_2 C_A^2}{\beta^2} \left\| z - u_h \right\|_{L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}^2$$

$$+ \frac{1}{4\epsilon_2} \left\| (u_h - \chi)^- \right\|_{L_\rho^2(\Theta; H^{\frac{1}{2}}(\Gamma_C))}^2 + \int_\Theta \langle \lambda_h^+, (u_h - \chi)^+ \rangle_{\Gamma_C} \rho \, dy$$

$$+ \frac{1}{4\epsilon_1} \left(C_{\tilde{A}} \left\| u_h - z \right\|_{L_\rho^2(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} + \left\| \lambda_h - \lambda_h^+ \right\|_{L_\rho^2(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} \right)^2 \tag{5.51}$$

5 Elliptic Stochastic Contact Problems

with $C_2 - \epsilon_2$, C_1, ϵ_1, $\epsilon_2 > 0$ and $C_1 + \frac{2C_2 C_A^2}{\beta^2} = \alpha - \epsilon_1$.

Proof. From (5.50), triangle inequality and $(a+b)^2 \le 2a^2 + 2b^2$ follows immediately

$$
\|\lambda_h - \lambda\|^2_{L^2_\rho(\Theta;\tilde{H}^{-\frac{1}{2}}(\Gamma_C))} \le \frac{C_A^2}{\beta^2} \|u - z\|^2_{L^2_\rho(\Theta;\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}
$$

$$
\le \frac{2C_A^2}{\beta^2} \|u - u_h\|^2_{L^2_\rho(\Theta;\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} + \frac{2C_A^2}{\beta^2} \|z - u_h\|^2_{L^2_\rho(\Theta;\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}.
$$

Consequently there holds for $C_2 > 0$

$$
(C_2 - \epsilon_2) \|\lambda_h - \lambda\|^2_{L^2_\rho(\Theta;\tilde{H}^{-\frac{1}{2}}(\Gamma_C))} + C_1 \|u - u_h\|^2_{L^2_\rho(\Theta;\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} \le \frac{2C_2 C_A^2}{\beta^2} \|z - u_h\|^2_{L^2_\rho(\Theta;\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}
$$

$$
- \epsilon_2 \|\lambda_h - \lambda\|^2_{L^2_\rho(\Theta;\tilde{H}^{-\frac{1}{2}}(\Gamma_C))} + (C_1 + \frac{2C_2 C_A^2}{\beta^2}) \|u - u_h\|^2_{L^2_\rho(\Theta;\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}
$$

If $C_2 - \epsilon_2 > 0$ and $C_1 > 0$ then the left hand side is a norm of the total error. If further $\epsilon_1, \epsilon_2 > 0$ and $C_1 + \frac{2C_2 C_A^2}{\beta^2} = \alpha - \epsilon_1$, then the last two terms on the right hand side are bounded by Lemma 5.14. A feasible set of constants is $C_1 = \frac{\alpha}{2}$, $\epsilon_1 = \frac{\alpha}{4}$, $C_2 = \frac{\alpha \beta^2}{8 C_A^2}$ and $\epsilon_2 = \frac{\alpha \beta^2}{16 C_A^2}$. $\quad\square$

It remains to derive an error estimation for the discretization error $\|u_h - z\|$. For that, additional notations, taken from Carstensen [10], are introduced.

$$
\psi = V^{-1}(K - \frac{1}{2})z
$$

$$
\psi_h^* := V^{-1}(K - \frac{1}{2})u_h
$$

$$
\psi_h := i_{hp} V_{hp}^{-1} i_{hp}^* (K - \frac{1}{2})u_h
$$

$$
S_h = W + (K' - \frac{1}{2}) i_{hp} V_{hp}^{-1} i_{hp}^* (K - \frac{1}{2})
$$

with the canonical embedding

$$
i_{hp}: \; V_{hp}^D \hookrightarrow H^{-\frac{1}{2}}(\Gamma_\Sigma) \tag{5.52}
$$

and its dual i_{hp}^*.

Lemma 5.16 (Galerkin Orthogonality). *For $v_h \in V_{hp} \otimes W_{kq}$ there holds*

$$
\int_\Theta \langle Sz - S_h u_h, v_h \rangle_{\Gamma_\Sigma} \rho \, dy = 0 \quad \forall v_h \in V_{hp} \otimes W_{kq}. \tag{5.53}
$$

Proof. Follows immediately from $V_{hp} \otimes W_{kq} \subset L^2_\rho(\Theta, \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$, (5.28) and (5.47). $\quad\square$

Lemma 5.17. *Let (u_h, λ_h) and z solve (5.28), (5.47) and $t \in L^2_\rho(\Theta; L^2(\Gamma_N))$. Then there holds the a posteriori estimate*

$$\|u_h - z\|^2_{L^2_\rho(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} + \|\psi_h - \psi\|^2_{L^2_\rho(\Theta; H^{-\frac{1}{2}}(\Gamma_\Sigma))} \le C \sum_{I \subset \mathcal{T}_{kq}} \sum_{E \subset \mathcal{T}_{hp}} \eta^2_h(I, E) \qquad (5.54)$$

with the for every space element E and stochastic element I given local indicator

$$\eta^2_h(I, E) := (1 + \frac{h_E}{p_E}) \|t - S_h u_h\|^2_{L^2_\rho(I; L^2(E \cap \Gamma_N))} + (1 + \frac{h_E}{p_E}) \|\lambda_h - S_h u_h\|^2_{L^2_\rho(I; L^2(E \cap \Gamma_C))}$$

$$+ h_E \left\| \frac{\partial}{\partial s}(V\psi_h - (K - \frac{1}{2})u_h) \right\|^2_{L^2_\rho(I; L^2(E))}.$$

Proof. Following the same arguments as in [10, Proposition 5.1] yields

$$C \left(\|u_h - z\|^2_{L^2_\rho(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} + \|\psi_h - \psi\|^2_{L^2_\rho(\Theta; H^{-\frac{1}{2}}(\Gamma_\Sigma))} \right)$$

$$\le \int_\Theta \left(\langle W(z - u_h), z - u_h \rangle_{\Gamma_\Sigma} + \langle V(\psi - \psi_h), \psi - \psi_h \rangle_{\Gamma_\Sigma} \right) \rho \, dy$$

$$= \int_\Theta \left(\langle Sz - S_h u_h, z - u_h \rangle_{\Gamma_\Sigma} + \langle V(\psi_h^* - \psi_h), \psi - \psi_h \rangle_{\Gamma_\Sigma} \right) \rho \, dy$$

$$=: A_1 + A_2.$$

The first term A_1 can be estimated using the Galerkin orthogonality of Lemma 5.16

$$A_1 := \int_\Theta \langle Sz - S_h u_h, z - u_h \rangle_{\Gamma_\Sigma} \rho \, dy = \int_\Theta \langle Sz - S_h u_h, z - u_h \rangle_{\Gamma_\Sigma} + \langle Sz - S_h u_h, u_h - v_h \rangle_{\Gamma_\Sigma} \rho \, dy$$

$$= \int_\Theta \langle Sz - S_h u_h, z - v_h \rangle_{\Gamma_\Sigma} \rho \, dy.$$

With $z - v_h \in L^2_\rho(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$ the equation (5.47) yields

$$\int_\Theta \langle Sz, z - v_h \rangle_{\Gamma_\Sigma} \rho \, dy = \int_\Theta \langle t, z - v_h \rangle_{\Gamma_N} + \langle \lambda_h, z - v_h \rangle_{\Gamma_C} \rho \, dy$$

and therewith

$$A_1 = \int_\Theta \langle t - S_h u_h, z - v_h \rangle_{\Gamma_N} \rho \, dy + \int_\Theta \langle \lambda_h - S_h u_h, z - v_h \rangle_{\Gamma_C} \rho \, dy$$

$$\le \sum_{I \subset \mathcal{T}_{kq}} \sum_{E \subset \mathcal{T}_{hq} \cap \Gamma_N} \|t - S_h u_h\|_{L^2_\rho(I; L^2(E))} \|z - v_h\|_{L^2_\rho(I; L^2(E))}$$

$$+ \sum_{I \subset \mathcal{T}_{kq}} \sum_{E \subset \mathcal{T}_{hq} \cap \Gamma_C} \|\lambda_h - S_h u_h\|_{L^2_\rho(I; L^2(E))} \|z - v_h\|_{L^2_\rho(I; L^2(E))}.$$

Let $\mathcal{P}_k$ be the standard L^2_ρ-projection onto W_{kq}, i.e. $\mathcal{P}_k : L^2_\rho(\Theta) \to W_{kq}$ such that

$$\int_\Theta (\mathcal{P}_k w - w) v \rho \, dy = 0 \quad \forall v \in W_{kq}.$$

In particular $\mathcal{P}_k$ is L_ρ^2-stable. Further, let i_h be the hp-Clément interpolation operator mapping onto V_{hp}, c.f. [62]. Then there holds

$$\|w - \mathcal{P}_k i_h w\|_{L_\rho^2(\Theta;L^2(E))} \leq \|w - i_h w\|_{L_\rho^2(\Theta;L^2(E))} + \|i_h w - \mathcal{P}_k i_h w\|_{L_\rho^2(\Theta;L^2(E))}$$

$$\leq C \frac{h_E}{p_E} \|w\|_{L_\rho^2(\Theta;H^1(\omega(E)))} + \|i_h w\|_{L_\rho^2(\Theta;L^2(E))}$$

$$\leq C \left(1 + \frac{h_E}{p_E}\right) \|w\|_{L_\rho^2(\Theta;H^1(\omega(E)))}$$

with $\omega(E)$ a net around E. By the L^2-stability and complex interpolation between L^2 and H^1 there holds

$$\|w - \mathcal{P}_k i_h w\|_{L_\rho^2(\Theta;L^2(E))} \leq C \left(1 + \frac{h_E}{p_E}\right)^{\frac{1}{2}} \|w\|_{L_\rho^2(\Theta;H^{\frac{1}{2}}(\omega(E)))}$$

Choosing $v_h = u_h + \mathcal{P}_k i_h(z - u_h)$ completes the estimate of A_1.

Since $u_h(y) \in V_{hp} \subset H_0^1(\Gamma_\Sigma)$ and $\psi_h \in V_{hp}^D \subset L^2(\Gamma_\Sigma)$ the mapping properties of V and K (cf. Lemma 2.1) yield

$$V(\psi_h - \psi_h^*) = V\psi_h - (K - \frac{1}{2})u_h \in H^1(\Gamma_\Sigma) \subset C^0(\Gamma_\Sigma).$$

Furthermore, $V(\psi_h - \psi_h^*)$ is orthogonal in $L^2(\Gamma_\Sigma)$ to V_{hp}^D, [15, Lemma 3.2.7]. Hence, for the characteristic function $\chi_E \in V_{hp}^D$ of an element $E \in \mathcal{T}_{hp}$ there holds

$$0 = \langle V(\psi_h - \psi_h^*), \chi_E \rangle_{\Gamma_\Sigma} = \int_E V(\psi_h - \psi_h^*) \, ds$$

and, therefore, the continuous function $V(\psi_h - \psi_h^*)$ has a zero on each boundary segment E. Since $V(\psi_h - \psi_h^*) \in H^1(\Gamma_\Sigma)$, the application of [10, Theorem 5.1] yields

$$A_2 := \int_\Theta \langle V(\psi_h - \psi_h^*), \psi - \psi \rangle_{\Gamma_\Sigma} \rho \, dy \leq \|V(\psi_h - \psi_h^*)\|_{L_\rho^2(\Theta;H^{\frac{1}{2}}(\Gamma_\Sigma))} \|\psi_h - \psi\|_{L_\rho^2(\Theta;H^{-\frac{1}{2}}(\Gamma_\Sigma))}$$

$$\leq \tilde{C} \left(\int_\Theta \sum_{E \subset \mathcal{T}_{hq}} h_E \left\| \frac{\partial}{\partial s}(V(\psi_h - \psi_h^*)) \right\|_{L^2(E)}^2 \rho \, dy \right)^{\frac{1}{2}} \|\psi_h - \psi\|_{L_\rho^2(\Theta;H^{-\frac{1}{2}}(\Gamma_\Sigma))}.$$

A combination of the upper results with the coercivity of W and V yields

$$C \left(\|u_h - z\|_{L_\rho^2(\Theta;\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}^2 + \|\psi_h - \psi\|_{L_\rho^2(\Theta;H^{-\frac{1}{2}}(\Gamma_\Sigma))}^2 \right)$$

$$\leq \sum_{I \subset \mathcal{T}_{kq}} \sum_{E \subset \mathcal{T}_{hq} \cap \Gamma_N} \left(1 + \frac{h_E}{p_E}\right)^{\frac{1}{2}} \|t - S_h u_h\|_{L_\rho^2(I;L^2(E))} \|z - u_h\|_{L_\rho^2(I;\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}$$

$$+ \sum_{I \subset \mathcal{T}_{kq}} \sum_{E \subset \mathcal{T}_{hq} \cap \Gamma_C} \left(1 + \frac{h_E}{p_E}\right)^{\frac{1}{2}} \|\lambda_h - S_h u_h\|_{L_\rho^2(I;L^2(E))} \|z - u_h\|_{L_\rho^2(I;\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))}$$

$$+ \tilde{C} \left(\int_\Theta \sum_{E \subset \mathcal{T}_{hq}} h_E \left\| \frac{\partial}{\partial s}(V(\psi_h - \psi_h^*)) \right\|_{L^2(E)}^2 \rho \, dy \right)^{\frac{1}{2}} \|\psi_h - \psi\|_{L_\rho^2(\Theta;H^{-\frac{1}{2}}(\Gamma_\Sigma))}.$$

The assertion follows with Young's inequality. $\qquad\square$

The assumption of $t \in L^2_\rho(\Theta; L^2(\Gamma_N))$ can be relaxed which leads to additional oscillation terms typically of higher order. Combining Lemma 5.15 and Lemma 5.17 yields the residual type error estimator.

Theorem 5.9. *Under the assumptions of Lemma 5.15 and Lemma 5.17 there exits a constant $C > 0$ such that*

$$C \left(\|\lambda_h - \lambda\|^2_{L^2_\rho(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))} + \|u - u_h\|^2_{L^2_\rho(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))} \right) \leq \sum_{I \subset \mathcal{T}_{kq}} \sum_{E \subset \mathcal{T}_{hp}} \eta^2_h(I, E)$$

$$+ \left\|(u_h - \chi)^-\right\|^2_{L^2_\rho(\Theta; H^{\frac{1}{2}}(\Gamma_C))} + \left\|\lambda_h - \lambda^+_h\right\|^2_{L^2_\rho(\Theta; \tilde{H}^{-\frac{1}{2}}(\Gamma_C))}$$

$$\tag{5.55}$$

$$+ \int_\Theta \left\langle \lambda^+_h, (u_h - \chi)^+ \right\rangle_{\Gamma_C} \rho \, dy.$$

The above error estimator is not local since the $H^{\frac{1}{2}}$ and the $\tilde{H}^{-\frac{1}{2}}$-norm are not local.

