Consistent FE-Analysis of Elliptic Variational Inequalities

Dissertation

zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften

vorgelegt von

Nicole Klein

aus Herdorf

eingereicht beim Department Mathematik der Fakultät IV

Universität Siegen

29.11.2012

Gutachter: Prof. Dr. Franz-Theo Suttmeier
Prof. Dr. Heribert Blum

Tag der mündlichen Prüfung: 14. März 2013

Gedruckt auf alterungsbeständigem holz- und säurefreiem Papier.

Abstract

Economical mesh structures are of great interest when simulating physical processes using the Finite Elemente Method. They are essential for a fast calculation producing results of high accuracy. In case of restricted problems, many a posteriori estimators which are the indicators for adaptive refinement turn out to be inconsistent in areas where the restriction takes place. The effort of the subject matter is to develop a method to overcome this problem by introducing saddle point formulations and using the Lagrangian multiplier to balance gaps in the error estimations. When dealing with sattle point problems there may arise the problem of unstable systems due to an injured inf-sup-condition, especially in the discrete case. We solve this problem using the Galerkin least squares method. In consequence we get additional terms which also have to be taken into account when developing the a posteriori estimators. To examine the general validity of this method we analyse problems of different type. That means linear and nonlinear problems with linear or nonlinear restrictions in the primal or dual variable, respectively. In all cases, the resulting adaptive mesh structures turn out out to be very efficient since they outline critical zones of the underlying problems which is confirmed by numerical tests.

Kurzfassung

In der Simulation von Fertigungsprozessen mit Hilfe der Finite Elemente Technik sind ökonomische Gitterstrukturen von großem Interesse da sie für eine schnelle Berechnung bei gleichzeitiger hoher Genauigkeit unverzichtbar sind. Bei der Behandlung von restringierten Problemen tritt allerdings häufig das Problem auf, dass die Fehlerschätzer, die als Indikatoren für eine adaptive Gitterstruktur dienen, in den restringierten Bereichen inkonsistent sind, was zu ineffizienten Verfeinerungen führt. Für die Optimierung der Fehlerschätzer führen wir in dieser Arbeit Sattelpunktformulierungen ein um mit Hilfe des Lagrangeparameters die entstehenden Inkonsistenzen auszugleichen. Bei der Verwendung von gemischten System kann hier das Problem der Instabilität durch eine verletzte inf-sup Bedingung auftreten, das wir durch die Verwendung des Galerkin least squares Ansatzes beheben. Hieraus resultieren allerdings zusätzliche Terme in der Problemformulierung, die bei der Entwicklung der a posteriori Schätzer ebenfalls berücksichtigt werden müssen. Um die allgemeine Gültigkeit unserer Methode zu untersuchen, analysieren wir Probleme mit unterschiedlichen Eigenschaften, das heißt lineare und nichtlineare Gleichungen mit linearen bzw. nichtlinearen Restriktionen in der primalen oder dualen Variable. In numerischen Tests stellt sich heraus, dass wir in allen Fällen effiziente Gitterstrukturen erzielen, die die kritischen Zonen der jeweiligen Probleme durch hohe Verfeinerung herausarbeiten.

Contents

1 Introduction

Production processes in industry are often based on central problems like contact, torsion and of course material laws. In case of contact for example, there are many applications including deformation processes (milling, deep drawing) or separation (beveling, cutting, planing). Numerical simulations often help to work efficiently because they replace complex experimental runs which cause a maximum of metal loss and hence high costs. In general, those simulations are based on the Finite Element Method. That means constructing a mathematical model of the real situation based on physical laws and then discretising the resulting variational equation or inequality by finite elements which are basic functions of an underlying mesh with a triangulation of triangle or quadrangle elements. The geometry of the workpieces is often very complex, so that getting good results by very fine meshes takes a lot of calculation time and memory capacity. Thus, adaptive methods are of great interest for production processes. In regions where the solution is smooth, a coarse mesh offers a fast calculation and suffices to get precise results. Fine mesh structures are necessary if there are critical zones or singularities. One can refine these zones on a heuristic level, however it would be very helpful if the program was factored in a criteria that makes it find critical zones itself and refine them automatically. That is the effort of residual based error estimators. There have already been many studies to develop such estimators (for an overview see for example [Ve96], [AO00]). In most cases we have to take care of restrictions in the variational inequality. The contact problem for example forbids a penetration of the workpiece into the obstacle. Taking these conditions into account the residual based error indicators often get inconsistent in the restricted regions due to a missing Galerkin orthogonality. That results in inefficient mesh structures. As an example we take an elastic membrane which is fixed at the boundaries and push it down by a body force f onto a planar obstacle. In regions of contact we find an over-refinement of the mesh since we do not expect errors worth mentioning there. The effect is a higher calculation time without getting solutions that are more precise compared to a triangulation with a coarse mesh in the contact zone.

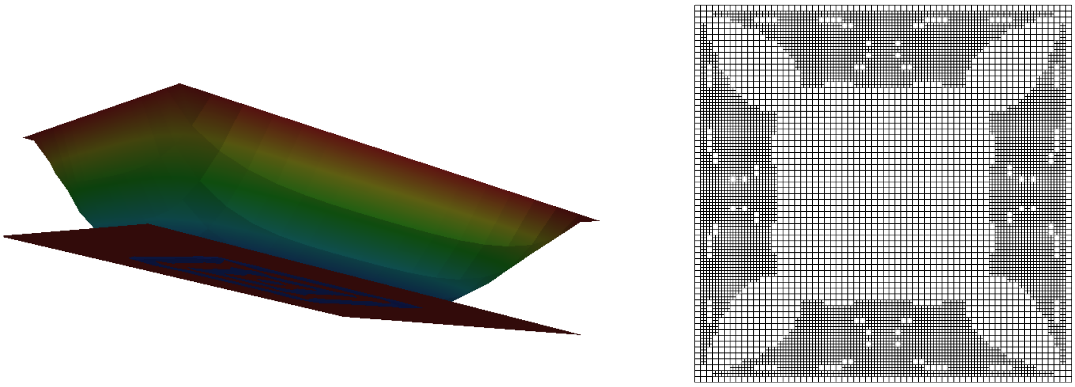


Figure 1.1: A membrane is pushed onto an obstacle (left). Inconsistent a posteriori error estimators cause an over-refinement in the contact zone (right).

The effort of the work at hand is to develop residual based error estimators which eliminate these inconsistencies. For this purpose we take principle examples of restricted problems and reformulate the variational inequalities resulting in saddle point problems. On this base we derive improved error estimators with the help of Lagrangian multipliers. The restrictions can now occur in different situations, so for a general study we will take an example for each situation and analyse the effects that are caused by the new estimators. The scheme in Figure 1.2 shows the different cases restrictions may occur.

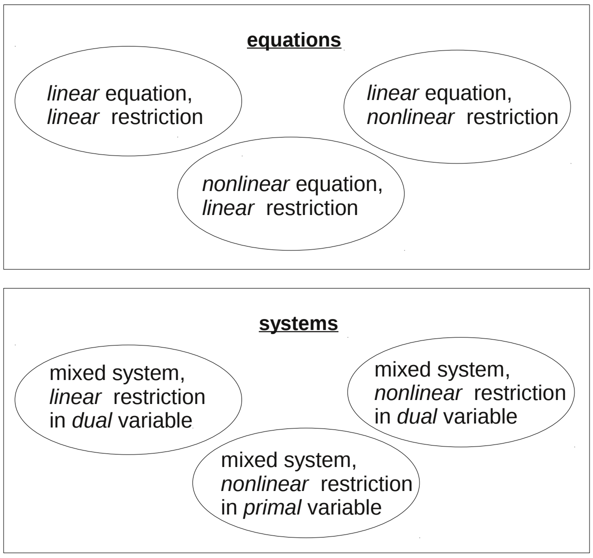

equations
linear equation, linear restriction
linear equation, nonlinear restriction
nonlinear equation, linear restriction
systems
mixed system, linear restriction in dual variable
mixed system, nonlinear restriction in dual variable
mixed system, nonlinear restriction in primal variable

Figure 1.2: Possible forms of restrictions on equations or systems.

The achievements are mesh structures that now concentrate on the critical zones which are in general areas between restricted and non-restricted regions. In case of the contact problem the over-refinement in the contact zone is balanced.

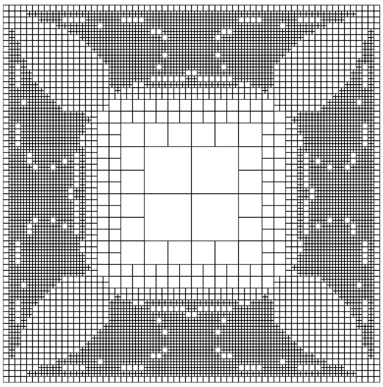


Figure 1.3: Adaptive mesh for the obstacle problem shown in Figure 1.1 (left) generated by an estimator which balances the inconsistencies by the use of a Lagrangian multiplier.

A problem that may arise when discretising mixed variational formulations is instability of the system. In most cases we get a solution due to the iterative solvers we use. However, the effect of the injured stability conditions are oscillations that influence the physical relevance of the resulting values. Again, taking the contact problem as an example, we observe oscillations in the Lagrangian parameter which has the physical meaning of a counter force to the body force f in regions of contact in a stable system. Therefore, a second assignment of this work is to stabilise the mixed discrete systems and to ensure its unique solvability. The technique used here is the Galerkin least squares method based on the studies of [HFB86]. Of course the additional terms that result from the stabilisation have an influence on the error estimators and have to be taken into account there. Figure 1.4 shows the Lagrangian multiplier in the contact zone having oscillations in the unstabilised system (left). On the right there are the resulting values of the parameter after stabilisation was performed.

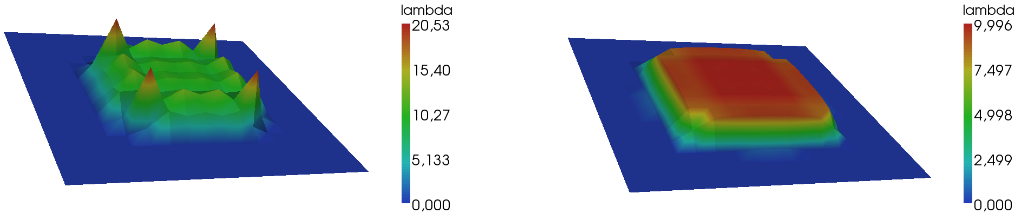


Figure 1.4: Left: Value of the dual variable in the unstabilised system showing non-physical oscillations. Right: Stabilised system by the least squares method. The dual parameter gets very smooth and retrieves its physical relevance since the body force used here was the constant value $f = -10$.

The structure of this work is the following:

In Chapter 2 some basics about restricted minimisation problems and their solvability are introduced. After the transfer to variational inequalities and their corresponding discrete systems is depicted, we give a short overlook of saddle point formulations and the theory of stable systems in the continuous as well as in the discrete case. Finally, we introduce the topic of Sobolev spaces and their main aspects.

The first example that is studied in Chapter 3 is the linear contact problem including a linear restriction. We present the problem within a variational inequality and examine existence and solvability. An a posteriori estimator is introduced. When formulating the mixed system we eliminate instabilities by presenting a least squares stabilisation. Again, an a posteriori estimator is developed which includes the Lagrangian multiplier and takes account of the stabilisation. It turns out to be consistent, which is confirmed by numerical tests at the end of the chapter. Furthermore, we present some solving algorithms and compare them in a benchmark.

Based on the example of Chapter 3, the situation in the fourth chapter describes a nonlinear contact problem, that means a linear restriction on the nonlinear equation. Once more we start with the introduction of the variational formulation and the existence proof. We establish the SQP-Method for solving nonlinear systems and show the derivation of an appropriate a posteriori estimator. This one is again compared with the estimator that we develop after presenting the corresponding saddle point formulation and the suitable least squares stabilisation of the nonlinear problem.

An example of a linear problem with a nonlinear restriction is the so called torsion problem which is studied in Chapter 5. We follow the scheme of the previous chapters and present the variational form of the problem and the existence results before introducing the corresponding saddle point formulation. The torsion problem is often presented with the restriction on a flow rule with constant yield condition. In addition to that we offer a theory with respect to a flow rule with a variable yield condition that may vary in space. The achieved error estimator turns out to produce efficient meshes which we compare to grids generated by known estimators. Furthermore, we give a short sketch of the construction of goal orientated estimators in case of the torsion problem.

Chapter 6 treats the aspect of plasticity. The primal and dual formulations of Strang's problem are studied which consist of a linear system with nonlinear restrictions on the primal variable. We point to the problems that occurred in the error analysis of the problem so far and start with formulating a mixed system for the primal formulation followed by a new a posteriori estimation. Following that line for the dual scheme, too, we first give a regularised version of the dual formulation to ensure existence of all components of the developed error estimator. In addition we require a stabilisation, which is again implemented by the least squares method. Numerical tests confirm the expected results and offer excellent mesh structures created by the new estimators.

As a last example we transfer the theory of the Lagrangian technique on the boundary of a domain in Chapter 7. A standard example in this case is the simplified Signorini problem which is again a contact situation but related to the boundary. Existence and uniqueness is ensured for the variational problem as well as for the stabilised saddle point formulation. It turns out that a consistent error estimator related to this problem works efficiently just as the utilized stabilisation does.

Parallel to this thesis we worked on a project in cooperation with the TU Dortmund. Several aspects of deep drilling were examined, especially in view of heat flux. Chapter 8 gives a short project-overview and the use of the presented methods therein.

At last, we give a short conclusion of the results we achieved in this work and mention some aspects and ideas for further studies.

Systems with nonlinear or linear restriction on the dual variable are not mentioned in this work. These cases have already been studied in the papers [HKS11], dealing with the special case of Bingham flow and [GHS10] which analyses the Stokes flow with cavitation. The latter is also presented in [Gi12] more precisely.

The numerical results are made in deal.II (Differential Equations Analysis Libary), which is a toolkit of C++ dealing with finite elements (for more information see www.dealii.org).

2 Basic principles

We introduce some basics we need in this work and start with the principle of restricted minimisation problems and their solvability. After introducing the variational inequalities and their corresponding discretisations, an equivalent formulation leads to the so called saddle point system which is also presented here followed by its theory of stable systems in the continuous as well as in the discrete case. As a last introducing topic we mention the Sobolev spaces and their main aspects. These principles among others can be found in Hackbusch [GR92], Braess [Br07], Céa [Ce78], Glowinski [Gl83] or [GLT76].

2.1 Elliptic minimisation problems

If we study physical problems we want them to take a state of best energetic situation. That means we look for the minimum u fulfilling

$$J(u) = \min_{v \in K} J(v)$$

of a functional $J : V \to \mathbb{R}$. v describes, for example, some kind of velocity or some other input values of a function space V. K can be defined by V itself, $K := V$, or by a subset, $K \subset V$, if we have restrictions on V, for instance a maximal displacement in a contact problem or a constraint of stress in an elasto-plastic problem. So choosing $K \subset V$ we have a restricted problem, otherwise an unrestricted one. We consider V to be a Hilbert space and by $\|v\|$ we denote the corresponding norm of v in V.

On V we take a continuous, symmetric bilinear form $a : V \times V \to \mathbb{R}$, which thus satisfies

$$|a(u,v)| \leq c\|u\| \, \|v\|, \quad u, v \in V,$$
$$a(u,v) = a(v,u), \quad u, v \in V$$

with a constant value $c > 0$. $a(\cdot,\cdot)$ is called elliptic if there is a constant $\alpha > 0$ such that

$$a(u,u) \geq \alpha\|u\|^2, \quad u \in V.$$

We take $K \subset V$ with K being a closed convex and nonempty subset of V. Let V' denote the dual of V and $\langle f, v\rangle := (f,v)$ the dual pairing between $f \in V'$ and $v \in K$. An elliptic minimisation problem of first kind is of the form:

$$\min_{v\in K} J(v) = \min_{v\in K}\{\frac{1}{2}a(v,v) - \langle f, v\rangle\}. \tag{2.1}$$

2.1.1 Existence and uniqueness of minimisation problems

One of the essential questions is about the existence of a solution for the minimisation problem given above. First we introduce a variational formulation of the problem and then make a statement about existence and uniqueness.

Definition 2.1. *Let V be a normed space with norm $\|\cdot\|$, then for a $w \in V$, J is Fréchet-differentiable in w if there exists a $J'(w) \in V'$ such that*

$$\forall v \in V : \;\; J(w+v) - J(w) = \langle J'(w), v\rangle + \varrho(v),$$

with a function $\varrho : V \to \mathbb{R}$ and $|\varrho(v)| = o(\|v\|)$ for $v \to 0$. $J'(w)$ is called the Fréchet-derivative in w.

Minimisation problems always have an equivalent variational formulation:

Theorem 2.1. *Suppose K is a closed, nonempty convex subset of a Banach space. $J : K \subset V \to \mathbb{R}$ is an in u Fréchet-differentiable convex functional. Then $u \in K$ is a minimum in J,*

$$J(u) \leq J(v) \quad \forall v \in K,$$

if and only if

$$u \in K \text{ and } \langle J'(u), v-u \rangle \geq 0 \quad \forall v \in K.$$

If $K = V$ we have

$$u \in K \text{ and } \langle J'(u), v \rangle = 0 \quad \forall v \in K.$$

If J is strict convex there exists at most one solution.

Let l be a continuous functional $l : V' \times V \to \mathbb{R}$ such that $l(v) = \langle f, v \rangle$. The functional $J(\cdot)$ of a minimisation problem of first kind is Fréchet-differentiable for all $w \in V$:

$$\begin{aligned} J(w+v) - J(w) &= \frac{1}{2}a(w+v, w+v) - l(w+v) - \frac{1}{2}a(w,w) - l(w) \\ &= a(w,v) - l(v) + \frac{1}{2}a(v,v). \end{aligned}$$

The continuity of the bilinear form $a(\cdot, \cdot)$ implies

$$a(v,v) \leq c\|v\|^2.$$

So we have $|a(v,v)| = o(\|v\|^2)$ for $\|v\| \to 0$. Therefore the Fréchet-derivative of J is

$$\langle J'(w), v \rangle = a(w,v) - l(v).$$

That implies

Theorem 2.2. *Let K be a closed convex nonempty subset of a Hilbert space V. Problem* (2.1) *is equivalent to the variational inequality:*
Find $u \in K$ such that

$$a(u, v-u) \geq l(v-u) \quad \forall v \in K. \tag{2.2}$$

Theorem 2.3 (Lions and Stampacchia [LS67])**.** *Problem* (2.2) *has a unique solution.*

A proof can also be found in Glowinski [Gl83]. Here, existence is proven by a fixed point problem for the more general case that the bilinearform $a(\cdot, \cdot)$ is not necessarily symmetric.

If we consider nonlinear problems of the form

$$u \in K, \quad \langle A(u) - f, v - u \rangle \geq 0 \quad \text{for } v \in K, \tag{2.3}$$

we introduce the following theorem that ensures the existence of $u \in K$ in (2.3). A proof can be found in [AH09]:

Theorem 2.4. *Let V be a real Hilbert space, and $K \subset V$ be nonempty, closed and convex. Assume $A : V \to V'$ is strongly monotone and Lipschitz continuous. Then for any $f \in V'$, the variational inequality* (2.3) *has a unique solution $u \in K$ which depends Lipschitz continuously on f.*

2.1.2 The discrete case

For numerical studies we need to approximate the continuous inequality by a discrete one. We keep all assumptions on V, K, l and a and we are now interested in an approximation of

$$a(u, v - u) \geq l(v - u) \quad \forall v, u \in K.$$

Like it is described in Glowinski [Gl83] we suppose that we have a parameter $h > 0$ converging to 0 and a family $\{V_h\}_h$ of closed subspaces of V. We are also given a family $\{K_h\}_h$ of closed convex nonempty subsets of V with $K_h \subset V_h \; \forall\, h$ (in general, we do not assume $K_h \subset K$) such that $\{K_h\}_h$ satisfies the following two conditions:

- If $\{v_h\}_h$ is such that $v_h \in K_h \; \forall h$ and $\{v_h\}_h$ is bounded in V, then the weak cluster points of $\{v_h\}_h$ belong to K.
- There exists $\chi \subset V, \overline{\chi} = K$ and $r_h : \chi \to K_h$ such that $\lim_{h \to 0} r_h v = v$ strongly in V, $\forall v \in \chi$.

The approximation of (2.2) is

$$a(u_h, v_h - u_h) \geq l(v_h - u_h) \quad \forall v_h, u_h \in K_h. \tag{2.4}$$

Theorem 2.1.1. *Problem* (2.4) *has a unique solution.*

Proof. In Theorem 2.3, taking V to be V_h and K to be K_h, we can perform the proof like in the continuous case. □

2.2 Approximation of elliptic variational inequalities

Systems (2.2) and (2.4) have unique solutions in K and K_h, respectively. In the theory of variational equalities the Céa-Lemma gives an abstract statement about error analysis. Due to the Galerkin orthogonality, to estimate the error of the Galerkin solution, it suffices to estimate the approximation error $\inf_{v_h \in V_h} \|u - v_h\|$. In case of variational inequalities the Galerkin orthogonality does not hold. Hence, we assume $K_h \subset K$ and formulate the generalised Céa-Lemma for inequalities:

Lemma 2.1. *Let $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ be a continuous and elliptic bilinearform on the Hilbert space V with $K \subset V$ being a convex, nonempty and closed subset of V and $l \in V'$ a continuous linear functional. There is a constant $c > 0$ independent of h and u, such that*

$$\|u - u_h\| \leq c\{\inf_{v_h \in K_h} [\|u - v_h\| + |a(u, v_h - u) - l(v_h - u)|^{\frac{1}{2}}]\}.$$

For a proof see [AH09]. The first part of the right hand side is the one describing the approximation error as usual. In addition there is a second term which is responsible for the inconsistency we will observe in the common a posteriori error estimates related to variational inequalities.

2.3 Lagrangian and Lagrange multipliers

In several cases it is more expedient to analyse the saddle point formulation of a problem with restrictions. The advantage is that we can look for a solution in V instead of a convex subset $K \subset V$. Appropriate projections on the Lagrangian multiplier ensuring the additional constraints on the solution are much easier to handle compared to the one on the primal variable in K. Our intention is to describe the condition of an element v belonging to the constraint set K by means of an inequality condition for a suitable functional of two arguments. Following Céa [Ce78], we introduce a cone Λ in an appropriate vector space and a suitable functional Φ on $V \times \Lambda$ in such a way that $\Phi(v, \mu) \leq 0$ is equivalent to the fact that v belongs to K. That means a transformation of (2.1) to a mini-max-problem for the functional

$$\mathcal{L}(v, \mu) = J(v) + \Phi(v, \mu) \quad \text{on } V \times \Lambda. \tag{2.5}$$

$\mathcal{L}$ is called the Lagrangian associated to problem (2.1). Under suitable hypothesis, if (u, λ) is the solution of (2.5), then u will be the solution of the minimisation problem (2.1). In Sections 2.3.1 and 2.3.2 we mainly follow [Ce78].

2.3.1 Saddle points

A pair $(u, \lambda) \in V \times \Lambda$ is called a saddle point of the functional $\mathcal{L} : V \times \Lambda \to \mathbb{R}$, if

$$\forall v \in V,\ \forall \mu \in \Lambda : \ \ \mathcal{L}(u, \mu) \leq \mathcal{L}(u, \lambda) \leq \mathcal{L}(v, \lambda).$$

In other words, $(u, \lambda) \in V \times \Lambda$ is a saddle point of $\mathcal{L}$ if the point u is a minimum for the functional

$$\mathcal{L}(\cdot, \lambda) : V \ni v \mapsto \mathcal{L}(v, \lambda) \in \mathbb{R},$$

and if the point λ is a maximum for the functional

$$\mathcal{L}(u, \cdot) : \Lambda \ni \mu \mapsto \mathcal{L}(u, \mu) \in \mathbb{R},$$

that means

$$\sup_{\mu \in \Lambda} \mathcal{L}(u, \mu) = \mathcal{L}(u, \lambda) = \inf_{v \in V} \mathcal{L}(v, \lambda).$$

Following [ET99] there holds

Theorem 2.5. *Let V and E be normed spaces with $\emptyset \neq \Lambda \subset E$ convex. Let $(u, \lambda) \in V \times \Lambda$ and the functional $\mathcal{L} : V \times \Lambda \to \mathbb{R}$ satisfy* *(i)* $\mathcal{L}_u := \mathcal{L}(u, \cdot)$ *concave and Fréchet-differentiable in* λ*,*

(ii) $\mathcal{L}_\lambda := \mathcal{L}(\cdot, \lambda)$ *convex and Fréchet-differentiable in* u*.*

Then (u, λ) *is saddle point of* $\mathcal{L}$ *if and only if*

$$\forall v \in V : \langle \mathcal{L}'_\lambda(u), v \rangle = 0 \tag{2.6}$$

$$\forall \mu \in \Lambda : \langle \mathcal{L}'_u(\lambda), \mu - \lambda \rangle \leq 0. \tag{2.7}$$

2.3.2 Saddle point formulation for minimisation problems

We assume a Hilbert space V, a set of constraints $K \subset V$ and a vector space E with a subset $\Lambda \subset E$ which is a cone with vertex at 0 and left invariant by the action of $\mathbb{R}^+$. Furthermore, there exists a mapping

$$\Phi : V \times \Lambda \to \mathbb{R}$$

such that

(i) the mapping $\Lambda \ni \mu \mapsto \Phi(v, \mu) \in \mathbb{R}$ is homogeneous of degree one, i.e.

$$\Phi(v, \varrho\mu) = \varrho\Phi(v, \mu) \quad \forall \varrho \geq 0, \tag{2.8}$$

(ii) a point $v \in V$ belongs to K if and only if

$$\Phi(v, \mu) \leq 0 \quad \forall \mu \in \Lambda. \tag{2.9}$$

Theorem 2.6. *Let* V *be a normed space and* K *be a subset of* V *such that we can find a cone* Λ *with vertex at 0 (in a suitable vector space) and a function* $\Phi : V \times \Lambda \to \mathbb{R}$ *satisfying (i) and (ii). Then there holds*

$$\inf_{v \in K} J(v) = \inf_{v \in V}(J(v) + \sup_{\mu \in \Lambda} \Phi(v, \mu)) = \inf_{v \in V} \sup_{\mu \in \Lambda}(J(v) + \Phi(v, \mu)). \tag{2.10}$$

Proof. Céa [Ce78] (Chapter V, Prop.1.1). □

Essential for the proof is the fact that for the Lagrangian functional, there holds

$$\sup_{\mu\in\Lambda}\Phi(v,\mu)=\begin{cases}0 & \text{if } v\in K,\\ +\infty & \text{if } v\notin K.\end{cases} \tag{2.11}$$

Definition 2.2. *The Lagrangian associated to the minimisation problem for J (with constraints defined by the set K) is the functional $\mathcal{L}: V\times\Lambda\to\mathbb{R}$ defined by*

$$\mathcal{L}(v,\mu)=J(v)+\Phi(v,\mu).$$

$\mu\in\Lambda$ is called the Lagrange multiplier.

We can prove: If (u,λ) is a saddle point for $\mathcal{L}$ then we have

$$\sup_{\mu\in\Lambda}\inf_{v\in V}\mathcal{L}(v,\mu)=\mathcal{L}(u,\lambda)=\inf_{v\in V}\sup_{\mu\in\Lambda}\mathcal{L}(v,\mu).$$

That implies the Lagrangian problem:
Find $(u,\lambda)\in V\times\Lambda$ such that

$$\mathcal{L}(u,\lambda)=\sup_{\mu\in\Lambda}\inf_{v\in V}\mathcal{L}(v,\mu) \tag{2.12}$$

is the associated dual problem of the minimisation problem (2.1).

Theorem 2.7. *If there exists a $\lambda\in\Lambda$ such that $(u,\lambda)\in V\times\Lambda$ is a saddle point for the Lagrangian associated to the minimisation problem, then u is a solution of the minimisation problem and λ is a solution of the dual problem* (2.12).

Proof. See Céa [Ce78] (Chapter V, Prop. 1.4). □

To get an equivalent variational formulation of the saddle point problem we obtain with the help of Theorem 2.5:

Theorem 2.8. *Let V and E be normed spaces with $\Lambda\subset E$ and $K\subset V$ convex. The functional $\Phi: V\times\Lambda\to\mathbb{R}$ is the Lagrangian functional. For $(u,\lambda)\in V\times\Lambda$ and the convex functional $J: V\to\mathbb{R}$ satisfying* *(i)* *J is in u Fréchet-differentiable.*

(ii) *$\Phi_u := \Phi(u, \cdot)$ concave and Fréchet-differentiable in λ.*

(iii) *$\Phi_\lambda := \Phi(\cdot, \lambda)$ convex and Fréchet-differentiable in u.*

u is the solution of (2.1) *if*

$$\forall v \in V : \langle J'(u) + \Phi'_\lambda(u), v\rangle = 0 \tag{2.13}$$

$$\forall \mu \in \Lambda : \langle \Phi'_u(\lambda), \mu - \lambda\rangle \leq 0. \tag{2.14}$$

In the test examples appearing in this work, the subset K will always be of the following form:

$$K := \{v \in V |\, \omega(v) \leq g\}$$

with g in a normed space E' and $\omega \in L(V, E')$. Then $g - \omega(v)$ is in a convex set $G \subset E'$ and the Lagrangian functional is chosen as

$$\Phi(v, \mu) := \langle \mu, \omega(v) - g\rangle. \tag{2.15}$$

In order to prove (2.15) to be a Lagrangian functional, $\Phi(\cdot, \cdot)$ has to fulfill (2.8) and (2.9) (see for example [Sc05]):

Lemma 2.2. *Let V be a normed space, $G \subset V$ a closed, convex cone and*

$$G' := \{\mu \in V' | \forall v \in G : \langle \mu, v\rangle \geq 0\}.$$

Then there holds

$$v \in G \quad \Leftrightarrow \quad \forall \mu \in G' : \langle \mu, v\rangle \geq 0.$$

To prove (2.8) for Φ, for $\alpha > 0$, $v \in V$ and $\mu \in \Lambda$ we can write

$$\Phi(v, \alpha\mu) = \langle \alpha\mu, \omega(v) - g\rangle = \alpha\langle \mu, \omega(v) - g\rangle = \alpha\Phi(v, \mu).$$

Following Lemma 2.2 $g - \omega(v) \in G$ and therefore $v \in K$ if and only if

$$0 \leq \langle \mu, g - \omega(v)\rangle = -\Phi(v, \mu) \quad \forall \mu \in G'. \tag{2.16}$$

Here it is $G' = \Lambda$, so all constraints are fulfilled.

Choosing the same notation as above we consider the problem of the following form: Find $(u, \lambda) \in V \times \Lambda$ such that

$$\mathcal{L}(v, \mu) := \frac{1}{2}a(v, v) - l(v) + b(v, \mu) - \langle \mu, g \rangle, \tag{2.17}$$

or in a more abstract way:

$$\mathcal{L} : V \times E \to V' \times E' \tag{2.18}$$

$$(u, \lambda) \mapsto (f, g), \tag{2.19}$$

where V and E are reflexive spaces, $\Lambda \subset E$ closed and convex and $g \in E'$. $b : V \times E \to \mathbb{R}$ is a continuous bilinear-form and $l(v) = \langle f, v \rangle$. If we want to ensure a unique saddle point $(u, \lambda) \in V \times \Lambda$ that fulfills (2.12), we have to comply with some additional conditions:

Theorem 2.9. *The saddle point problem* (2.17) *describes an isomorphism* $\mathcal{L} : V \times E \to V' \times E'$ *if and only if the following conditions are fulfilled:*

(i) The bilinear-form a is on $X := \{v \in V : b(v, \mu) = 0$ for $\mu \in E\}$ V-elliptic, i.e. there exists an $\alpha > 0$ such that

$$a(v, v) \geq \alpha \|v\|^2 \textit{ for } v \in X.$$

(ii) The bilinear-form b fulfills the inf-sup-condition:

$$\exists \beta > 0 : \inf_{\mu \in E} \sup_{v \in V} \frac{b(v, \mu)}{\|v\| \, \|\mu\|} \geq \beta, \quad v, \mu \neq 0.$$

If we define the continuous linear operator $B : V \to E'$ with

$$\langle Bv, \mu \rangle = b(v, \mu) \quad \forall v \in V, \ \mu \in E,$$

condition (ii) is equivalent to the statement that $Im(B)$ is closed in E' and B is surjective.

If the conditions of Theorem 2.9 are ensured, we can make a statement of stability for perturbation of the system:

Theorem 2.10. *If* $(u,\lambda) \in V \times \Lambda$ *is the solution of problem* (2.17) *under the conditions of Theorem 2.9 then there hold the stability estimates*

$$\|u\|_V \leq \frac{1}{\alpha}\|f\|_{V'} + \frac{1}{\beta}(\frac{c}{\alpha}+1)\|g\|_{E'},$$
$$\|\lambda\|_E \leq \frac{1}{\alpha}(\frac{c}{\beta}+1)\|f\|_{V'} + \frac{c}{\beta^2}(\frac{c}{\alpha}+1)\|g\|_{E'}.$$

2.3.3 The discrete saddle point problem

We approximate V by the finite dimensional subset V_h like described in Section 2.1.2. Let $J, I : V_h \to \mathbb{R}$ be convex functionals on V_h and $K_h = \{v_h | v_h \in V_h;\ I(v_h) \leq 0\}$. Then K_h is a convex set and our minimisation problem is defined by:
Find $u_h \in K_h$ such that

$$J(u_h) = \inf_{v_h \in K_h} J(v_h). \tag{2.20}$$

Let $\Lambda_h \subset \Lambda$ with $\Lambda_h = \{\mu_h |\, \mu_h \geq 0\}$ which is a cone with vertex in 0 in $\mathbb{R}$ and let

$$\Phi : V_h \times \Lambda_h \to \mathbb{R}$$

be defined by

$$\Phi(v_h, \mu_h) = \mu_h I(v_h).$$

Then the Lagrangian associated to problem (2.20) is

$$\mathcal{L}(v_h, \mu_h) = J(v_h) + \mu_h I(v_h),$$

or in an equivalent form:

$$\mathcal{L}(v_h, \mu_h) := \frac{1}{2}a(v_h, v_h) - \langle l, v_h\rangle + b(v_h, \mu_h) - \langle \mu_h, g\rangle \tag{2.21}$$

with a continuous bilinear-form $b : V_h \times E_h \to \mathbb{R}$ where $\Lambda_h \subset E_h$. Analogue to the continuous case we have a discrete form of the inf-sup-condition. We introduce

$$X_h := \{v_h \in V_h |\, b(v_h, \mu_h) = 0 \quad \forall \mu_h \in E_h\}.$$

Theorem 2.11. *A family of finite element spaces* V_h, E_h *fulfills the inf-sup-condition if there exist constants* $\alpha > 0$ *and* $\beta > 0$*, being independent of* h*, such that* *(i) The bilinear-form $a(\cdot,\cdot)$ is X_h-elliptic with an ellipticity constant $\alpha > 0$.*

(ii) There holds

$$\sup_{v_h \in V_h} \frac{b(v_h, \lambda_h)}{\|v_h\|} \geq \beta \|\lambda_h\| \quad \forall \lambda_h \in E_h.$$

If the inf-sup-condition is fulfilled a unique solution of (2.21) is ensured. That means we have a stable system.

Theorem 2.12. *If $(u_h, \lambda_h) \in V_h \times \Lambda_h$ is the solution to problem* (2.21) *under the conditions of Theorem 2.11 then there hold the stability estimates*

$$\begin{aligned} \|u_h\|_{V_h} &\leq \frac{1}{\alpha}\|f\|_{V_h'} + \frac{1}{\beta}(\frac{c}{\alpha}+1)\|g\|_{E_h'}, \\ \|\lambda_h\|_{E_h} &\leq \frac{1}{\alpha}(\frac{c}{\beta}+1)\|f\|_{V_h'} + \frac{c}{\beta^2}(\frac{c}{\alpha}+1)\|g\|_{E_h'}. \end{aligned}$$

Theorem 2.11 is a reliable indicator for stable systems, but in many cases it is very difficult to prove the condition. Therefore, we will now give an alternative requirement which is easier to verify and that ensures a saddle point of the Lagrangian formulation, but there is not necessarily given uniqueness of the Lagrangian parameter. There exist some qualifying hypotheses to ensure the existence of a saddle point. If one of these hypotheses is fulfilled Theorem 2.13 holds. We will only introduce one of these hypotheses because we will refer to this one in our studies.

Definition 2.3. *Slater hypothesis: There exists a vector $Z \in \mathbb{R}^n$ such that $I(Z) < 0$.*

Theorem 2.13. *Suppose the functionals J, I are convex and the Slater hypothesis holds. If problem* (2.20) *has a solution, i.e. there exists an $u_h \in K_h$ such that $J(u_h) = \inf_{v \in K_h} J(v)$, then the Lagrangian $\mathcal{L}$ has a saddle point.*

Later in this work we will come across some examples for systems that are not stable but solvable due to iterative solvers. However, the solution is not unique. In an unstable system the Lagrangian multiplier loses its physical relevance as we will see in resulting graphical outputs of the Lagrangian parameter. Oscillations appear when there is no stability given.

Proofs of the inf-sup-condition can be found in Braess [Br07] and Brezzi/Fortin [BF91] and a proof of Theorem 2.13 is shown in Céa [Ce78].