Remark 5.1. *The error estimator of Lemma 5.17 for the auxiliary problem is not suited for problems on open curves. On open curves, u lies in $\tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma)$ but in general not in $\tilde{H}^1(\Gamma_\Sigma)$, i.e. $Su \notin L^2(\Gamma_\Sigma)$. Although $u_h \in \tilde{H}^1(\Gamma_\Sigma)$ and therefore the residual $t(y) - S_h u_h(y)$ lies in $L^2(\Gamma_\Sigma)$, it does not lie in there uniformly. This problem does not appear on closed curves if the right hand side is sufficiently smooth.*

For the numerical experiments the $\tilde{H}^{-\frac{1}{2}}(\Gamma_C)$ is approximated by the $L^2(\Gamma_C)$-norm and v in the $H^{\frac{1}{2}}(\Gamma_C)$-norm by $\sqrt{\|v\|_{L^2(\Gamma_C)} \|v\|_{H^1(\Gamma_C)}}$ as in [70, Lemma IV.3]. For the local error indicator only the L^2-part of the $H^{\frac{1}{2}}(\Gamma_C)$-norm is used.

5.4 Numerical Experiments

For the numerical example, the space domain is $\Omega = [-0.5, 0.5]^2$ with $\Gamma_C = [-0.5, 0.5] \times \{-0.5\}$ and $\Gamma_N = \partial\Omega \backslash \Gamma_C$. $Y \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$ is uniformly distributed which corresponds to the density $\rho \equiv \frac{1}{2\sqrt{3}}$ and the stochastic domain $\Theta = [-\sqrt{3}, \sqrt{3}]$. For the first examples, the Neumann force is $t(x, y) = 1 + \cos(y)$ and the obstacle is $\chi(x, y) \equiv 1$. In particular, for the lowest order h-version ($p = 1$, $q = 0$) (5.36) is a conforming approach, i.e. $K_{hp,kq} \subset \tilde{K}$. Therefore, [59, Lemma 14] yields

$$\|u - u_h\|^2_{L^2_\rho(\Theta; H^{\frac{1}{2}}(\Gamma))} \leq C \left(\tilde{J}(u_h) - \tilde{J}(u) \right)$$

which allows to compute the approximation error numerically even in the absence of a known weak solution. The value $\tilde{J}(u) \approx 0.9158884$ has been obtained from Aitken extrapolation of a family of lowest order h-version. Figure 5.2 shows algebraic convergence with the estimated error being almost parallel to the error. This is surprising since the factor $(1 + h)$ in the residual error estimator of Lemma 5.17 is a non-optimal scaling
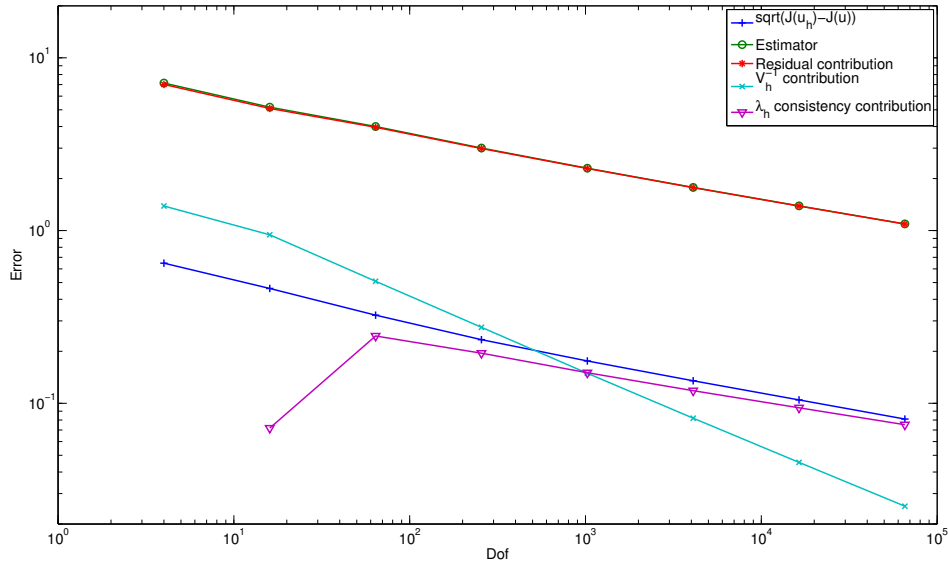
Figure 5.2: Error and error indicators for stochastic contact problem, conforming approach

factor. Furthermore, the h-scaling in front of the V^{-1}-approximation error contribution leads to a higher order convergence of this contribution in the error estimator compared to the consistency error of λ_h and the residual contribution, which is the dominating component of the error estimator. The consistency error decays at the same rate as the residual error contribution but is over one order of magnitude smaller. This may be further improved if the dual norm instead of the $L^2(\Gamma_C)$-norm would be used to compute this contribution. The remaining contributions are very small and displayed in Table 5.2.

For the second example, $t(x,y) = 0$, $\chi(x,y) = 1 - 3x_1$ and $\tilde{J}(u) \approx 0.3639245$. Again, the

dof	complementarity	obstacle in $H^{\frac{1}{2}}(\Gamma_C)$	obstacle in $L^2(\Gamma_C)$
4	0	$0.0013\cdot10^{-8}$	$0.0013\cdot10^{-8}$
16	0	$0.0013\cdot10^{-8}$	$0.0010\cdot10^{-8}$
64	0	$0.0016\cdot10^{-8}$	$0.0008\cdot10^{-8}$
256	$0.1292\cdot10^{-4}$	$0.0174\cdot10^{-8}$	$0.0046\cdot10^{-8}$
1024	$0.4662\cdot10^{-4}$	$0.3327\cdot10^{-8}$	$0.0533\cdot10^{-8}$
4096	$0.6974\cdot10^{-4}$	$0.6427\cdot10^{-8}$	$0.1367\cdot10^{-8}$
16384	$0.0762\cdot10^{-4}$	$0.0088\cdot10^{-8}$	$0.0008\cdot10^{-8}$
65536	$0.1567\cdot10^{-4}$	$0.0272\cdot10^{-8}$	$0.0021\cdot10^{-8}$

Table 5.2: Remaining error estimators for stochastic contact problem, conforming approach

lowest order h-version is conforming and the error decays at algebraic rate as displayed in Figure 5.3a. Now, the decay rate of the estimated error is less than of the error itself, which is expected from the non-optimal scaling factor $(1+h)$ in the residual estimator of Lemma 5.17. For the uniform p-version the method is no longer conforming. Therefore, the distance of the energy values is not an upper bound of the error in the energy norm since the consistency error is missing. This explains why the distance of the energy values decreases but not monotonically. The estimator is not able to capture the convergence at all since the residual in the $L^2(\Gamma)$-norm does not decrease. The reason is the missing p-scaling factor when going from the correct dual-norm for the residual to the $L^2(\Gamma)$-norm. For the adaptive h-version, both the error and the estimated error stagnate at a very high level. By Figure 5.3b the dominating contribution in the error estimator is again the residual of the auxiliary problem. Figure 5.4, which displays the adaptively generated meshes number six and 29, shows that the contact set is identified and that the meshes are symmetric to the x_2-axis as is expected from the x_2-symmetric obstacle. However, the meshes are extremely overrefined towards $\{0, -0.5\}$. In fact, the computed residual at the contact boundary is larger than at the Neumann boundary, with an error peak at tip of the obstacle at $\{0, -0.5\}$ where its kink is. In the two adjacent elements of this point the residual is given by $\lambda_h - S_h u_h$ and λ_h may not converge in the $L^2(\Gamma_C)$-norm but only in the dual norm due to the kink in the obstacle. Therefore, the residual is not uniformly in $L^2(\Gamma_C)$ and may not decay to zero in this norm. The hp-adaptive algorithm never did a p-refinement and is therefore identical to the h-adaptive method. Furthermore, the error contributions from the contact conditions defined in Lemma 5.15 decay at a sufficient rate or are several magnitudes smaller than the other contributions and, therefore, only the residual estimator for the auxiliary problem seems to be not suited for this kind of problems. Consequently, the separation of the error contributions should be kept but the error estimator for the auxiliary problem should be replaced.
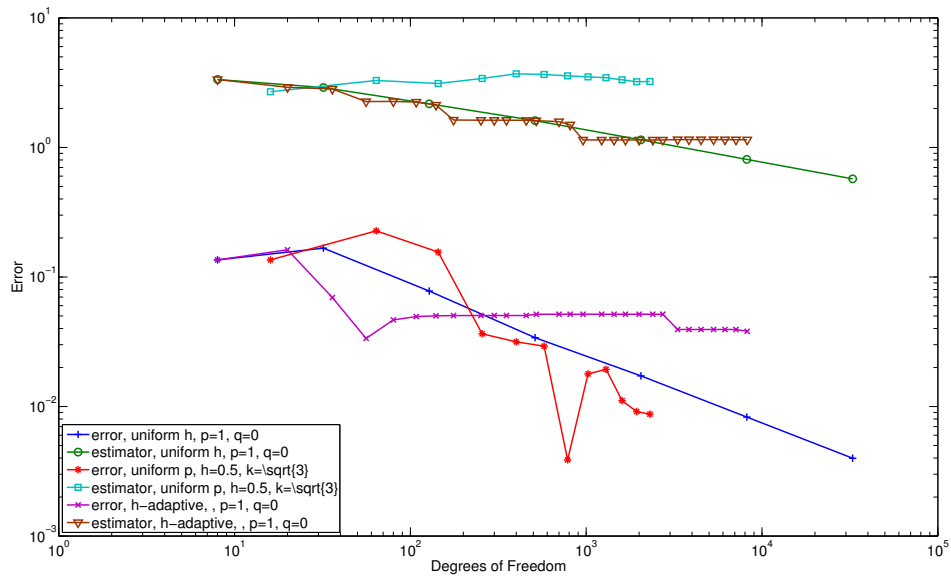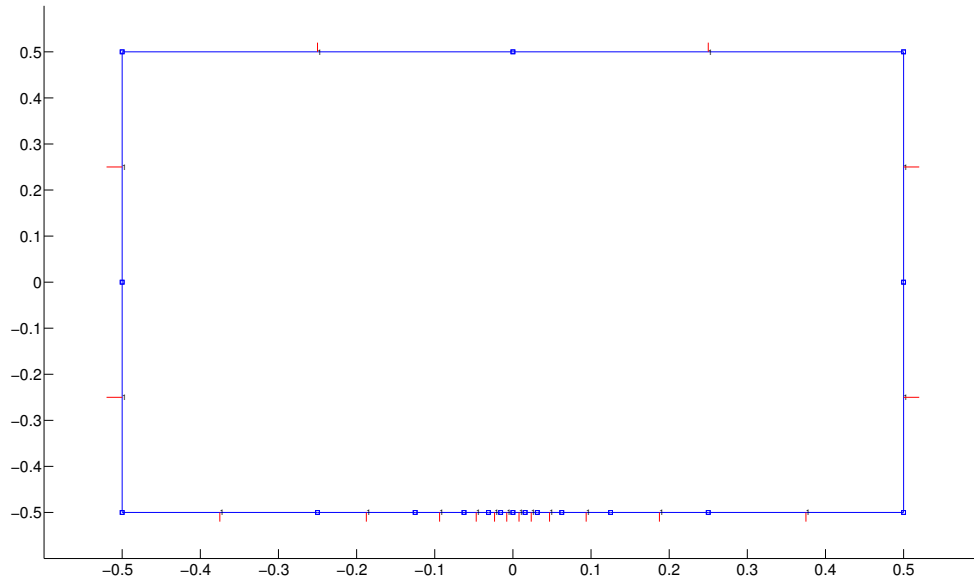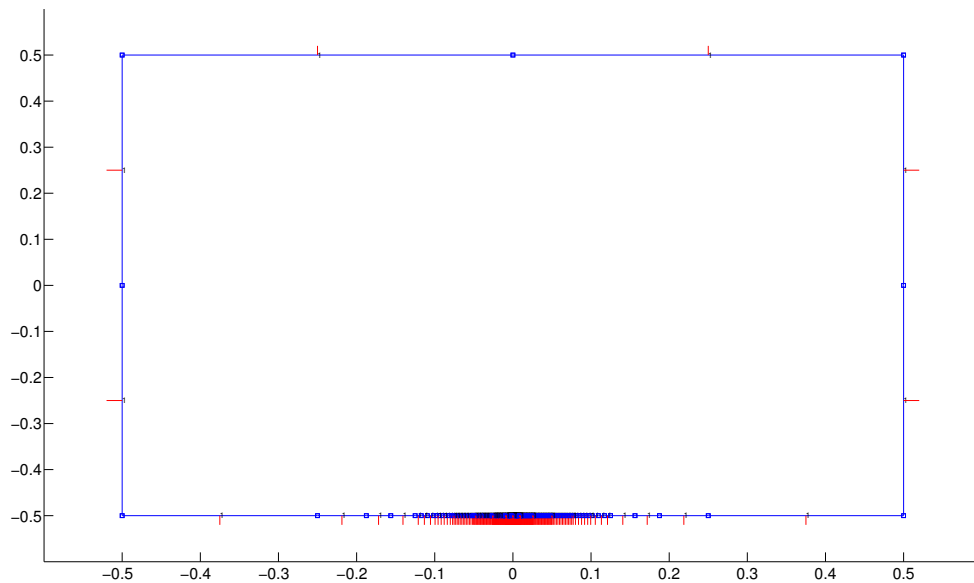
(a) Error and error estimator
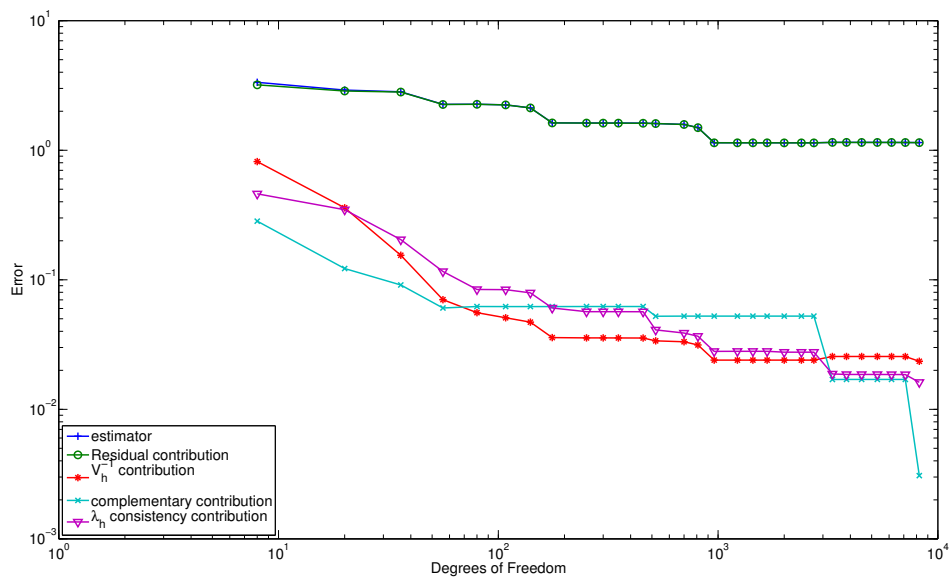
(b) Error estimator contributions, *h*-adaptive

Figure 5.3: Error and indicators for the second stochastic contact problem

(a) h-adaptive Mesh number 6

(b) h-adaptive Mesh number 29

Figure 5.4: Different meshes generated by the h-adaptive algorithm for the second stochastic contact problem

6 Hyperbolic Contact Problems with Friction

Many physical and technical applications deal with the elasto-dynamic, frictional contact of a body with a rigid foundation. From an economic's point of view, mathematical simulations of these problems are indispensable. However, the underlying hyperbolic differential operator is not symmetric and indefinite. Together with the unilateral contact and friction condition this significantly aggravates the mathematical analysis and the development of efficient algorithms.