2.4 Sobolev spaces

The Sobolev spaces are composed of the function space $L_2(\Omega)$. Let Ω be a bounded subset of $\mathbb{R}^d$. Then $L_2(\Omega)$ is defined as the space of functions where the integral over the square is finite:

Definition 2.4.1. *The space $L_2(\Omega)$ of a domain Ω is defined as the set of all measurable functions on Ω with*

$$L_2(\Omega) := \{v \mid \int_\Omega v^2 \, dx < \infty\},$$

in sense of Lebesgue.

$L_2(\Omega)$ is a Hilbert space with the (L_2-) inner product

$$(v, w)_0 := (v, w)_{L_2(\Omega)} = \int_\Omega vw \, dx$$

and the corresponding norm

$$\|v\|_0 := \|v\|_{L_2(\Omega)} = \left(\int_\Omega v^2 \, dx\right)^{\frac{1}{2}} = (v, v)^{\frac{1}{2}}.$$

Definition 2.4.2. *$u \in L_2(\Omega)$ has a weak derivative $v = \partial^\alpha u$ in $L_2(\Omega)$ if $v \in L_2(\Omega)$ and*

$$(\Phi, v)_0 = (-1)^{|\alpha|}(\partial^\alpha \Phi, u)_0 \quad \forall \Phi \in C_0^\infty(\Omega),$$

α being a multi-index.

$C^\infty(\Omega)$ is the space of smooth functions and $C_0^\infty(\Omega)$ is the subspace of functions being zero on the boundary of Ω.

Definition 2.4.3. *Let $m \in \mathbb{N} \cup \{0\}$. $H^m(\Omega)$ is the set of all functions $u \in L_2(\Omega)$ with weak derivatives $\partial^\alpha u$ for all $|\alpha| \le m$ in $L_2(\Omega)$. Furthermore, $H^m(\Omega)$ is a Hilbert space with the inner product*

$$(u, v)_m := (u, v)_{H^m(\Omega)} := \sum_{|\alpha| \le m} (\partial^\alpha u, \partial^\alpha v)_0$$

and the (Sobolev-) norm

$$\|u\|_m := \|u\|_{H^m(\Omega)} := \sqrt{\sum_{|\alpha|\leq m} \|\partial^\alpha u\|_0^2}. \tag{2.22}$$

$$|u|_m := \sqrt{\sum_{\alpha=m} \|\partial^\alpha u\|_0^2}$$

is called the half norm of H^m. $H_0^m(\Omega)$ is the completion of $C_0^\infty(\Omega)$ with respect to $\|\cdot\|_m$ in $L_2(\Omega)$ corresponding to (2.22)*.*

In general, the Sobolev space $W_p^l(\Omega)$, $l \geq 0$ integer, is the set of functions in $L^p(\Omega)$ which possesses all weak derivations up to order l, that belong to $L^p(\Omega)$, too. The corresponding norm is set by

$$\|u\|_{W_p^l(\Omega)} := \left[\int_\Omega \sum_{|\alpha|\leq l} |[D^\alpha u](x)|^p dx\right]^{\frac{1}{p}}. \tag{2.23}$$

In an analogous way, based on $L^\infty(\Omega)$, we define the space $W_\infty^l(\Omega)$.

Using L_2-functions, a point-wise restriction to the boundary is not reasonable. However, there exists a continuous linear mapping

$$\gamma : H^1(\Omega) \to L_2(\Gamma)$$

which is compatible to the point-wise restriction of functions $u \in C^0(\overline{\Omega}) \cap H^1(\Omega)$:

$$\forall u \in C^0(\overline{\Omega}) \cap H^1(\Omega) : \; u_{|\Gamma} = \gamma(u). \tag{2.24}$$

Theorem 2.14 (trace theorem)**.** *If Ω has a Lipschitz boundary Γ then there exists a constant $c > 0$ such that*

$$\|\gamma(u)\|_{L_2(\Gamma)} \leq c\|u\|_{H^1(\Omega)} \quad \forall u \in C^1(\overline{\Omega}).$$

For a proof see [Br07]. The trace operator γ generates new function spaces over the boundary Γ with the help of the underlying function spaces on Ω. To be more precise, we define $H^{\frac{1}{2}}(\Gamma) := \gamma(H^1(\Omega))$ with

$$H^{\frac{1}{2}}(\Gamma) := \{w \in L_2(\Gamma) : \exists v \in H^1(\Omega) \text{ with } w = \gamma(v)\}.$$

$H^{\frac{1}{2}}(\Gamma)$ denotes a Hilbert space, that is a proper dense subspace of $L_2(\Gamma)$ with the norm

$$\|w\|_{H^{\frac{1}{2}}(\Gamma)} = \inf\{\|v\|_{H^1(\Omega)} |\, v \in H^1(\Omega),\ w = \gamma(v)\}.$$

The dual space of $H^{\frac{1}{2}}(\Gamma)$ is denoted by $H^{-\frac{1}{2}}(\Gamma)$ with the norm defined by

$$\|g\|_{H^{-\frac{1}{2}}(\Gamma)} := \sup_{w \in H^{\frac{1}{2}}(\Gamma)} \frac{|g(w)|}{\|w\|_{H^{\frac{1}{2}}(\Gamma)}} \quad \forall g \in H^{\frac{1}{2}}(\Gamma).$$

A special group of Sobolev spaces on the boundary are those where the function values disappear on parts of it. So we define

$$H^1(\Omega, \Gamma_0) := \{v \in H^1(\Omega) | \gamma(v)|_{\Gamma_0} = 0\}$$

for a closed subset of the boundary $\Gamma_0 \subset \Gamma$ with $\int_{\Gamma_0} ds > 0$ and $H_0^1(\Omega) := H^1(\Omega, \Gamma)$. If Γ is sufficiently smooth, $H_0^1(\Omega)$ is proper dense in $L_2(\Omega)$. For $\Gamma_1 \subset \Gamma$ with $\Gamma_0 \cup \Gamma_1 = \Gamma$ and $\Gamma_0 \cap \Gamma_1 = \emptyset$ we define

$$H^{\frac{1}{2}}(\Gamma_1) := \gamma(H^1(\Omega, \Gamma_0)),$$

which is a closed subspace of $H^{\frac{1}{2}}(\Gamma)$ and denotes a Hilbert space with norm $\|\cdot\|_{H^{\frac{1}{2}}(\Gamma_1)}$ which is proper dense in $L_2(\Gamma_1)$. The dual space is $H^{-\frac{1}{2}}(\Gamma_1)$ with norm $\|\cdot\|_{H^{-\frac{1}{2}}(\Gamma_1)}$.

3 Obstacle problem

As a first test case we choose the linear contact problem including a linear restriction. We present the problem within a variational inequality and examine existence and solvability. When introducing an a posteriori estimator we notice it to be inconsistent in areas of contact. To overcome this problem we formulate the mixed system in order to develop an improved error estimator. The unique solvability of the continuous problem may be not given for the discrete case. So after discretising the system, we eliminate instabilities by presenting a least squares stabilisation which ensures uniqueness and hence gives a good basis for the investigation of the new a posteriori estimator including the Lagrangian multiplier and taking account of the stabilisation. This one turns out to be consistent, which is confirmed by numerical tests at the end of the chapter. Furthermore, we are interested in fast solvers, so we present different algorithms and compare them in a benchmark.

Simulations in mechanical engineering applications are very important for reducing the experimental effort. Often, models are of large complexity so that good mesh structures are necessary for economical calculations. Taking for example mechanical deformation where a workpiece is pressed into a form by an external force. An accurate mathematical modeling of the contact situation and a precise numerical simulation can help to process new concepts very efficiently and keep down the costs.

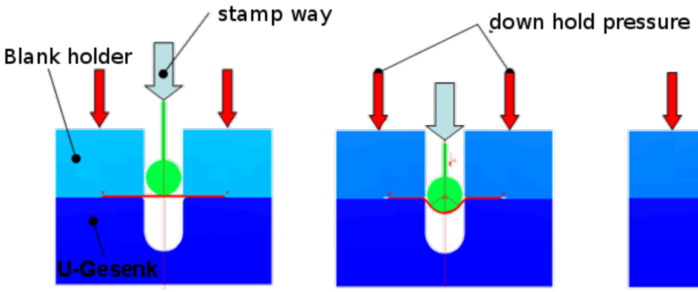


Figure 3.1: An example for mechanical deformation: One possible way of forming a tube out of a metal sheet is to press the sheet into an U-form as a first step and then bend the ends of the sheet around. Pressing the sheet into the U-form is a typical obstacle problem.

We will study the behaviour of the contact surface by a 2D obstacle problem. Taking a domain $\Omega = \mathbb{R}^2$ the corresponding mathematical model has the strong formulation

$$\begin{aligned} -\Delta u - f &\geq 0, \\ u - \Psi &\geq 0, \\ (u-\Psi)(-\Delta u - f) &= 0, \end{aligned} \tag{3.1}$$

see [Su08], with $u \in C^2(\Omega) \cap C(\bar{\Omega})$ and the boundary condition $u = 0$ on $\partial\Omega$. $f \in C(\Omega)$ represents the body force and $\Psi \in C^2(\Omega) \cap C(\bar{\Omega})$ denotes the obstacle. The second inequality in (3.1) is reasonable since we have zero boundary conditions and hence the solution u is also the displacement v of the system. Otherwise the gap would have to be calculated by $(u_0+v)-\Psi > 0$ where u_0 is the starting position of u.

3.1 Variational formulation

In order to use finite element techniques we rewrite system (3.1) to obtain a variational inequality:

$$u \in K: \quad (\nabla u, \nabla(\varphi - u)) \geq (f, \varphi - u) \quad \forall \varphi \in K, \tag{3.2}$$

where we set $V = H_0^1(\Omega)$ and $K = \{v \in V | v \geq \Psi \text{ a.e. in } \Omega\}$, with the obstacle $\Psi : \Omega \to \mathbb{R}$ and assume $f \in L_2(\Omega)$. Since K is nonempty, convex and closed, the problem is uniquely solvable by Theorem 2.3.

3.1.1 Discretisation

To use the finite element method we introduce decompositions $\mathbb{T}_h = \{T_i | 1 \leq i \leq N_h\}$ of Ω consisting of N_h quadrangular elements T_i, satisfying the usual conditions of shape regularity. The width of the mesh $\mathbb{T}_h$ is characterised in terms of a piecewise constant mesh size function $h = h(x) > 0$, where $h_T := h_{|T} = diam(T)$. So the approximated solution u_h of the discrete inequality is characterised by

$$u_h \in K_h : \quad (\nabla u_h, \nabla(\varphi - u_h)) \geq (f, \varphi - u_h) \quad \forall \varphi \in K_h, \tag{3.3}$$

where V_h is a finite element space on $\mathbb{T}_h$ and $K_h \subset V_h$ is a closed, convex subset, which is chosen as an appropriate discrete substitute of K. For our studies we use standard bilinear finite elements for discretisation. Ψ_h is the linear interpolant of Ψ and so K_h is given by $K_h = \{v \in V_h | v \geq \Psi_h \text{ in } \Omega\}$. For the sake of simplicity we assume Ψ to be polygonal and h is small enough such that we can set $\Psi = \Psi_h$. Here, unique solvability is guaranteed by Theorem 2.1.1.

We are especially interested in a posteriori error estimation. For getting a priori error bounds in case of various underlying problems see for example [Fa74] or [BHR77].

3.1.2 A posteriori error analysis

In the further process, we frequently use the standard interpolation operator $I_h : H^1(\Omega) \to V_h$ of Cléments type. For a function $\omega \in H^1(\Omega)$, the application of I_h is shortly denoted by $\omega_i := I_h \omega$. We set $\|\omega\|_B := \|\omega\|_{L_2(B)}$ for a subset $B \subset \Omega$. Dropping the index B, we assume $\|\omega\| := \|\omega\|_{L_2(\Omega)}$ to be the L_2-norm on Ω when $B = \Omega$. Furthermore we define $\omega_{x_j} := \cup\{T' \in \mathbb{T}_h | \, x_j \in T'\}$ if x_j is a vertex of a mesh cell and $\tilde{\omega}_T := \cup\{\omega_{x_j} \mid x_j \in T\}$. For the interpolation error, the estimates

$$\|\omega - \omega_i\|_T \le C_{i,T} h_T \|\nabla\omega\|_{\tilde{\omega}_T} \tag{3.4}$$

$$\|\omega - \omega_i\|_{\partial T} \le C_{i,\partial T} \sqrt{h_T} \|\nabla\omega\|_{\tilde{\omega}_T} \tag{3.5}$$

hold for all mesh cells $T \in \mathbb{T}_h$ (see for example [Br07]).

Following [Su08] in order to get an estimator measuring the approximation error, we set $e = u - u_h$ and start estimating $(\nabla e, \nabla e_i)$ by

$$\begin{aligned}(\nabla e, \nabla e_i) &= \underbrace{(f, e_i) - (\nabla u_h, \nabla e_i)}_{\overset{(3.3)}{\le} 0} + (\nabla u, \nabla(e_i - e)) - (f, e_i - e) + \underbrace{(\nabla u, \nabla e) - (f, e)}_{\overset{(3.2)}{\le} 0} \\ &\le (\nabla u, \nabla(e_i - e)) - (f, e_i - e) \end{aligned} \tag{3.6}$$

taking into account that $\Psi_h = \Psi$. Now we can easily estimate the error in the square of the energy norm:

$$\begin{aligned}(\nabla e, \nabla e) &= (\nabla e, \nabla(e - e_i)) + (\nabla e, \nabla e_i) \\ &\overset{(3.6)}{\le} (\nabla u, \nabla(e - e_i)) - (\nabla u_h, \nabla(e - e_i)) + (\nabla u, \nabla(e_i - e)) - (f, e_i - e) \\ &= (f, e - e_i) - (\nabla u_h, \nabla(e - e_i)).\end{aligned}$$

Cell-wise integration by parts results in

$$(\nabla e, \nabla e) \le \sum_{T \in \mathbb{T}_h} \omega_T \rho_T,$$

with local residuals ρ_T and weights ω_T defined by

$$\begin{aligned}\rho_T &:= h_T \|f + \Delta u_h\|_T + \frac{1}{2} h_T^{\frac{1}{2}} \|n \cdot [\nabla u_h]\,\|_{\partial T}, \\ \omega_T &:= \max\{h_T^{-1} \|e - e_i\|_T, h_T^{-\frac{1}{2}} \|e - e_i\|_{\partial T}\}.\end{aligned}$$

In ρ_T we exchanged half of the edge integral of cell T with the neighbour cell T' using that for the normal vectors there holds $n' = -n$ and define the jump of the normal derivative by

$$n \cdot [\nabla u_h] = [\partial_n u_h] := \partial_n u_h|_T + \partial_{n'} u_h|_{T'} = \partial_n u_h|_T - \partial_n u_h|_{T'} \tag{3.7}$$

which is zero on $\partial\Omega$. Using the interpolation estimates (3.4) and (3.5) yields the following estimate for the discretisation error in the energy norm:

Theorem 3.1.1. *For problem* (3.3)*, there holds the a posteriori error bound*

$$|e|_1^2 \leq C\eta(u_h) := C \sum_{T\in\mathbb{T}_h} \rho_T^2 \tag{3.8}$$

with local residuals ρ_T defined by

$$\rho_T := h_T\|f + \Delta u_h\|_T + \frac{1}{2}\sqrt{h_T}\|[\partial_n u_h]\|_{\partial T}.$$

Looking at Theorem 3.1.1 we notice that the error estimator is not consistent in the sense of the following definition:

Definition 3.1.2. *An error estimator $\eta(u_h)$ is called consistent, if it vanishes by replacing the discrete FE-approximation u_h by the solution u of the original problem, i.e. $\eta(u) = 0$.*

In regions of contact, there holds

$$\|\Delta u + f\| > 0, \tag{3.9}$$

which determines the actual unknown contact forces and is related to the second term of Lemma 2.1.

3.2 Saddle point problem

Inconsistency has a negative influence on meshstructures as we will see in Section 3.4.2. Economical meshes can only be received by consistent estimators. In physics, the contact force that causes the gap in (3.9) is called constraining force and can be discribed by Lagrangian multipliers. For this purpose, we present a mixed formulation of problem (3.2). In what follows, we set $a(\cdot,\cdot) = (\nabla\cdot,\nabla\cdot)$.

3.2.1 Saddle point formulation

The resulting Lagrangian formulation looks as follows (see Chapter 2):

Find a pair $(u, \lambda) \in V \times \Lambda$ with $\Lambda := \{q \in L_2(\Omega) | \, q \geq 0 \text{ a.e.}\}$ and

$$\begin{aligned}\mathcal{L}(u,\lambda) &= \inf_{\varphi \in V} \sup_{\omega \in \Lambda} \mathcal{L}(\varphi, \omega) \\ &= \inf_{\varphi \in V} \sup_{\omega \in \Lambda} \left\{ \frac{1}{2} a(\varphi,\varphi) - (f,\varphi) - (\omega, \varphi - \Psi) \right\},\end{aligned}$$

where $a(\varphi,\varphi) = (\nabla\varphi, \nabla\varphi)$.

Theorem 3.1 (regularity)**.** *If $f \in L^p(\Omega)$ $(2 \leq p < +\infty)$, we have*

$$u \in H_0^1 \cap W^{2,p}(\Omega),$$

with for $p \geq 1$

$$W^{2,p}(\Omega) = \{v | v, \frac{\partial v}{\partial x_i}, \frac{\partial^2 v}{\partial x_i \partial x_j} \in L^p(\Omega), i, j = 1, n\}.$$

For a proof see [LS92].

Theorem 3.2. *The pair (u, λ) with $\lambda = -\Delta u - f$ is the unique saddle point of $\mathcal{L}(v, \mu)$ on $V \times \Lambda$.*

To prove the theorem we follow [GLT76].

Proof. Existence:

Since we have

$$\sup_{\mu \in \Lambda} \int \Phi(\mu, u)\, dx = \sup_{\mu \in \Lambda} \int -\mu(u - \Psi)\, dx = \begin{cases} 0, & \text{if } u \geq \Psi \\ +\infty, & \text{otherwise} \end{cases}$$

(see (2.11)), from (3.1) we receive

$$\lambda = -\Delta u - f \tag{3.10}$$

with $\lambda(u - \Psi) = 0$ a.e. Here, $\lambda \in \Lambda = L^2_+(\Omega)$ due to Theorem 3.1. Furthermore,

$$u \geq \Psi \Rightarrow \int_\Omega \mu(u - \Psi)\, dx \geq 0 \quad \forall \mu \in L^2_+(\Omega), \tag{3.11}$$

and hence with $\mathcal{L}(v, \mu) = J(v) - (\mu, v)$, $J(\cdot)$ from (2.1), there holds:

$$\mathcal{L}(u, \mu) \leq \mathcal{L}(u, \lambda) \; (= J(u)) \quad \forall \mu \in L^2_+(\Omega).$$

With λ fixed, we consider the optimisation problem:

$$\min_{v \in V} \mathcal{L}(v, \lambda). \tag{3.12}$$

The solution of (3.12) is given by $-\Delta u_\lambda = f + \lambda$, $u_\lambda \in V$. Using Theorem 2.6 we know that u_λ is the unique optimal solution $u_\lambda = u$ and hence:

$$\mathcal{L}(u, \lambda) \leq \mathcal{L}(v, \lambda) \quad \forall v \in V.$$

We have shown that (u, λ) is a saddle point of $\mathcal{L}(v, \mu)$. By (3.11) and $\lambda(u - \Psi) = 0$ we observe

$$\int_\Omega (\mu - \lambda)(u - \Psi)\, dx \geq 0 \quad \forall \mu \in L^2_+(\Omega). \tag{3.13}$$

Uniqueness:

In order to prove uniqueness, we note that any saddle point $(u^*, \lambda^*) \in V \times \Lambda$ of $\mathcal{L}$ satisfies the relation (3.10). So every solution fulfills

$$\begin{aligned} &-\Delta u^* = f + \lambda^* \text{ on } \Omega \\ &u^* \in V \\ &\int_\Omega (\mu - \lambda^*)(u^* - \Psi) dx \geq 0 \quad \forall \mu \in \Lambda \\ &\lambda^* \in \Lambda. \end{aligned} \tag{3.14}$$

Setting $\mu = \lambda^*$ in (3.13) and $\mu = \lambda$ in the third relation of (3.14) and summing up the two equations we receive

$$\int_\Omega (\lambda^* - \lambda)(u^* - u) dx \leq 0.$$

Subtracting (3.10) from the appropriate equation in (3.14) there holds

$$\lambda^* - \lambda = -\Delta(u^* - u).$$

From the last two relations and using $u_{|\Gamma} = u^*_{|\Gamma} = 0$ we get

$$-\int_\Omega \Delta(u^* - u)(u^* - u) dx = \int_\Omega |\text{grad}(u^* - u)|^2 dx \leq 0$$

which gives $|u^* - u|_1 \leq 0$ and therefore $u^* = u$ and $\lambda^* = \lambda$. □

Derivation with respect to φ and ω leads to the saddle point problem:
Find a pair $(u, \lambda) \in V \times \Lambda$ fulfilling the mixed formulation

$$\begin{aligned} a(u, \varphi) - (\lambda, \varphi) &= (f, \varphi) && \forall \varphi \in V \\ (u, \omega - \lambda) &\geq (\Psi, \omega - \lambda) && \forall \omega \in \Lambda. \end{aligned} \tag{3.15}$$

Here, λ has a physical relevance. As it is different from zero only if there is contact, it describes a counter force to the external body force f.

The discrete version of (3.15) reads:
Find $(u_h, \lambda_h) \in V_h \times \Lambda_h$, such that

$$a(u_h, \varphi) - (\lambda_h, \varphi) = (f, \varphi) \qquad \forall \varphi \in V_h \tag{3.16}$$

$$(u_h, \omega - \lambda_h) \geq (\Psi, \omega - \lambda_h) \quad \forall \omega \in \Lambda_h := L_h \cap \Lambda \tag{3.17}$$

with $V_h = \{v \in H_0^1(\Omega) | v \text{ bilinear on } T \in \mathbb{T}_h\}$ and L_h consisting of either piecewise bilinear or constant functions on each cell T. In case of constant basis functions there holds

$$L_h = \{\omega |\, \omega = \sum_{T \in \mathbb{T}_h} \omega_T \Theta_T,\, \omega_T \in \mathbb{R}\} \tag{3.18}$$

with

$$\Theta = \text{characteristic function of } T,$$

and

$$\Lambda_h = \Lambda \cap L_h = \{\omega \in L_h |\, \omega_T \geq 0\, \forall T \in \mathbb{T}_h\}.$$

Now the advantage of choosing the Lagrangian method for finding consistent a posteriori estimators gets obvious. Locking at (3.16) and the first equation of (3.15) we notice that we gained equations for which the Galerkin orthogonality holds.

System (3.16)-(3.17) reads in matrix vector notation

$$Au_h + B\lambda_h = F \tag{3.19}$$

$$\forall \mu \in \Lambda_h : \quad (\mu - \lambda_h)^T (B^T u_h - \Psi) \leq 0 \tag{3.20}$$

with the usual stiffness matrix $A = (a_{ij})$ of elements $a_{ij} = a(v_j, v_i)$, $v \in V_h$ and right hand side $F_i = (f, v_i)$. We choose $\Lambda_h := \mathbb{R}^n_+$ and $B = (b_{i,j})$ with $b_{i,j} = (\omega_j, v_i), \omega \in \Lambda_h$ represents the pairing of λ_h and u_h.

3.2.2 Stabilisation

To ensure a stable system the inf-sup-condition (Theorem 2.11) has to be fulfilled by (3.16)-(3.17). Practical experience shows that using bilinear elements to discretise λ and u does not lead to a stable discretisation, see Section 3.4.1. It is very difficult to find a stable pair of finite element spaces in the sense that the discrete inf-sup-condition holds. Therefore, we want to stabilise the system. There are different methods for this purpose. One of them is to calculate the primal and the dual variable on different meshes. The mesh of λ_h has to be coarser than the one of u_h. This method is described in [Sc05]. For our studies we use the Galerkin least squares method by adding a consistent stabilisation term like proposed by Hughes et al. [HFB86] and theoretically analysed by Franca and Stenberg [FS91]. The advantage of this method is that the choice of finite element spaces that can be used is considerably enlarged and the stabilisation is easy to implement in an existing code. From (3.15) we know $\Delta u + \lambda + f$ to be equal to zero in case of $u \in C^2(\Omega) \cap C(\bar{\Omega})$. So the continuous system including consistent stabilisation terms reads

$$a(u,\varphi) - (\lambda,\varphi) + (u, \omega - \lambda) + (\Delta u + \lambda, \delta(\omega + \Delta\varphi) \geq (f,\varphi) + (\Psi, \omega - \lambda) - (f, \delta(\omega + \Delta\varphi)), \tag{3.21}$$

for all $(\varphi, \omega) \in V \times \Lambda$, where δ is a piecewise constant positive parameter function. In the discrete system u is approximated by bilinear functions, and since we have regular, parallel rectangles, the terms Δu_h and $\Delta\varphi$ vanish on every cell. If the grid is wrapped, neglecting Δu_h and $\Delta\varphi$ causes an interpolation error of the order $\mathcal{O}(\delta)$. Numerical tests and a comparison to studies about the MINI-Element show that δ ought to be chosen by $\delta = \gamma h^2$ with a positive constant γ, where on a cell T there holds $\delta = \delta_T := \delta_{|T} = \gamma h_T^2$. The approximation error can be neglected since our error estimator is of the order $\mathcal{O}(h)$ as we will see in Section 3.2.3. We receive

$$\begin{aligned} a(u_h, \varphi) - (\lambda_h, \varphi) = (f, \varphi) \quad \forall \varphi \in V_h \\ (u_h, \omega - \lambda_h) + \delta(\lambda_h, \omega) \geq (\Psi, \omega - \lambda_h) - \delta(f, \omega) \quad \forall \omega \in \Lambda_h. \end{aligned} \tag{3.22}$$

Let U_h be the discretisation of $L_2(\Omega)$. Setting

$$A_\delta(\{u_h, \lambda_h\}, \{\varphi, \omega\}) = F_\delta(\{\varphi, \omega\}) \quad \forall (\varphi, \omega) \in V_h \times U_h \tag{3.23}$$

with the bilinear form

$$A_\delta(\{u_h, \lambda_h\}, \{\varphi, \omega\}) := a(u_h, \varphi) - (\lambda_h, \varphi) + (u_h, \omega) + (\lambda_h, \delta\omega),$$

and right hand side

$$F_\delta(\{\varphi, \omega\}) := (f, \varphi) + (\Psi, \omega) - (f, \delta\omega),$$

the natural mesh-dependent norm corresponding to the bilinear-form A_δ is given by

$$|||\{u_h, \lambda_h\}|||_\delta^2 = |u_h|_1^2 + \|\delta^{\frac{1}{2}} \lambda_h\|^2.$$

Since

$$A_\delta(\{\varphi, \omega\}, \{\varphi, \omega\}) \geq c|||\{\varphi, \omega\}|||_\delta^2, \quad 0 < c \leq 1,$$

A_δ is X_h-elliptic with $X_h = V_h \times U_h$. Following the Lax-Milgram-Lemma (see e.g. [GR92] Chapter 3.3, Lemma 3.6), the unique solvability of the mixed problem is ensured.

Looking at (3.19)-(3.20), we get another mass matrix $C = (c_{i,j})$ with $c_{i,j} = \delta(\omega_i, \omega_j)$ which stabilises the system.

The matrix vector notation (3.19), (3.20) now looks as follows:

$$\begin{aligned} Au_h + B\lambda_h = F \\ \forall \mu \in \Lambda_h : \quad (\mu - \lambda_h)^T(B^T u_h - \Psi) - \delta(\mu - \lambda_h)^T(C\lambda_h + f) \leq 0. \end{aligned} \tag{3.24}$$

So what is derogatory to the system using a least squares stabilisation is the fact that we get another matrix-vector multiplication in our solver. It runs very robust, even on irregular meshes. The shape functions of the Lagrangian parameter can be chosen constant or even bilinear.

3.2.3 A posteriori error analysis

The reason for using the saddle point formulation of problem (3.1) is to develop an error estimator that is consistent in areas of contact, too. That will be put into practice by utilising the counter force λ to fill the gap in the residual term $\|\Delta u + f\|$. The additional stabilisation term causes an error which has to be taken into account by deriving the estimator.

Like in Section 3.1.2 we start with estimating $(\nabla e, \nabla e_i)$ by

$$\begin{aligned}
(\nabla e, \nabla e_i) &= (f, e_i) - (\nabla u_h, \nabla e_i) + (\nabla u, \nabla(e_i - e)) - (f, e_i - e) \\
&\quad + \underbrace{(\nabla u, \nabla e) - (f, e)}_{\leq 0} \\
&\overset{(3.16)}{\leq} -(\lambda_h, e_i) + (\nabla u, \nabla(e_i - e)) - (f, e_i - e).
\end{aligned}$$

The last term is estimated by ≤ 0 because we can test (3.2) with $\varphi = u_h$. Furthermore, we get

$$\begin{aligned}
(\nabla e, \nabla e_i) &\leq (\nabla u, \nabla(e_i - e)) - (f, e_i - e) - (\lambda_h, e_i + e - e) \\
&= (\nabla u, \nabla(e_i - e)) - (f, e_i - e) + (\lambda_h, e - e_i) - (\lambda_h, e) \\
&= (\nabla u, \nabla(e_i - e)) - (f, e_i - e) + (\lambda_h, e - e_i) \\
&\quad -(\lambda_h, u - u_h - \Psi + \Psi + \delta(f + \lambda_h) - \delta(f + \lambda_h)) \\
&= (\nabla u, \nabla(e_i - e)) - (f, e_i - e) + (\lambda_h, e - e_i) \\
&\quad +(\lambda_h, \Psi + \delta(f + \lambda_h) - u) + (\lambda_h, u_h - \Psi - \delta(f + \lambda_h)) \\
&= (\nabla u, \nabla(e_i - e)) - (f, e_i - e) + (\lambda_h, e - e_i) \\
&\quad +(\lambda_h, \Psi + \delta(f + (-\Delta u_h - f)) - u) + (\lambda_h, u_h - \Psi - \delta(f + \lambda_h)) \\
&= (\nabla u, \nabla(e_i - e)) - (f, e_i - e) + (\lambda_h, e - e_i) \\
&\quad +(\lambda_h, \Psi - \delta\Delta u_h - u) + (\lambda_h, u_h - \Psi - \delta(f + \lambda_h)) \\
&= (\nabla u, \nabla(e_i - e)) - (f, e_i - e) + (\lambda_h, e - e_i) \\
&\quad + \underbrace{(\lambda_h, \Psi - u)}_{\leq 0} + \underbrace{(\lambda_h, -\delta\{\Delta u_h + f + \lambda_h\} + u_h - \Psi)}_{=:N}.
\end{aligned}$$

Now we can measure the error in the energy-norm:

$$\begin{aligned}(\nabla e, \nabla e) &= (\nabla e, \nabla(e - e_i)) + (\nabla e, \nabla e_i)\\ &\leq (\nabla u, \nabla(e - e_i)) - (\nabla u_h, \nabla(e - e_i))\\ &\quad + (\nabla u, \nabla(e_i - e)) - (f, e_i - e) - (\lambda_h, e - e_i) + N\\ &= (f, e - e_i) - (\nabla u_h, \nabla(e - e_i)) + (\lambda_h, e - e_i) + N.\end{aligned}$$

Cell-wise integration by parts results in

$$(\nabla e, \nabla e) \leq \sum_{T \in \mathbb{T}} \omega_T \varrho_T + N,$$

with local residuals ϱ_T and weights ω_T defined by

$$\begin{aligned}\varrho_T &:= h_T \|f + \lambda_h + \Delta u_h\|_T + \frac{1}{2} h_T^{\frac{1}{2}} \|n \cdot \{\nabla u_h\}\|_{\partial T},\\ \omega_T &:= \max\{h_T^{-1}\|e - e_i\|_T,\ h_T^{-\frac{1}{2}}\|e - e_i\|_{\partial T}\},\end{aligned}$$

and

$$N = (\lambda_h, -\delta\{\Delta u_h + f + \lambda_h\} + u_h - \Psi) \tag{3.25}$$

using definition (3.7) for the jump terms $[\partial_n u_h]$. Next, one uses the interpolation estimates (3.4) and (3.5) and Young's inequality to get the following estimate:

$$\begin{aligned}(\nabla e, \nabla e) &\leq \sum_{T \in \mathbb{T}_h} \omega_T \varrho_T + N\\ &\leq C \sum_{T \in \mathbb{T}_h} \varrho_T \|\nabla e\|_T + N\\ &\leq C \sum_{T \in \mathbb{T}_h} (\varepsilon \varrho_T^2 + \frac{1}{4\varepsilon}\|\nabla e\|_T^2) + N\\ &= C\varepsilon \sum_{T \in \mathbb{T}_h} \varrho_T^2 + \frac{1}{4\varepsilon}\|\nabla e\|^2 + N.\end{aligned}$$

This results in the next theorem:

Theorem 3.2.1. *For the mixed FE-scheme* (3.22) *there holds the a posteriori error bound*

$$|e|_1^2 \leq \sum_{T \in \mathbb{T}_h} C \varrho_T^2 + cN \tag{3.26}$$

with

$$\varrho_T := h_T \|f + \lambda_h + \Delta u_h\|_T + \frac{1}{2}\sqrt{h_T}\|[\partial_n u_h]\|_{\partial T}$$

and N defined by (3.25) *where for interior interelement boundaries $[\partial_n u_h]$ denotes the jump of the normal derivative $\partial_n u_h$ which is zero on $\partial\Omega$.*

As we mentioned before the approximation error can be neglected because it is of the order $\mathcal{O}(h^2)$.

3.3 Solvers

Variational inequalities can be solved by different solving methods. Some of them are introduced here and we will later compare them with the help of test examples to figure out which one is the most efficient in different contact situations.

cgPSSOR

The cgPSSOR is a SSOR-preconditioned cg-algorithm (see [BBS04]) for solving the variational inequality (3.3). SSOR (Symmetric Successive Over-Relaxation) is, similar to Gauss-Seidel, a splitting method to approximate the solution of a linear system. The projections are performed after the forward and after the backward iteration, respectively. In contrast, the cg(conjugate gradient)-method is a modification of the gradient method and belongs to the Krylov space methods. For more information see for instance Braess [Br07]. The algorithm at hand fulfills one projected SSOR-step before evaluating u_h by a projected cg-iteration.

Algorithm 3.3.1.

Choose an initial u^0

```
    $u^1 = \text{pSSOR}(u^0)$
    for $j = 1, 2, ...$
        $\tilde{u}^{j+1} = \text{pSSOR}(u^j)$
        $d^{j-1} = u^j - u^{j-1}$
        $g^j = \tilde{u}^{j+1} - u^j$
        for $i = 1, 2, ..., \text{diam}(A)$
            if $(\max\{|g_i^j|, |d_i^{j-1}|\} > |\tilde{u}_i^{j+1} - \Psi_i|)$
                $d_i^{j-1} = 0$
                $g_i^j = 0$
        determine $d^j$ such that $\tilde{u}^{j+1} + d^j$ minimises the quadratic
        function (2.1) on the set
        $\tilde{u}^{j+1} + \text{span}\{d^{j-1}, g^j\}$
        $u^{j+1} = \tilde{u}^{j+1} + d^j$
        if $(u^{j+1} - \Psi < 0)$
            determine the largest number $\alpha$ such that $\tilde{u}^{j+1} + \alpha d^j - \Psi \geq 0$
            $u^{j+1} = \tilde{u}^{j+1} + \alpha d^j$
```

Here, pSSOR(v) calculates one projected SSOR-step. The condition rate of this solver is $O(\sqrt{\kappa})$ where κ is the condition of the system matrix.

Uzawa

To solve saddle point problems of a pair $(u, \lambda) \in V \times \Lambda$, a standard method is the Uzawa-algorithm which describes an alternation between a minimisation step in V and a maximisation step in Λ.

Following Braess [Br07] minimising a problem

$$J(u) = \frac{1}{2}u^T Au - f^T u$$

under restriction

$$Bu = g$$

leads to the system

$$\begin{array}{ccccc} Au & + & B^T\lambda & = & f, \\ Bu & & & = & g \end{array}$$

as described in (3.19)-(3.20). The Schur complement of the system is

$$BA^{-1}B^T\lambda = BA^{-1}f - g.$$

To solve these problems we introduce the following

Algorithm 3.3.2 (Uzawa's algorithm)**.**
Choose an initial iterate λ^0 *and* $\alpha > 0$.
For $s = 1, 2, ...$ *:*

$$\begin{aligned} Au^s &= f - B^T\lambda^{s-1}, \\ \lambda^s &= \max\left(0, \lambda^{s-1} + \alpha(Bu^s - g)\right). \end{aligned} \tag{3.27}$$

Here, α is assumed to be small enough. Looking at our stabilised problem we have a slightly different system (3.24) which includes another mass matrix C. Taking into account the additional term we get the extended version of the Uzawa-algorithm:

Algorithm 3.3.3.
Choose an initial iterate λ^0 *and* $\alpha > 0$.
For $s = 1, 2, ...$ *:*

$$\begin{aligned} Au^s &= f - B^T\lambda^{s-1}, \\ \lambda^s &= \max\left(0, \lambda^{s-1} + \alpha(Bu^s - g + C\lambda^{s-1})\right). \end{aligned} \tag{3.28}$$

The convergence of the system is ensured by

Theorem 3.3. *Let s_1 be the maximal eigenvalue of the Schur complement $S = B^T A^{-1} B - C$ and s_2 be the minimal eigenvalue. If the Schur complement is positive definite Uzawa's algorithm converges if, and only if*

$$0 < \alpha < \frac{2}{s_1}. \tag{3.29}$$

In addition, the optimal convergence parameter α is given by

$$\alpha_{opt} = \frac{2}{s_1 + s_2}. \tag{3.30}$$

A proof can be found in [Sa03].