6.1 Mixed Formulation for Hyperbolic Contact Problems with Tresca and Coulomb Friction

The hyperbolic contact problem (6.1) describes the elasto-dynamic frictional contact of an elastic body $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ over the time $I := (0, T]$ in the context of linear elasticity theory. The boundary $\overline{\partial \Omega} := \overline{\Gamma} = \overline{\Gamma_N} \cup \overline{\Gamma_D} \cup \overline{\Gamma_C}$ is decomposed into the disjoint Neumann, Dirichlet and contact parts. In particular, Γ_C is the boundary part which possibly comes into frictional contact with a smooth rigid foundation. For simplicity, $\overline{\Gamma_D} \cap \overline{\Gamma_C} = \emptyset$ is assumed whereas $\Gamma_D = \emptyset$ can be allowed.

$$\ddot{u} - \operatorname{div} \sigma(u) = f \qquad \text{in } \Omega \times I \qquad (6.1a)$$

$$\sigma(u) = \mathcal{A}\epsilon(u) \qquad \text{in } \Omega \times I \qquad (6.1b)$$

$$u = 0 \qquad \text{on } \Gamma_D \times I \qquad (6.1c)$$

$$\sigma(u)n = \hat{t} \qquad \text{on } \Gamma_N \times I \qquad (6.1d)$$

$$u_n \leq g, \ \sigma_n \leq 0, \ \sigma_n(u_n - g) = 0 \qquad \text{on } \Gamma_C \times I \qquad (6.1e)$$

$$|\sigma_t| \leq \mu_f, \ |\sigma_t| < \mu_f \Rightarrow \dot{u}_t = 0, \ |\sigma_t| = \mu_f \Rightarrow \exists\, \alpha \in \mathbb{R} : \sigma_t = \alpha^2 \dot{u}_t \qquad \text{on } \Gamma_C \times I \qquad (6.1f)$$

$$u(0) = u_0, \quad \dot{u}(0) = v_0 \qquad \text{in } \Omega. \qquad (6.1g)$$

Here, u, σ, ϵ, $\mathcal{A}$, f, $\hat{t}$, n, g and μ_f denote the displacement field, stress tensor, linearized strain tensor, elasticity tensor, volume force, surface force, outwards unit normal, non-negative gap function and the non-negative friction function, respectively. If $\mu_f = \mathcal{F}$, where $\mathcal{F}$ is the friction coefficient, the constraint (6.1f) models Tresca's friction law. This law allows non-zero tangential stress in the absence of contact which is undesirable but mathematically much easier than Coulomb's friction law, i.e. $\mu_f = \mathcal{F}|\sigma_n|$. To fully

describe the contact conditions a notion of the normal and tangential components of the displacement field and stress tensor is required. Let

$$u_n = u \cdot n, \quad u_t = u - u_n n, \quad \sigma_n = n \cdot \sigma(u) \cdot n \ \text{ and } \ \sigma_t = \sigma(u)n - \sigma_n n.$$

Up till now there exists no proof of existence for the unilateral elasto-dynamic contact problem. Cocou [18] and Cocou and Scarella [19] prove existence for dynamic contact problems assuming Kelvin-Voigt viscoelastic bodies. However, the elasto-dynamic contact problem cannot be viewed as a limit case of the viscoelastic problem, since in the limit the much needed boundedness of the velocity in the $H^1(\Omega)$-norm is lost. Hence, no weakly converging subsequence can be chosen. Other existence results hold if a normal compliance law is used as in the work of Martins and Oden [61] or frictionless unilateral boundary conditions for the wave equation [54, 44]. The two initial conditions in (6.1g) allow to rewrite (6.1) as a system of PDEs by introducing the new unknown $v := \dot{u}$ which represents the velocity field. Hence, (6.1a) and (6.1g) become

$$\dot{u} - v = 0 \qquad \text{in } \Omega \times I \tag{6.2a}$$

$$\dot{v} - \operatorname{div} \sigma(u) = f \qquad \text{in } \Omega \times I \tag{6.2b}$$

$$u(0) = u_0, \quad v(0) = v_0 \quad \text{in } \Omega. \tag{6.2c}$$

Even in the frictionless case, i.e. $\mu_f = 0$, a hp-FEM time discontinuous Galerkin (hp-TDG) variational inequality approach in which the normal pressure on Γ_C is eliminated yields a variational inequality with a non-symmetric, indefinite system matrix. To the best of the author's knowledge these can not yet be solved reliable even with projection-contraction methods. Therefore, a Lagrange multiplier $\lambda := -\sigma(u)n|_{\Gamma_C}$ representing the negative surface force on Γ_C is introduced. Similar to [7], let $X := H^1(I; L^2_{\Gamma_D}(\Omega)) \cap L^2(I; H^1_{\Gamma_D}(\Omega))$ with $H^1_{\Gamma_D}(\Omega)$ and $L^2_{\Gamma_D}(\Omega)$ the usual Sobolev spaces of vector valued functions vanishing on Γ_D. The space $H^1_{\Gamma_D}(\Omega)$ is equipped with the norm $||\chi||_1^2 := \sum_{i=1}^d \int_\Omega \chi_i^2 + (\nabla \chi_i)^2 \, dx$ and

$$H^1(I; L^2_{\Gamma_D}(\Omega)) := \left\{ v \in L^2(I; L^2_{\Gamma_D}(\Omega)) : \ \dot{v} \in L^2(I; L^2_{\Gamma_D}(\Omega)) \right\}.$$

Then, the mixed formulation is to find a triple $(u, v, \lambda) \in X \times X \times L(\lambda)$ satisfying

$$\int_I (\dot{u}, \varphi) - (v, \varphi) \, dt = 0 \qquad\qquad \forall \varphi \in X \tag{6.3a}$$

$$\int_I (\dot{v}, \chi) + a(u, \chi) + \langle \chi, \lambda \rangle_{\Gamma_C} \, dt = \int_I f(\chi) \, dt \qquad\qquad \forall \chi \in X \tag{6.3b}$$

$$\int_I \langle u_n, \mu_n - \lambda_n \rangle_{\Gamma_C} + \langle v_t, \mu_t - \lambda_t \rangle_{\Gamma_C} \, dt \leq \int_I \langle g, \mu_n - \lambda_n \rangle_{\Gamma_C} \, dt \quad \forall \mu \in L(\lambda) \tag{6.3c}$$

$$(u(x, 0), \chi) = (u_0, \chi), \quad (v(x, 0), \chi) = (v_0, \chi) \qquad\qquad \forall \chi \in L^2(\Omega). \tag{6.3d}$$

Here, $(\cdot, \cdot)$ and $\langle \cdot, \cdot \rangle_{\Gamma_C}$ denote the L^2-inner product over Ω and the duality pairing between the space $W := H^1_{\Gamma_D}(\Omega)|_{\Gamma_C}$ and its dual space $M := W'$, respectively. Therewith, the set of admissible Lagrange multipliers is

$$L(\lambda) := \left\{ \mu \in L^2(I; M) : \int_I \langle \mu, \eta \rangle_{\Gamma_C} \, dt \leq \int_I \langle \mu_f, \|\eta_t\| \rangle_{\Gamma_C} \, dt \ \forall \eta \in L^2(I; W) \text{ with } \eta_n \leq 0 \right\}.$$

$$\tag{6.4}$$

In the case of Tresca friction, i.e. $\mu_f = \mathcal{F}$, the Lagrange multiplier set is independent of λ. Furthermore, the bilinear form $a(\cdot, \cdot)$ and the linear form $f(\cdot)$ are given by

$$a(u, \chi) := \int_\Omega \mathcal{A}\epsilon(u) : \epsilon(\chi) \, dx \quad \text{and} \quad f(\chi) := \int_\Omega f\chi \, dx + \int_{\Gamma_N} \hat{t}\chi \, ds.$$

The weak formulation (6.3) can now be discretized in time by a time discontinuous Galerkin method. However, in the literature (see [12, 43, 53]) dynamic contact problems are often discretized in time by a finite difference scheme leading to a sequence of elliptic contact problems.

6.2 An hp-Time Discontinuous Galerkin Discretization

The common approach for simulations of dynamic contact problems is a finite difference scheme in time like Newmark. However, these lead to numerical instabilities. Khenous [43] explains this disadvantage by the fact that the nodes on the contact boundary have their own inertia which, even for energy conserving schemes, leads to instabilities. These instabilities result from the kinetic energy lost by a node being stopped on the contact boundary. Thus, energy conserving schemes make these nodes oscillate in order to keep their kinetic energy. Khenous proposes to redistribute the mass such that the contact nodes have no inertia whereas the crucial mass properties are maintained, i.e. conserving the total mass, the center of gravity and the inertia momentum. This method has been first introduced in [42] and a modified quadrature for a direct computation of this modified mass matrix is proposed in [32]. An analytic computation of the modification is derived in [2].

In the following an hp-time discontinuous Galerkin discretization is considered. Similar to the parabolic case let $\mathcal{T}_{hp}^s$ and $\mathcal{T}_{kq}^t$ be meshes enhanced with a polynomial degree distribution in space and in time, respectively. Recall that $\mathcal{T}_{kq}^t$ implies a decomposition of $I = (0, T]$ such that $\bar{I} = \bigcup_{n=1}^N \bar{I}_n$ with $I_n = (t_{n-1}, t_n)$ and the definitions for the one-sided limits and time jump

$$v_+^n := \lim_{0<s\to 0} v(t_n + s), \quad v_-^n := \lim_{0>s\to 0} v(t_n + s) \quad \text{and} \quad [v^n] := v_+^n - v_-^n.$$

Let

$$V_{hp} := \left\{ \chi \in H^1_{\Gamma_D}(\Omega) : \; \chi|_e \in \mathbb{P}_{pe}(e) \; \forall e \in \mathcal{T}_{hp}^s \right\} = \text{span}\left\{\phi_j\right\}_j,$$

$$P_{q_n}(I_n) := \left\{ \chi : I_n \to V_{hp} : \; \chi(t) = \sum_{i=0}^{q_n} \chi_i \vartheta_i(t), \; \chi_i \in V_{hp} \right\},$$

$$R_{q_n}(I_n) := \left\{ \chi : I_n \to W_{hp} : \; \chi(t) = \sum_{i=0}^{q_n} \chi_i \vartheta_i(t), \; \chi_i \in W_{hp} \right\} = P_{q_n}(I_n)|_{\Gamma_C \times I_n},$$

$$Q_{q_n}(I_n) := \left\{ \begin{array}{l} \mu : I_n \to M_{hp} : \; \mu(t) = \sum_{i=0}^{q_n} \mu_i \zeta_i(t), \; \mu_i \in M_{hp}, \\ \int_{I_n} \langle \mu, \eta \rangle_{\Gamma_C} \, dt \le \int_{I_n} \left\langle \mu_f^{hp}, \|\eta_t\|_h \right\rangle_{\Gamma_C} dt \text{ for } \eta \in R_{q_n}(I_n), \; \eta_{n,p_s,p_t} \le 0 \end{array} \right\}$$

where $\mu_f^{hp} = \mathcal{F}\lambda_n^{hp}$ in case of Coulomb friction and simply $\mu_f^{hp} = \mathcal{F}$ in case of Tresca friction. Here, ϕ_j and ϑ_j are Gauss-Lobatto-Lagrange basis functions. Let $G_{e,hp}$ be the affinely transformed Gauss-Lobatto points on the spatial element $e \in \mathcal{T}_{hp}^s$ of corresponding polynomial degree and $G_{hp} = \bigcup_e G_{e,hp}$ the global set which defines ϕ_j. Analogously, let G_{I_n,kq_n} be the affinely transformed Gauss-Lobatto points on the time element $I_n \in \mathcal{T}_{kq}^t$ which defines ϑ_j. For the ease of presentation the notation for acting on the whole of Ω or only on Γ_C does not differ in this chapter. Furthermore, W_{hp} is the trace space of V_{hp} restricted to Γ_C and M_{hp} is the dual space of W_{hp} spanned by biorthogonal basis functions ψ_j, i.e. $\int_{\Gamma_C} \psi_j \phi_i \, ds = \delta_{ij} \int_{\Gamma_C} \phi_i \, ds$. Analogously, ζ_j are the biorthogonal time basis functions to ϑ_j. Note that the discrete Lagrange multiplier Λ inherits the same mesh and polynomial degree distribution from its discrete primal variables U and V. In particular, U and V must have the same trace space on Γ_C for Lemma 6.2 to hold. For simplicity, U and V are sought in the same finite subspace $P_{q_n}(I_n)$. Furthermore, the discrete absolute value function on Γ_C is $\|\eta_t\|_h := \sum_{(p_s,p_t)\in G_{hp}\times G_{I_n,kq_n}} \|\vec{\eta}_{t,p_s,p_t}\| \phi_{p_s}\vartheta_{p_t}$ where $\vec{\eta}_{t,p_s,p_t}$ is the tangential component of the coefficient belonging to the basis function pair $(\phi_{p_s}, \vartheta_{p_t})$ in the linear combination of $\eta \in R_{q_n}(I_n)$ and $\|\cdot\|$ is the Euclidean norm. Therewith, the hp-TDG method is: For $1 \leq n \leq N$, let U_-^{n-1}, V_-^{n-1} be known and find $U = U|_{I_n} \in P_{q_n}(I_n)$, $V = V|_{I_n} \in P_{q_n}(I_n)$ and $\Lambda = \Lambda|_{I_n} \in Q_{q_n}(I_n)$ such that

$$\int_{I_n} (\dot{U}, \varphi) - (V, \varphi)dt + \left(\left[U^{n-1}\right], \varphi_+^{n-1}\right) = 0 \quad \forall\, \varphi \in P_{q_n}(I_n) \tag{6.5a}$$

$$\int_{I_n} (\dot{V}, \chi) + a(U, \chi) + \langle \chi, \Lambda\rangle_{\Gamma_C} \, dt + \left(\left[V^{n-1}\right], \chi_+^{n-1}\right) = \int_{I_n} f(\chi)dt \quad \forall\, \chi \in P_{q_n}(I_n) \tag{6.5b}$$

$$\int_{I_n} \langle U_n, \mu_n - \Lambda_n\rangle_{\Gamma_C} + \langle V_t, \mu_t - \Lambda_t\rangle_{\Gamma_C} \, dt \leq \int_{I_n} \langle g, \mu_n - \Lambda_n\rangle_{\Gamma_C} \, dt \quad \forall\, \mu \in Q_{q_n}(I_n). \tag{6.5c}$$

Here, U_-^0 and V_-^0 are approximations of u_0 and v_0, respectively. The use of biorthogonal basis functions allow the componentwise decoupling of the variational inequality constraint (6.5c) which is more complicated than in the previous chapters due to the friction component and the vector valued coefficients.