Using Uzawa's algorithm (Algorithm 3.3.3) for the linear obstacle problem we eliminate the first equation of the system which leads to the Richardson iteration

$$(BA^{-1}B^T - C)\lambda = BA^{-1}f - g \tag{3.31}$$

and we have to calculate

$$\lambda_k = \max\left(0, (1 - \alpha(BA^{-1}B^T - C))\lambda_{k-1} + \alpha(BA^{-1}f - g)\right). \tag{3.32}$$

The conditions for convergence of Uzawa's algorithm are fulfilled because A and C are symmetric and so is the Schur complement $S = B^T A^{-1} B - C$. With A and $\begin{pmatrix} A & B^T \\ B & C \end{pmatrix}$ being positive definite the Schur complement is positive definite, too. Let s_1 be the maximal eigenvalue of S and s_2 be the minimal eigenvalue. Then the condition of the matrix S is defined by $\kappa(S) = \frac{s_1}{s_2}$. If S is not well conditioned (κ is large) then the iterative method may converge very slowly. For a better illustration, in the following test case we also calculate the convergence factor which is, following [Sa03], defined by

$$E_k = \left(\frac{\|\lambda_{k+1} - \lambda_k\|}{\|\lambda_1 - \lambda_0\|}\right)^{\frac{1}{k}} = \left(\frac{\|(BA^{-1}f - g) - (BA^{-1}B^T)\lambda_k\|}{\|(BA^{-1}f - g) - (BA^{-1}B^T)\lambda_0\|}\right)^{\frac{1}{k}}, \tag{3.33}$$

with λ_0 and λ_k being the start value and the kth iterate of (3.32), and the corresponding convergence rate by

$$CR_k = -\ln(E_k). \tag{3.34}$$

Elements	condition	α_{opt}	E_k	CR_k	Uzawa steps (cg-it.)
16	2.15	271	0.3645	1.009	23(198)
64	6.19	1896	0.6692	0.4017	55(810)
256	22.32	9331	0.8735	0.1352	157(3744)
1024	86.12	39599	0.9629	0.038	541(23219)
4096	332.17	160864	0.9884	0.0117	1699(137497)
16384	1380.55	645977	0.9964	0.0036	5336(854203)
65536	5482.58	2.59e6	0.9988	0.0012	15063(4351832)

Table 3.1: Condition of the Schur matrix $S = B^T A^{-1} B - C$, convergence factor and convergence rate of Uzawa's algorithm for the linear obstacle problem with stabilisation. Here we set $\Psi = -0.25$ and stopped the algorithm when a residual of 1e-10 had been reached. The last column contains the iteration steps of Uzawa's algorithm and the sum of the inner cg-iterations in brackes.

Table 3.1 shows that there is slow convergence of the classical Uzawa-algorithm. The convergence factor is nearly 1 and the number of iterations nearly grows with h^{-2}. That motivates to develop a preconditioned Uzawa-algorithm.

Preconditioned Uzawa

Although working very robust, the Uzawa-algorithm needs a lot of iterations. There are many works dealing with preconditioning of the classical Uzawa-algorithm. See for example [Ca03], [Cu02]. Gimbel [Gi12] lists some improvements which will help us to make the Uzawa-algorithm run with very few outer iterations. First we have to find a good preconditioner A_p. It turns out that taking the diagonal entries of the Schur complement S, $A_p = \mathrm{diag}(BA^{-1}B^T + C)$, works very efficient. Furthermore, since the Schur complement is symmetric and positive definite, we include an SSOR- and a cg-step in every iteration loop. Let the SSOR-step pSSOR(v) be defined as follows:

Algorithm 3.3.4 (projected SSOR)**.**

```
for $k = 0, 1, 2, ...$
    $u^k := A^{-1}(f + B^T\lambda^k)$
    for $j = 1, ..., m$
        $\lambda_j^{k+\frac{1}{4}} := \lambda_j^k + \frac{1}{D_{jj}}\left(g_j - (Bu^k)_j - \sum_{i=1}^{j-1} C_{ji}\lambda_i^{k+\frac{1}{4}} - \sum_{i=j}^{m} C_{ji}\lambda_i^k\right)$
        if $(\lambda_j^{k+\frac{1}{4}} < 0)$    $\lambda_j^{k+\frac{1}{4}} = 0$
    $u^{k+\frac{1}{2}} := A^{-1}(f + B^T\lambda^{k+\frac{1}{4}})$
    for $j = m, ..., 1$
        $\lambda_j^{k+\frac{1}{2}} := \lambda_j^{k+\frac{1}{4}} + \frac{1}{D_{jj}}\left(g_j - (Bu^{k+\frac{1}{2}})_j - \sum_{i=1}^{j} C_{ji}\lambda_i^{k+\frac{1}{4}} - \sum_{i=j+1}^{m} C_{ji}\lambda_i^{k+\frac{1}{2}}\right)$
        if $(\lambda_j^{k+\frac{1}{2}} < 0)$    $\lambda_j^{k+\frac{1}{2}} = 0$
```

Finally the improved Uzawa-algorithm for (3.31)-(3.32) looks as follows:

Algorithm 3.3.5 (preconditioned Uzawa)**.**

Choose an initial λ^0

```
$D := \mathrm{diag}(BA^{-1}B^T + C)$
$\lambda^1 = \mathrm{pSSOR}(\lambda^0)$
for $k = 1, 2, ...$
    $\lambda^{k+\frac{1}{2}} = \mathrm{pSSOR}(\lambda^k)$
    $d^{k-1} = \lambda^k - \lambda^{k-1}$
    $g^k = \lambda^{k+\frac{1}{2}} - \lambda^k$
    for $i = 1, 2, ..., \mathrm{diam}(A)$
        if $(\max\{|g_i^k|, |d_i^{k-1}|\} > |\lambda^{k+\frac{1}{2}}|)$
            $d_i^{k-1} = 0$
            $g_i^k = 0$
    determine $d^k$ such that $\lambda^{k+1} + d^k$ minimises the quadratic
    function (2.1) on the set
    $\lambda^{k+\frac{1}{2}} + \mathrm{span}\{d^{k-1}, g^k\}$
    $\lambda^{k+1} = \lambda^{k+\frac{1}{2}} + d^k$
    if $(\lambda^{k+1} < 0)$
        determine the largest number $\alpha$ such that $\lambda^{k+\frac{1}{2}} + \alpha d^k \geq 0$
        $\lambda^{k+1} = \lambda^{k+\frac{1}{2}} + \alpha d^k$
```

To guarantee convergence the accuracy of A^{-1}, which can be calculated by a cg-solver or by a direct solver, must be less than square of the gained Uzawa residual. The improvement of this algorithm can be seen in Table 3.4 and 3.5. Figure 3.2 illustrates the better convergence, too.

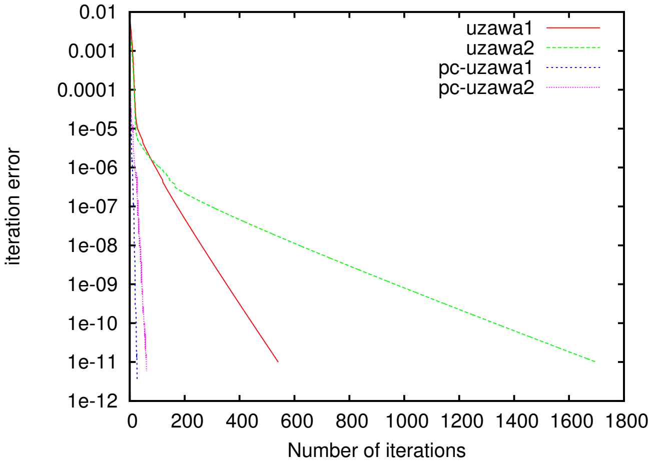


Figure 3.2: Comparison of convergence of classical Uzawa's algorithm (Alg. 3.3.3) and the preconditioned one (Alg. 3.3.5). Uzawa1 and pc-uzawa1 is calculated on 1024 cells and uzawa2 and pc-uzawa2 on 4096 cells. Obviously the preconditioned algorithm converges much faster than the classical one.

Penalty-method

The theory of penalty-terms is a standard method for nonlinear optimisation (see [GR92]). The main principle is to follow the problems restriction asymptotically with the help of an additional term (penalty-term) that consists of a penalty parameter and a measure of violation of the constraints. The measure of violation is zero in the region where constraints are fulfilled and increases the more the constraints are violated. We achieve a variational equality which only depends on a single parameter but has no restrictions in space anymore tending to the solution of the optimisation. In order to illustrate this method we assume that there is a continuous linear mapping $F : V \to V'$ and a continuous bilinear form $b : V \times W \to \mathbb{R}$. Moreover, we define $G \subset V$ as

$$G = \{v \in V \mid b(v, w) \leq g(w) \quad \forall w \in \mathcal{K}\}.$$

Here, $\mathcal{K} \subset W$ is a convex and closed cone in W and $g \in W'$. The associated mixed variational formulation

$$\begin{array}{lclcll} \langle Fu, v\rangle & + & b(v,p) & = & 0 & \forall v \in V \\ b(u, w-p) & & & \leq & g(w-p) & \forall w \in \mathcal{K} \end{array} \tag{3.35}$$

is now regularised by the following form:
Find $(u_\rho, p_\rho) \in V \times \mathcal{K}$:

$$\begin{array}{lclcll} \langle Fu_\rho, v\rangle & + & b(v,p_\rho) & = & 0 & \forall v \in V \\ b(u_\rho, w-p_\rho) & - & \frac{1}{\rho}(p_\rho, w-p_\rho) & \leq & g(w-p_\rho) & \forall w \in \mathcal{K}. \end{array} \tag{3.36}$$

$\rho > 0$ describes a constant parameter which is called penalty parameter. For the sake of simplicity we set $W = W'$ and define $B : V \to W'$ as follows:

$$(Bv, w) = b(v, w) \quad \forall v \in V,\ w \in W.$$

Then, the second equation of (3.36) can be written as

$$(p_\rho - \rho(Bu_\rho - g), w - p_\rho) \geq 0 \quad \forall w \in \mathcal{K}.$$

Let $P_\mathcal{K} : W \to \mathcal{K}$ denote the projection on the closed convex cone $\mathcal{K} \subset W$, so we have

$$p_\rho = \rho P_\mathcal{K}(Bu_\rho - g). \tag{3.37}$$

Consequently (3.36) can be formulated by (3.37) and

$$\langle Fu_\rho, v\rangle + \rho(P_\mathcal{K}(Bu_\rho - g), Bv) = 0 \quad \forall v \in V. \tag{3.38}$$

This is called the penalty-problem corresponding to the problem: Find $u \in G$:

$$\langle Fu, v - u\rangle \geq 0 \quad v \in G \tag{3.39}$$

with

$$G = \{v \in V \mid b(v, w) \leq g(w) \quad \forall w \in \mathcal{K}\}.$$

So we found a variational equation corresponding to (3.39). Solvability and convergence of this formulation are proven in [GR92].

Comparing (3.36) to (3.22) we notice that the stabilised obstacle problem already has the form of the penalty-formulation which is even consistent in our case. So we use this technique to solve the saddle point system by developing one variational equation out of it. We have $W = L_2(\Omega)$, so the cone $\mathcal{K}$ has the form

$$\mathcal{K} = \{w \in L_2(\Omega) \mid w \geq 0\}.$$

Setting $[\cdot]_+ : W \to \mathcal{K}$ with

$$[w]_+(x) := \max\{w(x), 0\},$$

the projection is defined as

$$P_{\mathcal{K}} w = [w]_+ \quad \forall w \in W.$$

Then our penalty-problem (3.38) looks as follows:

$$\int_\Omega \nabla u_\delta \nabla v \, dx - \frac{1}{\delta} \int_\Omega [\Psi - u_\delta - \delta f]_+ v \, dx = \int_\Omega f v \, dx \quad \forall v \in V.$$

Theorem 3.3.6. *Let $f \in L_\infty(\Omega)$ and $g \in W^2_\infty(\Omega)$ with $g|_\Gamma \leq 0$ and $\Omega \subset \mathbb{R}^2$ a convex polygon. For the obstacle problem*

$$\int_\Omega \nabla u \nabla (v - u) \, dx \geq \int_\Omega f(v - u) \, dx \quad \forall v \in G$$

with

$$G = \{v \in H^1_0(\Omega) \mid v \geq g\}$$

there exists a suitable corresponding penalty problem

$$\int_\Omega \nabla u_\rho \nabla v \, dx - \rho \int_\Omega [g - u_\rho]_+ v \, dx = \int_\Omega f v \, dx \quad \forall v \in H^1_0(\Omega). \tag{3.40}$$

Then (3.40) *has a unique solution $u_\rho \in H^1_0(\Omega)$ for all $\rho > 0$ and it holds the estimation*

$$||u - u_\rho||_{0,\infty} \leq (||g||_{2,\infty} + ||f||_{0,\infty})\rho^{-1}.$$

A proof can be found in [Gl83]. The resulting system has the form:

$$Au + D\Phi(u) = f, \tag{3.41}$$

where A is a $N \times N$ positive definite matrix, D is a diagonal matrix with positive diagonal elements d_i and where $u = (u_1, ..., u_N) \in \mathbb{R}^N$, $f \in \mathbb{R}^N$, $\Phi(u) \in \mathbb{R}^N$ with $(\Phi(u))_i = \Phi(u_i)$. An equation of this form can be solved by different methods. One is the Gradient Method (see Glowinski [Gl83], Chapter IV, Sec. 2.6):

Algorithm 3.1.
Choose a start value $u_0 \in \mathbb{R}^n$.
For $n = 0, 1, 2, ...$:

$$u^{n+1} = u^n - \alpha S^{-1}(Au^n + D\Phi(u^n) - f), \quad \alpha > 0. \tag{3.42}$$

Here, S is a symmetric positive definite matrix which can be chosen $S = Id \in \mathbb{R}^{N \times N}$. A better convergence speed allows the choice $S = A$ if A is symmetric and $S = \frac{A+A^*}{2}$ if $A \neq A^*$ with A^* being the adjoint of A. Although it needs less iterations than the Uzawa-algorithm there are faster alternatives. Choosing Newton's method we have to solve a linear system in every newton step:

Algorithm 3.2.
Choose a start value $u_0 \in \mathbb{R}^n$.
For $n = 0, 1, 2, ...$:

$$u^{n+1} = (A + D\Phi'(u^n))^{-1}(D\Phi'(u^n)u^n - D\Phi(u^n) + f), \tag{3.43}$$

where $\Phi'(v)$ denotes the diagonal matrix

$$\Phi'(v) = \begin{pmatrix} \Phi'(v_1) & & 0 \\ & \ddots & \\ 0 & & \Phi'(v_n) \end{pmatrix}. \tag{3.44}$$

$A + D\Phi'(v)$ is positive definite $\forall v \in \mathbb{R}^n$ because Φ is nondecreasing and thereby $\Phi' \geq 0$. Addicted to the start value it is possible that a damping is needed to reach convergence.

3.4 Numerical results

For our numerical tests we always take a surface area $\Omega = [0,1]^2$ and an external body force $f = -10$. The obstacle Ψ is varied throughout the examples.

3.4.1 Stability

Taking the test example with a smooth obstacle $\Psi = -0.25$ we compare displacement and Lagrangian multiplier in the stabilised and the unstabilised method for the numerical solution of the underlying problem. The subfigures of Figures 3.3 and 3.4 show the displacement in case of Q_1/Q_1 discretisation. In the unstabilised problem (Figure 3.3) the surface Ω soaks into the obstacle whereas in the stabilised one the surface lies exactly on the obstacle (Figure 3.4). The subfigures on the right show the contours of Ω in the contact zone.

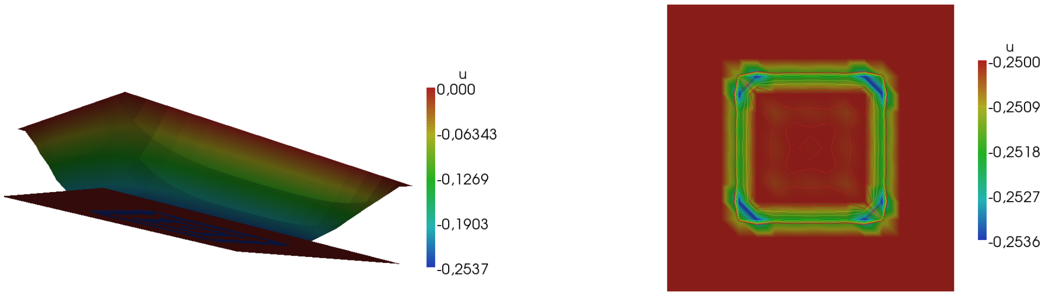


Figure 3.3: The obstacle is a smooth surface $\Psi = -0.25$ and $f = -10$. The left picture shows the displacement of the membrane which soaks into the obstacle in the unstabilised system. In the right picture there are the contours of the penetration and it can be seen that the obstacle condition is violated mostly at the boundary of the contact zone.

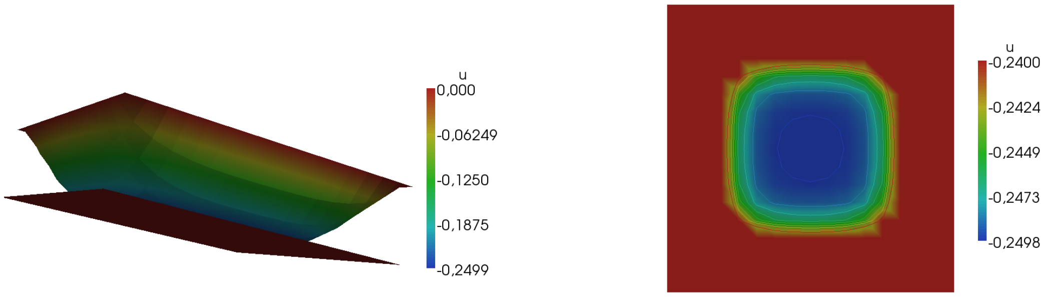


Figure 3.4: Having the same settings like in Figure 3.3, the stabilised system shows optimal values for the displacement.

Here we choose $\delta = \gamma h^2$ with $\gamma = 0.3$ and the finite element spaces $u_h, \lambda_h \in Q_1$ with

$$Q_1 := \{v \in C^0(\bar{\Omega}) |\ v_{|T} \in Q_1(T)\}.$$

On the left hand of Figure 3.5 there are the values of the Lagrangian multiplier, which has non-physical oscillations in the unstabilised case. Nevertheless, in the stable system (right) λ_h yields the physically expected values.

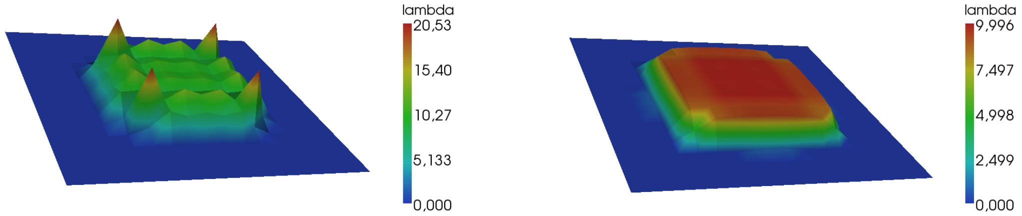


Figure 3.5: In the left picture there are the corresponding values of the Lagrangian multiplier to the unstable system in Figure 3.3 which are way too high. To adjust the force in the contact zone they ought to have a value of about 10 there. Furthermore, there are non-physical oscillations. The right picture shows the Lagrangian multiplier of the stabilised system (Figure 3.4) having the expected values.

We can see the same effect if we choose Elements $u_h \in Q_1$ and

$$\lambda_h \in Q_0 := \{v \in L_2(\Omega) |\ v_{|T} \in P_0(T)\}.$$

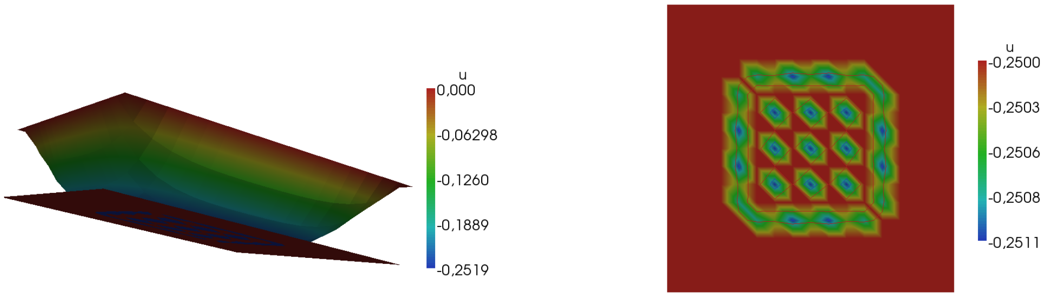


Figure 3.6: Choosing the same obstacle as in Figure 3.3 the left picture shows the displacement of the membrane which also soaks into the obstacle in the unstabilised system. The contours in the right point out the penetration.

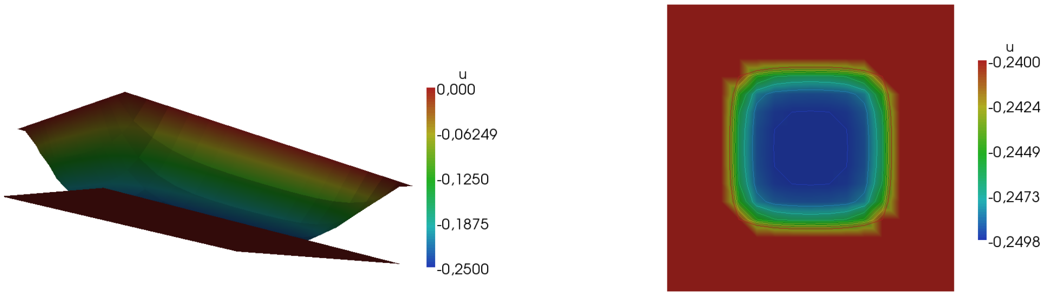


Figure 3.7: Optimal values in the stabilised system choosing $u_h \in Q_1$, $\lambda_h \in Q_0$.

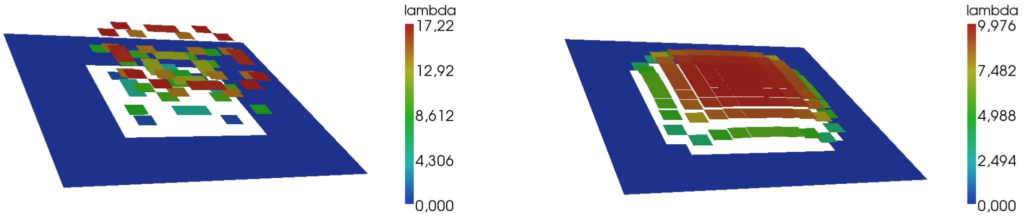
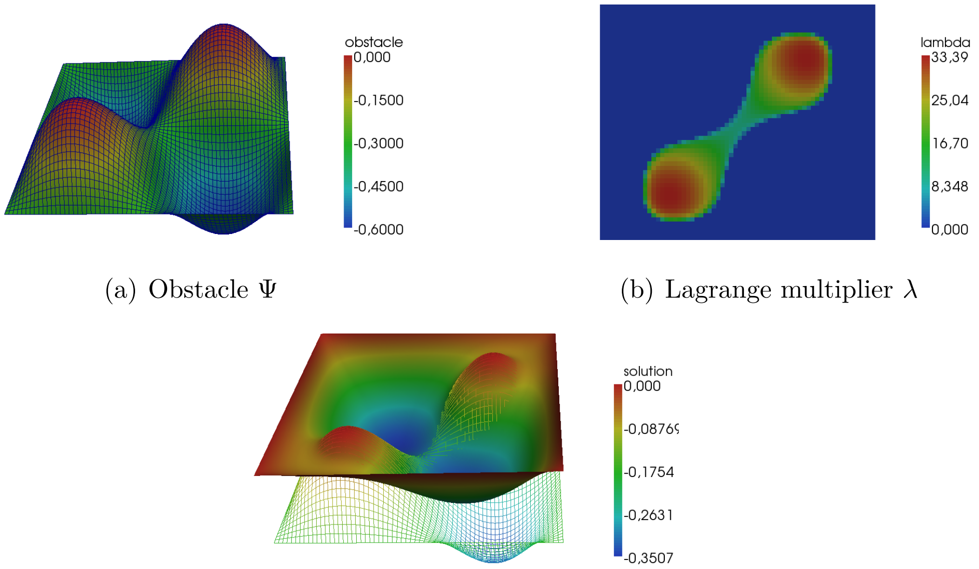


Figure 3.8: In the left picture there are the corresponding values of the Lagrangian multiplier in the unstabilised problem which are far too high. Furthermore, there are non-physical oscillations, too. The values of λ_h in the right picture belong to the stabilised system and show the expected values.

The values of the displacement of the unstable and the stabilised system can be seen in Figures 3.6 and 3.7 and the corresponding Lagrangian multipliers are shown in Figure 3.8.

Although being unstable the system converges to a solution since we use iterative solving methods which allow convergence in some cases. This would not be possible using direct solvers. However there are examples, i.e. if the obstacle gets more complicated, where the unstable system is not even calculable due to the injured inf-sup-condition. One example where the unstable system fails to converge against a solution is presented in Figure 3.9.



(a) Obstacle Ψ

(b) Lagrange multiplier λ

(c) Displacement u_h

Figure 3.9: This problem is only solvable if the system is stabilised. Figure (a) shows the obstacle $\Psi = 0.3\sin(\pi x)\sin(\pi y) - 0.3$ and (c) the displacement calculated by the Uzawa-algorithm. In Figure (b) there are the corresponding values of the Lagrangian multiplier.

3.4.2 Adaptivity

In this section we compare the two discussed error estimators (3.8) and (3.26). We choose a smooth obstacle with $\Psi = -0.25$ and an external body force $f = -10$ on a 2D surface $\Omega = [0,1]^2$. If we take a look at the grids created by adaptive mesh refinement we can see a great difference between estimator (3.8) and (3.26). While the grid created by (3.8) is well refined in the contact zone (Figure 3.10, left), estimator (3.26) yields coarse grids in the interior where $u_h = \Psi$ (Figure 3.10, right), which is more efficient because in the contact zone nothing happens that causes errors worth mentioning. Consequently, we save calculation time using the improved error estimator by reaching the same accurateness. The reason for creating these different meshes is the residual term $\|\Delta u_h + f\|$ in (3.8) which is large in the contact zone. In (3.26) $\|\Delta u_h + f + \lambda_h\|$ is a consistent term and therefore very small in the area of contact.

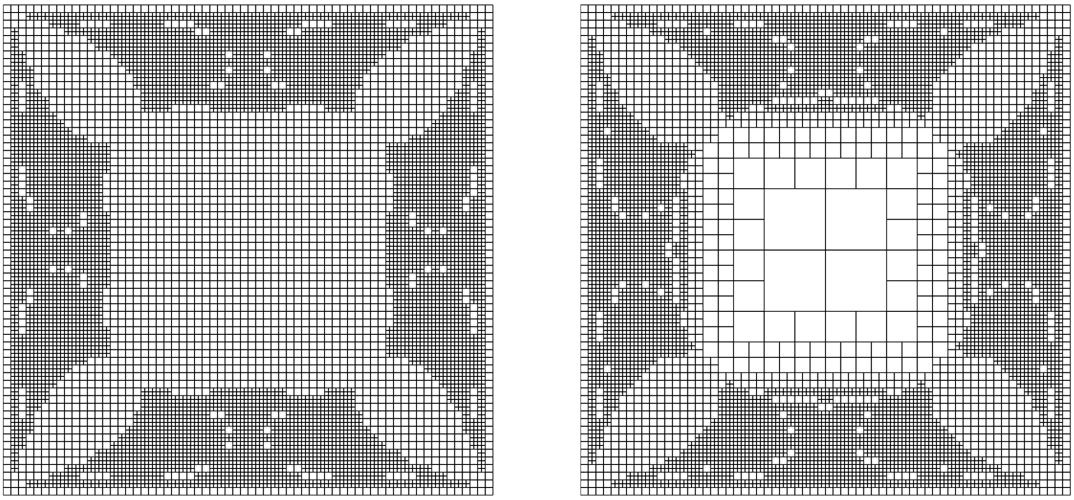


Figure 3.10: The grid on the left is the one without using the Lagrange multiplier in the error estimate. One can see the contact zone being refined very often even if this region is not error-prone. The mesh on the right has been created by estimator (3.26). Due to the term $\|\Delta u_h + f + \lambda_h\|$ there is almost no measurable error in the contact zone and so there is no dispensable refinement there.

# cells	e_{res}		e_{jump}		N		$\vert e\vert_1$	
64	1.32e+00	-	4.99e-01	-	3.01e-01	-	1.82e+00	-
256	7.13e-01	0.86	3.36e-01	0.57	1.02e-01	1.56	1.05e+00	0.79
1024	3.69e-01	0.98	1.84e-01	0.86	3.24e-02	1.66	5.53e-01	0.93
4096	1.86e-01	0.99	9.58e-02	0.94	1.22e-02	1.42	2.81e-01	0.97
16384	9.33e-02	0.99	4.88e-02	0.97	4.40e-03	1.48	1.42e-01	0.99
66536	4.67e-02	0.99	2.47e-02	0.98	1.56e-03	1.50	7.13e-02	0.99

Table 3.2: Convergence of the error terms of estimator (3.26). We set $e_{\text{res}} = (\sum_T h_T^2 \|\Delta u_h + \lambda_h + f\|^2)^{\frac{1}{2}}$, $e_{\text{jump}} = (\sum_T h_T \|n \cdot [\nabla u_h]\|^2_{\partial T})^{\frac{1}{2}}$ and $|e|_1$ denotes the complete error estimator. In the right columns there is always a value α determining the convergence order by $\mathcal{O}(h^\alpha)$ for every refinement step.

Global refinement steps show that the whole estimator has an optimal convergence rate of order $\mathcal{O}(h)$ which is proven in Table 3.2. Comparing the two estimators we find estimator (3.26) is giving a better convergence rate than (3.8). They are both upper bounds for the true error so estimator (3.26) turns out to be a more accurate one as shown in Figure 3.13.

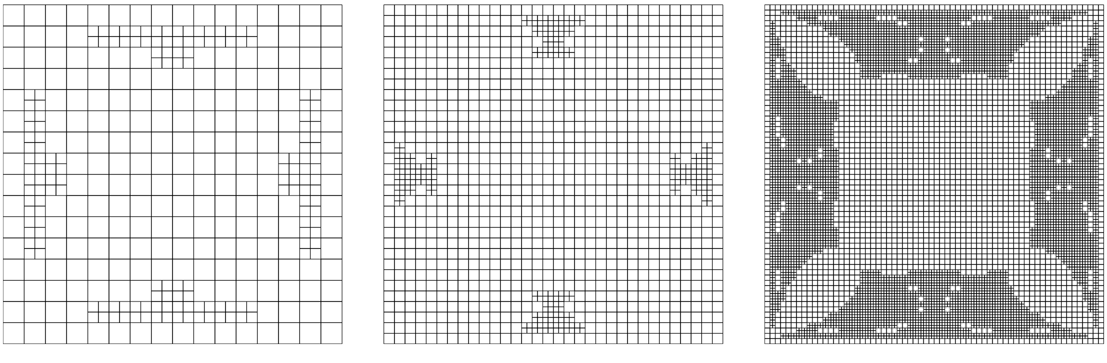


Figure 3.11: Sequence of grids created by estimator (3.8) offering a well refined contact zone.

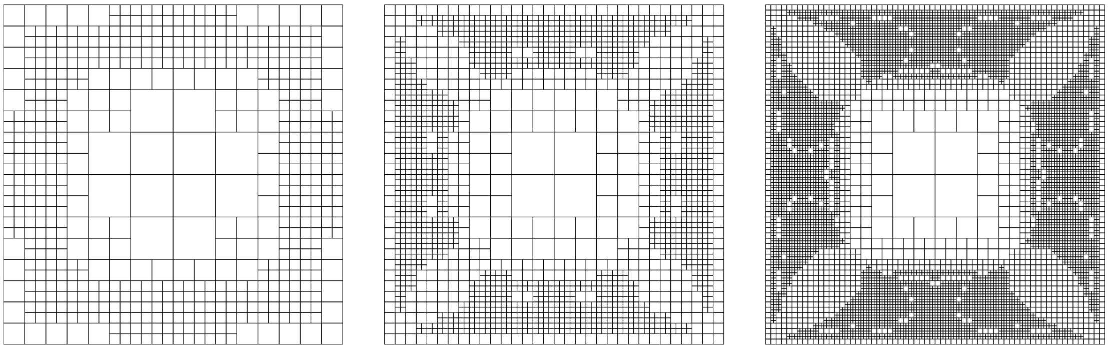


Figure 3.12: Sequence of grids created by estimator (3.26). Due to the consistent residual term there is almost no measurable error in the contact zone which involves more economical mesh structures.

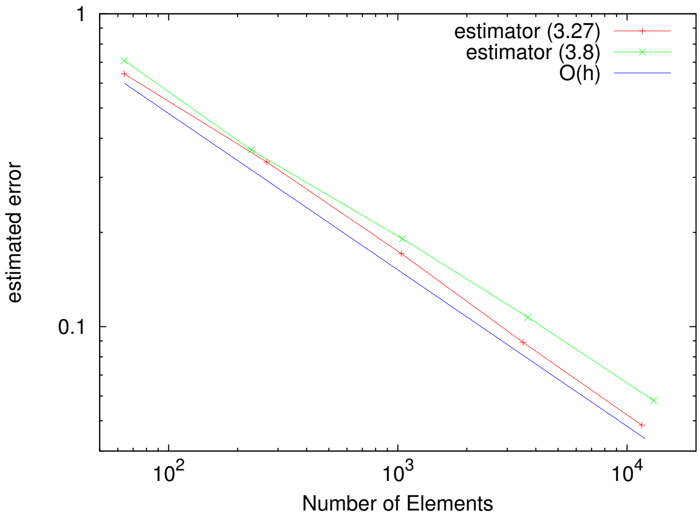


Figure 3.13: Estimator (3.26) offers a smaller error than estimator (3.8) at the same number of elements and it also tends to have a better rate of convergence than the inconsistent one. That is the achievement of the consistent residual term in (3.26).

Comparison of the terms $(\sum_T \|f+\Delta u_h\|_T^2)^{\frac{1}{2}}$ and $(\sum_T \|f+\lambda_h+\Delta u_h\|_T^2)^{\frac{1}{2}}$ in the contact zone shows the reason for over-refinement if we have no Lagrangian multiplier in the estimator: In areas of contact $(\sum_T \|f + \Delta u_h\|_T^2)^{\frac{1}{2}}$ gives the norm of f which is constant everywhere. In contrast, in $(\sum_T \|f + \lambda_h + \Delta u_h\|_T^2)^{\frac{1}{2}}$ the Lagrangian multiplier eliminates this inconsistency and reduces the error in every refinement step (see Figure 3.14, left). The improvement for the whole estimator is obvious if we look at Figure 3.13. However, one could expect it to be even better if we look at Figure 3.14. The fact that we use bilinear elements causes that Δu_h vanishes outside the contact zone, too. So here we only measure f in the norm which is no inconsistency but depends on the choice of the finite elements. Since $(\sum_T \|f+\lambda_h+\Delta u_h\|_T^2)^{\frac{1}{2}}$ tends to zero in the contact zone, the value of the norm in the whole area converges to the one that is measured in this norm outside the contact area (see Figure 3.14, right) which remains constant after a few steps.

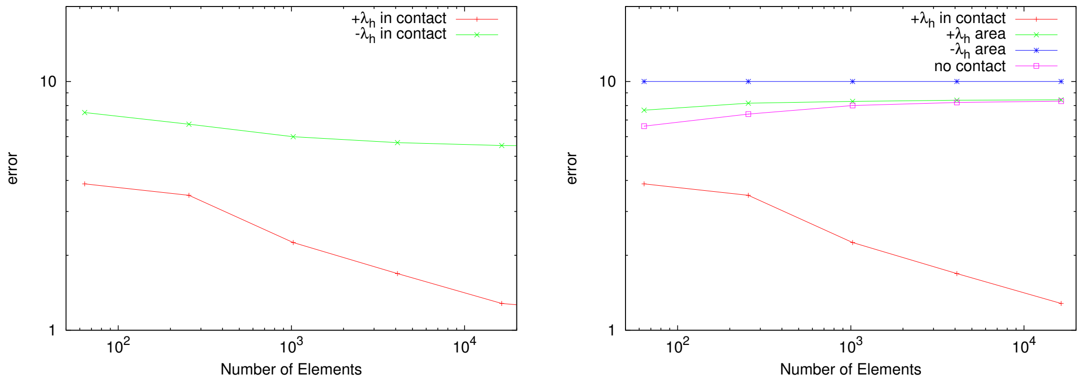


Figure 3.14: In areas of contact we compare $(\sum_T \|f+\lambda_h+\Delta u_h\|_T^2)^{\frac{1}{2}}$ to $(\sum_T \|f+\Delta u_h\|_T^2)^{\frac{1}{2}}$ within a global refinement (left). We find a reduction of the estimator term including λ_h whereas the other term is left nearly constant. Right: “+λ_h area“= $(\sum_T \|f+\lambda_h+\Delta u_h\|_T^2)^{\frac{1}{2}}$ in the whole area, “-λ_h area“= $(\sum_T \|f+\Delta u_h\|_T^2)^{\frac{1}{2}}$ in the whole area, ”no contact“= $(\sum_T \|f+\Delta u_h\|_T^2)^{\frac{1}{2}}$ in areas of no contact. Since the value of $(\sum_T \|f+\lambda_h+\Delta u_h\|_T^2)^{\frac{1}{2}}$ tends to zero in zones of contact, the value of the norm in the whole area converges to the value the norm reaches when neglecting the contact zone which is constant here since we use bilinear elements for u_h and hence Δu_h vanishes.

So just to show how efficient the use of the Lagrangian multiplier really is, we use biquadratic elements for u_h and compare the estimators again (Figure 3.15), being aware of the more complicated stabilisation situation. That means adding some mixed terms which is in fact no big issue to program. To overcome this problem we can also use Algorithm 3.3.1 and calculate λ_h by the residual (see Section 3.4.3).

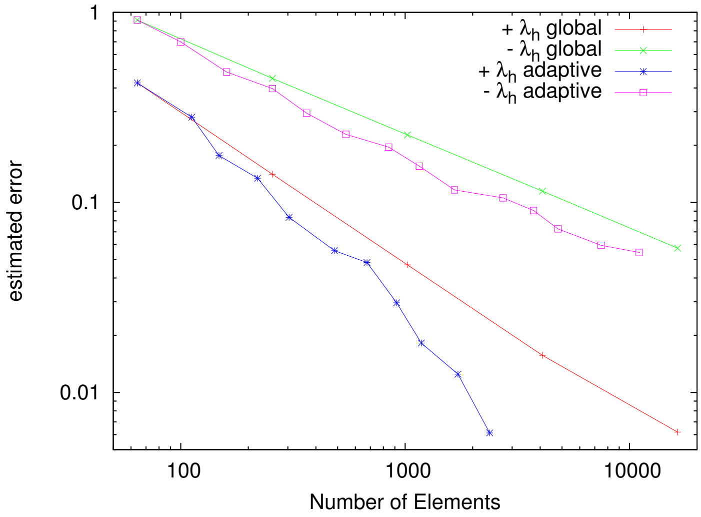


Figure 3.15: Comparison of the error estimators (3.26) and (3.8) for uniform and adaptive refinement using Q_2-Elements for the displacement u_h.

In Figure 3.16 the estimated convergence rates for uniform and adaptive refinement based on the a posteriori error estimate (3.26) are depicted.

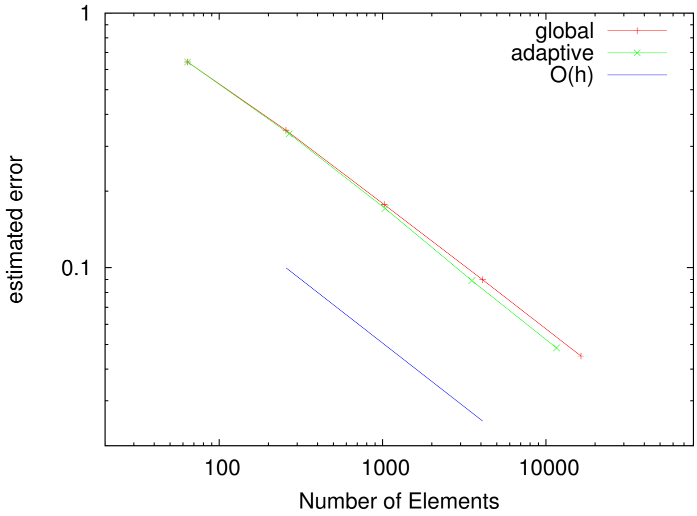


Figure 3.16: Comparison of the estimated convergence rate for adaptive and uniform refinement based on the estimator (3.26).

Since we consider a smooth obstacle here, the uniform approach leads to the optimal convergence rate of $\mathcal{O}(h)$. The adaptive refinement has only a slightly better convergence rate and also converges of optimal order. The advantage of adaptive refinement can be seen even better with a discontinuous obstacle.

# cells	$\lvert e\rvert_{1,\text{global}}$		# cells	$\lvert e\rvert_{1,\text{adaptive}}$	
64	0.880	-	64	0.880	-
256	0.530	0.73	304	0.465	0.82
1024	0.330	0.68	1432	0.230	0.91
4096	0.200	0.72	5776	0.116	0.98
16384	0.127	0.66	17824	0.072	0.85

Table 3.3: Table of convergence of the error estimator using a discontinuous obstacle with global and adaptive refinement. The right value α determines the convergence order by $\mathcal{O}(h^\alpha)$ for every refinement step. With global mesh refinement the error converges with about $\mathcal{O}(h^{\frac{2}{3}})$. In the adaptive case we have almost linear convergence.

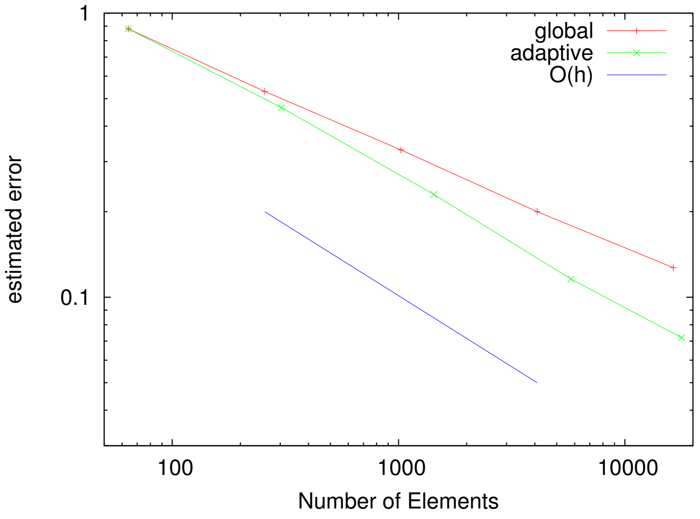


Figure 3.17: Comparison of the estimated convergence rate for adaptive and uniform refinement based on the estimator (3.26) using a discontinuous obstacle.

We choose

$$\Psi = \begin{cases} -0.25 \quad \text{if } x, y \in [0.5 - 0.125, 0.5 + 0.125] \\ -1 \quad \text{elsewhere} \end{cases} \tag{3.45}$$

Using global mesh refinement the singularity causes a slower descend of the convergence rate whereas the adaptive method almost reaches an optimal order of $\mathcal{O}(h)$.

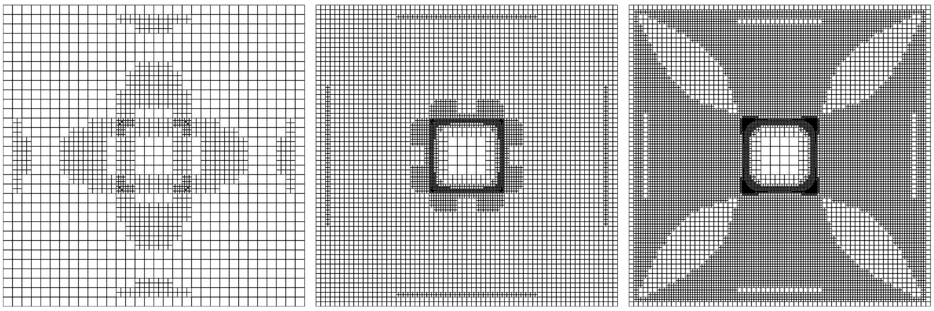


Figure 3.18: Sequence of grids created by estimator (3.26) using the discontinuous obstacle (3.45)

3.4.3 Solvers

We compare the presented solvers in two different contact situations. In the first one we set the obstacle to $\Psi = -0.25$ which leads to a large contact zone. This is a good assumption for the cGPSSOR-solver because many nodes are in contact and are simply projected onto the obstacle, a fact that makes the solver very efficient here. Although needing more iterations on the same number of elements ,the penalty-method has the same convergence order as the cGPSSOR which makes it efficient in cases of many elements, too. For this test example we had to damp the method with a damping factor $\alpha = 0.5$. A large contact zone induces that many entries of the Lagrange multiplier are nonzero. That results in many iterations of the Uzawa-algorithm and thus many cg-iterations. The preconditioned Uzawa-algorithm needs less iterations and has a better convergence rate than the original one. However, it can not keep up with the cGPSSOR algorithm.

Elements	Penalty	cGPSSOR	Uzawa	pc. Uzawa
64	53(953)	7	55(810)	7(201)
256	53(1381)	17	157(3744)	14(612)
1024	53(2347)	29	541(23219)	28(2024)
4096	52(4191)	65	1699(137497)	61(7948)
16384	54(8062)	142	5336(854203)	98(23685)
65536	51(14288)	238	15063(4351832)	218(104396)

Table 3.4: Comparison of the iteration number of different solution algorithms, $\Psi = -0.25$ and residual=1e-11. The columns contain the number of outer iteration steps of the different algorithms and in brackets there is the total number of cg-iterations. (Residual for inner solver: res=1e-16)

Elements	Penalty	cGPSSOR	Uzawa	pc. Uzawa
64	54(959)	16	17(240)	2(85)
256	53(1462)	30	147(3504)	4(225)
1024	53(2652)	57	489(20980)	5(457)
4096	55(5232)	104	1594(127417)	13(1824)
16384	51(9468)	208	4804(710805)	22(5436)
65536	51(18886)	416	13339(3799734)	40(18594)

Table 3.5: Comparison of the iteration number of different solution algorithms, $\Psi = -0.7$ and residual=1e-11. The columns contain the number of outer iteration steps of the different algorithms and in brackets there is the total number of cg-iterations. (Residual for inner solver: res=1e-16)

If we reduce the contact zone the solvers react differently. The cGPSSOR needs more iterations now since there are less projections. In contrast, the steps of the Uzawa-algorithm are reduced, especially in the pc Uzawa because the values of the Lagrange multiplier are zero in many nodes. Nevertheless, the cGPSSOR is the solver with minimum number of iterations. The advantage of the Uzawa and the pc Uzawa is that we are able to use a direct solver instead of a cg-solver for the inner problem. That makes the pc Uzawa attractive for solving problems with small contact zones if there are not too many elements. In that case a direct solver is much faster than an iterative one.

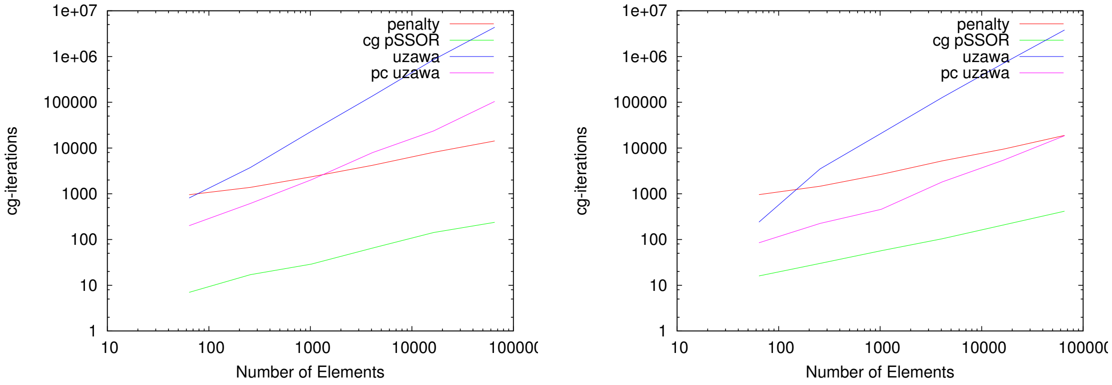
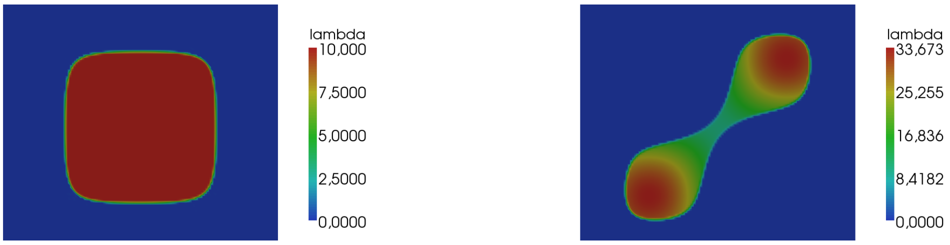


Figure 3.19: Comparison of the total number of cg-iterations that are used by different solvers where the obstacle is set to $\Psi = -0.25$ (left) and $\Psi = -0.7$ (right).

Comparison of the different solvers makes clear that the cGPSSOR is much more effective than Uzawa's algorithm. However, we need to know the Lagrangian multiplier for calculating our error estimator. There is another possibility to get these values by computing the residual $Ax - b$ since the Lagrange multiplier eliminates the inconsistency and therefore the gap of Ax and b.



(a) λ_h for $\Psi = -0.25$

(b) λ_h for $\Psi = 0.3\sin(\pi x)\sin(\pi y) - 0.3$

Figure 3.20: Lagrangian multiplier calculated by the residual of $Ax - b$. Figure (a) shows the Lagrangian multiplier for the obstacle $\Psi = -0.25$ and for (b) we chose the obstacle $\Psi = 0.3\sin(\pi x)\sin(\pi y) - 0.3$.

4 Nonlinear contact problems

In contrast to Chapter 3, we now analyse a nonlinear equation including a linear restriction. Therefore, we examine the contact problem again, which now underlies a nonlinear material law. We start with the introduction of the variational formulation and the proof of existence. For solving nonlinear restricted problems we establish the SQP-Method and give a statement about convergence. The problem of inconsistency returns by the derivation of an appropriate a posteriori estimator, which is again compared to the estimators that we develop after presenting the corresponding saddle point formulation and the suitable least squares stabilisation, which ensures the problem to be well posed. Here, two different ways of stabilisation are presented and both give adequate results. Numerical tests at the end of the chapter approve the theoretical achievements of stability and consistency. As a special example we discuss the minimal surface problem.

When we want to study contact problems in elasticity for example, we have to deal with different constitutive relations. In case of linearised elasticity we take the linear relation

$$\sigma = \mathcal{C}\varepsilon(u)$$

which is called Hooke's law, with the elasticity tensor $\mathcal{C}$ and σ defining the stress variable. We set $\mathbb{S}^d$ the space of second order symmetric tensors on $\mathbb{R}^d$. With $u : \Omega \to \mathbb{R}^d$, ε and σ take on values in $\mathbb{S}^d$ and hence $\mathcal{C}$ is a fourth order tensor. For small deformations, $\varepsilon(u)$ is defined by

$$\varepsilon(u) = \frac{1}{2}(\nabla u + (\nabla u)^T).$$

However, often we have to work with nonlinear stress-strain relations:

$$\sigma = \mathcal{F}(x, \varepsilon(u))$$

with a given nonlinear function $\mathcal{F}$ satisfying the following conditions:

- $\mathcal{F} : \Omega \times \mathbb{S}^d \to \mathbb{S}^d$.
- There exists $L > 0$ such that
$$\|\mathcal{F}(x, \varepsilon_1) - \mathcal{F}(x, \varepsilon_2)\| \leq L\|\varepsilon_1 - \varepsilon_2\| \;\; \forall \varepsilon_1, \varepsilon_2 \in \mathbb{S}^d, \text{ a.e. } x \in \Omega. \tag{4.1}$$
- There exists $m > 0$ such that
$$[\mathcal{F}(x, \varepsilon_1) - \mathcal{F}(x, \varepsilon_2)] : (\varepsilon_1 - \varepsilon_2) \geq m\|\varepsilon_1 - \varepsilon_2\|^2 \;\; \forall \varepsilon_1, \varepsilon_2 \in \mathbb{S}^d, \text{ a.e. } x \in \Omega. \tag{4.2}$$
- For any $\varepsilon \in \mathbb{S}^d$, $x \mapsto \mathcal{F}(x, \varepsilon)$ is measurable in Ω.
- The mapping $x \mapsto \mathcal{F}(x, 0) \in L_2(\Omega)^{d \times d}$.

A family of elasticity operators satisfying these conditions is called nonlinear Hencky materials (see [Ze88]). The equilibrium equation then reads

$$-\text{div}\, \sigma = f \text{ in } \Omega$$

with $\sigma = \mathcal{F}(\varepsilon(u))$. So in contrast to Chapter 3, where we have studied a completely linear problem, we now analyse applications in physics and mechanics with a material described by nonlinear constitutive laws.

Our intention is to check if the Lagrangian technique is still efficient in case of nonlinear systems. For the general study we take on a scalar function u.

4.1 Variational formulation

Again, taking a domain $\Omega \subset \mathbb{R}^2$, the strong formulation of the nonlinear obstacle problem reads

$$\begin{aligned} -\operatorname{div} \mathcal{F}(\nabla u) - f &\geq 0, \\ u - \Psi &\geq 0, \\ (u - \Psi)(-\operatorname{div} \mathcal{F}(\nabla u) - f) &= 0 \end{aligned} \tag{4.3}$$

with $u \in C^2(\Omega) \cap C(\bar{\Omega})$ and the Dirichlet boundary condition $u = 0$ on $\partial\Omega$. $f \in C(\Omega)$ represents the body force and $\Psi \in C^2(\Omega) \cap C(\bar{\Omega})$ describes the obstacle. Similar to (3.2), we have the variational inequality

$$u \in K : \quad (\mathcal{F}(\nabla u), \nabla(\varphi - u)) \geq (f, \varphi - u) \quad \forall \varphi \in K, \tag{4.4}$$

where $\mathcal{F}(\nabla u)$ is a nonlinear strong monotone and continuous operator which is zero if $u \equiv 0$ and that describes the material law. Like in (3.2) we set $V = H_0^1(\Omega)$ and $K = \{v \in V | v \geq \Psi \text{ a.e. in } \Omega\}$ and assume $f \in L_2(\Omega)$. Theorem 2.4 ensures a unique solution of (4.4).

Discretising (4.4) by bilinear finite elements we achieve

$$u_h \in K_h : \quad (\mathcal{F}(\nabla u_h), \nabla(\varphi - u_h)) \geq (f, \varphi - u_h) \quad \forall \varphi \in K_h, \tag{4.5}$$

with the discrete finite element spaces

$$V_h = \{v \in H_0^1(\Omega) |\, v \text{ bilinear on } T \in \mathbb{T}_h\},$$
$$K_h = \{v \in V_h |\, v \geq \Psi_h \text{ a.e. in } \Omega\}.$$

$K_h \subset V_h$ is a closed, convex and nonempty subset. Like it is described in Chapter 3, Ψ_h is the linear interpolant of Ψ and we assume $\Psi_h = \Psi$. Following [AH09], problem (4.5) has a unique solution, too.

4.1.1 SQP-Method

We want to solve inequality (4.5) with the help of the CG-PSSOR-Method (see Section 3.3). For a non-restricted problem one can use Newton's method for linearisation. That means, if we want to minimise a suitable two times differentiable functional $J(\cdot)$,

$$\min J(x), \quad x \in V, \tag{4.6}$$

we compose the quadratic approximation

$$q_i(x) := J(x^i) + \nabla J(x^i)^T (x - x^i) + \frac{1}{2}(x - x^i)^T \nabla^2 J(x^i)(x - x^i)$$

with an iteration index i. If the Hessian $\nabla^2 J(x^i)$ is positive definite, the solution x^{i+1} is defined by $\nabla q_i(x) = 0$. So we get the system

$$\nabla q_i(x) = \nabla J(x^i) + \nabla^2 J(x^i)(x - x^i)$$

and thus

$$x^{i+1} = x^i - \nabla^2 J(x^i)^{-1} \nabla J(x^i).$$

To avoid the calculation of $\nabla^2 J(x^i)^{-1}$ we solve the problem

$$\nabla^2 J(x^i) d^i = \nabla J(x^i)$$

and set

$$x^{i+1} = x^i - d^i.$$

Given a nonlinear problem with restrictions we have to use the SQP (Sequential Quadratic Programs)-Method. Using Newton's algorithm with restricted problems we can easily get a wrong solution. By setting the derivation of the quadratic approximation of $J(x)$ to zero, the algorithm may converge to a solution that does not exist in the convex cone we are searching in. In contrast, the SQP-Algorithm minimises the quadratic linearised problem under the given restrictions. We study a special variant of SQP-Methods, which is called the Levitin-Polyak-Method and can be found in [LP66] or [GT97]. Our problem is characterised by the minimisation of

$$J(x_0) + \nabla J(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 J(x_0)(x - x_0) \to \min, \quad x \in K.$$

Here, $J(x_0)$ is locally quadratic approximated under the restrictions of the problem. To get an equation in x we rewrite the Taylor approximation as follows:

$$
\begin{aligned}
J(x) &\approx J(x_0) + \nabla J(x_0)^T(x-x_0) + \frac{1}{2}(x-x_0)^T\nabla^2 J(x_0)(x-x_0)\\
&= J(x_0) + \nabla J(x_0)^T x - \nabla J(x_0)^T x_0 + \frac{1}{2}x^T\nabla^2 J(x_0)(x-x_0)\\
&\quad - \frac{1}{2}x_0^T\nabla^2 J(x_0)(x-x_0)\\
&= \nabla J(x_0)^T x + \frac{1}{2}x^T\nabla^2 J(x_0)x - \frac{1}{2}x^T\nabla^2 J(x_0)x_0 - \frac{1}{2}x_0^T\nabla^2 J(x_0)x + C\\
&= \nabla J(x_0)^T x + \frac{1}{2}x^T\nabla^2 J(x_0)x - x_0^T\nabla^2 J(x_0)x + C\\
&= \left(\nabla J(x_0)^T - x_0^T\nabla^2 J(x_0)\right)x + \frac{1}{2}x^T\nabla^2 J(x_0)x + C
\end{aligned}
$$

with $C = J(x_0) - \nabla J(x_0)^T x_0 + \frac{1}{2}x_0^T\nabla^2 J(x_0)x_0$ including all terms that are independent of x. We can neglect C because it vanishes by using our minimisation techniques. Let $a(\psi;\varphi)$ be a suitable semilinearform with right hand side $F(\varphi)$, then there holds:

$$
\begin{aligned}
\nabla J(x_0)^T x &= a(x_0;x) - F(x)\\
x^T\nabla^2 J(x_0)x &= \frac{\partial}{\partial x}(a(x,x) - F(x))(x_0)
\end{aligned}
$$

where $\frac{\partial}{\partial x}a(x,x)(x_0)$ denotes the derivation of the semilinearform $a(\cdot;\cdot)$ and hence its linearisation with arguments (x,x) at the iteration point x_0. Then we have to minimise

$$
\begin{aligned}
&(\nabla J(x_0) - x_0^T\nabla^2 J(x_0))x + \frac{1}{2}x^T\nabla^2 J(x_0)x\\
=&a(x_0;x) - F(x) - \frac{\partial}{\partial x}(a(x_0,x) - F(x))(x_0) + \frac{1}{2}\frac{\partial}{\partial x}(a(x,x) - F(x))(x_0)\\
=&a(x_0;x) - F(x) - \frac{\partial}{\partial x}a(x_0,x)(x_0) + \frac{\partial}{\partial x}F(x)(x_0) + \frac{1}{2}\frac{\partial}{\partial x}a(x,x)(x_0)\\
&- \frac{1}{2}\frac{\partial}{\partial x}F(x)(x_0)\\
=&a(x_0;x) - F(x) - \frac{\partial}{\partial x}a(x_0,x)(x_0) + F(x_0) + \frac{1}{2}\frac{\partial}{\partial x}a(x,x)(x_0) - \frac{1}{2}F(x_0).
\end{aligned}
$$

Again, we neglect the terms that are independent of x and achieve

$$
\frac{1}{2}\frac{\partial}{\partial x}a(x,x)(x_0) - \left(\frac{\partial}{\partial x}a(x_0,x)(x_0) - a(x_0;x) + F(x)\right) \to \min, \quad x \in K.
$$

To obtain a numerical solution we discretise the equation in space with $x_h \in K_h$

and receive

$$\frac{1}{2}\frac{\partial}{\partial x_h}a(x_h,x_h)(x_{h,0})-\left(\frac{\partial}{\partial x_h}a(x_{h,0},x_h)(x_{h,0})-a(x_{h,0};x_h)+F(x_h)\right)\to \min, x_h\in K_h,$$

which has to be minimised in the SQP-Algorithm.

Algorithm 4.1 (Levitin-Polyak)**.**

1 *Choose a start value* $x_h^0 \in K_h$ *and set the iteration index* $i = 0$*. Let* γ *be the error tolerance.*

2 *Minimise the linearised problem*

$$\frac{1}{2}\frac{\partial}{\partial x_h}a(x_h,x_h)(x_h^i)-\left(\frac{\partial}{\partial x_h}a(x_h^i,x_h)(x_h^i)-a(x_h^i;x_h)+F(x_h)\right)\to \min,$$

$x_h \in K_h$.

3 *Set* $\varepsilon = \|a(x_h;\varphi_h) - F(\varphi_h)\|$ *for* $\varphi_h \in K_h$*.*

4 *If* $\varepsilon > \gamma$ *set* $i = i+1$*,* $x_h^i = x_h$ *and go to step 2, else finish.*

4.1.2 Convergence of the Levitin-Poljak-Method

Unfortunately, this method is not convergent in general. The starting value has to be taken out of a sufficiently small neighbourhood of the solution we are looking for, but then it shows quadratic convergence:

Theorem 4.1.1. *For the iterated sequence* $\{x_k\}$ *of the Levitin-Poljak-Method, there exists a* $\varrho > 0$ *so that for any* $x_0 \in K \cap U_\varrho(\tilde{x})$ *the sequence* $\{x_k\}$ *converges to the local solution* $\tilde{x}$ *of* (4.6)*. With a constant* $c > 0$ *there holds the estimation*

$$\|x_{k+1}-\tilde{x}\| \le c\|x_k-\tilde{x}\|^2 \quad \textit{for } k = 1,2,...$$

In order to get the algorithm more stable with regard to the choice of the start value, we can use a damping method similar to the damped Newton-Method $x_h^{i+1} = x_h^i + \Delta x_h^i$.

Using such a damping step, we can even form a statement about global convergence (see [Ha77]) if we use a certain step wide $\alpha_i \in (0,1]$:

$$x_h^{i+1} = x_h^i + \alpha_i \Delta x_h^i.$$

In contrast to Newton's method, Δx_h^i is a linear combination of the previous and the new solution x_h^i and x_h^{i+1}

$$\Delta x_h^i := x_h^{i+1} - x_h^i.$$

We iterate x_h^{i+1} and so we modify our update-step to get the new solution $\tilde{x}_h^{i+1}$ from the old solution x_h^i and the new iterated one:

$$\tilde{x}_h^{i+1} = (1-\alpha_i)x_h^i + \alpha_i x_h^{i+1} \quad \text{for } 0 < \alpha_i \leq 1.$$

Algorithm 4.2 (Damped SQP-Method)**.**

1 Choose a start value $x_h^0 \in K_h$ and set the iteration index $i = 0$, $\varepsilon = a(x_h^0; \varphi_h) - F(\varphi_h)$, $\varphi_h \in K_h$. Let γ be the error tolerance.

2 Minimise the linearised problem

$$\frac{1}{2}\frac{\partial}{\partial x_h} a(x_h, x_h)(x_h^i) - \left(\frac{\partial}{\partial x_h} a(x_h^i, x_h)(x_h^i) - a(x_h^i; x_h) + F(x_h)\right) \to \min,$$

$x_h \in K_h$.

3 Set $k = 0$ and $\alpha_0^i = 1$.

4 Set $\tilde{x}_h^{i+1} = (1-\alpha_k^i)x_h^i + \alpha_k^i x_h$.

5 If $\|a(\tilde{x}_h^{i+1}; \varphi_h) - F(\varphi_h)\| < \varepsilon$ set $x_h^{i+1} = \tilde{x}_h^{i+1}$ and go to step 6, else set $\alpha_{k+1}^i = \frac{1}{2}\alpha_k^i$, $k = k+1$ and go to step 4.

6 Set $\varepsilon = a(x_h^{i+1}; \varphi_h) - F(\varphi_h)$.

7 If $\varepsilon > \gamma$ set $i = i+1$ and go to step 2, else finish.

4.1.3 A posteriori error analysis

To be able to use adaptive meshes we develop an a posteriori error estimator for problem (4.5).

For a strong monotone operator there holds:

$$\gamma \|\nabla(u - u_h)\|^2 \leq (\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla u - \nabla u_h)$$

with $\gamma > 0$.

We denote with $e_i := I_h e$ the interpolation of e on the finite element space and start with the following estimation:

$$\begin{aligned}
(\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e_i) &= \underbrace{(f, e_i) - (\mathcal{F}(\nabla u_h), \nabla e_i)}_{\overset{(4.5)}{\leq} 0} + (\mathcal{F}(\nabla u), \nabla(e_i - e)) \\
&\quad -(f, e_i - e) + \underbrace{(\mathcal{F}(\nabla u), \nabla e) - (f, e)}_{\overset{(4.4)}{\leq} 0} \\
&\leq (\mathcal{F}(\nabla u), \nabla(e_i - e)) - (f, e_i - e)
\end{aligned}$$

This allows us to achieve the following:

$$\begin{aligned}
\gamma \|\nabla(u - u_h)\|^2 &\leq (\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e - \nabla e_i) + (\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e_i) \\
&\leq (\mathcal{F}(\nabla u), \nabla(e - e_i)) - (\mathcal{F}(\nabla u_h), \nabla(e - e_i)) \\
&\quad + (\mathcal{F}(\nabla u), \nabla(e_i - e)) - (f, e_i - e) \\
&= (f, e - e_i) - \sum_T (\mathcal{F}(\nabla u_h), \nabla e - \nabla e_i)_T \\
&= (f, e - e_i) - \sum_T \Big[- (\text{div } \mathcal{F}(\nabla u_h), e - e_i)_T \\
&\quad + \int_{\partial T} (n \cdot \mathcal{F}(\nabla u_h)) \cdot (e - e_i) \, d\Gamma \Big] \\
&\leq \sum_T \Big(\|f + \text{div } \mathcal{F}(\nabla u_h)\|_T \|e - e_i\|_T + \frac{1}{2} \|n \cdot [\mathcal{F}(\nabla u_h)]\|_{\partial T} \|e - e_i\|_{\partial T} \Big)
\end{aligned}$$

There holds

$$\gamma(\nabla e, \nabla e) \leq \sum_{T \in \mathbb{T}_h} \omega_T \varrho_T, \tag{4.7}$$

with local residuals ϱ_T and weights ω_T defined by

$$\begin{aligned}\varrho_T &:= h_T \| f + \text{div } \mathcal{F}(\nabla u_h) \|_T + \frac{1}{2} h_T^{\frac{1}{2}} \| n \cdot [\mathcal{F}(\nabla u_h)] \, \|_{\partial T}, \\ \omega_T &:= \max\{h_T^{-1} \| e - e_i \|_T, h_T^{-\frac{1}{2}} \| e - e_i \|_{\partial T}\},\end{aligned}$$

where on the interior interelement boundaries $n \cdot [\mathcal{F}(\nabla u_h)]$ denotes the jump of $\mathcal{F}(\nabla u_h)$ at the element faces in normal direction.

The interpolation estimates (3.4) and (3.5) yield the following estimate for the discretisation error in the energy norm:

Theorem 4.1.2. *For problem* (4.5) *there holds the a posteriori error bound*

$$|e|_1^2 \leq C \sum_{T \in \mathbb{T}_h} \varrho_T^2 \tag{4.8}$$

with local residuals ϱ_T defined by

$$\varrho_T := h_T \| f + \text{div } \mathcal{F}(\nabla u_h) \|_T + \frac{1}{2} h_T^{\frac{1}{2}} \| n \cdot [\mathcal{F}(\nabla u_h)] \, \|_{\partial T}.$$

Like in the linear case we achieve an inconsistent estimator since in regions of contact there is an unknown contact force which causes a gap $|f + \text{div}\, \mathcal{F}(\nabla u_h)| > 0$.

4.2 Saddle point problem

In order to avoid the estimator's inconsistency, we introduce the Lagrangian formulation which looks very similar to the linear case.

4.2.1 Saddle point formulation

For the problem at hand there holds:
Find a pair $(u, \lambda) \in V \times \Lambda := \{q \in L_2(\Omega) |\, q \geq 0 \text{ a.e.}\}$ with

$$\begin{aligned}\mathcal{L}(u, \lambda) &= \inf_{\varphi \in V} \sup_{\omega \in \Lambda} \mathcal{L}(\varphi, \omega) \\ &= \inf_{\varphi \in V} \sup_{\omega \in \Lambda} \{J(\nabla \varphi) - (f, \varphi) - (\omega, \varphi - \Psi)\},\end{aligned}$$

with the nonlinear Fréchet-differentiable functional $J(\cdot)$.

Derivation with respect to φ and ω leads to the saddle point system:

$$\begin{array}{llcccl} u \in V: & (\mathcal{F}(\nabla u), \nabla \varphi) & - & (\lambda, \varphi) & = & (f, \varphi) & \forall \varphi \in V \\ \lambda \in \Lambda: & (u, \omega - \lambda) & & & \geq & (\Psi, \omega - \lambda) & \forall \omega \in \Lambda \end{array} \tag{4.9}$$

where $\mathcal{F}(\cdot)$ denotes the Fréchet-derivative of $J(\cdot)$. Existence can be proven like in the linear case. For uniqueness we note that there holds:

$$\lambda = -\mathrm{div}\, \mathcal{F}(\nabla u) - f \tag{4.10}$$

and furthermore

$$\int_\Omega (\mu - \lambda)(u - \Psi) dx \geq 0 \quad \forall \mu \in \Lambda. \tag{4.11}$$

Any other saddle point $(u^*, \lambda^*) \in V \times \Lambda$ of $\mathcal{L}$ fulfills

$$\begin{aligned} &- \mathrm{div}\, \mathcal{F}(\nabla u^*) = f + \lambda^* \text{ on } \Omega \\ &u^* \in V \\ &\int_\Omega (\mu - \lambda^*)(u^* - \Psi) dx \geq 0 \quad \forall \mu \in \Lambda \\ &\lambda^* \in \Lambda. \end{aligned} \tag{4.12}$$

Setting $\mu = \lambda^*$ in (4.11) and $\mu = \lambda$ in the third relation of (4.12) and summing up the two equations we receive

$$\int_\Omega (\lambda^* - \lambda)(u^* - u) dx \leq 0. \tag{4.13}$$

Using (4.10) there holds

$$\lambda^* - \lambda = -\mathrm{div}\, \mathcal{F}(\nabla u^*) + \mathrm{div}\, \mathcal{F}(\nabla u)$$

and hence by (4.13)

$$-\int_\Omega \mathrm{div}\, (\mathcal{F}(\nabla u^*) - \mathcal{F}(\nabla u))(u^* - u) dx = \int_\Omega (\mathcal{F}(\nabla u^*) - \mathcal{F}(\nabla u))(\nabla u^* - \nabla u) dx \leq 0.$$

Since we have a strong monotone operator we achieve

$$\gamma \|\nabla u^* - \nabla u\|^2 \leq \int_\Omega (\mathcal{F}(\nabla u^*) - \mathcal{F}(\nabla u))(\nabla u^* - \nabla u) dx \leq 0,$$

which gives $|u^* - u|_1 \leq 0$ and therefore $u^* = u$ and $\lambda^* = \lambda$.