Lemma 6.1. *The convex set $Q_{qn}(I_n)$ is equivalent to*

$$Q_{qn}(I_n) = \left\{ \begin{array}{l} \mu = \sum_{p_s,p_t \in G_{hp}\times G_{I_n,kq_n}} \mu_{p_t,p_s}\psi_{p_s}(x)\zeta_{p_t}(t) : \ \mu_{n,p_s,p_t} \geq 0, \\ \|\mu_{t,p_s,p_t}\| \leq \mathcal{F}\Lambda_{n,p_s,p_t} \ \forall\, (p_s,p_t) \in G_{hp}\times G_{I_n,kq_n} \end{array} \right\}. \tag{6.6}$$

Proof. This proof follows the ideas in the proof of [37, Lemma 2.3] generalized to a higher order approach and to a time dependent problem. Recall that $D_{p_s}^s = \int_{\Gamma_C} \phi_{p_s}dx > 0$ and $D_{p_t}^t = \int_{I_n} \vartheta_{p_t} \, dt > 0$ by Lemma 4.6. Then, write $\chi \in R_{q_n}(I_n)$ as its linear combination

$$\chi = \sum_{p_t,p_s} \chi_{p_s,p_t}\phi_{p_s}(x)\vartheta_{p_t}(t) = \sum_{p_t,p_s} \left(\chi_{n,p_s,p_t} n_{p_s} + \vec{\chi}_{t,p_s,p_t}\right)\phi_{p_s}(x)\vartheta_{p_t}(t)$$

with ϕ_{p_s}, ϑ_{p_t} the scalar nodal basis functions and with the coefficients split into normal and tangential components. Hence, the normal part of χ can be written as $\chi_n =$

$\sum_{p_t,p_s} \chi_{n,p_s,p_t} n_{p_s} \phi_{p_s} \vartheta_{p_t}$. Furthermore, there holds

$$Q_{q_n}(I_n) \ni \mu = \sum_{i=0}^{q_n} \mu_i \zeta_i(t) = \sum_{p_t \in G_{I_n,kq_n}} \mu_{p_t} \zeta_{p_t}(t) = \sum_{p_t \in G_{I_n,kq_n}, p_s \in G_{hp}} \mu_{p_s,p_t} \psi_{p_s}(x) \zeta_{p_t}(t)$$

$$= \sum_{p_t,p_s} \left(\mu_{n,p_s,p_t} n_{p_s} + \vec{\mu}_{t,p_s,p_t} \right) \psi_{p_s}(x) \zeta_{p_t}(t).$$

where the coefficients are also split into normal and tangential components. Let $\chi \in R_{q_n}(I_n)$ with $\chi_{n,p_s,p_t} \leq 0$. Hence, inserting the linear combinations of μ and χ in

$$\int_{I_n} \langle \mu, \chi \rangle_{\Gamma_C} \; dt \leq \int_{I_n} \langle \mathcal{F}\lambda_n, \|\vec{\chi}_t\|_h \rangle_{\Gamma_C} \; dt, \tag{6.7}$$

exploiting the biorthogonality of the basis functions and using that $\mathcal{F}$ is independent of x and t yields

$$\sum_{p_s,p_t} \mu_{p_s,p_t} \chi_{p_s,p_t} D_{p_s}^s D_{p_t}^t \leq \mathcal{F} \sum_{p_s,p_t} \lambda_{n,p_s,p_t} \|\vec{\chi}_{t,p_s,p_t}\| D_{p_s}^s D_{p_t}^t. \tag{6.8}$$

Since both χ_{n,p_s,p_t} and $\vec{\chi}_{t,p_s,p_t}$ are arbitrary and $D_{p_s}^s D_{p_t}^t$ is positive the above inequality (6.8) reduces to

$$\mu_{n,p_s,p_t} \chi_{n,p_s,p_t} + \vec{\mu}_{t,p_s,p_t} \vec{\chi}_{t,p_s,p_t} \leq \mathcal{F}\lambda_{n,p_s,p_t} \|\vec{\chi}_{t,p_s,p_t}\| \quad \forall (p_s,p_t) \in G_{hp} \times G_{I_n,kq_n}. \tag{6.9}$$

Recall that $\chi_{n,p_s,p_t} \leq 0$. Choosing $\chi_{t,p_s,p_t} = 0$ in (6.9) yields $\mu_{n,p_s,p_t} \geq 0$. Finally, choosing $\chi_{n,p_s,p_t} = 0$ and $\chi_{t,p_s,p_t} = \mu_{t,p_s,p_t}$ yields $\|\mu_{t,p_s,p_t}\| \leq \mathcal{F}\lambda_{n,p_s,p_t} = \mathcal{F}|\lambda_{n,p_s,p_t}|$. For the opposite direction simply note that with μ_{p_s,p_t} as in (6.6) and $\chi_{n,p_s,p_t} \leq 0$

$$\mu_{p_s,p_t} \chi_{p_s,p_t} = \mu_{n,p_s,p_t} \chi_{n,p_s,p_t} + \vec{\mu}_{t,p_s,p_t} \vec{\chi}_{t,p_s,p_t} \leq \|\vec{\mu}_{t,p_s,p_t}\| \|\vec{\chi}_{t,p_s,p_t}\| \leq \mathcal{F}\lambda_{n,p_s,p_t} \|\vec{\chi}_{t,p_s,p_t}\|$$

yields (6.9). Summing over all points and exploiting the biorthogonality yields the assertion. $\qquad\square$

Lemma 6.2. *If U and V have the same trace mesh on Γ_C then the constraint (6.5c) is equivalent to the system*

$$u_{n,p_s,p_t} D_{p_s}^s D_{p_t}^t \leq g_{p_s,p_t} := \int_{I_n} \int_{\Gamma_C} g\psi_{p_s}(x)\zeta_{p_t}(t) \; dx \; dt \tag{6.10a}$$

$$\lambda_{n,p_s,p_t} \geq 0 \tag{6.10b}$$

$$\lambda_{n,p_s,p_t}\left(U_{n,p_s,p_t} - \frac{g_{p_s,p_t}}{D_{p_s}^s D_{p_t}^t}\right) = 0 \tag{6.10c}$$

$$\begin{cases} |\lambda_{t,p_s,p_t}| \leq \mathcal{F}\lambda_{n,p_s,p_t} \\ |\lambda_{t,p_s,p_t}| < \mathcal{F}\lambda_{n,p_s,p_t} \Rightarrow v_{t,p_s,p_t} = 0 \\ |\lambda_{t,p_s,p_t}| = \mathcal{F}\lambda_{n,p_s,p_t} \Rightarrow \exists\, \alpha \in \mathbb{R} : \lambda_{t,p_s,p_t} = \alpha^2 v_{t,p_s,p_t} \end{cases} \tag{6.10d}$$

for all points $(p_s,p_t) \in G_{hp} \times G_{I_n,kq_n}$ where u_{p_s,p_t}, v_{p_s,p_t} and λ_{p_s,p_t} are the expansion coefficients of U, V and Λ, respectively. The additional index n and t denotes the normal and tangential component, respectively.

Proof. This proof follows the idea of the proof of [37, Lemma 2.6] generalized to a higher order approach and to a time dependent problem. First, write u, v, λ and μ as linear combinations

$$u = \sum_{p_t \in G_{I_n,kq_n}} \sum_{p_s \in G_{hp}} u_{p_s,p_t} \phi_{p_s}(x) \vartheta_{p_t}(t), \quad v = \sum_{p_t,p_s} v_{p_s,p_t} \phi_{p_s}(x) \vartheta_{p_t}(t) \tag{6.11}$$

$$\lambda = \sum_{p_t,p_s} \lambda_{p_s,p_t} \psi_{p_s}(x) \zeta_{p_t}(t), \qquad \mu = \sum_{p_t,p_s} \mu_{p_s,p_t} \psi_{p_s}(x) \zeta_{p_t}(t). \tag{6.12}$$

Inserting these linear combinations into (6.5c) yields, due to the biorthogonality of the basis functions and μ being arbitrary after dividing by the positive factor $D_{p_s}^s D_{p_t}^t$, the following: Find $\lambda \in Q_{q_n}(I_n)$ such that for all $p_s \in G_{hp}$, $p_t \in G_{I_n,kq_n}$ and for all $\mu \in Q_{q_n}(I_n)$ there holds

$$u_{n,p_s,p_t} (\mu_{n,p_s,p_t} - \lambda_{n,p_s,p_t}) + \vec{v}_{t,p_s,p_t} \left(\vec{\mu}_{t,p_s,p_t} - \vec{\lambda}_{t,p_s,p_t} \right) \le \frac{g_{p_s,p_t}}{D_{p_s}^s D_{p_t}^t} (\mu_{n,p_s,p_t} - \lambda_{n,p_s,p_t}). \tag{6.13}$$

The exploited biorthogonality only holds if u and v have the same trace space on Γ_C. Lemma 6.1 already supplies $\lambda_{n,p_s,p_t} \ge 0$ for all points $(p_s, p_t) \in G_{hp} \times G_{I_n,kq_n}$. If $\mu_{n,p_s,p_t} = \lambda_{n,p_s,p_t} + \eta_{n,p_s,p_t}$ with $\eta_{n,p_s,p_t} > 0$, μ, $\eta \in Q_{q_n}(I_n)$ and $\vec{\mu}_{t,p_s,p_t} = \vec{\lambda}_{t,p_s,p_t}$ is chosen then inequality (6.13) yields (6.10a). Whereas choosing $\mu_{n,p_s,p_t} = 0$ and $\mu_{n,p_s,p_t} = 2\lambda_{n,p_s,p_t}$ with $\vec{\mu}_{t,p_s,p_t} = \vec{\lambda}_{t,p_s,p_t}$ yields (6.10c). In order to prove the frictional condition (6.10d) note that Lemma 6.1 already implies that $\vec{\mu}_{t,p_s,p_t}, \vec{\lambda}_{t,p_s,p_t} \in \mathcal{B}(\mathcal{F}\lambda_{n,p_s,p_t})$, i.e. they lie in a $(d-1)$-dimensional ball of center zero and radius $\mathcal{F}\lambda_{n,p_s,p_t}$. For $\mu_{n,p_s,p_t} = \lambda_{n,p_s,p_t}$ inequality (6.13) reduces to

$$\vec{v}_{t,p_s,p_t} \left(\vec{\mu}_{t,p_s,p_t} - \vec{\lambda}_{t,p_s,p_t} \right) \le 0 \qquad \vec{\mu}_{t,p_s,p_t}, \vec{\lambda}_{t,p_s,p_t} \in \mathcal{B}(\mathcal{F}\lambda_{n,p_s,p_t}). \tag{6.14}$$

Assume that the first case holds, i.e. $\left\| \vec{\lambda}_{t,p_s,p_t} \right\| < \mathcal{F}\lambda_{n,p_s,p_t}$. Hence, there exists a ball $\mathcal{B}_{\vec{\lambda}_{t,p_s,p_t}}(\epsilon)$ with center $\vec{\lambda}_{t,p_s,p_t}$ and radius $\epsilon > 0$ such that $\mathcal{B}_{\vec{\lambda}_{t,p_s,p_t}}(\epsilon) \subset \mathcal{B}(\mathcal{F}\lambda_{n,p_s,p_t})$. Therewith, for any $\vec{\tilde{\mu}}_{t,p_s,p_t} \in \mathcal{B}_{\vec{\lambda}_{t,p_s,p_t}}(\epsilon)$ with $\left\| \vec{\tilde{\mu}}_{t,p_s,p_t} \right\| = \epsilon$ there holds $\vec{\mu}_{t,p_s,p_t}^{\pm} = \vec{\lambda}_{t,p_s,p_t} \pm \vec{\tilde{\mu}}_{t,p_s,p_t} \in \mathcal{B}(\mathcal{F}\lambda_{n,p_s,p_t})$. Inserting these two μ into (6.14) yields

$$\pm\vec{v}_{t,p_s,p_t} \vec{\tilde{\mu}}_{t,p_s,p_t} \le 0 \quad \Rightarrow \vec{v}_{t,p_s,p_t} = 0.$$

For last case, i.e. $\left\| \vec{\lambda}_{t,p_s,p_t} \right\| = \mathcal{F}\lambda_{n,p_s,p_t}$, define the half-space, c.f. Figure 6.1,

$$\mathcal{H}(\vec{v}_{t,p_s,p_t}) = \left\{ \vec{\eta}_{t,p_s,p_t} \in \mathbb{R}^{d-1} : \vec{v}_{t,p_s,p_t} \left(\vec{\eta}_{t,p_s,p_t} - \vec{\lambda}_{t,p_s,p_t} \right) > 0 \right\}.$$

Assume there exists no $\alpha \in \mathbb{R}$ such that $\vec{\lambda}_{t,p_s,p_t} = \alpha^2 \vec{v}_{t,p_s,p_t}$. Then, the intersection of the half-space with the ball $\mathcal{H}(\vec{v}_{t,p_s,p_t}) \cap \mathcal{B}(\mathcal{F}\lambda_{n,p_s,p_t}) \ne \emptyset$ is not empty. Consequently, there holds for all $\vec{\mu}_{t,p_s,p_t} \in \mathcal{H}(\vec{v}_{t,p_s,p_t}) \cap \mathcal{B}(\mathcal{F}\lambda_{n,p_s,p_t})$ that $\vec{v}_{t,p_s,p_t} \left(\vec{\mu}_{t,p_s,p_t} - \vec{\lambda}_{t,p_s,p_t} \right) > 0$ which is a contradiction to (6.14).