4.2.2 Stabilisation

Discretising the problem again results in an unstable system:

$$\begin{array}{llcll} u_h \in V_h : & (\mathcal{F}(\nabla u_h), \nabla\varphi) \;-\; (\lambda_h, \varphi) & = & (f, \varphi) & \forall \varphi \in V_h \\ \lambda_h \in \Lambda_h : & (u_h, \omega - \lambda_h) & \geq & (\Psi, \omega - \lambda_h) & \forall \omega \in \Lambda_h \end{array} \tag{4.14}$$

with Λ_h consisting of cellwise bilinear or constant functions on each cell of $\mathbb{T}_h$. We know from (4.3):

$$\operatorname{div} \mathcal{F}(\nabla u) + \lambda + f = 0$$

in case of $u \in C^2(\Omega) \cap C(\bar{\Omega})$. Using this knowledge we extend system (4.9) by adding the resulting consistent term and after discretisation we obtain for $(u_h, \lambda_h) \in V_h \times \Lambda_h$:

$$\begin{aligned} (\mathcal{F}(\nabla u_h), \nabla\varphi) - (\lambda_h, \varphi) + (u_h, \omega - \lambda_h) + (\operatorname{div} \mathcal{F}(\nabla u_h) + \lambda_h, \delta(-\operatorname{div} \mathcal{F}(\nabla\varphi) + \omega)) \\ \geq (f, \varphi) + (\Psi, \omega - \lambda_h) - (f, \delta(-\operatorname{div} \mathcal{F}(\nabla\varphi) + \omega)) \end{aligned} \tag{4.15}$$

for all $(\varphi, \omega) \in V_h \times \Lambda_h$. We set

$$A_\delta(\{u_h, \lambda_h\}, \{\varphi, \omega\}) = F_\delta(\{\varphi, \omega\}) \quad \forall(\varphi, \omega) \in V_h \times U_h, \tag{4.16}$$

with

$$\begin{aligned} A_\delta(\{u_h, \lambda_h\}, \{\varphi, \omega\}) :=& (\mathcal{F}(\nabla u_h), \nabla\varphi) - (\lambda_h, \varphi) + (u_h, \omega) \\ &+ (\operatorname{div} \mathcal{F}(\nabla u_h) + \lambda_h, \delta(-\operatorname{div} \mathcal{F}(\nabla\varphi) + \omega)) \end{aligned}$$

and right hand side

$$F_\delta(\{\varphi, \omega\}) := (f, \varphi) + (\Psi, \omega) - (f, \delta(-\operatorname{div} \mathcal{F}(\nabla\varphi) + \omega)),$$

where $\delta > 0$ is a piecewise constant parameter function and U_h denotes the discretisation of $L_2(\Omega)$. The natural mesh-dependent norm corresponding to A_δ is given by

$$|||\{u_h, \lambda_h\}|||_\delta^2 := |u_h|_1^2 + \|\delta^{\frac{1}{2}} \lambda_h\|^2.$$

For A_δ there holds

$$\begin{aligned}
A_\delta(\{\varphi,\omega\},\{\varphi,\omega\}) &= (\mathcal{F}(\nabla\varphi),\nabla\varphi) - (\omega,\varphi) + (\varphi,\omega) \\
&\quad + (\delta(\operatorname{div}\mathcal{F}(\nabla\varphi) + \omega), -\operatorname{div}\mathcal{F}(\nabla\varphi) + \omega) \\
&= (\mathcal{F}(\nabla\varphi),\nabla\varphi) + (\delta(\operatorname{div}\mathcal{F}(\nabla\varphi) + \omega), -\operatorname{div}\mathcal{F}(\nabla\varphi) + \omega) \\
&= (\mathcal{F}(\nabla\varphi),\nabla\varphi) - (\delta\operatorname{div}\mathcal{F}(\nabla\varphi), \operatorname{div}\mathcal{F}(\nabla\varphi)) + \|\delta^{\frac{1}{2}}\omega\|^2 \\
&\geq (\mathcal{F}(\nabla\varphi),\nabla\varphi) - \|\delta^{\frac{1}{2}}\operatorname{div}\mathcal{F}(\nabla\varphi)\|^2 + \|\delta^{\frac{1}{2}}\omega\|^2.
\end{aligned}$$

Since we have bilinear elements for φ, $\operatorname{div}(\cdot)$ is linear and bounded and therefore continuous:

$$\|\operatorname{div}\mathcal{F}(\nabla\varphi_1) - \operatorname{div}\mathcal{F}(\nabla\varphi_2)\| \leq M\|\mathcal{F}(\nabla\varphi_1) - \mathcal{F}(\nabla\varphi_2)\| \tag{4.17}$$

with a positive constant M. Hence we get

$$A_\delta(\{\varphi,\omega\},\{\varphi,\omega\}) \geq (\mathcal{F}(\nabla\varphi),\nabla\varphi) - M\delta^*\|\mathcal{F}(\nabla\varphi)\|^2 + \|\delta^{\frac{1}{2}}\omega\|^2$$

where $\delta^* = \max_T \delta_T$.

Due to the strong monotonicity of $\mathcal{F}(\cdot)$ and the Lipschitz continuity we can do the following estimation with constants m and L from (4.2) and (4.1):

$$\begin{aligned}
A_\delta(\{\varphi,\omega\},\{\varphi,\omega\}) &\geq m\|\nabla\varphi\|^2 - M\delta^* L\|\nabla\varphi\|^2 + \|\delta^{\frac{1}{2}}\omega\|^2 \\
&= (m - M\delta^* L)\|\nabla\varphi\|^2 + \|\delta^{\frac{1}{2}}\omega\|^2.
\end{aligned}$$

Theorem 4.1. *If* $0 < \delta^* < \frac{m}{M\,L}$*, with* $\delta^* = \max_T \delta_T$*,* M *from* (4.17) *and* m *and* L *denoting the constants taken from the strong monotonicity* (4.2) *and the Lipschitz continuity* (4.1)*,* A_δ *is coercive with a positive constant* c*:*

$$A_\delta(\{\varphi,w\},\{\varphi,w\}) \geq c|||\{\varphi,w\}|||_\delta^2.$$

So for δ sufficient small stability is guaranteed. Now we have an arbitrary operator $\mathcal{F}(\cdot)$ fulfilling the conditions of Theorem 2.4, which can be very complex and relating to the mixed terms arising due to the stabilisation, the equation might be extensive for programming. So inspired by [Be95] we introduce an alternative way of stabilisation with the confinement that Λ_h can only contain cellwise constant elements: Find $(u_h, \lambda_h) \in V_h \times \Lambda_h$ such that:

$$\begin{array}{rcccl} (\mathcal{F}(\nabla u_h), \nabla\varphi) & - & (\lambda_h, \varphi) & = & (f, \varphi) \quad \forall\varphi \in V_h \\ (u_h, \omega - \lambda_h) & + & \sum_T \delta_T \sum_{\Gamma\subset T}([\lambda_h]_\Gamma, [\omega - \lambda_h]_\Gamma)_\Gamma & \geq & (\Psi, \omega - \lambda_h) \quad \forall\omega \in \Lambda_h \end{array} \tag{4.18}$$

where $[\cdot]_\Gamma$ denotes the jump over an element edge Γ. This is also a consistent stabilisation since the jump terms vanish for the continuous solution λ. The term $\sum_T \delta_T \sum_{\Gamma\subset T}([\lambda_h]_\Gamma, [\omega]_\Gamma)_\Gamma$ has the meaning of a weighted discrete Laplacian.

We set

$$(\delta[\lambda_h], [\omega]) = \sum_T \delta_T \sum_{\Gamma\subset T}([\lambda_h]_\Gamma, [\omega]_\Gamma)_\Gamma$$

and

$$\|\delta^{\frac{1}{2}}[\lambda_h]\|^2 = (\delta\,[\lambda_h], [\lambda_h]).$$

Existence can also be proven by defining a mesh-dependend (semi-)norm

$$|\!|\!|\{u_h, \lambda_h\}|\!|\!|_\delta^2 := |u_h|_1^2 + \|\delta^{\frac{1}{2}}[\lambda_h]\|^2.$$

With a redefinition of

$$A_\delta(\{u_h, \lambda_h\}, \{\varphi, \omega\}) := (\mathcal{F}(\nabla u_h), \nabla\varphi) - (\lambda_h, \varphi) + (u_h, \omega) + (\delta\,[\lambda_h], [\omega])$$

and right hand side

$$F_\delta(\{\varphi, \omega\}) := (f, \varphi) + (\Psi, \omega)$$

we can easily see

$$A_\delta(\{\varphi, \omega\}, \{\varphi, \omega\}) \geq c|\!|\!|\{\varphi, \omega\}|\!|\!|_\delta^2, \quad 0 < c < 1.$$

Here $\|\delta^{\frac{1}{2}}[\lambda_h]\|^2$ is a seminorm so we have uniqueness only for the primal variable.

Again we want to solve the system with the help of Uzawa's algorithm (Algorithm 3.3.2). However, we have to handle a nonlinear equation.

We linearise the system using the SQP-Method and have to solve Uzawa's algorithm in every SQP-step:

Algorithm 4.3.

1 *Choose a start value* u_h^0 *and set* $i = 0$. *Let* $\gamma > 0$ *be the error tolerance of the SQP-Algorithm and* $\varepsilon > 0$ *be the one of Uzawa's algorithm.*

2 *Linearise the equation*

$$A(u_h^i)u_h = f - B^T\lambda_h$$

with u_h^i *describing the point of approximation.*

3 *Choose a start value* λ_h^0 *and* u_h^0 *for Uzawa's algorithm, set* $k = 1$.

4 *Solve the linearised system*

$$\begin{aligned} (A - A')(u_h^i)u_h^k &= f - B^T\lambda_h^{k-1} - A'(u_h^i)u_h^i \\ \lambda_h^k &= \max\{0, \lambda_h^{k-1} + \alpha(Bu_h^k - C\lambda_h^{k-1} - g)\}. \end{aligned}$$

5 *If* $\frac{\|u_h^k - u_h^{k-1}\|}{\|u_h^0 - u_h^k\|} > \varepsilon$, *set* $k = k+1$ *and go to step 4, else set* $u_h^{i+1} = u_h^k$, $\lambda_h = \lambda_h^k$ *and go to step 6.*

6 *If* $\|A(u_h^{i+1})u_h^{i+1} - f + B^T\lambda_h\| > \gamma$, *set* $i = i+1$ *and go to step 2, else finish.*

4.2.3 A posteriori error analysis

In Section 4.1.3 we recognised that the a posteriori error estimator we derived without using Langrange techniques is not consistent. This problem can be eliminated by the following approach using the Lagrangian multiplier. We start with an estimator for the stabilised system (4.15) and rewrite (4.14) in the following way:

Find $(u_h, \lambda_h) \in V_h \times \Lambda_h$ such that:

$$\begin{aligned}
&(\mathcal{F}(\nabla u_h), \nabla\varphi) - (\operatorname{div}\mathcal{F}(\nabla u_h), \delta \operatorname{div}\mathcal{F}(\nabla\varphi)) - (\lambda_h, \varphi) - (\lambda_h, \delta \operatorname{div}\mathcal{F}(\nabla\varphi)) \\
&\qquad\qquad = (f, \varphi) + (f, \delta \operatorname{div}\mathcal{F}(\nabla\varphi)) \qquad \forall\varphi \in V_h \\
&(u_h, \omega - \lambda_h) + (\operatorname{div}\mathcal{F}(\nabla u_h), \delta\omega) + (\lambda_h, \delta\omega) \\
&\qquad\qquad \geq (\Psi, \omega - \lambda_h) - (f, \delta\omega) \qquad \forall\omega \in \Lambda_h.
\end{aligned}$$

For the estimation there holds

$$\begin{aligned}
(\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e_i) &= (\mathcal{F}(\nabla u), \nabla e_i) - (f, e_i) - (\lambda_h, e_i) \\
&\quad - (\operatorname{div}\mathcal{F}(\nabla u_h) + f + \lambda_h, \delta \operatorname{div}\mathcal{F}(\nabla e_i)) \\
&= (\mathcal{F}(\nabla u), \nabla e_i - \nabla e + \nabla e) - (f, e_i) - (\lambda_h, e_i) \\
&\quad - (\operatorname{div}\mathcal{F}(\nabla u_h) + f + \lambda_h, \delta \operatorname{div}\mathcal{F}(\nabla e_i)) \\
&= -(f, e_i) - (\lambda_h, e_i) + (\mathcal{F}(\nabla u), \nabla e_i - \nabla e) + (\mathcal{F}(\nabla u), \nabla e) \\
&\quad - (\operatorname{div}\mathcal{F}(\nabla u_h) + f + \lambda_h, \delta \operatorname{div}\mathcal{F}(\nabla e_i)) \\
&\leq -(f, e_i) - (\lambda_h, e_i) + (\mathcal{F}(\nabla u), \nabla e_i - \nabla e) + (f, e) \\
&\quad - (\operatorname{div}\mathcal{F}(\nabla u_h) + f + \lambda_h, \delta \operatorname{div}\mathcal{F}(\nabla e_i)).
\end{aligned}$$

The last inequality originates from testing (4.4) with u_h. Then we continue with

$$\begin{aligned}
(\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e) &= (\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e - \nabla e_i + \nabla e_i) \\
&\leq (\mathcal{F}(\nabla u), \nabla(e - e_i)) - (\mathcal{F}(\nabla u_h), \nabla(e - e_i)) \\
&\qquad + (\mathcal{F}(\nabla u), \nabla(e_i - e)) + (f, e - e_i) - (\lambda_h, e_i) \\
&\qquad - (\operatorname{div}\mathcal{F}(\nabla u_h) + f + \lambda_h, \delta \operatorname{div}\mathcal{F}(\nabla e_i)) \\
&= (f, e - e_i) - (\mathcal{F}(\nabla u_h), \nabla(e - e_i)) - (\lambda_h, e_i) \\
&\qquad - (\operatorname{div}\mathcal{F}(\nabla u_h) + f + \lambda_h, \delta \operatorname{div}\mathcal{F}(\nabla e_i)).
\end{aligned} \tag{4.19}$$

Looking at the last terms we go on estimating as follows:

$$\begin{aligned}
-(\lambda_h, e_i) & -(\operatorname{div} \mathcal{F}(\nabla u_h) + f + \lambda_h, \delta \operatorname{div} \mathcal{F}(\nabla e_i)) \\
& \leq \quad -(\lambda_h, e_i - e + e) \\
& \qquad + \sum_T \delta_T \| \operatorname{div} \mathcal{F}(\nabla u_h) + f + \lambda_h \|_T \, \| \operatorname{div} \mathcal{F}(\nabla e_i) \|_T \\
& \overset{(4.17)}{\leq} \quad (\lambda_h, e - e_i) - (\lambda_h, e) \\
& \qquad + M \sum_T \delta_T \| \operatorname{div} \mathcal{F}(\nabla u_h) + f + \lambda_h \|_T \, \| \mathcal{F}(\nabla e_i) \|_T \\
& \overset{(4.1)}{\leq} \quad (\lambda_h, e - e_i) - (\lambda_h, u - u_h + \Psi - \Psi) \\
& \qquad + LM \sum_T \delta_T \| \operatorname{div} \mathcal{F}(\nabla u_h) + f + \lambda_h \|_T \, \| \nabla e_i \|_T \\
& \leq \quad (\lambda_h, e - e_i) + (\lambda_h, \Psi - u) + (\lambda_h, u_h - \Psi) \\
& \qquad + LM \sum_T \delta_T \| \operatorname{div} \mathcal{F}(\nabla u_h) + f + \lambda_h \|_T \, \| \nabla e \|_T \\
& \overset{(\Psi \leq u)}{\leq} \quad (\lambda_h, e - e_i) + (\lambda_h, u_h - \Psi) \\
& \qquad + LM \sum_T \delta_T \| \operatorname{div} \mathcal{F}(\nabla u_h) + f + \lambda_h \|_T \, \| \nabla e \|_T .
\end{aligned}$$

Again, we use the strong monotonicity of the operator $\mathcal{F}(\cdot)$ which allows the following estimation, setting $c = LM$:

$$\begin{aligned}
\gamma \| \nabla (u - u_h) \|^2 & \leq (\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e) \\
& \leq (f, e - e_i) - (\mathcal{F}(\nabla u_h), \nabla (e - e_i)) + (\lambda_h, e - e_i) \\
& \quad + (\lambda_h, u_h - \Psi) + c \sum_T \delta_T \| \operatorname{div} \mathcal{F}(\nabla u_h) + f + \lambda_h \|_T \, \| \nabla e \|_T
\end{aligned}$$

and by partial integration we achieve:

$$\begin{aligned}\gamma\|\nabla(u-u_h)\|^2 &\le (f+\lambda_h, e-e_i) + (\lambda_h, u_h-\Psi)\\ &\quad + c\sum_T \delta_T\| \operatorname{div}\mathcal{F}(\nabla u_h) + f + \lambda_h\|_T\,\|\nabla e\|_T\\ &\quad - \sum_T \Big[(-\operatorname{div}\,\mathcal{F}(\nabla u_h), e-e_i)_T + \int_{\partial T} n\cdot\mathcal{F}(\nabla u_h)\cdot(e-e_i)d\Gamma\Big]\\ &\le (\lambda_h, u_h-\Psi) + c\sum_T \delta_T\| \operatorname{div}\mathcal{F}(\nabla u_h) + f + \lambda_h\|_T\,\|\nabla e\|_T\\ &\quad + \sum_T \Big[\|f+\lambda_h + \operatorname{div}\,\mathcal{F}(\nabla u_h)\|_T\|e-e_i\|_T\\ &\qquad + \frac{1}{2}\|n\cdot[\mathcal{F}(\nabla u_h)]\|_{\partial T}\|e-e_i\|_{\partial T}\Big].\end{aligned}$$

There holds

$$\gamma(\nabla e,\nabla e) \le \sum_{T\in\mathbb{T}_h} \omega_T\varrho_T + N,$$

with local residuals ϱ_T and weights ω_T defined by

$$\begin{aligned}\varrho_T &:= h_T\|f+\lambda_h+\operatorname{div}\,\mathcal{F}(\nabla u_h)\|_T + \frac{1}{2}h_T^{\frac{1}{2}}\|n\cdot[\mathcal{F}(\nabla u_h)]\,\|_{\partial T},\\ \omega_T &:= \max\{h_T^{-1}\|e-e_i\|_T, h_T^{-\frac{1}{2}}\|e-e_i\|_{\partial T}\},\end{aligned}$$

and

$$\tilde{N} = (\lambda_h, u_h-\Psi) + c\sum_T \delta_T\|\operatorname{div}\mathcal{F}(\nabla u_h) + f + \lambda_h\|_T\,\|\nabla e\|_T.$$

After using estimates (3.4), (3.5) for ω_T and Young's inequality like presented in Section 3.2.3 we achieve

Theorem 4.2.1. *For problem* (4.15) *there holds the a posteriori error bound*

$$\gamma(\nabla e,\nabla e) \le C\sum_{T\in\mathbb{T}_h} \varrho_T^2 + N, \tag{4.20}$$

with local residuals ϱ_T *defined by*

$$\varrho_T := (h_T+\delta_T)\|f+\lambda_h+\operatorname{div}\,\mathcal{F}(\nabla u_h)\|_T + \frac{1}{2}h_T^{\frac{1}{2}}\|n\cdot[\mathcal{F}(\nabla u_h)]\,\|_{\partial T},$$

where for interior interelement boundaries $n \cdot [\mathcal{F}(\nabla u_h)]$ denotes the jump of $\mathcal{F}(\nabla u_h)$ in normal direction and

$$N = (\lambda_h, u_h - \Psi).$$

We derive a second estimator for system (4.18):

$$\begin{aligned}(\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e_i) &= (\mathcal{F}(\nabla u), \nabla e_i) - (f, e_i) - (\lambda_h, e_i)\\ &= (\mathcal{F}(\nabla u), \nabla e_i - \nabla e + \nabla e) - (f, e_i) - (\lambda_h, e_i)\\ &= -(f, e_i) - (\lambda_h, e_i) + (\mathcal{F}(\nabla u), \nabla e_i - \nabla e) + (\mathcal{F}(\nabla u), \nabla e)\\ &\le -(f, e_i) - (\lambda_h, e_i) + (\mathcal{F}(\nabla u), \nabla e_i - \nabla e) + (f, e).\end{aligned}$$

Again, in the last inequality we test (4.4) with u_h. Then we continue with

$$\begin{aligned}(\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e) &= (\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e - \nabla e_i + \nabla e_i)\\ &\le (\mathcal{F}(\nabla u), \nabla(e - e_i)) - (\mathcal{F}(\nabla u_h), \nabla(e - e_i))\\ &\quad + (\mathcal{F}(\nabla u), \nabla(e_i - e)) + (f, e - e_i) - (\lambda_h, e_i)\\ &= (f, e - e_i) - (\mathcal{F}(\nabla u_h), \nabla(e - e_i)) - (\lambda_h, e_i).\end{aligned} \tag{4.21}$$

Looking at the last term of (4.21) again, we now go on estimating:

$$\begin{aligned}(\lambda_h, e_i) &= (\lambda_h, e_i - e + e)\\ &= -(\lambda_h, e - e_i) + (\lambda_h, e)\\ &= -(\lambda_h, e - e_i) + (\lambda_h, u - u_h + \Psi - \Psi + u_h - u_h)\\ &= -(\lambda_h, e - e_i) - (\lambda_h, \Psi - u) - (\lambda_h, u_h - \Psi) + (\lambda_h, u_h) - (\lambda_h, u_h).\end{aligned} \tag{4.22}$$

Testing (4.18) with $\omega = 0$ there holds by (4.21) and (4.22)

$$\begin{aligned}
(\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e) \leq & (f, e - e_i) - (\mathcal{F}(\nabla u_h), \nabla(e - e_i)) + (\lambda_h, e - e_i) \\
& + \underbrace{(\lambda_h, \Psi - u)}_{\leq 0} + (\lambda_h, u_h - \Psi) - (\lambda_h, u_h) + (\lambda_h, u_h) \\
\leq & (f, e - e_i) - (\mathcal{F}(\nabla u_h), \nabla(e - e_i)) + (\lambda_h, e - e_i) \\
& + (\lambda_h, u_h - \Psi) - (\lambda_h, u_h) + (\Psi, \lambda_h) \\
& - \sum_T \delta_T \sum_{\Gamma \subset T} ([\lambda_h]_\Gamma, [\lambda_h]_\Gamma)_\Gamma \\
\leq & (f, e - e_i) - (\mathcal{F}(\nabla u_h), \nabla(e - e_i)) + (\lambda_h, e - e_i) \\
& + (\lambda_h, u_h - \Psi) + \underbrace{(\lambda_h, \Psi - u_h)}_{\leq 0} + \sum_T \delta_T \sum_{\Gamma \subset T} \|[\lambda_h]\|_\Gamma^2 \\
\leq & (f, e - e_i) - (\mathcal{F}(\nabla u_h), \nabla(e - e_i)) + (\lambda_h, e - e_i) \\
& + (\lambda_h, u_h - \Psi) + \sum_T \delta_T h_T (\max_{\Gamma \subset T} |\, [\lambda_h]_\Gamma \,|)^2.
\end{aligned}$$

By the strong monotonicity and integration by parts we have

$$\begin{aligned}
\gamma \|\nabla(u - u_h)\|^2 & \leq (\mathcal{F}(\nabla u) - \mathcal{F}(\nabla u_h), \nabla e) \\
& \leq (f, e - e_i) - (\mathcal{F}(\nabla u_h), \nabla(e - e_i)) + (\lambda_h, e - e_i) \\
& \quad + (\lambda_h, u_h - \Psi) + \sum_T \delta_T h_T (\max_{\Gamma \subset T} |\, [\lambda_h]_\Gamma \,|)^2 \\
& = (f + \lambda_h, e - e_i) + (\lambda_h, u_h - \Psi) + \sum_T \delta_T h_T (\max_{\Gamma \subset T} |\, [\lambda_h]_\Gamma \,|)^2 \\
& \quad - \sum_T \Big[(-\operatorname{div} \mathcal{F}(\nabla u_h), e - e_i)_T + \int_{\partial T} n \cdot \mathcal{F}(\nabla u_h) \cdot (e - e_i) d\Gamma \Big] \\
& \leq \sum_T \Big[\|f + \lambda_h + \operatorname{div} \mathcal{F}(\nabla u_h)\|_T \|e - e_i\|_T \\
& \qquad + \frac{1}{2} \|n \cdot [\mathcal{F}(\nabla u_h)]\|_{\partial T} \|e - e_i\|_{\partial T} \Big] \\
& \quad + (\lambda_h, u_h - \Psi) + \sum_T \delta_T h_T (\max_{\Gamma \subset T} |\, [\lambda_h]_\Gamma \,|)^2.
\end{aligned}$$

Finally, there holds

$$\gamma(\nabla e, \nabla e) \leq \sum_{T \in \mathbb{T}_h} \omega_T \varrho_T + N,$$

with local residuals ϱ_T and weights ω_T defined by

$$\begin{aligned}\varrho_T &:= h_T\|f+\lambda_h+\mathrm{div}\ \mathcal{F}(\nabla u_h)\|_T + \frac{1}{2}h_T^{\frac{1}{2}}\|n\cdot[\mathcal{F}(\nabla u_h)]\,\|_{\partial T},\\ \omega_T &:= \max\{h_T^{-1}\|e-e_i\|_T, h_T^{-\frac{1}{2}}\|e-e_i\|_{\partial T}\},\end{aligned}$$

and

$$N=(\lambda_h,u_h-\Psi)+\sum_T \delta_T h_T(\max_{\Gamma\subset T}|\,[\lambda_h]_\Gamma\,|)^2. \tag{4.23}$$

After using estimates (3.4), (3.5) and Young's inequality we achieve

Theorem 4.2.2. *For problem* (4.18) *there holds the a posteriori error bound*

$$\gamma(\nabla e,\nabla e)\le C\sum_{T\in\mathbb{T}_h}\varrho_T^2+N, \tag{4.24}$$

with local residuals ϱ_T *defined by*

$$\varrho_T := h_T\|f+\lambda_h+\mathrm{div}\ \mathcal{F}(\nabla u_h)\|_T + \frac{1}{2}h_T^{\frac{1}{2}}\|n\cdot[\mathcal{F}(\nabla u_h)]\,\|_{\partial T},$$

where for interior interelement boundaries $n\cdot[\mathcal{F}(\nabla u_h)]$ *denotes the jump of* $\mathcal{F}(\nabla u_h)$ *in normal direction and* N *defined by* (4.23).

4.3 Numerical results

As a test example we consider the minimal surface problem with the nonlinear operator $\mathcal{F}(\nabla u)=\frac{1}{\sqrt{1+|\nabla u|^2}}\nabla u$. $\mathcal{F}(\cdot)$ fulfills the conditions of Theorem 2.4 so the variational problem is well defined and has a unique solution (see for example [Kl09] or [LS71] and the references there in).

4.3.1 Adaptivity

Choosing the domain $\Omega = [-1,1]^2$, the obstacle $\Psi = -0.15$ and a body force $f = 5 \cdot \exp(-9(x^2+y^2))$, we perform some global refinement on our test example to find out the convergence rate of estimator (4.24). Table 4.1 presents a linear convergence in this case.

Remark 4.3.1. *We have a constant obstacle here, so the divergence-terms in system* (4.15) *almost vanish in the contact zone. Since they also get very small elsewhere, we can neglect them for our test example and stabilise with the C-Matrix* $(\lambda_h, \delta\omega)$ *and the corresponding right hand side* $(f, \delta\omega)$.

Remark 4.3.2. *Using* Q_1/Q_0*-Elements, estimator* (4.20) *and* (4.24) *offer very similar solutions using the corresponding stable systems, so that we show tables and figures for only one of them in each case.*

# cells	e_{res}		e_{jump}		N		$\lvert e\rvert_1$	
64	3.956e-01	-	1.245e-01	-	6.867e-02	-	5.152e-01	-
256	2.033e-01	0.96	7.880e-02	0.66	2.970e-02	1.21	2.796e-01	0.88
1024	9.902e-02	1.04	4.334e-02	0.86	1.267e-02	1.23	1.415e-01	0.99
4096	4.835e-02	1.04	2.238e-02	0.96	5.487e-03	1.21	7.045e-02	1.01
16384	2.419e-02	1.00	1.141e-02	0.97	2.329e-03	1.24	3.547e-02	0.99
65536	1.213e-02	1.00	5.766e-03	0.98	1.045e-03	1.16	1.782e-02	0.99

Table 4.1: We set $e_{\text{res}} = (\sum_T h_T^2 \|\Delta u_h + \lambda_h + f\|_T^2)^{\frac{1}{2}}$, $e_{\text{jump}} = (\sum_T h_T \|n \cdot [\mathcal{F}(\nabla u_h)]\|_{\partial T}^2)^{\frac{1}{2}}$ and $|e|_1$ denotes the complete error estimator. In the right columns there is always a value α determining the convergence order by $\mathcal{O}(h^\alpha)$ for every refinement step. Global refinements lead to a descend of $\mathcal{O}(h)$ for the whole estimator (4.24).

By comparing estimator (4.8) to (4.24) within adaptive refinement, we achieve the same effect as in the linear case. The estimator without the Lagrangian multiplier causes an over-refinement in the contact zone, whereas the improved estimator concentrates on the transition region (Figure 4.1).

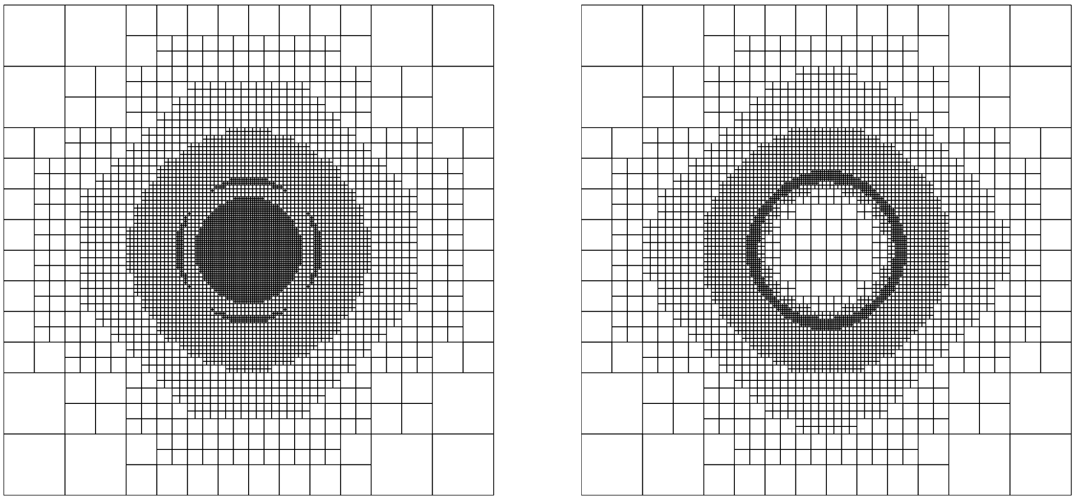


Figure 4.1: Adaptive refinement by different estimators for the minimal surface problem with a constant obstacle. The left grid has been refined by estimator (4.8) and shows an over-refinement of the contact zone similar to the linear case. The estimator including the Lagrangian multiplier (4.24) produces a coarse grid in the contact zone and a very fine one at the transition area, where the material that is pushed down by a body force touches the obstacle.

The reason for this effect gets clear if we look at Figure 4.5. There, we compare the inconsistent term of (4.8) to the corresponding term in (4.24) and (4.20), respectively and observe that the residual error in the contact zone rests constant in the inconsistent case which causes the over-refinement. In areas of contact $(\sum_T \|f + \operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ gives the norm of f, which is constant in every refinement step whereas in $(\sum_T \|f + \lambda_h + \operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ the Lagrangian multiplier eliminates this inconsistency and reduces the error uniformly (see Figure 4.5).

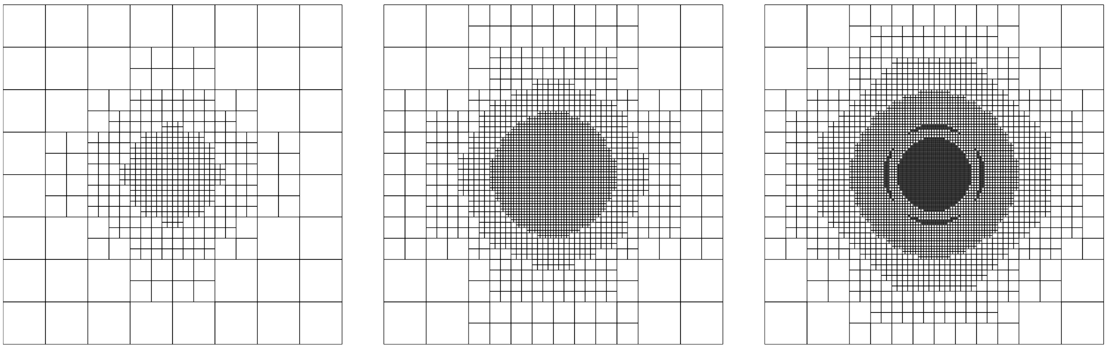


Figure 4.2: Sequence of grids created by estimator (4.8) offering a well refined contact zone.

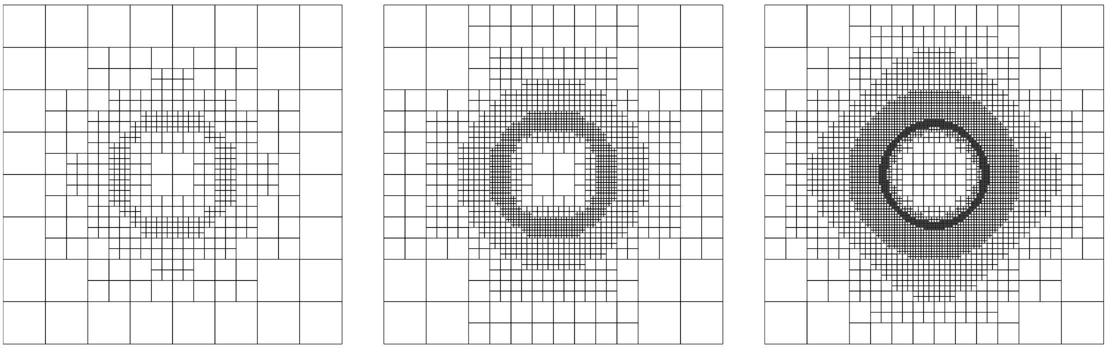


Figure 4.3: Sequence of grids created by estimator (4.24). Due to the consistent residual term there is almost no measurable error in the contact zone which involves more economical mesh structures.

The grids are more economical since we get the desired accurateness with less degrees of freedom than in the inconsistent case, and the estimator offers a sharper error bound which can be seen in Figure 4.4. The improvement will be even more obvious if we use biquadratic elements for u_h and compare the estimators again (Figure 4.7). By the use of bilinear elements, $\operatorname{div} \mathcal{F}(\nabla u_h)$ almost vanishes outside the contact zone, too. So here we only measure f in the norm which is no inconsistency but depends on the choice of the finite elements.

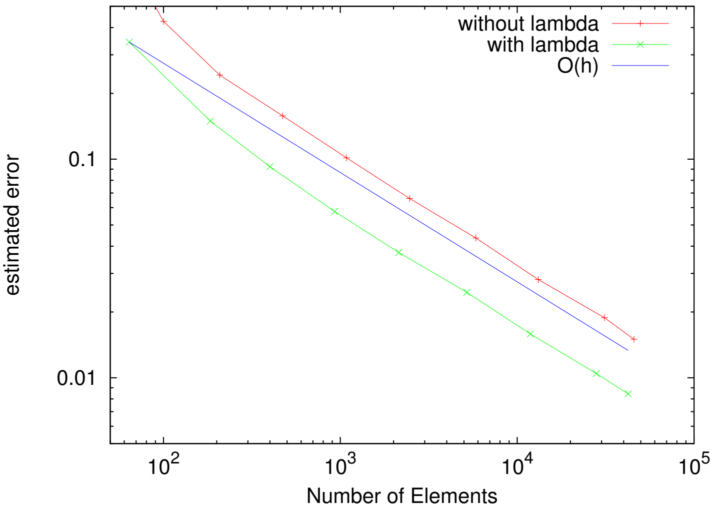


Figure 4.4: Comparison of estimator (4.8) and (4.24). Since both estimators are upper bounds of the true error, the Lagrange technique serves a sharper estimation.

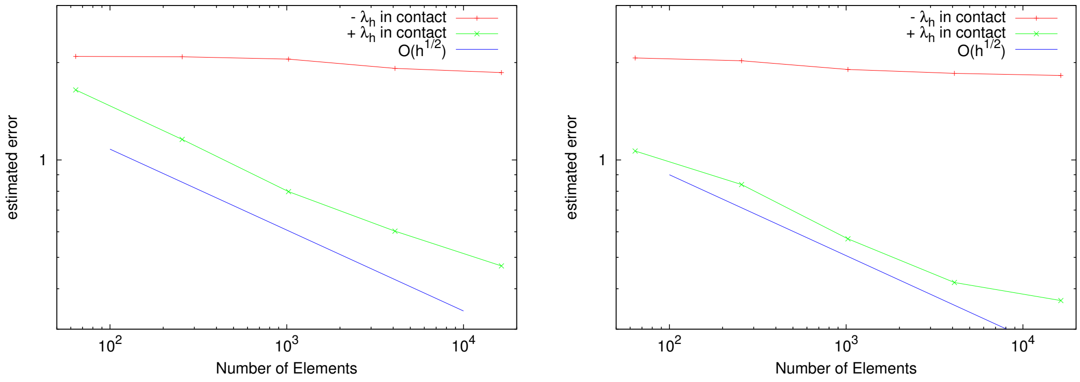


Figure 4.5: We compare $(\sum_T \|f+\lambda_h+\operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ to $(\sum_T \|f+\operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ by using the stabilisation term $\sum_T \delta_T \sum_{\Gamma\subset T}([\lambda_h]_\Gamma,[\omega]_\Gamma)$ with Q_1/Q_0-Elements (left) and stabilisation (4.15) for Q_1/Q_1-Elements (right). The effectivity of the Lagrangian multiplier can be observed since we get a consistent residual term.

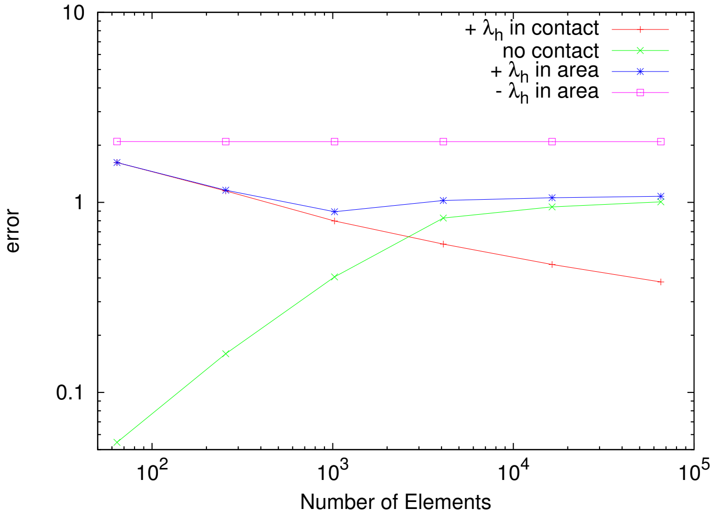


Figure 4.6: We compare $(\sum_T \|f+\lambda_h+\operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ to $(\sum_T \|f+\operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ within a global refinement. “+λ_h area“= $(\sum_T \|f+\lambda_h+\operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ in the whole area, “-λ_h area“= $(\sum_T \|f+\operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ in the whole area, ”no contact“= $(\sum_T \|f+\operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ in areas of no contact. Since the value of $(\sum_T \|f+\lambda_h+\operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ tends to zero in zones of contact the value of the norm in the whole area converges to the value the norm reaches when neglecting the contact zone which is constant here since we use bilinear elements for u_h and hence $\operatorname{div}\mathcal{F}(\nabla u_h)$ almost vanishes.

Since $(\sum_T \|f+\lambda_h+\operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ tends to zero in the contact zone, the value of the norm in the whole area converges to the one that is measured in this norm outside the contact area (see Figure 4.6) which rests constant after a few steps. Using biquadratic elements the norm $(\sum_T \|f+\lambda_h+\operatorname{div}\mathcal{F}(\nabla u_h)\|_T^2)^{\frac{1}{2}}$ tends to zero in the whole area and hence the estimator gets a better convergence rate.

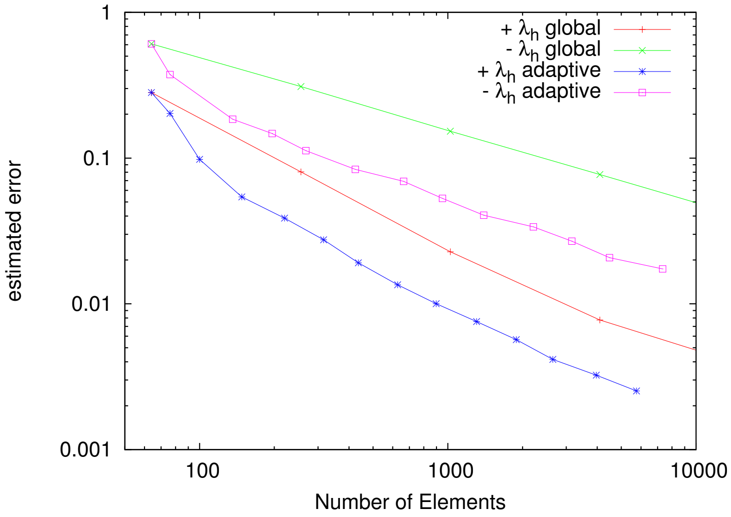


Figure 4.7: Comparison of estimator (4.8) and (4.20) using Q_2-Elements for the displacement.

The effectivity of estimator (4.24) can also be seen in the comparison of global and adaptive refinement. Here, we test with a discontinuous obstacle $\Psi = -0.08$ if $x < 0$, $y < 0$ or $x > 0$, $y > 0$ and $\Psi = -0.2$ elsewhere.

# cells	$\lvert e\rvert_{1,\text{global}}$		# cells	$\lvert e\rvert_{1,\text{adaptive}}$	
64	8.554e-01	-	64	8.554e-01	-
256	4.958e-01	0.79	232	2.879e-01	1.69
1024	2.898e-01	0.77	1108	1.277e-01	1.04
4096	1.747e-01	0.73	4276	6.594e-02	0.98
16384	1.083e-01	0.69	17428	3.310e-02	0.99

Table 4.2: Table of convergence of estimator (4.24) using global and adaptive refinement and a discontinuous obstacle. In the right columns there is always a value α determining the convergence order by $\mathcal{O}(h^{\alpha})$ for every refinement step. In the adaptive case we have a linear convergence whereas global refinement leads to a slower descend of the estimator.

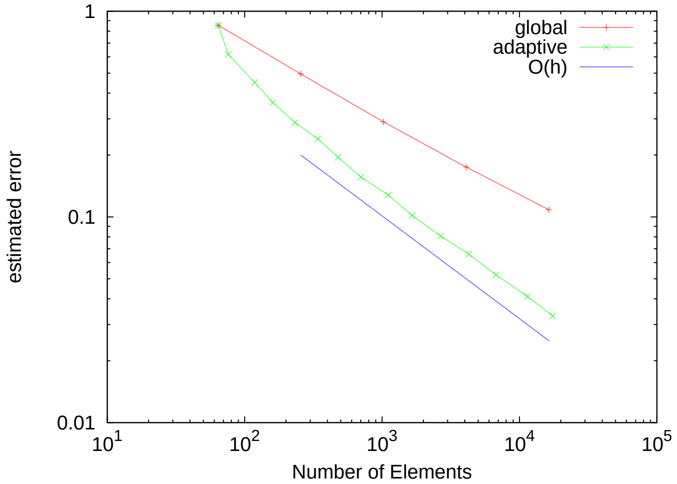
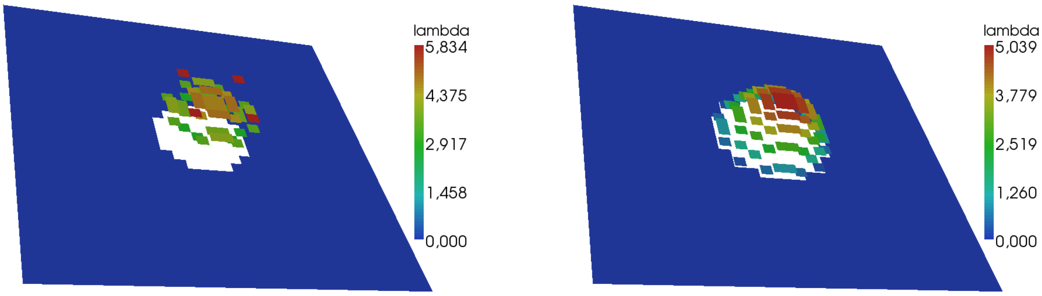


Figure 4.8: Using adaptive refinement in case of a discontinuous obstacle we see a better convergence than in the global case.

A sequence of global refinements only leads to a convergence order of about $\mathcal{O}(h^{\frac{3}{4}})$ whereas the adaptive refinement still reaches a rate of $\mathcal{O}(h)$.

4.3.2 Stability

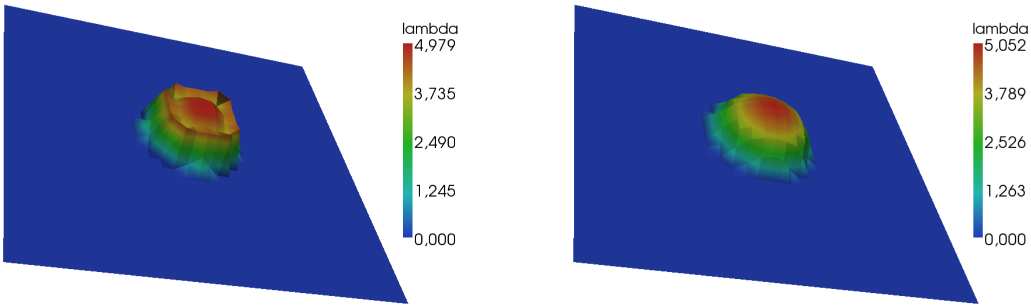
We calculate our test example with two different choices of finite element combinations and differ between the stabilised and the unstabilised system. Using Q_1/Q_0-Elements for (u_h, λ_h) without applying the least squares terms, we observe oscillations in the values of the Lagrangian multiplier as we can see in Figure 4.9(a). Furthermore, the solution u_h penetrates the obstacle a little bit. The same effects can be seen using Q_1/Q_1-Elements (Figure 4.10(a), Figure 4.11).



(a) Unstabilised: Lagrangian multiplier λ_h

(b) Stabilised: Lagrangian multiplier λ_h

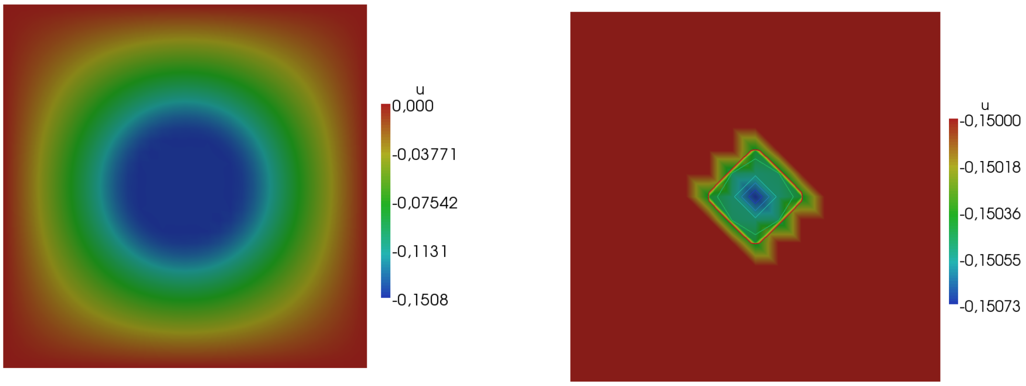
Figure 4.9: Using Q_1-Elements for displacement and Q_0-Elements for the Lagrangian multiplier there are oscillations in the latter for an unstabilised system (a). After stabilising we achieve the expected values for λ_h (b). Both variants of stabilisation ((4.15) and (4.18)) show the same effect.



(a) Unstabilised: Lagrangian multiplier λ_h

(b) Stabilised: Lagrangian multiplier λ_h

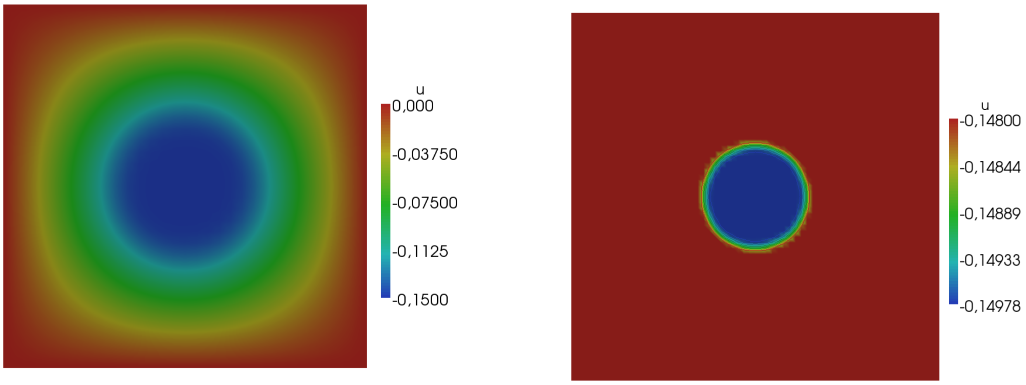
Figure 4.10: With Q_1/Q_1-Elements we observe the same effect like in the case using piecewise constant elements for the Lagrangian multiplier. If we leave the system unstabilised there are non-physical oscillations (a) in contrast to the stabilised system (b).



(a) Unstabilised: Displacement u_h

(b) Unstabilised: Contours of u_h

Figure 4.11: The values of the displacement u_h seek a little bit into the obstacle which can be seen on the scale of subfigure (a) and looking at the contours of the unstabilised problem in subfigure (b).



(a) Stabilised: Displacement u_h

(b) Stabilised: Contours of u_h

Figure 4.12: The obstacle is set by $\Psi = -0.15$ and not penetrated by u_h solving the stabilised system. Here we used Q_1/Q_1-Elements.

If we activate the stabilising terms λ_h gets very smooth and the imperfections concerning λ_h and u_h vanish, which can be observed in Figure 4.9(b) for a piecewise constant λ_h and in Figure 4.10(b) for Λ_h consisting of piecewise bilinear elements. The displacement u_h of the stabilised system is shown in Figure 4.12. Again, λ_h has the physical meaning of a counter force balancing f in the contact zone. A fitting procedure offers $\delta = 0.36h^2$ to be a good choice for both finite element combinations using stabilisation (4.15). When stabilising with the jump terms of λ_h we propose $\delta = 0.0002h$.

5 Torsion problem

The torsion problem which is studied in this chapter is an example for a linear problem with a nonlinear restriction. We give a short introduction how to achieve the variational form of the problem and present the existence results before introducing the corresponding saddle point formulation. The torsion problem is often presented with the restriction on a constant yield condition. In addition to that, we offer a theory with respect to a variable yield condition that may vary in space. On this base we develop an a posteriori estimator that turns out to produce efficient meshes which we compare to grids generated by known estimators. Furthermore, we give a short explanation of the construction of goal orientated estimators in case of the torsion problem. At the end of the chapter the new estimator is examined by numerical tests.

The construction of numerical algorithms for solving discrete problems of contact type like the obstacle problem in Chapter 3 is straight-forward, since the restrictions only apply to the solution itself. In the present case, we have to handle point-wise restrictions for the gradient of the solution. A standard example for variational inequalities with gradient constraints is the torsion problem which is described in [Gl83]. In what follows we basically conform to [Gl83],[Su08] and [DL76]. The physical situation behind the torsion problem is a cylindrical bar $\overline{\Omega} \in \mathbb{R}^3$, bounded by two plane sections Γ_0, Γ_1. The curved surface area is denoted by $\Gamma_2 = \partial\overline{\Omega}\backslash(\Gamma_0 \cup \Gamma_1)$. We assume that this bar is made up of an isotropic elastic perfectly plastic material whose plasticity yield is given by the Von Mises criterion. Starting from a zero-stress initial state, an increasing torsion moment is applied to the bar by twisting one of its ends and fix the other one.

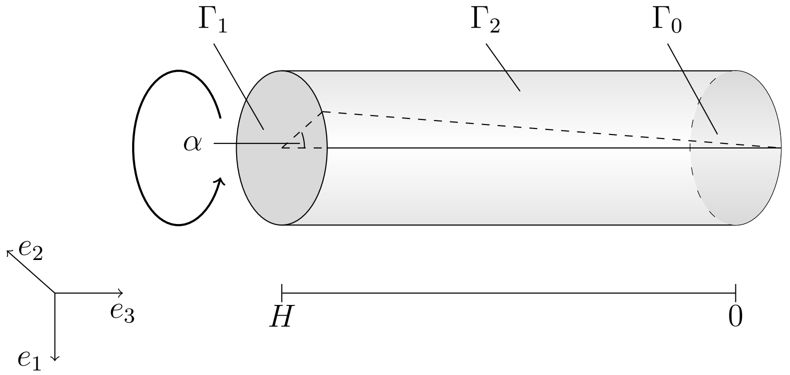


Figure 5.1: A cylindrical bar is twisted on Γ_1 and fixed on Γ_0. α denotes the angle of rotation.

The result is a displacement U which depends on the angle of rotation $\alpha > 0$. Let n be the outward normal of $\partial\overline{\Omega}$. The corresponding classical notation then reads

$$\begin{gathered} T \cdot n = 0 \quad \text{on } \Gamma_2 = \partial\overline{\Omega} \backslash (\Gamma_0 \cup \Gamma_1) \\ T_{33} = 0 \quad \text{on } \Gamma_0, \Gamma_1 \\ U_1 = U_2 = 0 \quad \text{on } \Gamma_0 \\ U_1 = -\alpha H x_2, \quad U_2 = \alpha H x_1 \quad \text{on } \Gamma_1, \end{gathered}$$

with a given $\alpha > 0$ and H denotes the height of the bar. Displacement U and the corresponding stress tensor T are affected by the volume force $F = 0$ which acts in $\overline{\Omega} \in \mathbb{R}^3$. Along the portion Γ_0 of the boundary the first two components of the displacement vector U are fixed, whereas on Γ_1 we have a prescribed torsion. The following pictures in Figure 5.2 show a milling process and its model situation which is an example of mechanical work where torsion appears.

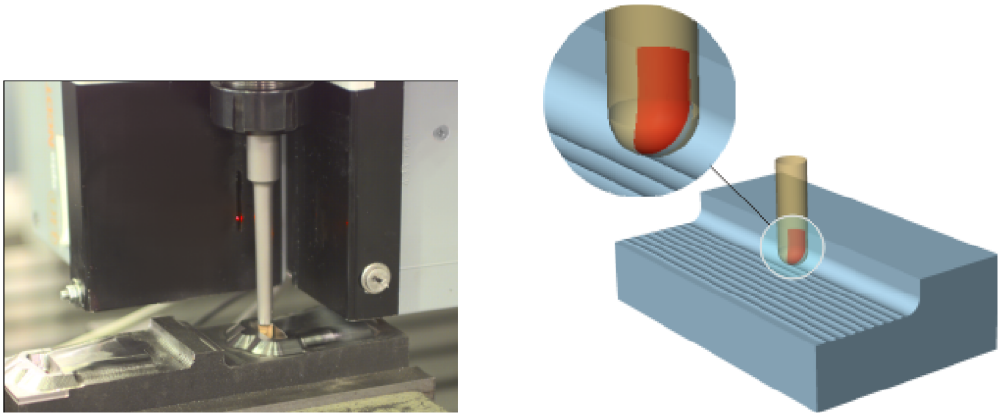


Figure 5.2: Snapshot of a milling process (left) and its model situation (right) taken from [Su08].

Twisting the bar results in an inner stress which is called shear stress. The tensor

$$T = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix}$$

describes the complete stress of the bar. We can reduce the number of unknowns by using physical laws. For kinematic deformation we assume that there is no strain along the main axis of the bar. So the diagonal of the stress tensor which describes the normal stresses disappears. In addition the stress tensor is symmetric so that there are only three unknown entries left: T_{12}, T_{13} and T_{23}. Torsion induces stress orthogonal to the normal of the surface, which in our case is the e_3-direction. Therefore T_{12} is zero, too. As described in [DL76] there are only two nonzero components for the stress left:

$$T_{ij}(x) = \begin{cases} T_{ij}(x_1, x_2) & \text{if } (ij = 13) \text{ or } (ij = 23) \\ 0 & \text{otherwise.} \end{cases}$$

Under these conditions the equilibrium equations $\operatorname{div} T + F = 0$ reduce to

$$\partial_1 T_{13} + \partial_2 T_{23} = 0.$$

That means there exists a $\Theta = \Theta(x_1, x_2)$ such that

$$T_{13} = \partial\Theta/\partial x_2, \quad T_{23} = -\partial\Theta/\partial x_1. \tag{5.1}$$

The equations $T_{ij}n_j = 0$ on Γ_2 can now be written as

$$d\Theta/ds = 0,$$

where s is a curvilinear abscissa on Γ_0. That means $\Theta =$ constant on the boundary of Γ_0 in $\mathbb{R}^2$ and since (5.1) defines Θ up to a constant, we can choose $\Theta \in H_0^1(\Gamma_0)$, in order to fulfill $T_{i3} \in L_2(\Omega)$. Taking account of the Von Mises yield criterion, $\{T_{ij}\} \in K_\sigma$, where K_σ is the set of admissible stresses, can then be written with reference to

$$\Theta \in K = \{v |\, v \in H_0^1(\Gamma_0), |\text{grad}\, v| \leq 1 \text{ a.e. on } \Gamma_0\}.$$

Following [DL76] (Chapter V, sec. 6.3) the field Θ associated with $\{T_{ij}\}$ by (5.1) minimises on K the functional

$$\frac{1}{2}\int_{\Gamma_0} |\text{grad }\ v|^2 dx_1 dx_2 + \mu\alpha \int_{\Gamma_0} (x_1 \partial v/\partial x_1 + x_2 \partial v/\partial x_2)\, dx_1 dx_2,$$

with the shear module μ. Using Greens formula we achieve

$$\frac{1}{2}\int_{\Gamma_0} |\text{grad }\ v|^2 dx_1 dx_2 - 2\mu\alpha \int_{\Gamma_0} v\, dx_1 dx_2.$$

Γ_0 is a 2D surface and for the further studies we set $\Omega := \Gamma_0$.

5.1 Variational formulation

Setting $a(\varphi, \psi) := (\nabla\varphi, \nabla\psi)$ and $f = 2\mu\alpha$, where the constant parameter $\mu > 0$ denotes the shear module, the torsion problem can now be written as:
Find $u \in K$, such that

$$a(u, v - u) \geq (f, v - u) \quad \forall v \in K, \tag{5.2}$$

where

$$K = \{v \in H_0^1(\Omega)| \ |\nabla v| \leq 1 \text{ a.e. in } \Omega\}.$$

This is equivalent to the minimisation problem:
Find $u \in K$ such that

$$J(u) < J(w) = \frac{1}{2}a(w, w) - (f, w) \quad \forall w \in K. \tag{5.3}$$

Theorem 5.1.1. *Problem* (5.2) *has a unique solution.*

For a proof see [Gl83].

The stress vector σ can be expressed by u by the relation $\sigma = \nabla u$, so that u has the physical meaning of a stress potential, and we can obtain σ once the solution of (5.3) is known.

$u \in K$ is approximated by $u_h \in K_h \subset K$ with

$$a(u_h, v - u_h) \geq (f, v - u_h) \quad \forall v \in K_h, \tag{5.4}$$

where

$$K_h = \{v \in V_h |\, |\nabla v| \leq 1\}, \quad V_h = \{v \in C(\overline{\Omega}) |\, v \text{ bilinear on } T \in \mathbb{T}_h\} \cap H_0^1(\Omega). \tag{5.5}$$

Since K_h is a closed convex nonempty subset of V_h, there holds the

Proposition 5.1.2. *The approximate problem* (5.4) *has a unique solution.*

The proof is analogue to the continuous case.

5.2 Saddle point problem

Yet, to find good a posteriori estimates for the torsion problem is quite difficult. One is given by [LY00] in the following form:

$$|u - u_h|_1^2 \leq C \sum_T (h_T^2 \|f + \Delta u_h\|_T^2 + h_T \|n \cdot [\nabla u_h]\|_{\partial T}^2 + \text{“terms of higher order”}). \tag{5.6}$$

This estimator is not consistent. We take for example a quadratic domain $\Omega = [-1, 1]^2$ like described in Section 5.3 and a constant body force. Torsion appears and the workpiece begins to plastify, starting at the boundaries.

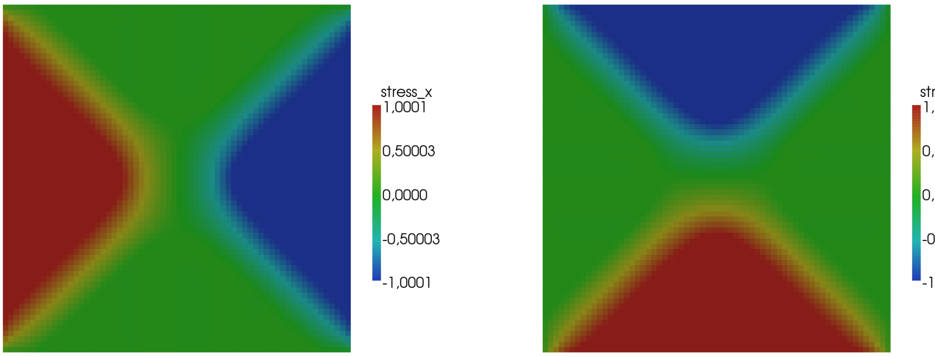


Figure 5.3: Stress in x- and y-direction of a torsion problem.

At each boundary there exists the main stress in x-direction or in y-direction so that in regions where $|\nabla u| = 1$ there holds $\Delta u \approx 0$ and hence $|f + \Delta u| > 0$. Furthermore, we will observe an over-refinement in the elastic zones (Section 5.3, Figure 5.7). That effect appears since the error of the jump terms is almost zero in regions of plasticity. Using cellwise bilinear elements for u_h the residual term rests constant for all cells and there is no further estimation for the plastic fraction. The advantage of the Lagrangian formalism applied to this problem is that an a posteriori error estimator can be derived that gives us very good mesh structures by refining adaptively.

5.2.1 Saddle point formulation

In order to present such an estimator, we introduce the Lagrangian functional

$$\mathcal{L}(v, \omega) = \frac{1}{2} a(v, v) - (f, v) + \frac{1}{2} \int_\Omega \omega((\nabla v)^2 - 1)\, dx \tag{5.7}$$

for $v \in V := H_0^1(\Omega)$ and $\omega \in \Lambda = \{q \in L^\infty(\Omega) |\, q \geq 0 \text{ a.e.}\}$. Following Céa [Ce78], we know that if there exists a saddle point (u, λ) then u is the solution of the minimisation functional (5.3). Existence of the saddle point is proven for the physical case of a constant f for example in [Br72]. We choose f in such a way that $\mathcal{L}(\cdot, \cdot)$ has a solution in $V \times \Lambda$. For the discretisation we choose K_h and V_h like in (5.5) and

$$\Lambda_h = \Lambda \cap L_h = \{\omega \in L_h |\, \omega_T \geq 0\, \forall T \in \mathbb{T}_h\}$$

with L_h consisting of cellwise constant functions like described in (3.18). Existence of the saddle point $(u_h, \lambda_h) \in V_h \times \Lambda_h$ for the discrete problem is guaranteed by Proposition 3.5 in [Gl83]. Like in the continuous case u_h is then the solution of problem (5.4). Derivation of (5.7) with respect to v and ω lead to the problem:
Find a pair $(u, \lambda) \in V \times \Lambda$ fulfilling the mixed formulation

$$\begin{aligned} ((1+\lambda)\nabla u, \nabla v) &= (f, v) \quad \forall v \in V \\ (|\nabla u|^2, \omega - \lambda) &\leq (1, \omega - \lambda) \quad \forall \omega \in \Lambda. \end{aligned} \tag{5.8}$$

Its discrete version reads

$$\begin{array}{lrcll} u_h \in V_h : & ((1+\lambda_h)\nabla u_h, \nabla v) & = & (f, v) & \forall v \in V_h \\ \lambda_h \in \Lambda_h : & (|\nabla u_h|^2, \omega - \lambda_h) & \leq & (1, \omega - \lambda_h) & \forall \omega \in \Lambda_h. \end{array}$$

Remark 5.2.1. *As we will see in Section 5.3 (Figure 5.14) the torsion problem discretised by Q_1/Q_0-Elements runs very robust. However, the system is not stable. A possible way of stabilisation is to add the consistent matrix, including the jumps of the Lagrangian multiplier over an element edge, to the linearised system (see Section 4.2.2). For linearisation we rewrite the system in an implicite way:*
Find $(u_h, \lambda_h) \in V_h \times \Lambda_h$ such that:

$$\begin{array}{ccccc} (\nabla u_h, \nabla v) & + & (\lambda_h \beta, \nabla v) & = & (f, v) & \forall v \in V_h \\ (\beta \nabla u_h, \omega - \lambda_h) & - & \sum_T \sum_{\Gamma \subset T} h_T([\lambda_h]_\Gamma, \delta[\omega - \lambda_h]_\Gamma)_\Gamma & \leq & (1, \omega - \lambda_h) & \forall \omega \in \Lambda_h \end{array} \tag{5.9}$$

with a constant parameter $\delta > 0$. Here, $\beta^i = \nabla u_h^{i-1}$ is iterated by a fixed point iteration. In every iteration step we have a system, that is unique solvable in u_h. For a better convergence of the fixed point iteration we can use a linear extrapolation of β^i by $2\beta^i - \beta^{i-1}$.

5.2.2 Variable yield condition

Up to now we have assumed that our workpiece has a flow rule with constant yield condition. In physics the yield condition of a material often depends for example on temperature which can be a function of space. Figure 5.4[1] shows a possible stress-strain diagram for an aluminium alloy at different temperatures.

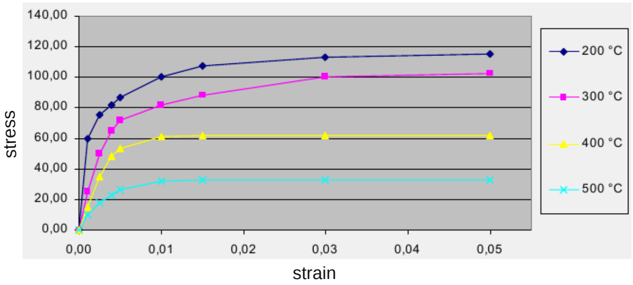


Figure 5.4: Stress-strain diagram using the Von Mises yield condition for an aluminium alloy at different temperatures.

Our intention is to study a torsion problem with a variable yield condition. Inspired by Kunze and Rodriguez [KR00] this can be realised by a non-constant gradient constraint:
Find a scalar function u on Ω being a minimiser of the functional

$$J(\varphi) = \frac{1}{2}(\nabla\varphi, \nabla\varphi) - (f, \varphi)$$

over the set

$$K_g = \{\varphi \in H_0^1(\Omega) |\ |\nabla\varphi| \leq g(x) \text{ a.e. in } \Omega\},$$

with given data

$$g(x) > 0 \text{ a.e. in } \Omega.$$

We gain the variational formulation

$$u \in K_g : \quad a(u, \varphi - u) \geq (f, \varphi - u) \quad \forall\varphi \in K_g. \tag{5.10}$$

Since K is a nonempty, closed and convex set, problem (5.10) has a unique solution.

[1] Taken from: [MJ05]

Proof: In order to use Theorem 2.3, we have to prove K_g to be nonempty, closed and convex. With $v = 0 \in K_g$ since $g(x) > 0$ a.e. in Ω, K_g is nonempty. K_g is a convex set if every convex combination of elements in K_g is an element of K_g, too. Therefore, we have to show $(\varepsilon u + (1-\varepsilon)v)$ being an element of K for all $v, u \in K$, $\varepsilon \in [0,1]$. Taking $u, v \in H_0^1(\Omega)$, $\varepsilon u + (1-\varepsilon)v$ is in $H_0^1(\Omega)$. We have to show $|\nabla(\varepsilon u + (1-\varepsilon)v)| \leq g$:

$$\begin{aligned}|\nabla(\varepsilon u + (1-\varepsilon)v)| &= |\varepsilon \nabla u + (1-\varepsilon)\nabla v| \\ &\leq \varepsilon|\nabla u| + (1-\varepsilon)|\nabla v| \\ &\leq \varepsilon g + (1-\varepsilon) g = g.\end{aligned}$$

At last we have to prove the closure of the subset. Therefore, the limit of every convergent sequence in K_g again has to be in K_g. Let $(v_n)_{n\in\mathbb{N}}$ be a convergent sequence in K_g with limit v:

$$0 = \lim_{n\to\infty} \|v_n - v\|^2_{H_0^1(\Omega)} = \lim_{n\to\infty} (\|v_n - v\|^2_{L_2(\Omega)} + |v_n - v|^2_{H_0^1(\Omega)})$$

Both, $\|v_n - v\|^2_{L_2(\Omega)}$ and $|v_n - v|^2_{H^1(\Omega)}$ have to converge to 0 because both norms are non-negative. That means that v_n converges to v in the H^1-seminorm and therefore

$$\lim_{n\to\infty} \nabla v_n = \nabla v \quad \text{a.e.}$$

There holds $v_n \in K_g \ \forall n$, so we have $v \in K_g$, too.

□

Again we introduce

$$\begin{aligned}V_h &:= \{\varphi \in H_0^1(\Omega) |\, \varphi \text{ bilinear on } T \in \mathbb{T}_h\} \quad \text{and} \\ K_{g,h} &:= \{\varphi \in V_h| \ |\nabla\varphi| \leq g_h\},\end{aligned}$$

where g_h is a cellwise constant approximation of g, defined by

$$g_{h|T} = \operatorname{ess}\inf_{x\in T} g(x).$$

Then the discrete version of (5.10) reads:
Find $u_h \in K_{g,h}$ such that

$$a(u_h, \varphi - u_h) \geq (f, \varphi - u_h) \quad \forall \varphi \in K_{g,h}, \tag{5.11}$$

where existence and uniqueness can again be proven analogously to the continuous case.

The Lagrangian formulation is defined by

$$\mathcal{L}(\varphi, \omega) = \frac{1}{2} a(\varphi, \varphi) - (f, \varphi) + \frac{1}{2} \int \omega((\nabla \varphi)^2 - g^2) \tag{5.12}$$

for $\varphi \in V := H_0^1(\Omega)$ and $\omega \in \Lambda = \{q \in L^\infty(\Omega) | \, q \geq 0 \text{ a.e.}\}$. For our discrete problem we choose V_h, $K_{h,g}$ as above and introduce $\Lambda_h \subset \Lambda$ by

$$\Lambda_h = \{\omega \in \Lambda | \, \omega \text{ constant over each } T \in \mathbb{T}_h\}.$$

Theorem 5.2.2. *The Lagrangian $\mathcal{L}$ has a saddle point $\{u_h, \lambda_h\}$ in $V_h \times \Lambda_h$ where u_h is the solution of* (5.11)*.*

Proof: Following [Ce78](Chapter 5) existence of the saddle point (u_h, λ_h) can be proven if V_h and L_h are finite dimensional and if the assumptions of Theorem 2.13 are fulfilled. So we have to show that there exists an element v_h of V_h where the constraints are strictly satisfied in a neighborhood U_h: $|\nabla w_h|^2 - g_h^2 < 0$ for all $w_h \in U_h$. We define $\tilde{g}_h = \operatorname{ess\,inf}_{x \in \Omega} g(x)$. Then $\tilde{g}_h > 0$ and for $v_h = 0$ there holds $|\nabla v_h|^2 - \tilde{g}_h^2 = -\tilde{g}_h^2 < 0$. The functional $\mathcal{F} : V_h \to \mathbb{R}$ defined by $w_h \mapsto |\nabla w_h|^2 - \tilde{g}_h^2$ is continuous in 0 and hence there exists a neighborhood of v_h where the constraint is strictly fulfilled. □

Remark 5.2.3. *The boundedness of $\|\lambda_h\|_{L^\infty}$ is given by the relation:*

$$((1 + \lambda_h)\nabla u_h, \nabla u_h) = \sum_{T \in \Omega \setminus B_h} \int_T (\nabla u_h)^2 dx + \sum_{T \in B_h} \int_T (1 + \lambda_h) dx = (f, u_h)$$

where B_h is defined by

$$B_h = \{T \in \mathbb{T} | \lambda_h > 0 \text{ on } T\},$$

implying that $|\nabla u_h| = 1$ on B_h.

Derivations of the Lagrangian functional lead to the system:

Find a pair $(u_h, \lambda_h) \in V_h \times \Lambda_h$ fulfilling

$$\begin{aligned} ((1+\lambda_h)\nabla u_h, \nabla\varphi) &= (f, \varphi) \quad \forall \varphi \in V_h \\ (|\nabla\varphi|^2, \omega - \lambda) &\leq (g_h^2, \omega - \lambda) \quad \forall \omega \in \Lambda_h. \end{aligned} \tag{5.13}$$

We use Uzawa's algorithm to solve the discrete torsion problem:

1. Choose an initial iterate λ_h^0, $\alpha > 0$ and a residual res.
2. Solve the linear problem $\int_\Omega (1+\lambda_h)\nabla u_h \nabla v \, dx = (f, v) \quad \forall v \in V_h$.
3. Update: $\lambda_h^{s+1} = \max(0, \lambda_h^s + \alpha(|\nabla u_h^s|^2 - g_h^2))$ on each cell.
4. If $\frac{\|\lambda_h^{s+1} - \lambda_h^s\|}{\|\lambda_h^s\|} > \text{res}$: Set $s = s+1$ and go back to 2,
   else finish.

The Lagrangian formulation allows us to rewrite the a posteriori error estimator in the following way:

5.2.3 A posteriori error analysis

We consider a more general problem for the error estimation:

Minimise

$$J(\varphi) = \frac{1}{2}(\nabla\varphi, \nabla\varphi) - (f, \varphi), \quad \varphi \in H_0^1 \tag{5.14}$$

over the set

$$K_{g,\xi} = \{\varphi \in H_0^1 | \ |\nabla\varphi - \xi(x)| \leq g(x) \text{ a.e. in } \Omega\}, \tag{5.15}$$

with given data $\xi \in L_2(\Omega)$ and $g(x) > 0$ a.e. in Ω.