To prove the opposite direction, (6.10a) is multiplied by μ_{n,p_s,p_t}, followed by a subtraction of (6.10c) and addition of the tangential part $\vec{v}_{t,p_s,p_t} \left(\vec{\mu}_{t,p_s,p_t} - \vec{\lambda}_{t,p_s,p_t} \right) \le 0$. The

latter trivially holds by (6.10d). Summing over all points and exploiting the biorthogonality yields together with Lemma 6.1 the assertion. □
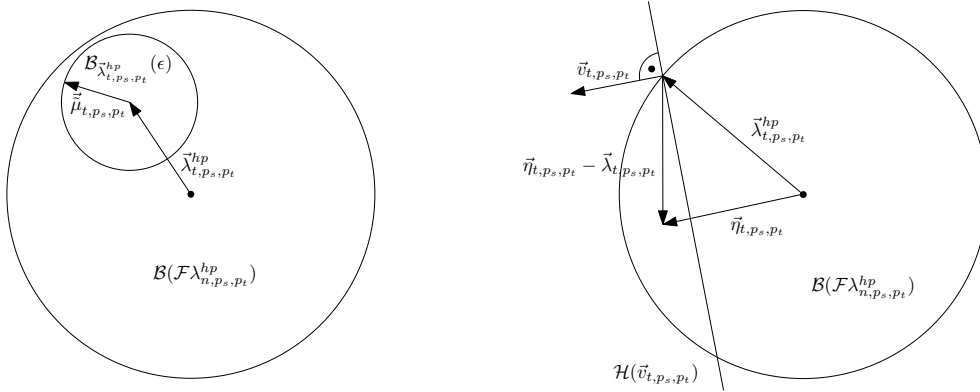

Figure 6.1: Illustration of $\mathcal{B}_{\vec{\lambda}_{t,p_s,p_t}}(\epsilon) \subset \mathcal{B}(\mathcal{F}\lambda_{n,p_s,p_t})$ and of $\mathcal{H}(\vec{v}_{t,p_s,p_t})$, based on [37, Figure 2.1]

Lemma 6.3. *If $\mathcal{F} = 0$ and the volume and surface forces are time independent, then the discrete mixed formulation (6.5) is energy dissipative for $p = 1$ and $q = 0$, i.e.*

$$J(u(t), \dot{u}(t)) \geq J(u(t'), \dot{u}(t')) \qquad \forall\, 0 \leq t \leq t' \leq T \tag{6.15}$$

with the system's energy defined by

$$J(u(t), \dot{u}(t)) := \frac{1}{2}\left(\dot{u}(t), \dot{u}(t)\right) + \frac{1}{2}a(u(t), u(t)) - f(u(t)). \tag{6.16}$$

Proof. Follows in [2, Lemma 7]. □

6.3 Iterative Solvers for the Mixed Method

The componentwise decoupling of the weak contact constraints allows the formulation of a semi-smooth Newton (SSN) method and of an Uzawa algorithm to solve the discrete problem (6.5) iteratively. For the first approach the point-wise condition (6.10) is rewritten into finding the roots of two semi-smooth non-linear complementarity functions (NCF). The Uzawa algorithm exploits a point-wise projection of the Lagrange multiplier. The CPU-time of both solvers is reduced by an (almost) block diagonalization of the global system matrix.

6.3.1 Semi-Smooth Newton

For the complementarity problem in the normal component u_{n,p_s,p_t} the penalized Fischer-Burmeister NCF is used as in the elliptic stochastic case of Section 5.2.3. For the ease

of presentation the restriction to the contact boundary is omitted. The NCF for the frictional condition differs between Tresca friction, i.e. (6.10d) without Λ_{n,p_s,p_t}, and Coulomb friction. The function C_T is chosen for Tresca friction and C_C for Coulomb friction as studied in [37, 16, 17].

$$C_T(v,\lambda) = \max\left\{\mathcal{F}, \left\|\vec{\lambda}_{t,p_s,p_t} + c_t\vec{v}_{t,p_s,p_t}\right\|\right\}\vec{\lambda}_{t,p_s,p_t} - \mathcal{F}(\vec{\lambda}_{t,p_s,p_t} + c_t\vec{v}_{t,p_s,p_t}) \tag{6.17}$$

$$C_C(v,\lambda) = \max\left\{\mathcal{F}\lambda_{n,p_s,p_t}, \left\|\vec{\lambda}_{t,p_s,p_t} + c_t\vec{v}_{t,p_s,p_t}\right\|\right\}\vec{\lambda}_{t,p_s,p_t} - \mathcal{F}\lambda_{n,p_s,p_t}(\vec{\lambda}_{t,p_s,p_t} + c_t\vec{v}_{t,p_s,p_t}) \tag{6.18}$$

Here $c_t > 0$ is an arbitrary constant.

Remark 6.1. *In [16, 17] the term $\mathcal{F}\max\{0,\lambda_{n,p_s,p_t}\}$ and in [37] the term $\mathcal{F}\max\{0,\lambda_{n,p_s,p_t} + c_t(u_{n,p_s,p_t} - g_{p_s,p_t})\}$ are used instead of $\mathcal{F}\lambda_{n,p_s,p_t}$. All three formulations are equivalent since $\lambda_{n,p_s,p_t} \geq 0$ due to $\varphi_\mu = 0$, but the function (6.18) should be the easiest and most efficient of them as it contains the same types but fewer non-linear terms.*

Lemma 6.4.   *1. For $\mathcal{F} > 0$, the pair (v,λ) satisfies (6.10d) for Tresca friction if and only if $C_T(v,\lambda) = 0$ with an arbitrary constant $c_t > 0$.*

   2. For $\mathcal{F} = 0$, (v,λ) satisfies (6.10d) if and only if $C_T(v,\lambda) = 0$ for two pairwise different $c_t > 0$.

   3. A similar result holds for Coulomb friction with $C_C(v,\lambda) = 0$.

Proof. For $\mathcal{F} > 0$ and Tresca friction this is the result of [37, Theorem 5.1].
For $\mathcal{F} = 0$ equation (6.10d) yields $\lambda_{t,p_s,p_t} = 0$. These together immediately imply $C_T = 0$. For the opposite direction, note that C_T reduces to $C_T(v,\lambda) = \left\|\vec{\lambda}_{t,p_s,p_t} + c_t\vec{v}_{t,p_s,p_t}\right\|\vec{\lambda}_{t,p_s,p_t}$. Assume $\left\|\vec{\lambda}_{t,p_s,p_t} + c_t\vec{v}_{t,p_s,p_t}\right\| > 0$. Then $C_T = 0$ immediately implies $\vec{\lambda}_{t,p_s,p_t} = 0$ and condition (6.10d) with $\alpha = 0$. Now assume $\left\|\vec{\lambda}_{t,p_s,p_t} + c_t\vec{v}_{t,p_s,p_t}\right\| = 0$. Then (6.10d) can not be deduced from $C_T = 0$ alone. However, if $C_T = 0$ also for $0 < \tilde{c}_t \neq c_t$ then either for $\tilde{c}_t$ the first case holds or the fact that $\left\|\vec{\lambda}_{t,p_s,p_t} + \tilde{c}_t\vec{v}_{t,p_s,p_t}\right\| = 0$ yields $\vec{\lambda}_{t,p_s,p_t} = -\tilde{c}_t\vec{v}_{t,p_s,p_t}$. Thus, $\vec{\lambda}_{t,p_s,p_t} = \vec{v}_{t,p_s,p_t} = 0$ can be deduced, which yields the assertion. The proof for C_C is analogous. $\qquad\square$

Remark 6.2. *C_T and C_C are not NCFs if $\mathcal{F} = 0$ and $\mathcal{F}\lambda_{n,p_s,p_t} = 0$, respectively.*

If the conditions of Remark 6.2 do not hold then with the NCFs φ_μ and C the problem (6.5a), (6.5b) and (6.10) is equivalent to finding the root of

$$0 \stackrel{!}{=} F(u,v,\lambda) = \begin{pmatrix} M_p u - Mv - f_u \\ M_p v + Au + D\lambda - f_v \\ \varphi_\mu(u,\lambda) \\ C(v,\lambda) \end{pmatrix} \tag{6.19}$$

where the first two lines are the matrix representation of (6.5a) and (6.5b). In the numerical experiments the non-linear problem (6.19) is solved using the following damped SSN algorithm. To define a suitable stopping criterion and line search the non-negative merit function

$$\Psi(u, v, \lambda) := \frac{1}{2} F(u, v, \lambda)^2 \tag{6.20}$$

is defined. In particular, solving (6.19) is equivalent to finding a minimizer triple of Ψ. Due to the use of the NCF C_T and C_C, Ψ is no longer continuously differentiable as in the previous chapters.

Algorithm 6.1. *(SSN algorithm for elasto-dynamic frictional contact)*

---

1. *Choose initial solution u^0, $v^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^{n_C}$, $\beta \in (0, 1)$, $\sigma \in (0, \frac{1}{2})$, tol > 0.*

2. *For $k = 0, 1, 2, \ldots$ do*

   a) *If $\left\| \nabla \Psi(u^k, v^k, \lambda^k) \right\| <$ tol or $\left\| \Psi(u^k, v^k, \lambda^k) \right\| <$ tol then stop.*

   b) *Compute subdifferential $H_k \in \partial F(u^k, v^k, \lambda^k)$ and find direction $d^k \in \mathbb{R}^{2n+n_C}$ with $d^k = \left(d_u^k, d_v^k, d_\lambda^k \right)$ such that*

   $$H_k d^k = -F(u^k, v^k, \lambda^k). \tag{6.21}$$

   c) *Compute search length $t_k := \max \left\{ \beta^l : l = 0, 1, 2, \ldots \right\}$ such that*

   $$\Psi(u^k + t_k d_u^k, v^k + t_k d_v^k, \lambda^k + t_k d_\lambda^k) \leq \Psi(u^k, v^k, \lambda^k) + \sigma t_k \nabla \Psi(u^k, v^k, \lambda^k) d^k.$$

   d) *Update solution vectors and goto step 2.*

   $$u^{k+1} = u^k + t_k d_u^k, \quad v^{k+1} = v^k + t_k d_v^k, \quad \lambda^{k+1} = \lambda^k + t_k d_\lambda^k$$

In case of Tresca friction, the following subdifferential

$$H_k = \begin{pmatrix} M_p & -M & 0 \\ A & M_p & D \\ \frac{\partial \varphi(u^k, \lambda^k)}{\partial u} & 0 & \frac{\partial \varphi(u^k, \lambda^k)}{\partial \lambda} \\ 0 & \frac{\partial C(v^k, \lambda^k)}{\partial v} & \frac{\partial C(v^k, \lambda^k)}{\partial \lambda} \end{pmatrix}$$

with $\frac{\partial \varphi(u, \lambda)}{\partial u}$, $\frac{\partial \varphi(u, \lambda)}{\partial \lambda}$ as in Section 5.2.3 and

$$\frac{\partial C_T(v, \lambda)}{\partial v} = \begin{cases} -\mathcal{F} c_t T & \text{, if } |\lambda_t + c_t v_t| \leq \mathcal{F} \\ c_t(\lambda_t - \mathcal{F})T & \text{, if } \lambda_t + c_t v_t > \mathcal{F} \\ -c_t(\lambda_t + \mathcal{F})T & \text{, otherwise} \end{cases}$$

$$\frac{\partial C_T(v, \lambda)}{\partial \lambda} = \begin{cases} 0 & \text{, if } |\lambda_t + c_t v_t| \leq \mathcal{F} \\ (\lambda_t + |\lambda_t + c_t v_t| - \mathcal{F})T & \text{, if } \lambda_t + c_t v_t > \mathcal{F} \\ (-\lambda_t + |\lambda_t + c_t v_t| - \mathcal{F})T & \text{, otherwise} \end{cases}$$

is chosen where N and T are the algebraic representations of the normal and tangential vector in $2d$, respectively.