For the discretisation there holds

$$K_{h,g,\xi} = \{\varphi \in V_h | \ |\nabla\varphi - \xi_h| \leq g_h\} \tag{5.16}$$

where ξ_h and g_h are cellwise constant approximations of ξ and g, defined by

$$g_{h|T} = \operatorname{ess}\inf_{x \in T} g(x), \tag{5.17}$$

$$|\nabla u_h - \xi| \leq |\nabla u_h - \xi_h|. \tag{5.18}$$

We start estimating

$$a(e,e_i) = a(u,e_i-e)-(f,e_i-e)+\underbrace{(f,e_i)-a(u_h,e_i)}_{I}+\underbrace{a(u,e)-(f,e)}_{\leq 0}. \tag{5.19}$$

Since we made the assumptions (5.17) and (5.18) we can test u_h in (5.10), replacing K_g by $K_{g,\xi}$ and hence the last two terms can be estimated by zero. For term I we get

$$(f,e_i)-a(u_h,e_i) = (\lambda_h(\nabla u_h-\xi_h),\nabla(e_i-e))+\underbrace{(\lambda_h(\nabla u_h-\xi_h),\nabla e)}_{II}. \tag{5.20}$$

Furthermore, term II can be estimated by

$$\begin{aligned}
(\lambda_h(\nabla u_h-\xi_h),\nabla e) &= (\lambda_h(\nabla u_h-\xi_h),\nabla u-\xi_h+\xi_h-\nabla u_h)\\
&\leq \sum_T\int_T \lambda_h|\nabla u_h-\xi_h|\,|\nabla u-\xi+\xi-\xi_h|\,dx\\
&\quad +(\lambda_h(\nabla u_h-\xi_h),(\xi_h-\nabla u_h))\\
&\leq \sum_T\int_T \lambda_h|\nabla u_h-\xi_h|(|\nabla u-\xi|+|\xi-\xi_h|)\,dx\\
&\quad +(\lambda_h(\nabla u_h-\xi_h),(\xi_h-\nabla u_h))\\
&\overset{(5.15)}{\leq} \sum_{T\in B_h}\int_T \lambda_h g_h(g+|\xi-\xi_h|)\,dx-\sum_T\int_T\lambda_h(\nabla u_h-\xi_h)^2\,dx\\
&= \sum_{T\in B_h}\int_T(\lambda_h g_h(g-g_h)+\lambda_h g_h|\xi-\xi_h|)dx.
\end{aligned}$$

Collecting these results we obtain for the square of the energy error

$$\begin{aligned}
a(e,e) &= a(e,e-e_i)+a(e,e_i)\\
&\leq (\nabla u,\nabla(e-e_i))-(\nabla u_h,\nabla(e-e_i))+(\nabla u,\nabla(e_i-e))-(f,e_i-e)\\
&\quad -(\lambda_h(\nabla u_h-\xi_h),\nabla(e-e_i))+\sum_{T\in B_h}\int_T\lambda_h g_h(g-g_h)+\lambda_h g_h|\xi-\xi_h|\,dx\\
&\leq (f,e-e_i)-(\nabla u_h,\nabla(e-e_i))-(\lambda_h(\nabla u_h-\xi_h),\nabla(e-e_i))\\
&\quad +\sum_{T\in B_h}\int_T\lambda_h g_h(g-g_h)+\lambda_h g_h|\xi-\xi_h|\,dx.
\end{aligned}$$

Cellwise integration by parts results in

$$(\nabla e, \nabla e) \leq \sum_{T \in \mathbb{T}} \omega_T \varrho_T + N,$$

with

$$N = \sum_{T \in B_h} \int_T \lambda_h g_h (g - g_h) + \lambda_h g_h |\xi - \xi_h| \, dx$$

and the local residuals ϱ_T and weights ω_T defined by

$$\begin{aligned} \varrho_T &:= h_T ||f + \nabla \lambda_h \nabla u_h + (1 + \lambda_h) \Delta u_h - \nabla(\lambda_h \xi_h)||_T \\ &+ \frac{1}{2} h_T^{\frac{1}{2}} ||n \cdot [(1 + \lambda_h) \nabla u_h - \lambda_h \xi_h] \, ||_{\partial T}, \\ \omega_T &:= \max\{h_T^{-1} ||e - e_i||_T, \; h_T^{-\frac{1}{2}} ||e - e_i||_{\partial T}\}, \end{aligned}$$

where for interior interelement boundaries $n \cdot [(1 + \lambda_h) \nabla u_h - \lambda_h \xi_h]$ denotes the jump of $\{(1 + \lambda_h) \nabla u_h - \lambda_h \xi_h\}$ over the element faces.

Next, one uses the interpolation estimates (3.4), (3.5)

$$\omega_T \leq C_{i,T} ||\nabla e||_T,$$

and with the help of Hoelder- and Young's inequality we yield the estimation measuring the discretisation error in the energy norm:

Theorem 5.2.4. *Problem* (5.14) - (5.18) *with nonlinear gradient constraints has the a posteriori error estimator*

$$|e|_1^2 \leq C \sum_{T \in \mathbb{T}} \varrho_T^2 + N, \tag{5.21}$$

with

$$\varrho_T := h_T ||f + \nabla \lambda_h \nabla u_h + (1 + \lambda_h) \Delta u_h - \nabla(\lambda_h \xi_h)||_T + \frac{1}{2} h_T^{\frac{1}{2}} ||n \cdot [(1 + \lambda_h) \nabla u_h - \lambda_h \xi_h] \, ||_{\partial T},$$

where for interior interelement boundaries $n \cdot [(1 + \lambda_h) \nabla u_h - \lambda_h \xi_h]$ *denotes the jump* $\{(1 + \lambda_h) \nabla u_h - \lambda_h \xi_h\}$ *over the element faces and*

$$N = c \sum_{T \in B_h} \int_T \lambda_h g_h (g - g_h) + \lambda_h g_h |\xi - \xi_h| \, dx.$$

Setting $\xi = \xi_h = 0$ we receive the estimator for the torsion problem (5.13).

5.2.4 DWR-Method

A great advantage of formulation (5.8) is the fact that we now have an equation which allows us to apply the Dual-Weight-Residual Method. The resulting local error indicators generate economical meshes which are tailored according to the particular goal of the computation as well as singularities of the domain.

We solve the primal problem like before and perform a postprocess on λ_h for example by using the L_2-projection $P_{L_2} : Q_0 \to Q_1$. On a heuristic level we replace the primal problem by

$$u \in V : \quad ((1+\tilde{\lambda})\nabla u, \nabla\varphi) = (f, \varphi) \quad \forall \varphi \in V.$$

Let $J(\cdot)$ be an arbitrary linear error functional defined on V, $\tilde{\lambda} := P_{L_2}(\lambda_h)$ is the L_2-projection of λ_h and $z \in V$ the solution of the corresponding dual problem

$$((1+\tilde{\lambda})\nabla\varphi, \nabla z) = J(\varphi) \quad \forall \varphi \in V. \tag{5.22}$$

We set $\tilde{z}_h \in \tilde{V}_h$ the finite element approximation of z using higher order finite elements, for example $\tilde{z}_{h|T} \in Q_2$ and achieve

$$((1+\tilde{\lambda})\nabla\varphi, \nabla\tilde{z}_h) = J(\varphi) \quad \forall \varphi \in \tilde{V}_h.$$

Taking $\varphi = e$ in (5.22) there holds

$$\begin{aligned} J(e) &= ((1+\tilde{\lambda})\nabla(u-u_h), \nabla z) \\ &= ((1+\tilde{\lambda})\nabla(u-u_h), \nabla(z-\tilde{z}_h)) + ((1+\tilde{\lambda})\nabla(u-u_h), \nabla\tilde{z}_h) \\ &\leq |1+\tilde{\lambda}|\, \|\nabla(u-u_h)\|\, \|\nabla(z-\tilde{z}_h)\| + ((1+\tilde{\lambda})\nabla(u-u_h), \nabla\tilde{z}_h). \end{aligned}$$

From a priori estimations we know for a convex, polygonal domain Ω there holds

$$\begin{aligned} \|\nabla(u-u_h)\| &\leq ch, \\ \|\nabla(z-\tilde{z}_h)\| &\leq ch^2. \end{aligned}$$

Calculating adaptively on a domain including singularities we almost reach those convergence rates, too. The last part can be further estimated:

$$\begin{aligned}((1+\tilde{\lambda})\nabla(u-u_h),\nabla\tilde{z}_h) &= ((1+\tilde{\lambda})\nabla(u-u_h),\nabla(\tilde{z}_h-I_h\tilde{z}_h))\\ &= (f,\tilde{z}_h-I_h\tilde{z}_h)-((1+\tilde{\lambda})\nabla u_h,\nabla(\tilde{z}_h-I_h\tilde{z}_h))\\ &= \sum_{T\in\mathbb{T}_h}\{(f,\tilde{z}_h-I_h\tilde{z}_h)_T\\ &\quad+(\nabla(1+\tilde{\lambda})\nabla u_h+(1+\tilde{\lambda})\Delta u_h,\tilde{z}_h-I_h\tilde{z}_h)_T\\ &\quad-\frac{1}{2}(n\cdot[(1+\tilde{\lambda})\nabla u_h],\tilde{z}_h-I_h\tilde{z}_h)_{\partial T}\}\end{aligned}$$

where $I_h\tilde{z}_h$ is the interpolation of $\tilde{z}_h$ on $\tilde{V}_h$. So for the discretisation error measured in form of a linear functional $J(\cdot)$ one receives:

$$\begin{aligned}J(u)-J(u_h)\le\sum_{T\in\mathbb{T}_h}&\{(f+\nabla(1+\tilde{\lambda})\nabla u_h+(1+\tilde{\lambda})\Delta u_h,\tilde{z}_h-I_h\tilde{z}_h)_T\\ &-\frac{1}{2}(n\cdot[(1+\tilde{\lambda})\nabla u_h],\tilde{z}_h-I_h\tilde{z}_h)_{\partial T}\}+\text{“terms of higher order”}.\end{aligned}\tag{5.23}$$

5.3 Numerical results

As a test example we consider the torsion problem on $\Omega=[-1,1]^2$ choosing a constant force $f=5$ and a discretisation using Q_1-Elements for the displacement u_h and cellwise constant functions for the Lagrange parameter λ_h.

To test convergence of the true error we consider a two-dimensional problem taken from [GLT76] and we set

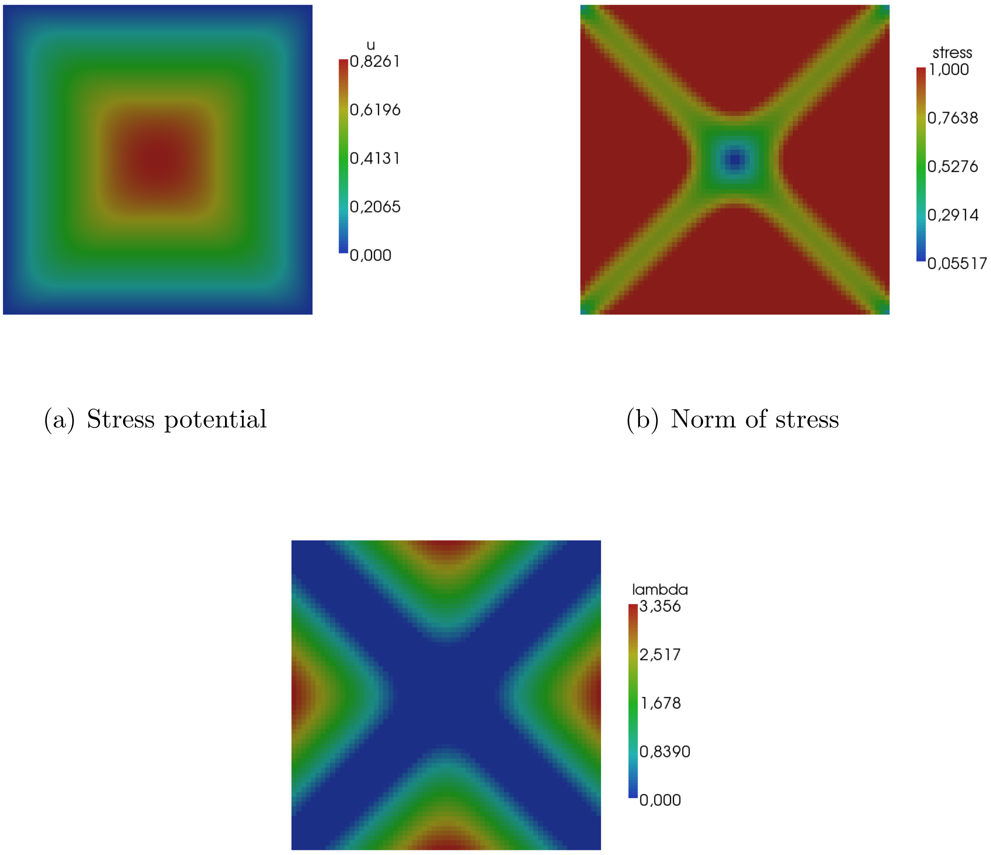
$$\begin{aligned}\Omega&=\{x|x_1^2+x_2^2<R^2\}\\ f&=c\end{aligned}$$

with $c=5$ and $R=1$. Then setting $r=(x_1^2+x_2^2)^{\frac{1}{2}}$, the solution u of (5.2) is given by

$$u(x)=\frac{c}{4}(R^2-r^2),\qquad \text{if } c\le\frac{2}{R};$$

for $c > \frac{2}{R}$:

$$\begin{aligned} u(x) &= R - r, && \text{if } \frac{2}{c} \leq r \leq R, \\ u(x) &= \frac{c}{4}\left[(R^2 - r^2) - (R - \frac{2}{c})^2\right], && \text{if } 0 \leq r \leq \frac{2}{c}. \end{aligned}$$



(a) Stress potential

(b) Norm of stress

(c) Lagrangian multiplier

Figure 5.5: Solution of the torsion problem on a domain $\Omega = [-1, 1]^2$ choosing a constant force $f = 5$ and a discretisation of Q_1-Elements for the stress potential u_h and cellwise constant functions for the Lagrange parameter λ_h.

The error norms show the expected results as we can see in Table 5.1.

# cells	H_1-error		L_2-error	
20	4.186e-01	-	1.048e-01	-
80	2.130e-01	0.98	2.728e-02	1.94
320	1.069e-01	0.99	7.002e-03	1.96
1280	5.356e-02	1.00	1.706e-03	2.04
5120	2.680e-02	1.00	4.363e-04	1.97
20480	1.340e-02	1.00	1.082e-04	2.01

Table 5.1: The H_1- and L_2-error of our test problem show the expected error rates. In the right columns there is always a value α determining the convergence order by $\mathcal{O}(h^\alpha)$ for every refinement step. The H_1-error shows a $\mathcal{O}(h)$-behaviour whereas the error in the L_2-norm has a rate of $\mathcal{O}(h^2)$.

For simulating a variable yield condition as described in Section 5.2.2 we write the yield condition as a function of x in Ω and give two examples to show the resulting stress condition.

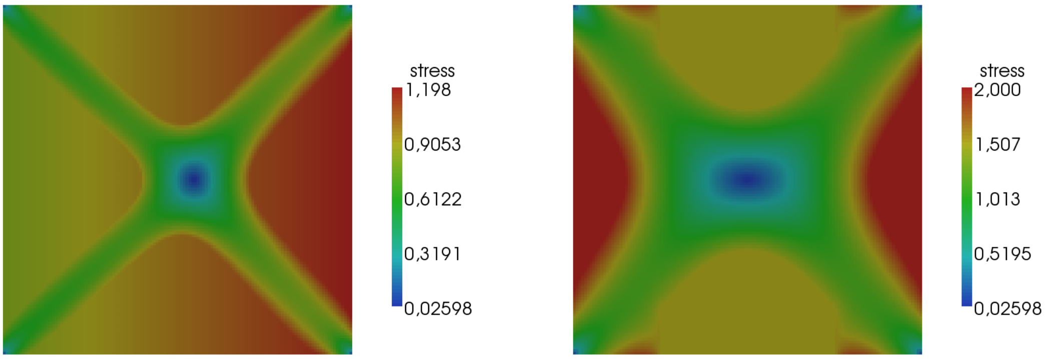


Figure 5.6: Torsion problem with a variable yield condition $g(x) = 0.2x + 1$ (left) and $g(x) = 1.5$ if $|x| < 0.5$ and $f(x) = 2$ elsewhere (right).

5.3.1 Adaptivity

For testing the rate of convergence of estimator (5.21) we explicate some global refinement steps. Table 5.2 offers a linear convergence in this case.

# cells	e_{res}		e_{jump}		N		$\lvert e\rvert_1$	
16	2.236e+00	-	8.326e-01	-	2.543e-01	-	3.017e+00	-
64	1.118e+00	1.00	5.359e-01	0.63	1.354e-01	0.91	1.645e+00	0.87
256	5.590e-01	1.00	2.865e-01	0.90	6.704e-02	1.01	8.435e-01	0.96
1024	2.795e-01	1.00	1.483e-01	0.95	3.344e-02	1.00	4.272e-01	0.98
4096	1.398e-01	1.00	7.515e-02	0.98	1.671e-02	1.00	2.148e-01	0.99
16384	6.988e-02	1.00	3.770e-02	0.99	8.351e-03	1.00	1.076e-01	1.00
65536	3.494e-02	1.00	1.886e-02	1.00	4.175e-03	1.00	5.380e-02	1.00

Table 5.2: Convergence of estimator (5.21). We set $e_{\text{res}} = (\sum_T h_T^2 ||f+\nabla\lambda_h\nabla u_h+(1+\lambda_h)\Delta u_h||_T^2)^{\frac{1}{2}}$, $e_{\text{jump}} = (\sum_T h_T ||n\cdot[(1+\lambda_h)\nabla u_h]\,||_{\partial T}^2)^{\frac{1}{2}}$ and $|e|_1$ denotes the complete estimator (5.21). In the right columns the value α determines the convergence order by $\mathcal{O}(h^\alpha)$. We get linear convergence in every component and thus the whole estimator converges with $\mathcal{O}(h)$.

Comparing (5.6) to estimator (5.21), we find improvements in the latter related to the residual term and the term including jump errors. That leads to different mesh structures. As the solution is smooth in the elastic parts and also not error-prone in regions of plasticity we would expect the transition areas to be the most critical parts. Looking at Figure 5.7, estimator (5.6) shows an over-refinement in the elastic zone whereas estimator (5.21) offers a grid with a dense mesh at the transition zone between elastic and plastic regions which is more reasonable (Figure 5.8).

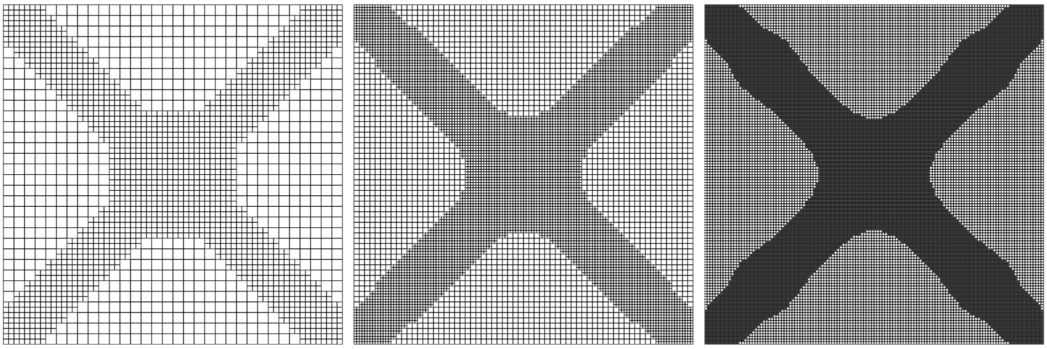


Figure 5.7: A sequence of grids created by estimator (5.6). It shows an over-refinement in the elastic regions.

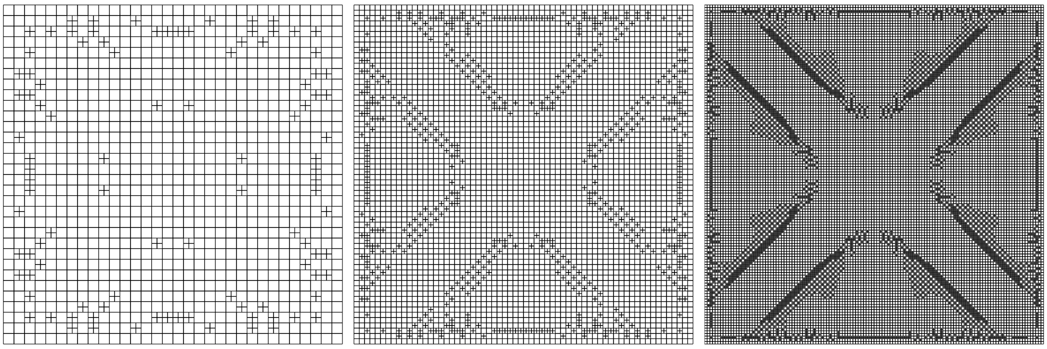


Figure 5.8: A sequence of grids created by estimator (5.21). The zone where elastic and plastic areas meet is always well refined. Instead of refining the elastic part, the estimator concentrates more on the plastic zones.

Estimator (5.6) causes an over-refinement in the elastic zones since the estimation of the jump terms is very small where $|\nabla u_h| = 1$ which causes an unbalance in the estimation. Figure 5.9 shows that the Lagrange multiplier compensates this effect and the term $\|n \cdot [(1+\lambda_h)\nabla u_h]\,\|$ describes an estimation of optimal order in regions of plasticity.

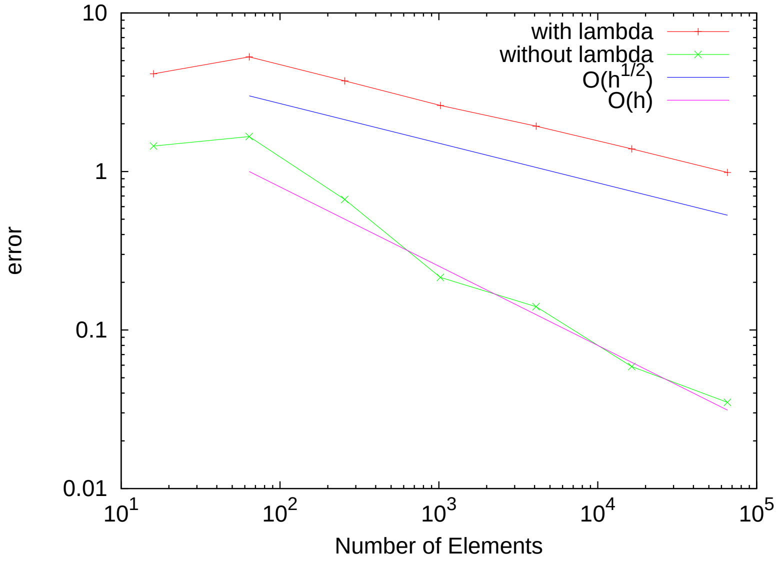
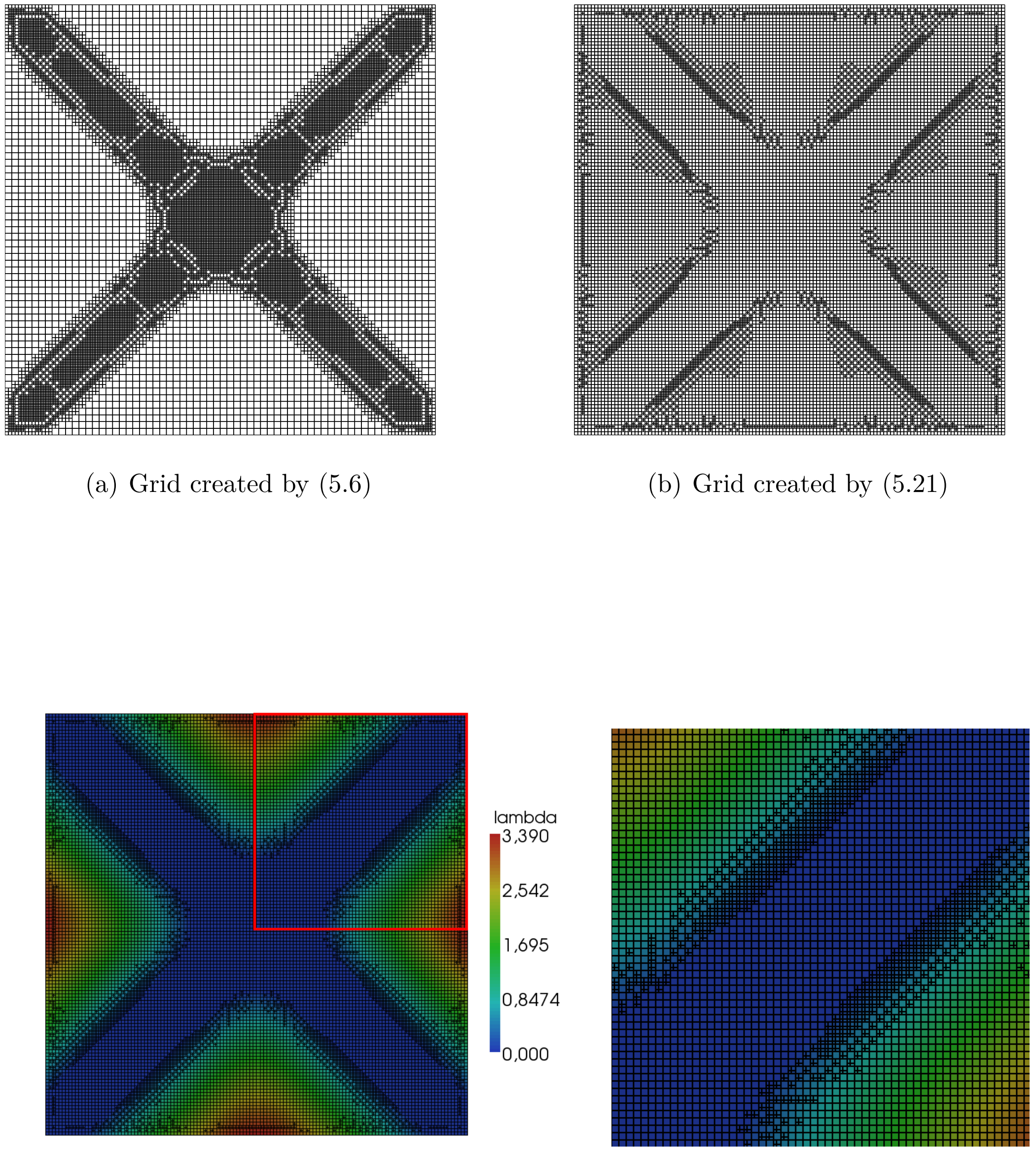


Figure 5.9: Calculation of $(\sum_T \|n \cdot [\nabla u_h]\|^2_{\partial T})^{\frac{1}{2}}$ and $(\sum_T \|n \cdot [(1+\lambda_h)\nabla u_h]\,\|^2_{\partial T})^{\frac{1}{2}}$ in regions of plasticity. Without the Lagrange multiplier the error of the jumps descends with $\mathcal{O}(h)$. Thus zones of plasticity are not well refined. In contrast, estimator (5.21) characterises a well balanced estimator in all regions, showing a $\mathcal{O}(h^{\frac{1}{2}})$-behaviour.



(a) Grid created by (5.6)

(b) Grid created by (5.21)

(c) Solution for the Lagrangian multiplier

(d) Zoom to a refinement zone

Figure 5.10: Mesh (a) is created by estimator (5.6). It shows an over-refinement in the elastic regions. In contrast the second one (b) created by estimator (5.21) is a mesh mostly refined in the zone where elastic and plastic areas meet. The background colours of picture (c) show the Lagrangian multiplier and we can easily see that the refinement appears in the transition zone since the values of the Lagrangian multiplier only exist where $|\nabla u_h| = 1$. Picture (d) is a zoom to the red marked zone in (c).

The effectivity of adaptive refinement can be seen in Figure 5.12. Using uniform refinements, the estimated error only shows a slow descent whereas in the adaptive case a linear convergence order is conserved. We point out that for a cell number of about 36000 the uniform refinement offers an error of about 4.46e-02. The same error is gained with only a tenth of cells using adaptive refinement.

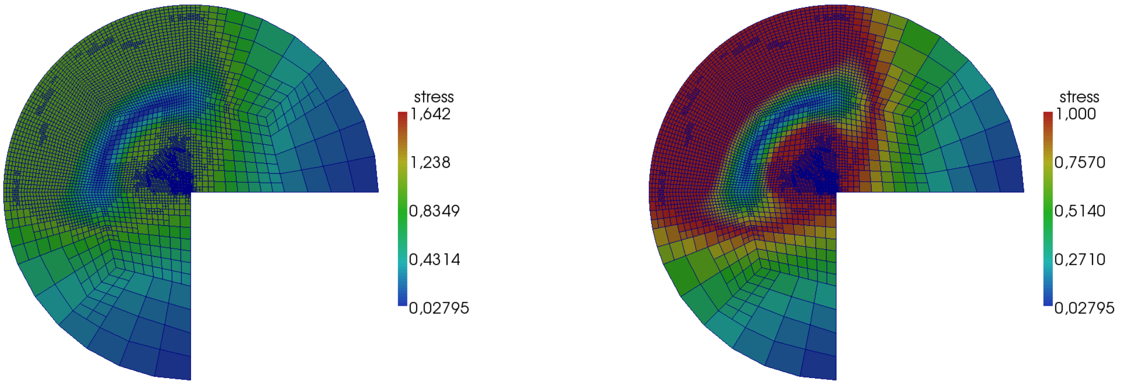


Figure 5.11: $\|\sigma_h\|$ on a three-quarter circle with midpoint $(0,0)$ and radius 1 when using a constant right hand side $f = 8$ if $(x+0.5)^2 + (y-0.5)^2 < 0.5$ and $f = 0$ elsewhere and a constant yield condition $g(x,y) = 1$. The left figure shows the true stresses. The colour scale of the right one is scaled to 1 to show the regions where the yield stress is reached.

In Figure 5.11 we find the stress in the singularity to be extremely high. On that account we expect the most significant error in this corner. The effectivity of adaptive refinement can be observed here since the singularity is refined very well and therefore the error is confined efficiently as we can see in Figure 5.12.

# cells	$\|e\|_{1,\text{global}}$		# cells	$\|e\|_{1,\text{adaptive}}$	
144	1.780e-01	-	144	1.780e-01	-
576	1.115e-01	0.67	726	9.000e-02	0.84
2304	7.587e-02	0.56	2493	5.043e-02	0.94
9216	5.634e-02	0.43	8838	2.631e-02	1.03
36864	4.456e-02	0.34	13323	2.158e-02	0.97

Table 5.3: Convergence table for the refinement of the area illustrated in Figure 5.11 and a constant body force $f = -5$. Using global refinement steps, we observe a slow convergence. On the contrary the adaptive case shows very efficient results. Again, in the right columns there is a value α determining the convergence order by $\mathcal{O}(h^\alpha)$.

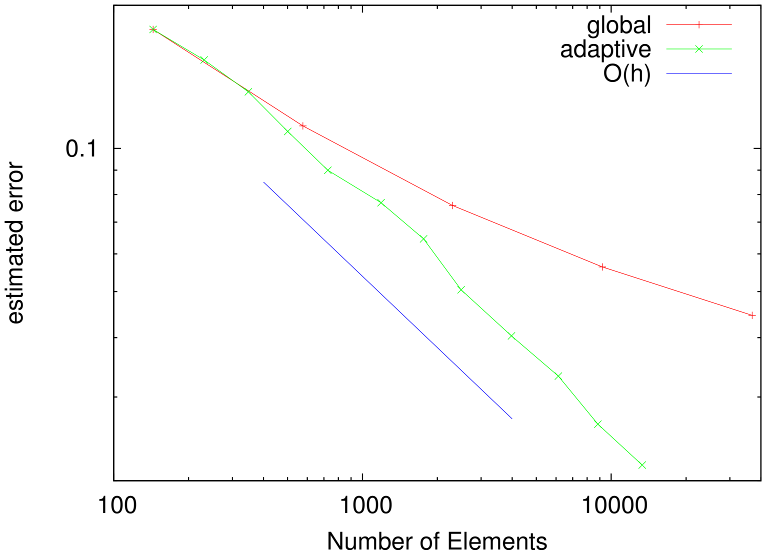


Figure 5.12: Convergence of global and adaptive refinement calculated on the domain of Figure 5.11.

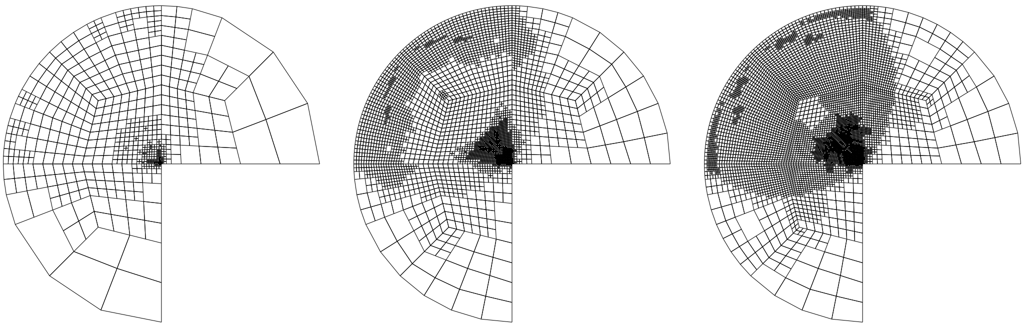
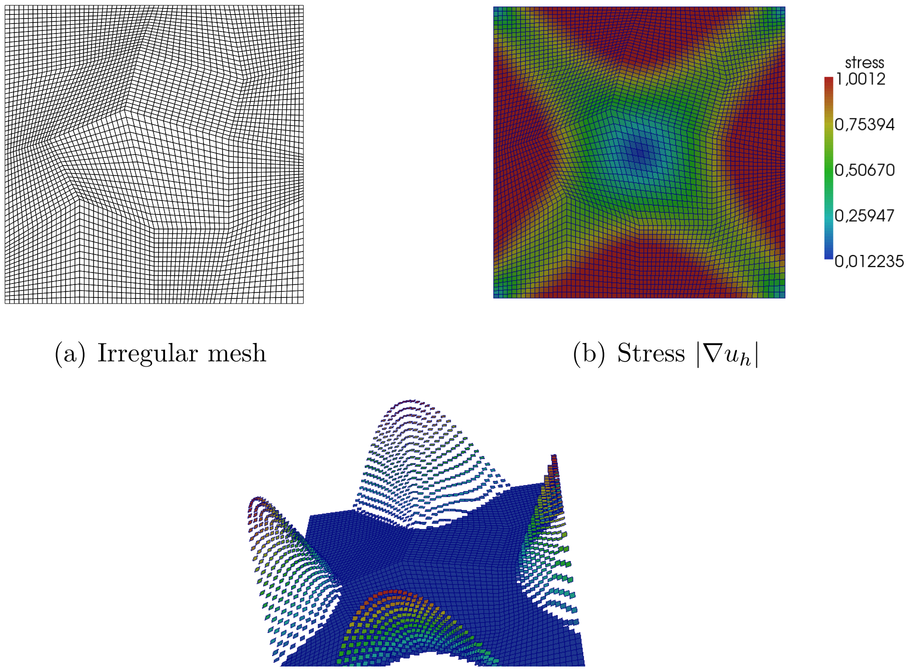


Figure 5.13: Sequence of grids by refining the three-quarter circle of Figure 5.11 adaptively using estimator (5.21).

The discretisation with Q_1/Q_0-Elements is not stable. However, we get smooth results for the dual variable, too, which can be seen calculating the stress on an irregular mesh:



(a) Irregular mesh

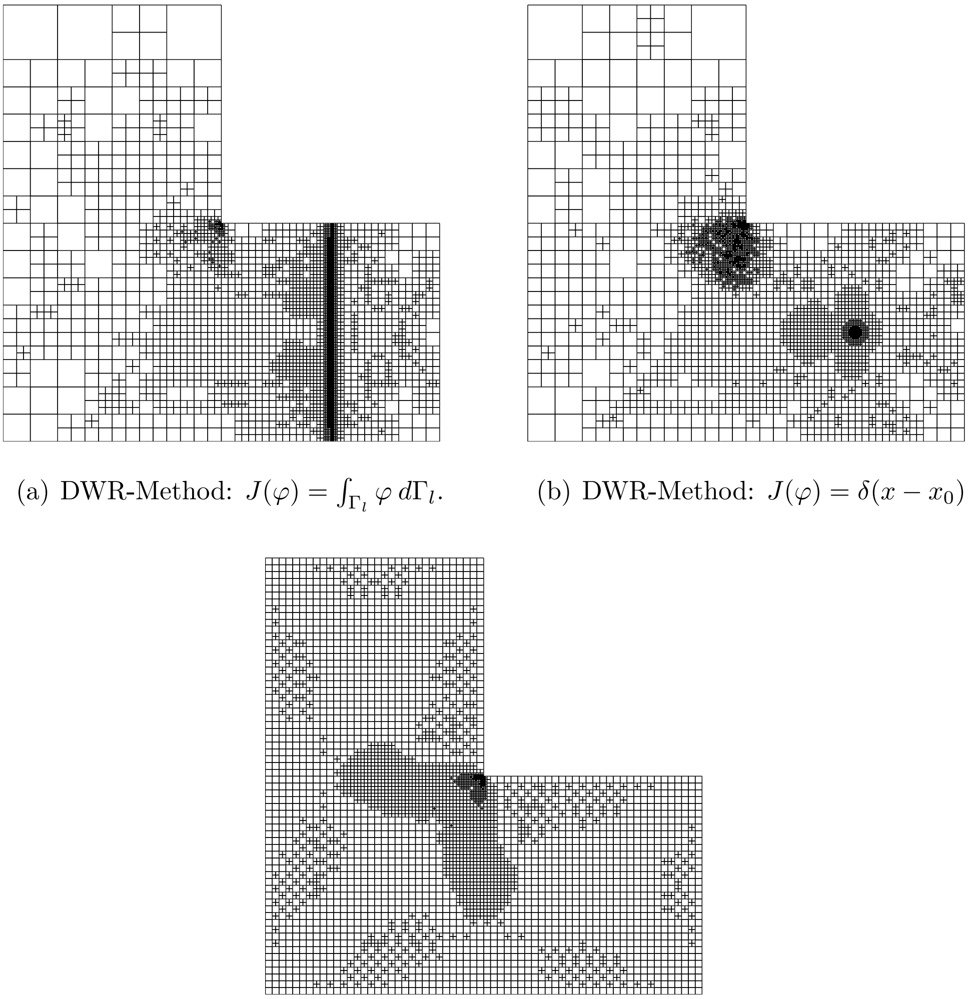
(b) Stress $|\nabla u_h|$

(c) Lagrange multiplier λ_h

Figure 5.14: Irregular mesh (a) and stress of the torsion problem calculated on the irregular mesh (b). The Lagrangian multiplier looks very smooth (c) so that we can expect the system to run very robust.

5.3.2 DWR-Method

For showing the efficiency of the Dual-Weight-Redidual-Method we consider an L-shape domain $\Omega = [-1,1]^2 \backslash (0,1]^2$ and as a functional we integrate the stress potential over a line l in the area with $x = 0.5$ or measure it in a single point.



(a) DWR-Method: $J(\varphi) = \int_{\Gamma_l} \varphi \, d\Gamma_l$.

(b) DWR-Method: $J(\varphi) = \delta(x - x_0)$.

(c) Adaptive refinement by estimator (5.21)

Figure 5.15: In (a) the functional $J(\varphi)$ describes an integral over the line with $x = 0.5$ and in (b) there is a point value $x_0 = (0.5, -0.5)$ the quantity of interest. As a refinement indicator we used the DWR-method and hence estimator (5.23). (c): Adaptive refinement by estimator (5.21).