Remark 6.3. *The subdifferentials $\frac{\partial C_T}{\partial \lambda}(0,0)$ and $\frac{\partial C_T}{\partial v}(0,0)$ vanish if $\mathcal{F} = 0$. Hence, H_k is not invertible. However, for $\mathcal{F} = 0$ the original friction condition is equivalent to a linear complementarity function which is implemented as a special case if both subdifferentials vanish.*

The most expensive part in the SSN algorithm is the computation of the direction d^k. The system matrix H_k is sparse, indefinite, non-symmetric, ill-conditioned and increases rapidly in size if the polynomial degree in time is raised due to the tensor product structure of the space and time discretization. In [75] numerical experiments have shown that the hp-TDG system matrix for a simple parabolic problem can be block diagonalized in $\mathbb{C}$ if orthonormalized time basis functions are used. The size of such a block is the number of spatial degrees of freedom (dof). Exploiting the invertibility of the standard $1d$-time mass matrix M_2, this idea can be generalized to arbitrary basis functions. Due to the use of a Lagrange multiplier space spanned by biorthogonal basis functions, H_k can only be almost block diagonalized in the sense that the part associated with the primal variables u and v has a block structure but at the cost of additional entries in the part of the Lagrange multiplier tested with the primal variables. However, the number of additional entries is benign since the Lagrange multiplier acts only on the trace of u and v. For a mathematical description the following three abbreviations are defined to reflect the time influence.

$$[M_1]_{ij} = \int_{I_n} \vartheta'_j \vartheta_i \, dt + \vartheta_j^+(t_{n-1})\vartheta_i^+(t_{n-1}), \quad [M_2]_{ij} = \int_{I_n} \vartheta_j \vartheta_i \, dt, \quad [M_3]_{ij} = \delta_{ij} \int_{I_n} \vartheta_j \, dt$$

First assume that the differential system is only discretized in time and that the Lagrange multiplier is known. Hence, after multiplying with M_2^{-1} from the left it becomes

$$M_2^{-1}M_1 \begin{pmatrix} u \\ v \end{pmatrix} + L \begin{pmatrix} u \\ v \end{pmatrix} = M_2^{-1}f - M_2^{-1}M_3\lambda \tag{6.22}$$

where L is the linear spatial operator. Numerical experiments have shown that $M_2^{-1}M_1$ is diagonalizable in $\mathbb{C}$ and that there exits a matrix $Q \in \mathbb{C}^{(q+1)\times(q+1)}$ such that

$$Q^{-1}M_2^{-1}M_1Q = T = diag\left(\mu_0, \ldots, \mu_q\right). \tag{6.23}$$

Multiplying (6.22) with Q^{-1} from the left, changing the order of L and Q^{-1} since L is independent of time and inserting the identity QQ^{-1} yields

$$Q^{-1}M_2^{-1}M_1QQ^{-1} \begin{pmatrix} u \\ v \end{pmatrix} + LQ^{-1} \begin{pmatrix} u \\ v \end{pmatrix} = Q^{-1}M_2^{-1}f - Q^{-1}M_2^{-1}M_3\lambda.$$

With the change of variables, $\begin{pmatrix} x \\ y \end{pmatrix} = Q^{-1} \begin{pmatrix} u \\ v \end{pmatrix}$, the previous equation is equivalent to the system

$$\mu_j \begin{pmatrix} x_j \\ y_j \end{pmatrix} + L \begin{pmatrix} x_j \\ y_j \end{pmatrix} + \left[Q^{-1}M_2^{-1}M_3\lambda\right]_j = \left[Q^{-1}M_2^{-1}f\right]_j \quad j = 0, \ldots, q$$

with μ_j defined in (6.23). If the Lagrange multiplier is not known the term $\left[Q^{-1}M_2^{-1}M_3\lambda\right]_j$ couples the equations for different $j's$. In the next step the standard spatial discretization is introduced. Here $\otimes$ stands for a matrix-block-vector multiplication $[Q \otimes f]_j = \sum_{k=0}^{q}[Q]_{jk}\,\vec{f}_k$ where $\vec{f}_k$ is the kth vector block of size equal to the spatial dof. Hence, the system of linear equations $(j = 0, \ldots, q)$

$$\left[\mu_j\begin{pmatrix} M & 0 \\ 0 & M \end{pmatrix} + \begin{pmatrix} 0 & -M \\ A & 0 \end{pmatrix}\right]\begin{pmatrix} x_j \\ y_j \end{pmatrix} + \left[Q^{-1}M_2^{-1}M_3 \otimes B\lambda\right]_j = \left[Q^{-1}M_2^{-1} \otimes \vec{f}\right]_j$$

is obtained. If x and y are known, the change of variables can be reversed using $u = Q\otimes x$ and $v = Q \otimes y$. Thus, the change of $\frac{\partial \varphi}{\partial u}$ and $\frac{\partial C}{\partial v}$ to

$$Q \otimes \frac{\partial \varphi}{\partial u} \cdot x = f_\varphi, \quad \text{and} \quad Q \otimes \frac{\partial C}{\partial v} \cdot y = f_C$$

is straight forward. Only $\frac{\partial \varphi}{\partial \lambda}$, $\frac{\partial C}{\partial \lambda}$ remain unchanged. With the abbreviations

$$\hat{X}_{jj} = \begin{pmatrix} \mu_j M & -M \\ A & \mu_j M \end{pmatrix}, \qquad \hat{N}_{ij} = \begin{pmatrix} [Q]_{ij}\left[\frac{\partial \varphi}{\partial u}\right]_j & 0 \\ 0 & [Q]_{ij}\left[\frac{\partial C}{\partial v}\right]_j \end{pmatrix}$$

$$\hat{B}_{ij} = \begin{pmatrix} 0 \\ \left[Q^{-1}M_2^{-1}M_3\right]_{ij} B \end{pmatrix}, \quad \hat{L}_{jj} = \begin{pmatrix} \left[\frac{\partial \varphi}{\partial \lambda}\right]_j \\ \left[\frac{\partial C}{\partial \lambda}\right]_j \end{pmatrix}$$

the algebraic representation of the almost block diagonalized system equals

$$\begin{pmatrix} \hat{X}_{00} & \cdots & 0 & \hat{B}_{00} & \cdots & \hat{B}_{0q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{X}_{qq} & \hat{B}_{q0} & \cdots & \hat{B}_{qq} \\ \hat{N}_{00} & \cdots & \hat{N}_{0q} & L_{00} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \hat{N}_{q0} & \cdots & \hat{N}_{qq} & 0 & \cdots & L_{qq} \end{pmatrix} \begin{pmatrix} x^0 \\ y^0 \\ \vdots \\ x^q \\ y^q \\ \lambda^0 \\ \vdots \\ \lambda^q \end{pmatrix} = \begin{pmatrix} \left[Q^{-1}M_2^{-1} \otimes \vec{f}\right]_0 \\ \vdots \\ \left[Q^{-1}M_2^{-1} \otimes \vec{f}\right]_q \\ [f_\varphi]_0 \\ [f_C]_0 \\ \vdots \\ [f_\varphi]_q \\ [f_C]_q \end{pmatrix}. \qquad (6.24)$$

In general, the time polynomial is smaller than 15 and thus the diagonalizing and inverting of the time matrices is very fast and only needs to be carried out once per time step. As pointed out earlier this is not a block diagonalization but the population structure of the global system matrix has been significantly improved as can be seen in Figure 6.2. This can be exploited by both direct and iterative solvers. The experimental speed up is displayed in Figure 6.3. In case of $q = 3$, this strategy leads to a reduction of the computational time for Matlab's direct solver by a factor of two whereas for $q = 6$ this factor is already eleven. In general the number of non-zero entries is almost reduced by the factor $q + 1$.
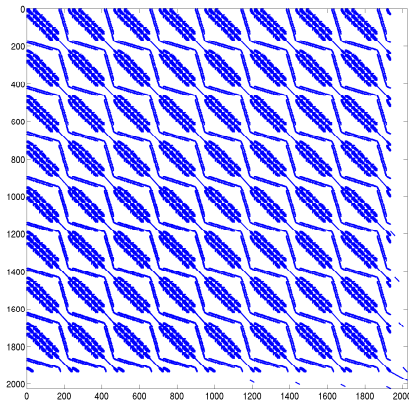
(a) 265296 nonzero entries
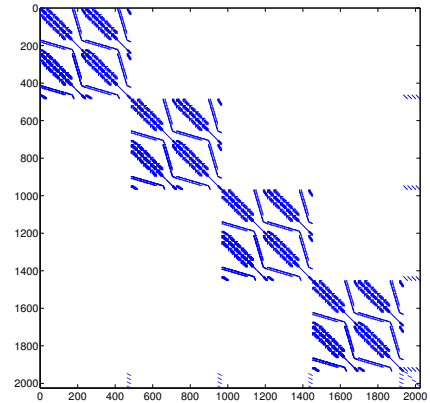
(b) 66852 nonzero entries

Figure 6.2: Sparsity pattern of direct (left) vs. almost block diagonalized (right) approach for $q = 3$, uniform spatial mesh $p = 2$, 968+968 primal dof, 88 Lagrangian dof
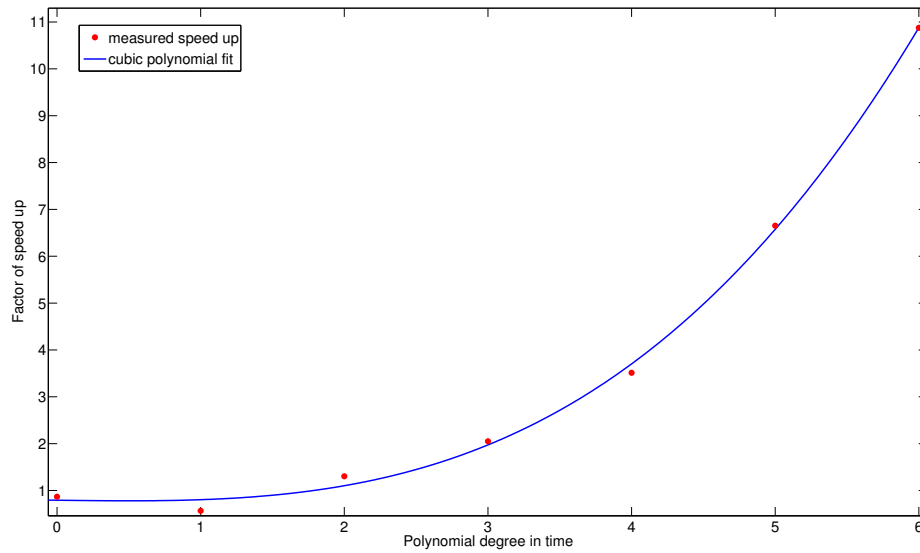

Figure 6.3: Experimental speed up of almost block diagonalization for uniform spatial mesh with 25 elements and $p = 2$, 968+968 primal dof, 88 Lagrangian dof

6.3.2 Uzawa

The problem (6.5a), (6.5b) and (6.10) can also be solved by the following highly parallel

Algorithm 6.2. *(Uzawa algorithm for elasto-dynamic frictional contact)*

1. *Choose initial solution $\lambda^0 \in \mathbb{R}^{n_C}$, $c_n > 0$, $c_t > 0$, $tol > 0$*

2. *For $k = 0, 1, 2, \ldots$ do*

   a) *Find displacement u^k and velocity v^k by solving (6.5a), (6.5b) with λ^k given.*

   b) *Update Lagrange multiplier by pointwise projection.*

   $$\lambda_{n,ps,pt}^k = \max\left\{0, \lambda_{n,p_s,p_t} + c_n(u_{n,p_s,p_t} - g_{p_s,p_t})\right\} \tag{6.25}$$

   $$\lambda_{t,ps,pt}^k = \mathcal{F}\frac{\lambda_{t,p_s,p_t} + c_t v_{t,p_s,p_t}}{\max\left\{\mathcal{F}, |\lambda_{t,p_s,p_t} + c_t v_{t,p_s,p_t}|\right\}} \tag{6.26}$$

   c) *Stop if $\left\|\lambda^{k-1} - \lambda^k\right\| < tol\left\|\lambda^{k-1}\right\|$ or $\left\|\lambda^{k-1} - \lambda^k\right\| < tol$, else goto step 2.*

The advantage of this algorithm is that λ appears as a known function in (6.5a) and (6.5b). Thus, the corresponding system matrix can be block diagonalized and the subproblems solved in parallel. The projection of λ can be carried out pointwise with only marginal computational cost. However, the convergence rate heavily depends on the damping parameters and is in general very slow.

6.4 Numerical Experiments

As a numerical example, a multiple impact of a square with a smooth rigid foundation is simulated. More precisely, the reference configuration of the linear elastic body is $\Omega = [-1, 1]^2$ with $E = 2000$ and $\nu = 0.3$ and the time interval is $I = (0, 0.9)$. The gap function is $g(x) = 1 - \frac{1}{2}\cos x$, the time independent volume force is $f = (0, -20)^T$ and the surface force on Γ_N is zero everywhere. The measure of Γ_D is zero and the locations of Γ_N and Γ_C are as displayed in Figure 6.4. The initial conditions are $u = \dot{u} = 0$ and in case of Tresca friction the friction coefficient is $\mu_f = \mathcal{F} = 0.5$.

Since the volume and surface forces are time independent the system must not lose energy in the frictionless case. The energy is defined as in (6.16). Figure 6.5 displays the energy of the system which is the same as the loss of energy due to the initial conditions. The lowest order h-version is energy dissipative as predicted by Lemma 6.3. Most but not all of the energy is lost during the two impact faces, i.e. between 0.2 and 0.4 and between 0.6 and 0.75. For the coarsest mesh (blue line) too much energy is lost for the square to rebounce. Moreover, for refined h-meshes the energy loss during the first impact becomes much smaller and the square rebounces higher. Therefore, the second

$$\Gamma_N$$

$$\Omega = [-1, 1]^2$$

$$\Gamma_C$$

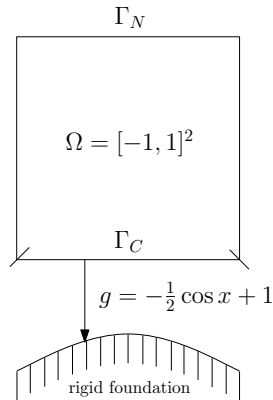$$g = -\tfrac{1}{2}\cos x + 1$$

rigid foundation

Figure 6.4: Reference configuration of the linear elastic body

impact occurs at a later time. For the *hp*-method, the mesh is geometrically graded towards the boundary of Ω to resolve the contact set and the reflection of the shock wave. The shock wave results from the sudden chance in velocity at the time of impact and is responsible for the rebouncing of the object. The time mesh is geometrically graded toward the beginning and ending of the impact from the right. These time points are experimentally determined by the solution of the *h*-version. The *hp*-method is not energy dissipative as it artificially generates energy peaks. However, much fewer degrees of freedom, and therefore CPU time, are needed to obtain an equally good approximation in terms of energy loss compared to the lowest order *h*-version. Since no analytic solution is available, the error is measured by the distance of the norm values, i.e. $\left(\|u_{0,\infty}\|^2_{L^2I;H^1(\Omega)} - \|u_{hp}\|^2_{L^2I;H^1(\Omega)}\right)^{\frac{1}{2}}$. Figure 6.6a shows the superiority of the *hp*-method and also to the extremely slow convergence of the uniform *h*-version. The same observations are made for Tresca frictional case, displayed in Figure 6.6b

The numerical experiment visualized in Figure 6.7 indicates that the semi-smooth Newton method converges locally Q-quadratic in the frictionless case and locally super linear in the Tresca frictional case. In both cases, the SSN algorithm is by far superior to the Uzawa algorithm, which, although not visible, converges extremely slowly. As in the parabolic case, Figure 6.8a shows that the choice of the parameter μ in the penalized Fischer-Burmeister NCF has only very little influence on the average number of SSN iterations per time step for the *hp*-method. However, the number of iterations can be reduced significantly if the initial solution for the iterative solver is obtained from extrapolation. This can be done in a similar way to the parabolic case as displayed in Figure 6.8b.
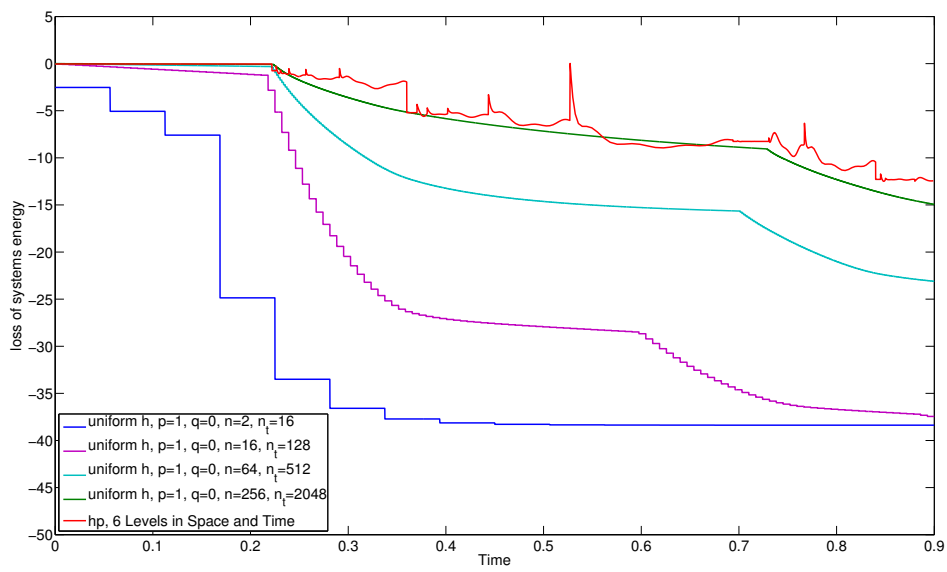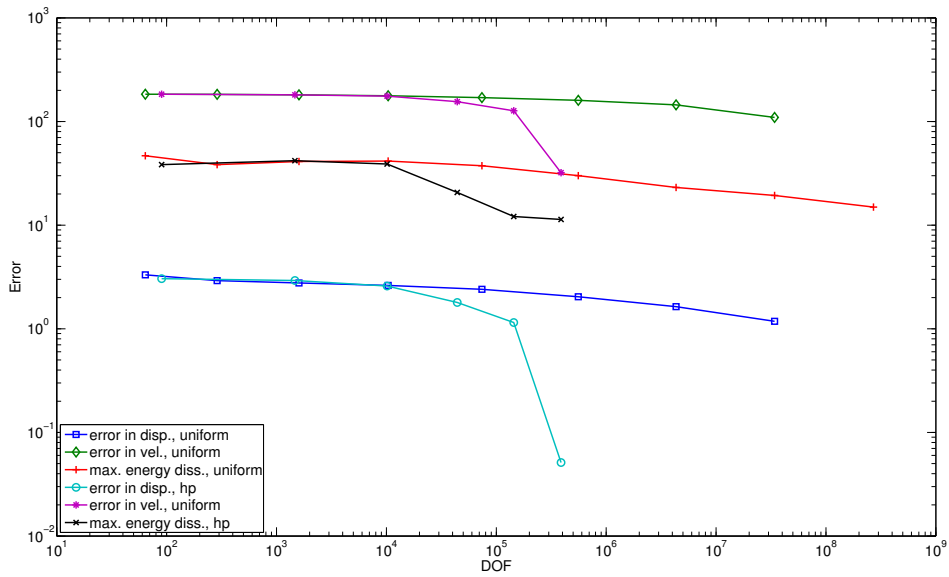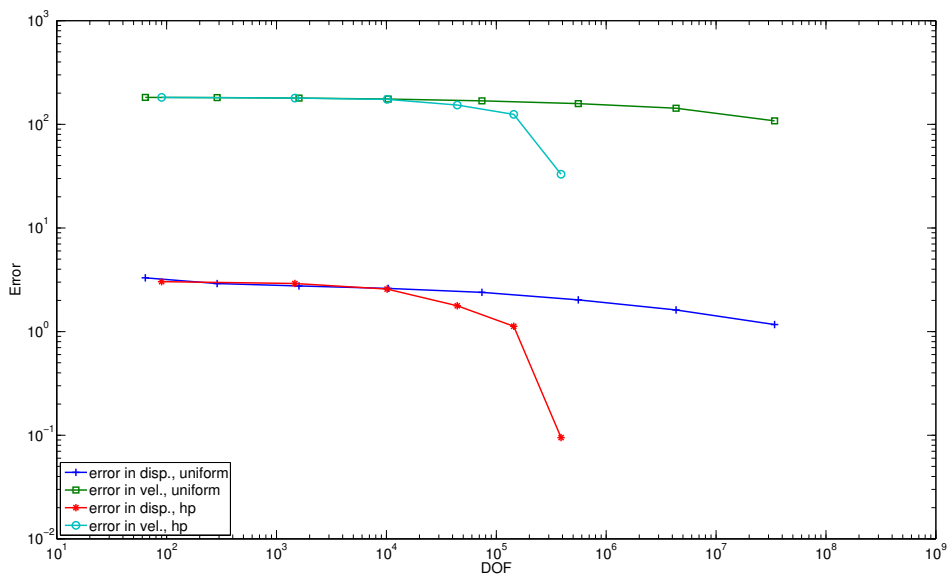
Figure 6.5: Evolution of the system's energy over time

(a) Frictionless

(b) Tresca friction $\mathcal{F} = 0.5$

Figure 6.6: Distance of the norm values for elasto-dynamic contact

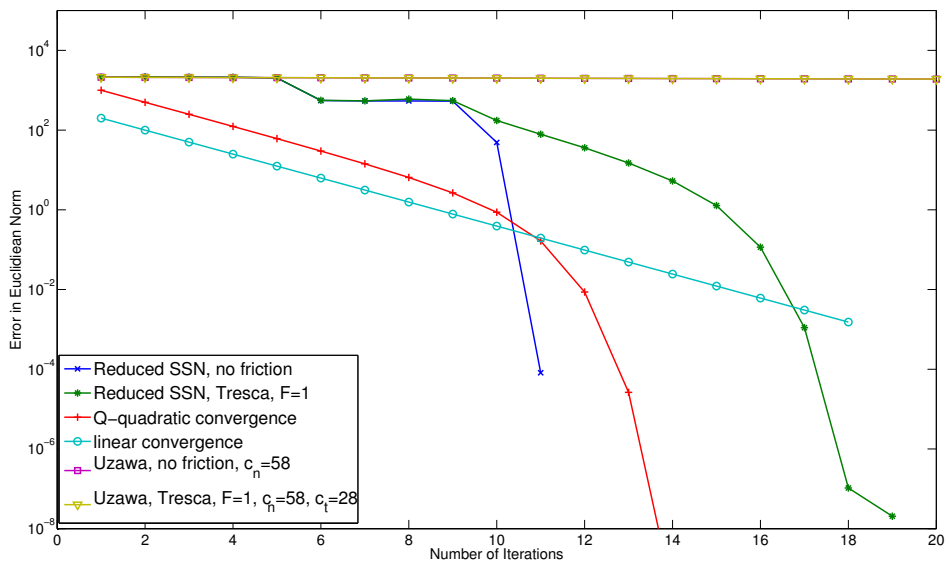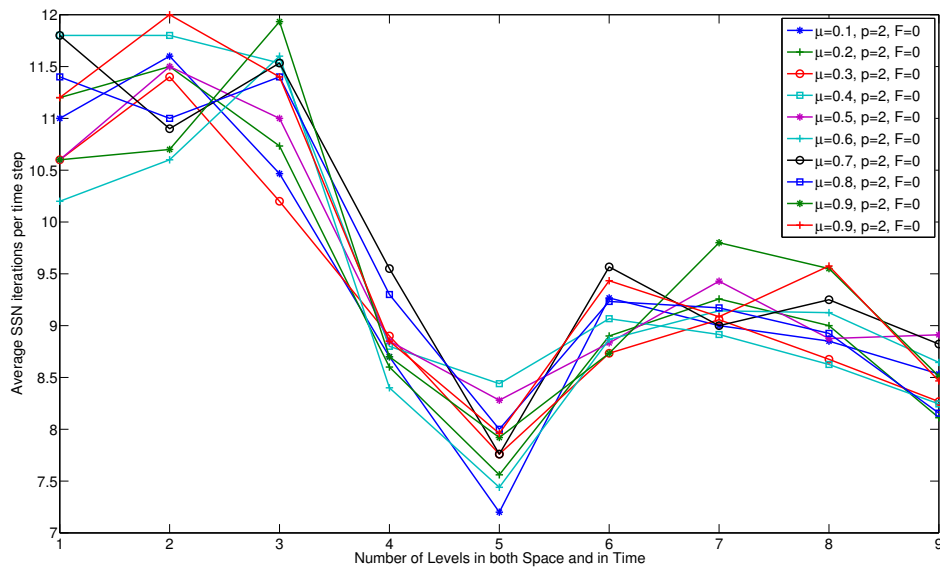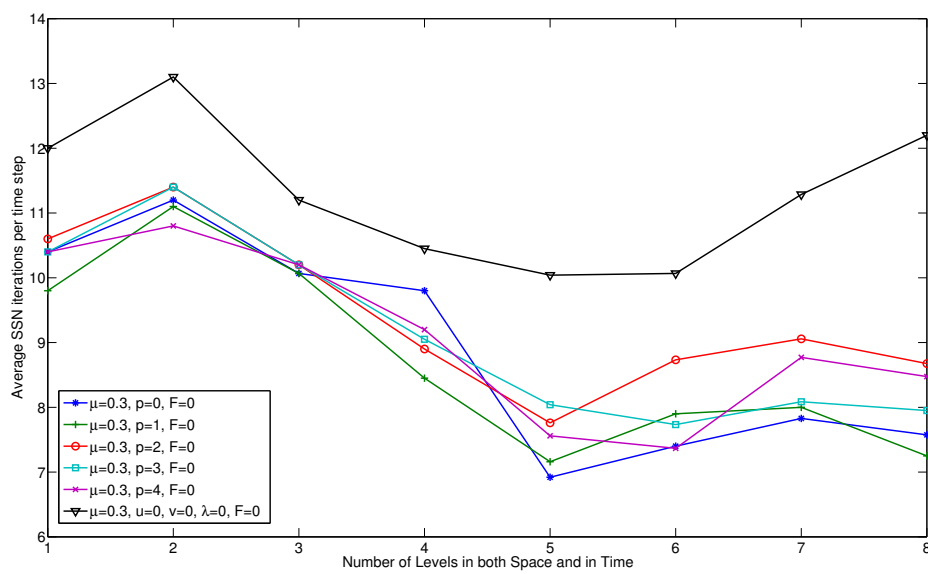Figure 6.7: Error reduction for different solvers per iteration step

(a) variation of μ, frictionless

(b) variation of init. sol., frictionless

Figure 6.8: Average number of SSN iterations per time step for hp-method

7 Conclusion

In this dissertation an approximation strategy is presented and analyzed which is suitable for different classes of differential operators and discretization methods for both contact and obstacle problems. In particular, a hp-FEM interior penalty discontinuous Galerkin (IPDG) method is used for a non-symmetric, elliptic obstacle problem, a hp-FEM IPDG/time discontinuous Galerkin (TDG) method for a parabolic obstacle problem, a hp-BEM Galerkin method for an exterior, elliptic stochastic contact problem and an hp-FEM Galerkin for a linear elasto-dynamic frictional contact problem. Common to all discretization methods is a mixed formulation in which the non-penetration condition and the friction condition are weakly enforced by a single variational inequality constraint. It is shown that on a continuous level, this mixed formulation is equivalent to a variational inequality formulation in which the non-penetration condition is strongly enforced. All but one of the considered differential operators are not symmetric and, therefore, the variational inequality formulation is not equivalent to a constraint minimization problem. Hence, only Uzawa and projection-contraction solvers are guaranteed to converge. Therefore, an efficient iterative solver is constructed, which reduces the computational time for the solvers from three hours to six seconds in a 1D parabolic case. The key for that is to write the discrete Lagrange multiplier for the mixed method as a linear combination of basis functions which are globally biorthogonal to the basis functions of the primal variable. If the primal basis functions are Gauss-Lobatto-Lagrange basis functions on an interval in 1D, rectangular in 2D, and cube in 3D, the system matrix of the variational inequality constraint becomes a positive definite diagonal matrix. This allows to rewrite the discrete sign and non-penetration conditions into a complementarity problem for each coefficient of the solution's vectors. Employing a vector-valued penalized Fischer-Burmeister nonlinear complementarity function reformulates the discrete mixed problem into finding the unique root of a strongly semi-smooth function. This in turn is solved by a semi-smooth Newton (SSN) method for which local Q-quadratic convergence is proven. The iterative solver has therewith optimal convergence properties for a Newton method and only requires coercivity, but not symmetry of the differential operator. The challenging part is the local construction of basis functions which are globally biorthogonal, even on irregular meshes, which naturally arise in hp-adaptivity. In DG methods, they must be biorthogonal to L^2-basis functions and, therefore, a discontinuous extension of affinely transformed local basis biorthogonal functions with zero is sufficient. For $H^{\frac{1}{2}}(\Gamma_C)$-conforming primal basis functions, the same assembling algorithm can be used for the dual as for the primal variable. Furthermore, it is proven that the discrete variational

inequality formulation is equivalent to the discrete mixed method if the strong discrete non-penetration condition is a one-sided box constraint on the expansion coefficients with the constraining coefficients obtained from a mortar projection of the obstacle. Therewith, the efficient SSN algorithm is a fast iterative solver for both the mixed and variational inequality formulations.

Typically, the solution of contact and obstacle problems is of reduced regularity which is induced by the a priori unknown free boundary. For the BEM approximation of the stochastic problem, which is a conforming approach in the primal variable, a residual a posteriori error estimator is derived. For this, the residual estimator from [10] is extended to an auxiliary problem in a $L^2_\rho(\Theta; \tilde{H}^{\frac{1}{2}}(\Gamma_\Sigma))$-setting and a separating of the approximation error into a discretization error of the auxiliary problem and into an error arising from the contact conditions as in [8] is used, yet generalized to arbitrary obstacles. For non-conforming methods like DG, this approach is not possible. Therefore, a hierarchical a posteriori error estimator for the IPDG variational inequality formulation is derived, since the discrete variational inequality and mixed methods are equivalent. Due to the mesh dependent energy norm, a $p-(p+1)$ estimator requires less CPU-time than a $h-\frac{h}{2}$ estimator and less time to implement. Using the local analyticity estimate from [36] to decide weather an element is h- or p-refined, the numerical experiments to the elliptic obstacle case showed the aspired exponential convergence. Even in the parabolic case the convergence rates have been improved significantly compared to uniform meshes. The further experiments to the stochastic and hyperbolic cases show the variety of applicability of the general scheme presented and analyzed in this dissertation.

The fundamental a priori error estimates for the obstacle problems with DG and the numerical experiments encourage additional research on the approximation properties of the discrete finite element spaces. Further research should be done on the discrete inf-sup-condition for a p-version IPDG with biorthogonal basis functions. Therewith, error estimates which include the error in the Lagrange multiplier may be proven. Investigations on how the subdifferentials H_k can be preconditioned may lead to a further reduction of computational time.

Bibliography

[1] I. Babuška, R. Tempone, and G. E. Zouraris, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, Siam J. Numer. Anal, 42 (2004), pp. 800–825.

[2] L. Banz, *Elasto-dynamic contact problems*, master's thesis, Brunel University West London, September 2008.

[3] L. Banz, A. Schröder, and E. P. Stephan, *hp-adaptive mixed-fem with biorthogonal basis for elliptic obstacle problems*. in preparation.

[4] S. Bartels and C. Carstensen, *Averaging techniques yield reliable a posteriori finite element error control for obstacle problems*, Numer. Math., 99 (2004), pp. 225–249.

[5] C. Bernardi and Y. Maday, *Polynomial interpolation results in Sobolev spaces*, J. Comput. Appl. Math., 43 (1992), pp. 53–80.

[6] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, J. Polit. Economy, 81 (1973), pp. 637–654.

[7] H. Blum, T. Jansen, A. Rademacher, and K. Weinert, *Finite elements in space and time for dynamic contact problems*, Internat. J. Numer. Methods Engrg., 76 (2008), pp. 1632–1644.

[8] D. Braess, *A posteriori error estimators for obstacle problems–another look*, Numer. Math., 101 (2005), pp. 415–421.

[9] H. Brézis, *Problèmes unilatéraux*, J. Math. Pures Appl. (9), 51 (1972), pp. 1–168.

[10] C. Carstensen, *A posteriori error estimate for the symmetric coupling of finite elements and boundary elements*, Computing, 57 (1996), pp. 301–322.

[11] C. Carstensen and E. Stephan, *Adaptive coupling of boundary elements and finite elements*, Modélisation mathématique et analyse numérique, 29 (1995), pp. 779–817.

[12] V. Chawla and T. Laursen, *Energy consistent algorithms for frictional contact problems*, Internat. J. Numer. Methods Engrg., 42 (1998), pp. 799–827.

[13] B. Chen, X. Chen, and C. Kanzow, *A penalized Fischer-Burmeister NCP-*

Bibliography

function: Theoretical investigation and numerical results, Citeseer, 1997.

[14] ———, *A penalized Fischer-Burmeister NCP-function*, Math. Program., 88 (2000), pp. 211–216.

[15] A. Chernov, *Nonconforming boundary elements and finite elements for interface and contact problems with friction - hp-version for mortar, penalty and Nitsche's methods*, PhD thesis, Universität Hannover, 2006.

[16] P. Christensen, A. Klarbring, J. Pang, and N. Strömberg, *Formulation and comparison of algorithms for frictional contact problems*, Internat. J. Numer. Methods Engrg., 42 (1998), pp. 145–173.

[17] P. Christensen and J. Pang, *Reformulation–nonsmooth, piecewise smooth, semismooth and smoothing methods, chapter frictional contact algorithms based on semismooth newton methods*, 1999.

[18] M. Cocou, *Existence of solutions of a dynamic Signorini's problem with nonlocal friction in viscoelasticity*, Z. Angew. Math. Phys., 53 (2002), pp. 1099–1109.

[19] M. Cocou and G. Scarella, *Analysis of a dynamic unilateral contact problem for a cracked viscoelastic body*, Z. Angew. Math. Phys., 57 (2006), pp. 523–546.

[20] M. Costabel, *Boundary integral operators on Lipschitz domains: Elementary results*, SIAM J. Math. Anal., 19 (1988), p. 613.

[21] R. Dautary and J. Lions, *Mathematical analysis and numerical methods for science and technology. Evolution problems I*, vol. 5, Springer-Verlag, 1992.

[22] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Program., 75 (1996), pp. 407–439.

[23] W. Dörfler, *A convergent adaptive algorithm for poisson's equation*, SIAM J. Numer. Anal., 33 (1996), pp. 1106–1124.

[24] W. Dörfler and R. Nochetto, *Small data oscillation implies the saturation assumption*, Numerische Mathematik, 91 (2002), pp. 1–12.

[25] B. Eaves, *On the basic theorem of complementarity*, Math. Program., 1 (1971), pp. 68–75.

[26] C. Erath, S. Ferraz-Leite, S. Funken, and D. Praetorius, *Energy norm based a posteriori error estimation for boundary element methods in two dimensions*, Appl. Numer. Math., 59 (2009), pp. 2713–2734.

[27] R. Falk, *Error estimates for the approximation of a class of variational inequalities*, Mathematics of computation, 28 (1974), pp. 963–971.

[28] S. Ferraz-Leite, C. Ortner, and D. Praetorius, *Convergence of simple adaptive Galerkin schemes based on h- h/2 error estimators*, Numer. Math., 116 (2010), pp. 291–316.

[29] A. Fischer, *A special Newton-type optimization method*, Optimization, 24 (1992), pp. 269–284.

[30] ——, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Math. Program., 76 (1997), pp. 513–532.

[31] R. Glowinski, *Numerical methods for nonlinear variational problems*, (1984).

[32] C. Hager, S. Hüeber, and B. Wohlmuth, *A stable energy conserving approach for frictional contact problems based on quadrature formulas*, Internat. J. Numer. Methods Engrg., 73 (2007), pp. 205–225.

[33] D. Han, *Solving linear variational inequality problems by a self-adaptive projection method*, Appl. Math. Comput., 182 (2006), pp. 1765–1771.

[34] P. Hild and P. Laborde, *Quadratic finite element methods for unilateral contact problems*, Appl. Numer. Math., 41 (2002), pp. 401–421.

[35] P. Houston, D. Schötzau, and P. Thomas, *Energy norm a posteriori error estimation of hp-adaptive discontinuous Galerkin methods for elliptic problems*, Math. Models Methods Appl. Sci., 17 (2007), pp. 33–62.

[36] P. Houston and E. Süli, *A note on the design of hp-adaptive finite element methods for elliptic partial differential equations*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 229–243.

[37] S. Hüeber, *Discretization techniques and efficient algorithms for contact problems*, PhD thesis, Universität Stuttgart, 2008.

[38] S. Hüeber and B. Wohlmuth, *A primal–dual active set strategy for non-linear multibody contact problems*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 3147–3166.

[39] K. Ito and K. Kunisch, *Parabolic variational inequalities: The Lagrange multiplier approach*, J. Math. Pures Appl. (9), 85 (2006), pp. 415–449.

[40] C. Kanzow, *Inexact semismooth Newton methods for large-scale complementarity problems*, Optim. Methods Softw., 19 (2004), pp. 309–325.

[41] A. Keese, *Numerical solution of systems with stochastic uncertainties: a general purpose framework for stochastic finite elements*, PhD thesis, Technische Universität Braunschweig, 2004.

[42] H. Khenous, *Problémes de contact unilatéral avec frottenment de Coulomb en élastostatique et élastodynamique. Etude mathématique et résolution numérique*,

PhD thesis, INSA de Toulouse, France, 2005.

[43] H. KHENOUS AND ET AL., *Mass redistribution method for finite element contact problems in elastodynamics*, Eur. J. Mech. A-Solids, 27 (2008), pp. 918–932.

[44] J. KIM, *A boundary thin obstacle problem for a wave equation*, Comm. Partial Differential Equations, 14 (1989), pp. 1011–1026.

[45] D. KINDERLEHRER AND G. STAMPACCHIA, *An introduction to variational inequalities and their applications*, Society for Industrial Mathematics, 2000.

[46] S. KOU AND H. WANG, *Option pricing under a double exponential jump diffusion model*, Management Science, (2004), pp. 1178–1192.

[47] A. KREBS, *On solving nonlinear variational inequalities by p-version finite elements*, PhD thesis, Universität Hannover, 2004.

[48] A. KREBS AND E. STEPHAN, *A p-version finite element method for nonlinear elliptic variational inequalities in 2D*, Numerische Mathematik, 105 (2007), pp. 457–480.

[49] K. KUNISCH, *Semi-smooth Newton method for non-differentiable optimization problems.* online, February 19th 2008. Lipschitz Lectures.

[50] K. KUNISCH AND F. RENDL, *An infeasible active set method for quadratic problems with simple bounds*, SIAM J. Optim., 14 (2003), pp. 35–52.

[51] P. LABORDE AND Y. RENARD, *Fixed point strategies for elastostatic frictional contact problems*, Math. Methods Appl. Sci., 31 (2008), pp. 415–441.

[52] B. LAMICHHANE AND B. WOHLMUTH, *Biorthogonal bases with local support and approximation properties*, Math. Comp., 76 (2007), pp. 233–249.

[53] T. LAURSEN AND V. CHAVLA, *Design of energy conserving algorithms for frictionless dynamic contact problems*, Int. J. Numer. Meth. Eng., 40 (1997), pp. 863–886.

[54] G. LEBEAU AND M. SCHATZMAN, *A wave problem in a half-space with a unilateral constraint at the boundary*, J. Differential Equations, 53 (1984), pp. 309–361.

[55] J. LIONS, *Quelques méthodes de résolution des problemes aux limites non linéaires*, vol. 76, Dunod Paris, 1969.

[56] J. LIONS AND G. STAMPACCHIA, *Variational inequalities*, Communications on Pure and Applied Mathematics, 20 (1967), pp. 493–519.

[57] M. MAISCHAK, *The analytical computation of the Galerkin elements for the Laplace, Lamé and Helmholtz equation in 2D-BEM.* Technical report to Maiprogs.

[58] ——, *The analytical computation of the Galerkin elements for the Laplace, Lamé and Helmholtz equation in 3D-BEM.* Technical report to Maiprogs.

[59] M. MAISCHAK AND E. STEPHAN, *Adaptive hp-versions of BEM for Signorini problems*, Appl. Numer. Math., 54 (2005), pp. 425–449.

[60] O. L. MANGASARIAN, *Equivalence of the complementarity problem to a system of nonlinear equations*, SIAM J. Appl. Math., 31 (1976), pp. 89–92.

[61] J. A. C. MARTINS AND J. T. ODEN, *Existence and uniqueness results for dynamic contact problems with nonlinear normal and friction interface laws*, Nonlinear Anal., 11 (1987), pp. 407–428.

[62] J. MELENK AND B. WOHLMUTH, *On residual-based a posteriori error estimation in hp-FEM*, Adv. Comput. Math., 15 (2001), pp. 311–331.

[63] K. MOON, R. NOCHETTO, T. VON PETERSDORFF, AND C. ZHANG, *A posteriori error analysis for parabolic variational inequalities*, Mathematical Modelling and Numerical Analysis, 41 (2007), pp. 485–511.

[64] L. QI, *C-differentiability, C-differential operators and generalized Newton methods*, Research report, School of Mathematics, The University of New South Wales, Sydney, Australia (January 1996), (1996).

[65] L. QI AND H. JIANG, *Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations*, Mathematics of Operations Research, 22 (1997), pp. 301–325.

[66] B. RIVIÈRE, *Discontinuous Galerkin methods for solving elliptic and parabolic equations: theory and implementation*, Siam, 2008.

[67] B. RIVIÈRE, M. WHEELER, AND V. GIRAULT, *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. Part I*, Comput. Geosci., 3 (1999), pp. 337–360.

[68] ——, *A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems*, SIAM J. Numer. Anal., 39 (2002), pp. 902–931.

[69] D. SCHÖTZAU AND C. SCHWAB, *Time discretization of parabolic problems by the hp-version of the discontinuous Galerkin finite element method*, SIAM J. Numer. Anal., 38 (2001), pp. 837–875.

[70] A. SCHRÖDER, *Fehlerkontrollierte adaptive h-und hp-Finite-Elemente-Methoden für Kontaktprobleme*, PhD thesis, Universität Dortmund, 2006.

[71] A. SCHRÖDER, H. BLUM, A. RADEMACHER, AND H. KLEEMANN, *Mixed fem of higher order for contact problems with friction*, J. Numer. Anal. Model., 8 (2011), pp. 302–323.

[72] O. STEINBACH, *Numerische Näherungsverfahren für elliptische Randwertprobleme:*

Bibliography

Finite Elemente und Randelemente, Vieweg + Teubner, 2003.

[73] H. VAN TREES, *Detection, estimation, and modulation theory, Part I*, vol. 218, New York: Wiley, 1968.

[74] R. VERFÜRTH, *A review of a posteriori error estimation and adaptive mesh-refinement techniques*, vol. 1, Wiley-Teubner, 1996.

[75] T. WERDER, K. GERDES, D. SCHÖTZAU, AND C. SCHWAB, *hp-Discontinuous Galerkin time stepping for parabolic problems*, Comp.Methods Appl. Mech. Engrg., 190 (2001), pp. 6685–6708.

[76] B. WOHLMUTH, *A mortar finite element method using dual spaces for the Lagrange multiplier*, SIAM J. Numer. Anal., 38 (2001), pp. 989–1012.

[77] ——, *Discretization methods and iterative solvers based on domain decomposition*, Springer Verlag, 2001.

[78] E. ZEIDLER, *Nonlinear functional analysis and its applications, part II/A: Nonlinear monotone operators*, Springer Verlag, 1989.

[79] S.-P. ZHU, *An exact and explicit solution for the valuation of american put options*, Quantitative Finance, 6 (2006), pp. 229–242.

Curriculum Vitae

Personal Data:

Name:	Lothar Banz
Born:	12th August 1985, Hannover, Germany
Citizenship:	German

Education:

since 12/2008		Ph.D. Student at the Institute of Applied Mathematics, Leibniz University Hannover, Germany
2004-2010		Economics/Business studies, Leibniz University Hannover, Germany
	10/2010	Diplom in Economics/Business studies ($\sim$ Master)
2007-2008		Mathematics studies, Brunel University West London, England
	11/2008	Master in Mathematics
2004-2007		Mathematics studies, Leibniz University Hannover
	09/2007	Bachelor in Mathematics
2003-2004		First year undergraduate studies in Mathematics and Economics/Business studies for high school students, Leibniz University Hannover
1991-2004		School attendance in Hannover, Germany
2001-2002		School attendance in Albury, New South Wales, Australia

Employment:

since 12/2008	Research Staff at the Institute of Applied Mathematics, Leibniz University Hannover, Germany
08/2011	Visiting lecturer, Universidad del Norte, Columbia

Award/Voluntary Service:

since 01/2010		Vice delegate at the Graduate Academy, Leibniz University Hannover
	11/2008	John Gegory Memorial Price, Brunel University West London
04/2006-03/2007		Vice President of AIESEC Hannover