It is obvious that the estimator of the DWR-Method concentrates on the singularity just as the quantity of interest whereas estimator (5.21) mainly refines the singularity. Getting a better knowledge of functionals that give for example information about the stress situation in one point is of great interest for engineering processes.

6 Strang's problem

In this chapter we treat the aspect of plasticity. The primal and dual formulations of Strang's problem are studied which consist of a linear system with nonlinear restrictions on the primal variable. There have been efforts in error analysis of this problem before and we point to the problems that occurred. We first concentrate on the primal mixed fomulation and start with expressing the plasticity restriction by another Lagrangian multiplier which helps us to develop a consistent a posteriori estimator. When we concentrate on the dual mixed system, we first give a regularised version of the dual formulation to ensure existence of all components of the developed error estimator. In addition we require a stabilisation, which is again put into practice by the least squares method. Numerical tests confirm stability and offer excellent mesh structures created by the new estimators as well as optimal convergence order.

The mathematical model of anti-plane shear or Strang's problem describes a nonlinear relation between a stress vector $\sigma = (\sigma_1, \sigma_2)$ and a scalar displacement u that arise when an external load is acting on a bounded domain $\Omega \subset \mathbb{R}^2$. The nonlinear relation, the so called plastic behaviour of the material, is taken into account by the restriction $|\sigma| \leq 1$.

As described in [DL76], the problem stems from the flow theory of elastic perfectly plastic materials. Like in Chapter 4, in general ε and σ take on values in $\mathbb{S}^d$ where for small deformations, $\varepsilon(u)$ is defined by

$$\varepsilon(u) = \frac{1}{2}(\nabla u + (\nabla u)^T).$$

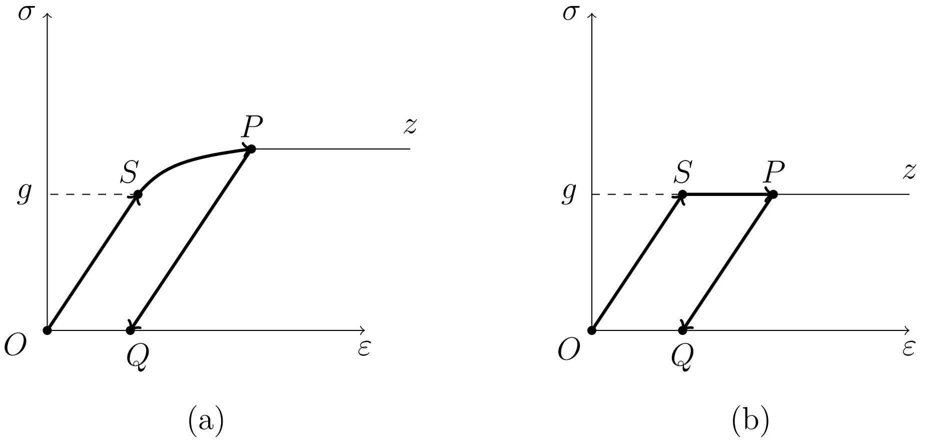


Figure 6.1: (a): Stress-strain relation including work hardening. (b): Stress-strain relation for an elastic perfectly plastic material.

We consider a metal rod and plot the stress-strain relation in a graph, where ε is marked off as abscissa and σ as ordinate (Figure 6.1 (a)). When a force is acting on the material, ε increases, starting from zero. The effect is that σ increases too and (ε, σ) describes a straight line segment from the origin O. When σ reaches a suitable value g, (ε, σ) starts to describe a curve, starting form S. The more the metal is loaded, the curve becomes closer to a parallel to the abscissa. Up to the point S, where (ε, σ) is linear, we have an elastic behaviour, whereas the arc Sz describes the plastic region which provokes that within relaxation the material does not go back to its origin state. Furthermore, PQ turns out to be linear and reversible when decreasing ε and we observe $OS < PQ$ as long as the arc Sz is not a half-line parallel to $O\varepsilon$. This effect is called work hardening. It increases the yield stress g by permitting a zone of linear reversible behaviour with a greater amplitude than that obtained starting by the natural state. For our studies we consider a perfectly plastic material (Figure 6.1 (b)). In this case the stress σ never passes the threshold g, which is independent of the amount of strain.

Following [Jo76/1], the general problem reads

$$\operatorname{div}\sigma = -f, \quad \varepsilon(\dot{u}) = A : \dot{\sigma} + \lambda \text{ in } \Omega,$$
$$\lambda : (\tau - \sigma) \leq 0 \ \forall \tau \text{ with } \mathcal{F}(\tau) \leq 0, \quad \lambda : \dot{\sigma} = 0 \text{ in } \Omega,$$
$$\dot{u} = 0 \text{ on } \Gamma_u, \quad \sigma \cdot n = g \text{ on } \Gamma_\sigma,$$

where $\dot{u}$ is the time derivative of u and A is a fourth order tensor. Here, $\mathcal{F}$ denotes the flow rule which can be defined by the Von Mises yield function $\mathcal{F}(\tau) = |\sigma| - g$, $g > 0$ and λ describes the plastic growth[1]. Setting in two dimensions

$$\begin{aligned} K_1 &= \{\tau \in H^{div}(\Omega) |\, \mathcal{F}(\tau) \leq 0,\ \tau \cdot n = g \text{ on } \Gamma_\sigma\}, \\ K_2 &= \{\tau \in (L_2(\Omega))^2 |\, \mathcal{F}(\tau) \leq 0\}, \\ V &= H_0^1(\Omega), \end{aligned}$$

with

$$H^{div}(\Omega) = \{\tau \in (L_2(\Omega))^2 | \ \operatorname{div}\tau \in L_2(\Omega)\}$$

and $v = \dot{u}$, we can formulate the dual mixed system by:
Find a pair $(v, \sigma) : I \to L_2(\Omega) \times K_1$ satisfying

$$(A\dot{\sigma}, \tau - \sigma) + (v, \operatorname{div}\tau - \operatorname{div}\sigma) \geq 0 \quad \forall \tau \in K_1, \tag{6.1}$$
$$-(\operatorname{div}\sigma, \varphi) = (f, \varphi) \quad \forall \varphi \in L_2(\Omega), \tag{6.2}$$
$$\sigma(0) = 0,$$

and the corresponding primal formulation:
Find a pair $\{v, \sigma\} : I \to V \times K_2$, such that

$$(A\dot{\sigma}, \tau - \sigma) - (\nabla v, \tau - \sigma) \geq 0 \quad \forall \tau \in K_2,$$
$$(\sigma, \nabla\varphi) = (f, \varphi) + (g, \varphi)_{\Gamma_\sigma} \quad \forall \varphi \in V,$$
$$\sigma(0) = 0.$$

[1] Here, λ is a second order tensor describing the Lagrangian multiplier for the yield condition.

Here, I is the time interval $I := [0, T]$. Furthermore, λ is the part of the strain rate $\varepsilon(v)$ due to the plastic flow. In the purely elastic case, $\mathcal{F}(\sigma) < 0$, there holds $\lambda = 0$, otherwise, if $\mathcal{F}(\sigma) = 0$, we have

$$\begin{aligned} \lambda = 0, \quad & \text{if } \sigma = g,\ \dot{\sigma} < 0, \\ \lambda \geq 0, \quad & \text{if } \sigma = g,\ \dot{\sigma} = 0. \end{aligned}$$

Collapse occurs when there is no admissible stress. That means, the external loads cannot be kept in equilibrium without exceeding the yield condition. To analyse Strang's problem [St79], we neglect the rate dependence and consider an infinitely long vertical pipe with square cross-section Ω, filled with a plastic material. If some force f is acting vertically, it is balanced by the shear stresses $\sigma_1 = \sigma_{xz}$ and $\sigma_2 = \sigma_{yz}$ and the equilibrium law is given by

$$\operatorname{div} \sigma = -f.$$

The Von Mises yield criterion is satisfied if $|\sigma| \leq 1$. The fixed square results in zero boundary conditions on the displacement.

6.1 Variational formulation

Due to the considerations above the classical form of Strang's problem is given by

$$\begin{gathered} -\operatorname{div} \sigma = f, \quad \sigma = \Pi \nabla u \quad \text{in } \Omega, \\ u = 0 \quad \text{on } \partial\Omega, \end{gathered} \tag{6.3}$$

where Π denotes the pointwise projection onto the circle with radius 1. We introduce

$$\begin{aligned} H &:= (L_2(\Omega))^2, \\ \Pi H &:= \{\tau \in H, |\tau| - 1 \leq 0\}, \end{aligned}$$

where $|\tau|^2 = \tau_1^2 + \tau_2^2$. Now, similar to the approach above (or see Johnson [Jo78]), we formulate the primal mixed system of (6.3) by:

Find $\{\sigma, u\} \in \Pi H \times V$ such that

$$(\sigma, \tau - \sigma) - (\nabla u, \tau - \sigma) + (\sigma, \nabla\varphi) \geq (f, \varphi) \quad \forall\{\tau, \varphi\} \in \Pi H \times V.$$

For our discretisation we choose $u_h \in V_h$ where $V_h \subset H_0^1(\Omega)$ contains standard bilinear shape functions. The stresses are constructed of elementwise constant functions of $\Pi H_h \subset \Pi H$:

$$(\sigma_h, \tau - \sigma_h) - (\nabla u_h, \tau - \sigma_h) + (\sigma_h, \nabla\varphi) \geq (f, \varphi) \quad \forall\{\tau, \varphi\} \in \Pi H_h \times V_h. \tag{6.4}$$

Existence and uniqueness for the stress σ in the continuous case have been proven, e.g. by Johnson [Jo76/1]. For the discrete system the choice of piecewise bilinear functions for φ and τ consisting of piecewise constant functions results in a stable system on triangular elements. On quadrilateral elements it is unknown whether the LBB condition is fulfilled, but numerical tests show the system runs very robust (see [Dr11]).

6.1.1 A posteriori error analysis

A posteriori error estimates controlling the stress in the L_2-norm have been developed by Johnson and Hansbo ([JH91]):

$$\|\sigma - \sigma_h\|^2 \leq \sum_{j=1}^{2} \|hC_j^i R_j(\sigma_h)\|^2_{L_2(\Omega_h^e)} + \sum_{j=1}^{2} \|h|\varepsilon(u_h)|C_j^i R_j(\sigma_h)\|_{L^1(\Omega_h^p)} \tag{6.5}$$

with certain interpolation constants C_j^i and Ω_h^e, Ω_h^p denoting the parts where the discrete solution behaves elastic and plastic respectively. Furthermore, there holds

$$R_1(\sigma_h) = |\text{div}\sigma_h + f| \quad \text{on } T \in \mathbb{T}_h,$$
$$R_2(\sigma_h) = \max_{E \subset \partial T} \sup_E \frac{1}{2} \frac{|[\sigma_h n_E]|}{h_T} \quad \text{on } T \in \mathbb{T}_h,$$

where an element edge is denoted by E, its normal vector by n_E and $[\cdot]$ is the jump across E. This estimate appears to be suboptimal because on Ω_h^p we get at most a convergence order $\mathcal{O}(h^{\frac{1}{2}})$. In addition, inconsistency appears in the plastic regions due to the term $R_1(\sigma_h)$.

For an optimal error estimator there have to be improvements that allow a better estimation in the plastic regions. We rewrite the problem to a saddle point formulation from which we derive an error estimator localised to the critical parts.

6.2 Saddle point formulation for the primal mixed system

In preparation of the announced error analysis, we follow [Su10] and introduce the Lagrangian functional

$$\mathcal{L}(\tau, \varphi, \omega) = \frac{1}{2}\left\{(\tau, \tau) + \int_\Omega \omega(\tau^2 - 1)dx\right\} + (f, \varphi) - (\tau, \nabla\varphi) \tag{6.6}$$

for triples $(\tau, \varphi, \omega) \in H_h \times V_h \times \Lambda_h$, where we assume $\Lambda_h \subset \Lambda$ with

$$\begin{aligned}\Lambda_h &= \{\omega \in \Lambda \mid \omega \text{ constant on each } T \in \mathbb{T}_h\},\\ \Lambda &= \{q \in L^\infty(\Omega) \mid q \geq 0 \text{ a.e. in } \Omega\}.\end{aligned}$$

Appropriate derivations allow an equivalent characterisation of the solution of (6.6):

$$((1 + \lambda_h)\sigma_h, \tau) = (\nabla v_h, \tau) \quad \forall \tau \in H_h \tag{6.7}$$

$$(\sigma_h, \nabla\varphi) = (f, \varphi) \quad \forall \varphi \in V_h \tag{6.8}$$

$$(\sigma_h^2 - 1, \omega - \lambda_h) \geq 0 \quad \forall \omega \in \Lambda_h. \tag{6.9}$$

Remark 6.2.1. v_h *in* (6.7) *does not necessarily coincide with* u_h *in* (6.4).

When introducing

$$B_h = \{T \in \mathbb{T}_h \mid \lambda_h > 0 \text{ on } T\},$$

implying that $|\sigma_h| = 1$ on B_h, we can conclude from

$$(f, v_h) = (\sigma_h, \nabla v_h) = ((1 + \lambda_h)\sigma_h, \sigma_h) = \sum_{T \in \Omega \setminus B_h} \int_T (\sigma_h)^2 dx + \sum_{T \in B_h} \int_T (1 + \lambda_h) dx$$

that $\|\lambda_h\|_{L^\infty}$ is bounded.

Like described in [Su10], following the line of arguments presented in Glowinski [Gl83], Proposition 3.5, for a similar elasto-plastic torsion problem, the existence of a saddle point $(\sigma_h, v_h, \lambda_h) \in H_h \times V_h \times \Lambda_h$ of (6.6) is guaranteed. In addition, the stress component σ_h is the solution of the original discrete problem (6.4).
Again, we use standard Uzawa-type schemes to solve the discrete problem (6.7)-(6.9) by the following iterative algorithm:

1. Choose an initial iterate λ_h^0, $\rho > 0$ and a residual res.

2. Solve $\int_\Omega (1 + \lambda_h^\nu)\sigma_h \tau \, dx - (\tau, \nabla v_h) + (\sigma_h, \nabla \varphi) = (f, \varphi) \quad \forall (\tau, \varphi) \in H_h \times V_h$.

3. Update: $\lambda_h^{\nu+1} = \max(0, \lambda_h^\nu + \rho(|\sigma_h^\nu|^2 - 1))$ on each cell.

4. If $\frac{\|\lambda_h^{\nu+1} - \lambda_h^\nu\|}{\|\lambda_h^{\nu+1} - \lambda_h^0\|} >$ res: Set $\nu = \nu + 1$ and go back to 2,
   else finish.

The iteration of λ_h costs a lot of time. We now have a system of three unknowns instead of two. Getting knowledge of this values can also be done by calculating the residual of the primal system. This reduces the system that has to be iterated to two unknowns. From the Lagrangian formulation we know

$$(1 + \lambda_h)\sigma_h = \nabla u_h.$$

Solving (6.4) and calculating the residual by evaluating

$$\lambda_h(i) = \sqrt{\frac{|\nabla \tilde{u}_h(i)|^2}{|\tilde{\sigma}_h(i)|^2}} - 1, \quad |\tilde{\sigma}_h(i)| > 0 \tag{6.10}$$

for every degree of freedom i of λ_h, where $\nabla \tilde{u}_h$ and $\tilde{\sigma}_h$ are the averaged values of ∇u_h and σ_h on a cell T, we nearly get the exact values for λ_h. Although this is not the mathematical correct way, it is a fast alternative to iterate (6.7)-(6.9).

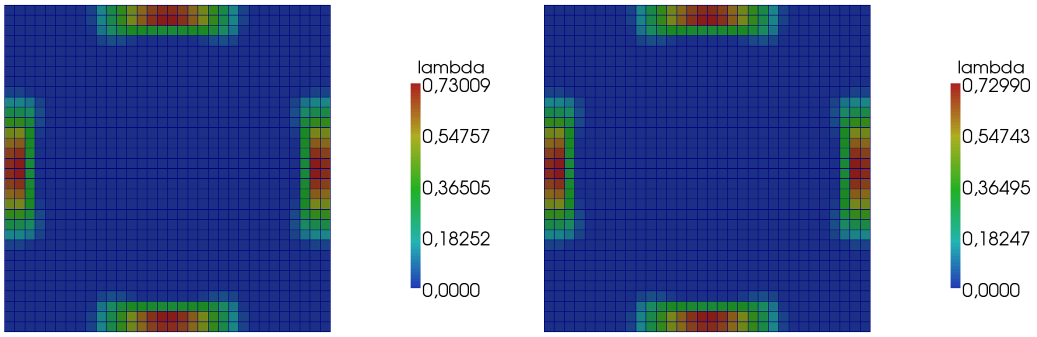


Figure 6.2: Left: λ_h iterated by (6.7)-(6.9). Right: λ_h calculated by (6.10) after iterating (6.4).

cells	residual	iteration
64	0.08	0.17
256	0.69	2.14
1024	7.26	41.65
4096	104.57	1111.79

Table 6.1: Time in s for calculating λ_h by using the residual (6.10) and by iterating (6.7)-(6.9).

6.2.1 A posteriori error analysis

To achieve an a posteriori error estimator, which especially outlines the critical zones of the problem, we use equations (6.7) and (6.8). From (6.7) we have

$$\sigma_h = \frac{1}{1+\lambda_h}\nabla v_h.$$

For case of simplicity we set

$$\kappa_h := \frac{1}{1+\lambda_h}.$$

Then we only have to solve a linear equation

$$(\kappa_h \nabla u_h, \nabla \varphi) = (f, \varphi) \quad \forall \varphi \in V_h$$

with an update $|\kappa_h \nabla v_h| \leq 1$ in every Uzawa iteration step. After the iteration has finished we get our Lagrangian parameter in every degree of freedom i from

$$\lambda_h(i) = \begin{cases} \frac{1}{\kappa_h(i)} - 1, & \text{if } \kappa_h(i) > 0, \\ 0, & \text{if } \kappa_h(i) = 0. \end{cases}$$

In what follows we set

$$a(u, \varphi) = (\kappa_h \nabla u, \nabla \varphi) \quad \forall \varphi \in V$$

and

$$s(u, \varphi) = ((\kappa - \kappa_h) \nabla u, \nabla \varphi) \quad \forall \varphi \in V.$$

Then our linear equation is written as

$$a(u, \varphi) + s(u, \varphi) = (f, \varphi) \quad \forall \varphi \in V. \tag{6.11}$$

$s(\cdot, \cdot)$ can be separated into two parts:

$$s(\cdot, \cdot) = s_h(\cdot, \cdot) + s_N(\cdot, \cdot),$$

where

$$s_h(u, \varphi) = \sum_{T \in A \subset \mathbb{T}_h} s(u, \varphi)_{|T}, \quad s_N(u, \varphi) = \sum_{T \in \mathbb{T}_h \setminus A} s(u, \varphi)_{|T}.$$

Here, A is the set of cells where $s(\cdot, \cdot)$ is active in our calculations, that means we calculate with the continuous solution κ. On the rest of the cells $s(\cdot, \cdot)$ is not active and we calculate with the discrete value κ_h. Furthermore, there holds the equation

$$a(u_h, \varphi) + s_h(u_h, \varphi) = (f, \varphi) \quad \forall \varphi \in V_h. \tag{6.12}$$

Remark 6.2.2. *Since we do not know the continuous solution of κ, we will set $s_h(\cdot, \cdot) = 0$ at the end of our estimation. $s_h(\cdot, \cdot)$ is to be understood as an auxiliary variable.*

It follows by subtracting (6.12) from (6.11) for all $\varphi \in V_h$:

$$\begin{aligned} 0 &= a(u - u_h, \varphi) + s(u, \varphi) - s_h(u_h, \varphi) \\ &= a(u - u_h, \varphi) + s(u, \varphi) - \{s_h(u_h, \varphi) + s_N(u_h, \varphi)\} + s_N(u_h, \varphi) \\ &= a(u - u_h, \varphi) + s(u - u_h, \varphi) + s_N(u_h, \varphi). \end{aligned} \tag{6.13}$$

We make the assumption that there exists a solution of the continuous problem with $\lambda \in \Lambda$. Then there exists an upper bound $\|\lambda\|_\infty < M$ and hence $\kappa > \gamma = \frac{1}{1+M}$. Therefore we get:

$$(\kappa \nabla e, \nabla e)_\Omega = \int_\Omega \kappa \nabla e \cdot \nabla e \, dx > \gamma \int_\Omega \nabla e \cdot \nabla e \, dx.$$

Using this result we can start our estimation:

$$\begin{aligned} \gamma \|\nabla u - \nabla u_h\|^2 &\leq a(e,e) + s(e,e) \\ &\overset{(6.13)}{=} a(e,e) + s(e,e) - \{a(e,e_i) + s(e,e_i) + s_N(u_h,e_i)\} \\ &= a(u-u_h, e-e_i) + s(u-u_h, e-e_i) - s_N(u_h,e_i) \\ &= (f, e-e_i) - a(u_h, e-e_i) - s_h(u_h, e-e_i) \\ &\quad -s_N(u_h, e-e_i) - s_N(u_h, e_i) \\ &= (f, e-e_i) - a(u_h, e-e_i) - s_h(u_h, e-e_i) \qquad (6.14) \\ &\quad -s_N(u_h, e). \end{aligned}$$

The last term can be estimated by Young's inequality:

$$\begin{aligned} s_N(u_h,e)_T &= ((\kappa - \kappa_h)\nabla u_h, \nabla e)_T \\ &\leq \|(\kappa-\kappa_h)\nabla u_h\|_T \, \|\nabla e\|_T \\ &\leq \frac{1}{2\gamma}\|(\kappa-\kappa_h)\nabla u_h\|_T^2 + \frac{\gamma}{2}\|\nabla e\|_T^2 \end{aligned}$$

and so we have

$$s_N(u_h,e) \leq \frac{1}{2\gamma} \sum_T \|(\kappa-\kappa_h)\nabla u_h\|_T^2 + \frac{\gamma}{2}\|\nabla e\|^2.$$

Taking result (6.14) and setting $s_h(u_h, e-e_i) = 0$ (because we never use the exact solution of κ_h) we achieve

$$\begin{aligned}\frac{\gamma}{2}\|\nabla u - \nabla u_h\|^2 \leq\ & (f, e - e_i) - a(u_h, e - e_i) - s_h(u_h, e - e_i)\\ &+ \frac{1}{2\gamma}\sum_T \|(\kappa - \kappa_h)\nabla u_h\|_T^2\\ =\ & \sum_T \{(f + \kappa_h \Delta u_h + \nabla\kappa_h \nabla u_h, e - e_i)_T - \frac{1}{2}(n \cdot [\kappa_h \nabla u_h], e - e_i)_{\partial T}\}\\ &+ \frac{1}{2\gamma}\sum_T \|(\kappa - \kappa_h)\nabla u_h\|_T^2.\end{aligned}$$

Finally the error estimator results in

$$\begin{aligned}\frac{\gamma}{4}\|\nabla e\|^2 \leq & \frac{2}{\gamma}C^2 \sum_T \{h^2 \|f + \kappa_h \Delta u_h + \nabla\kappa_h \nabla u_h\|_T^2 + \frac{1}{4}h\|n \cdot [\kappa_h \nabla u_h]\|_{\partial T}^2\}\\ &+ \frac{1}{2\gamma}\sum_T \|(\kappa - \kappa_h)\nabla u_h\|_T^2.\end{aligned}$$

Similar to the Z^2-techniques proposed by Zienkiewicz and Zhu [ZZ87], we introduce $\mathcal{M}(\kappa_h)$ which is a (superconvergent) approximation of κ and achieve the following theorem:

Theorem 6.2.3. *For Strang's problem* (6.7)-(6.9) *there holds the a posteriori error bound*

$$\begin{aligned}\|\nabla e\|^2 \leq & \frac{8}{\gamma^2}C^2 \sum_T \{h^2 \|f + \kappa_h \Delta u_h + \nabla\kappa_h \nabla u_h\|_T^2 + \frac{1}{4}h\|n[\kappa_h \nabla u_h]\|_{\partial T}^2\}\\ &+ \frac{2}{\gamma^2}\sum_T \|(\mathcal{M}(\kappa_h) - \kappa_h)\nabla u_h\|_T^2\end{aligned} \tag{6.15}$$

with

$$\kappa_h = \frac{1}{1 + \lambda_h}$$

and where for interior interelement boundaries $n \cdot [\kappa_h \nabla u_h]$ denotes the jump of the normal derivative $\kappa_h \nabla u_h$.

6.3 The dual mixed system

Using the primal mixed formulation of Strang's problem we considered $u \in V = H_0^1(\Omega)$. In practice there exist examples where a solution of u can have discontinuities or is not uniquely defined. The regularity result $u \in H^1$ is only achieved under additional conditions, described in [Jo78], which assume an arbitrary small linear hardening. Therefore, it may be more convenient to use the dual formulation of the problem which requires weaker conditions on u. Before we introduce the dual mixed formulation, we follow Johnson and Hansbo [JH91] in order to give a regularised version of the problem which has better regularity results and thus will help us later on when processing error estimations.

6.3.1 Regularisation of Strang's problem

The original problem is defined by:
Find $\sigma \in P_f^{div}$ such that

$$J(\sigma) \leq J(\tau) \quad \forall \tau \in P_f^{div} \tag{6.16}$$

with the following definitions:

$$\begin{aligned}
J(\tau) &= \frac{1}{2}\|\tau\|^2, \quad \|\tau\|^2 = \int_\Omega |\tau|^2 dx,\\
P_f &= P \cap H_f,\\
P_f^{div} &= P \cap H_f^{div},\\
P &= \{\tau \in H : |\tau(x)| \leq 1 \text{ a.e. in } \Omega\},\\
H_f^{div} &= H_f \cap H^{div},\\
H_f &= \{\tau \in H : \ -\text{div } \tau = f \text{ in } \Omega\},\\
H^{div} &= \{\tau \in H : \ \text{div}\, \tau \in L_2(\Omega)\},\\
H &= [L_2(\Omega)]^2,\\
\tilde{V} &= L_2(\Omega),\\
V &= H_0^1(\Omega).
\end{aligned}$$

We introduce a regularisation of problem (6.16) with a regularisation parameter $\mu > 0$:
Find $\sigma_\mu \in H_f^{div}$ such that

$$J_\mu(\sigma_\mu) \leq J_\mu(\tau) \quad \forall \tau \in H_f^{div}, \tag{6.17}$$

where

$$J_\mu(\tau) = \frac{1}{2}\|\tau\|^2 + \frac{1}{2\mu}\|\tau - \pi\tau\|^2.$$

The convex functional $\Phi_\mu(\tau) = \frac{1}{2\mu}\|\tau - \pi\tau\|^2$ has the monotone Gateaux derivative $\Phi'_\mu(\tau) = \frac{1}{\mu}(\tau - \pi\tau)$ (see [JH91]) and $\pi\tau(x)$ denotes a projection of $\tau(x)$ onto the unit disc $\{r \in \mathbb{R}^2 : \, |r| \leq 1\}$ defined by

$$\pi\tau(x) = \begin{cases} \tau(x), & \text{if } |\tau(x)| \leq 1, \\ \frac{\tau(x)}{|\tau(x)|}, & \text{if } |\tau(x)| > 1. \end{cases}$$

This problem has a unique solution since σ_μ is the minimum of a convex functional.
Safe load hypothesis: There exists $\chi \in H_f^{div}$ and $\delta > 0$ with $|\chi(x)| \leq 1 - \delta$, a.e. $x \in \Omega$.

The mixed formulation of the regularised problem is then given by:
Find a pair $(\sigma_\mu, u_\mu) \in H \times V$ such that:

$$\begin{aligned} \sigma_\mu + \tfrac{1}{\mu}(\sigma_\mu - \pi\sigma_\mu) &= \nabla u_\mu \qquad && \text{in } \Omega, \\ -\operatorname{div}\sigma_\mu &= f && \text{in } \Omega, \end{aligned}$$

with the unique solution $(\sigma_\mu, u_\mu) \in H \times V$. Alternatively the problem is mentioned as:
Find $(\sigma_\mu, u_\mu) \in H^{div} \times \tilde{V}$ such that

$$(\sigma_\mu, \tau) + (\mu^{-1}(\sigma_\mu - \pi\sigma_\mu), \tau) + (u_\mu, \operatorname{div}\tau) - (\operatorname{div}\sigma_\mu, \varphi) = (f, \varphi)$$

$\forall(\tau, \varphi) \in H^{div} \times \tilde{V}$, respectively. For all $\mu > 0$ the mixed problem has a unique solution $(\sigma_\mu, u_\mu) \in H^{div} \times \tilde{V}$ with σ_μ satisfying (6.17). If the safe load hypothesis is ensured and $P_f^{div} \neq \emptyset$, it is shown in [Jo76/1] that there exists a solution $\sigma \in P_f^{div}$ of problem (6.16) and σ_μ tends weakly to σ in H as μ tends to zero.

With respect to the safe load hypothesis there holds

$$\|\frac{1}{\mu}(\sigma_\mu - \pi\sigma_\mu)\|_{L_1(\Omega)} + \|\nabla u_\mu\|_{L_1(\Omega)} \leq C \tag{6.18}$$

with a constant C independent of μ, and hence there exists an $u \in BV(\Omega) \equiv \{v \in L_2(\Omega) : \nabla v \in [M(\overline{\Omega})]^2\}$, where $M(\overline{\Omega})$ is the set of bounded measures on $\overline{\Omega}$, such that u_μ converges to u weak star in $BV(\Omega)$ and such that u is a displacement corresponding to the stress solution σ of (6.16) (see [JH91]).

6.3.2 Saddle point formulation for the dual mixed system

In what follows we make the assumption that there are no discontinuities in the displacement which may appear if there is an accumulated slip in plastic regions. We introduce

$$K := \{\tau \in H^{div}(\Omega)\Big| \; |\tau| - 1 \leq 0\},$$

and the variational inequality for the dual mixed problem:
Find a pair $(\sigma, u) \in K \times \tilde{V}$ satisfying

$$(\sigma, \tau - \sigma) + (u, \operatorname{div} \tau - \operatorname{div} \sigma) - (\operatorname{div} \sigma, \varphi) \geq (f, \varphi) \quad \forall (\tau, \varphi) \in K \times \tilde{V}$$

with a body force $f \in L^\infty(\Omega)$. Written in a block system, the problem looks like follows:

$$\begin{array}{lcllll} (\sigma, \tau - \sigma) & + & (u, \operatorname{div} \tau - \operatorname{div} \sigma) & \geq & 0 & \forall \tau \in K \\ -(\operatorname{div} \sigma, \varphi) & & & = & (f, \varphi) & \forall \varphi \in \tilde{V}. \end{array}$$

When we introduce the corresponding finite element spaces

$$\begin{aligned} K_h &:= \{\tau \in K(\Omega) | \; \tau \text{ bilinear on } T \in \mathbb{T}_h\}, \\ \tilde{V}_h &:= \{\varphi \in \tilde{V}(\Omega) | \; \varphi \text{ constant on } T \in \mathbb{T}_h\}, \end{aligned}$$

the discrete version reads

$$\begin{array}{lcllll} (\sigma_h, \tau - \sigma_h) & + & (u_h, \operatorname{div} \tau - \operatorname{div} \sigma_h) & \geq & 0 & \forall \tau \in K_h \\ -(\operatorname{div} \sigma_h, \varphi) & & & = & (f, \varphi) & \forall \varphi \in \tilde{V}_h. \end{array} \tag{6.19}$$

6.3.3 Stabilisation

However, (6.19) does not lead to a stable system. As described in [Be95] we stabilise the system by adding the jump of the displacement u_h over an element edge $\Gamma = \partial T$:

$$\begin{array}{lcll} (\sigma_h, \tau - \sigma_h) + (\operatorname{div}\sigma_h, \delta_2 \operatorname{div}\tau) & + & (u_h, \operatorname{div}\tau - \operatorname{div}\sigma_h) & \geq -(f, \delta_2 \operatorname{div}\tau) \\ -(\operatorname{div}\sigma_h, \varphi) & + & \sum_T \sum_{\Gamma \subset T} \delta_{1,T}([u_h]_\Gamma, [\varphi]_\Gamma)_\Gamma & = (f, \varphi). \end{array} \tag{6.20}$$

The stabilising term $\sum_T \sum_{\Gamma \subset T} \delta_{1,T}([u_h]_\Gamma, [\varphi]_\Gamma)_\Gamma$ has the meaning of a weighted discrete Laplacian. Consistency is still satisfied because the jump terms vanish for a continuous displacement which is ensured by the assumption above.

The second stabilisation term $(\operatorname{div}\sigma_h, \delta_2 \operatorname{div}\tau)$ simplyfies the stabilisation proof since for u_h we now have ellipticity on the whole space $H^{div}(\Omega)$ for $\delta_2 > 0$ and not only on the kernel. By setting

$$(\delta_1[u_h], [\varphi]) := \sum_T \sum_{\Gamma \subset T} \delta_{1,T}([u_h]_\Gamma, [\varphi]_\Gamma)_\Gamma \tag{6.21}$$

and

$$\|\delta_1^{\frac{1}{2}}[u_h]\|^2 = (\delta_1[u_h], [u_h]), \tag{6.22}$$

we define

$$\begin{aligned} A_\delta((\sigma_h, u_h), (\tau, \varphi)) &:= (\tau, \sigma_h) + (\mathrm{div}\tau, \delta_2 \mathrm{div}\sigma_h) \\ &\quad + (\mathrm{div}\tau, u_h) - (\mathrm{div}\sigma_h, \varphi) + (\delta_1[u_h], [\varphi]) \\ F_\delta(\tau, \varphi) &:= -(\mathrm{div}\tau, \delta_2 f) + (f, \varphi) \end{aligned}$$

with

$$A_\delta((\sigma_h, u_h), (\tau, \varphi)) = F_\delta(\tau, \varphi) \quad \forall (\varphi, \tau) \in \tilde{V}_h \times U_h \tag{6.23}$$

where U_h is the unrestricted space $U_h = H^{div}(\Omega) \cap \{v \in C(\overline{\Omega}) | v \text{ bilinear on } T \in \mathbb{T}_h\}$.

The mesh dependend (semi-)norm is defined by

$$|||(\sigma_h, u_h)|||_\delta := \|\sigma_h\|_0^2 + \|\delta_2^{\frac{1}{2}} \mathrm{div}\sigma_h\|_0^2 + \|\delta_1^{\frac{1}{2}}[u_h]\|_{\partial T}^2$$

and even by setting $\delta_2 = 1$:

$$|||(\sigma_h, u_h)|||_\delta = \|\sigma_h\|^2_{H_{\text{div}}} + \|\delta_1^{\frac{1}{2}}[u_h]\|^2_{\partial T}.$$

It is obvious that A_δ is positive definite since

$$A_\delta(\{\tau,\varphi\},\{\tau,\varphi\}) \geq c|||\{\tau,\varphi\}|||^2_\delta, \quad 0 < c \leq 1.$$

Remark 6.3.1. *An advantage of the stabilisation we used in* (6.20) *is the positive definite mass matrix* $(\delta_1[u_h],[\varphi])$*. This matrix allows us to evaluate the Schur complement in* σ_h *of the system*

$$\begin{array}{lcrcl} (\tau,\sigma_h) + (\text{div}\tau, \delta_2\text{div}\sigma_h) & + & (\text{div}\tau, u_h) & = & -(\text{div}\tau, \delta_2 f) \\ -(\text{div}\sigma_h, \varphi) & + & (\delta_1[u_h],[\varphi]) & = & (f,\varphi). \end{array}$$

The Schur complement reads

$$(B^T M^{-1} B - A)\sigma_h = B^T M^{-1} F - G$$

with

$$\begin{array}{lll} M_{i,j} = (\delta_1[u_{h,j}],[\varphi_i]), & F_i = (f_i,\varphi_i), & G_i = -(\text{div}\,\tau_i, \delta_2 f_i), \\ B_{i,j} = -(\text{div}\,\sigma_{h,j},\varphi_i), & B^T_{i,j} = (\text{div}\,\tau_j, u_{h,i}), & A_{i,j} = (\tau_j,\sigma_{h,i}) + (\text{div}\,\tau_j, \delta_2\,\text{div}\,\sigma_{h,j}). \end{array}$$

Using a cgPSSOR method, we can solve σ_h *directly without calculating* u_h*. This is a faster variant than the Uzawa algorithm, which calculates* u_h *and* σ_h *in every iteration step.*

6.3.4 A posteriori error analysis

We will provide an error estimator by using the appropriate dual mixed formulation of the regularised problem, with regularisation parameter $\mu > 0$:
Find $(\sigma_\mu, u_\mu) \in H^{div} \times \tilde{V}$ such that

$$(\sigma_\mu,\tau) + (\mu^{-1}(\sigma_\mu - \pi\sigma_\mu),\tau) + (u_\mu, \text{div}\,\tau) - (\text{div}\,\sigma_\mu,\varphi) = (f,\varphi) \tag{6.24}$$

$\forall(\tau,\varphi) \in H^{div} \times \tilde{V}$, which has a unique solution.

Lemma 6.1. *There holds the estimate*

$$\Phi_\mu(\sigma_\mu) = \frac{1}{2\mu}|\sigma_\mu - \pi\sigma_\mu|^2 \leq C.$$

A proof can be found in [Jo76/1].

For the error there holds

$$\|\sigma - \sigma_h\| \leq \|\sigma - \sigma_\mu\| + \|\sigma_\mu - \sigma_h\|.$$

As a first step we prove the strong convergence of σ_μ:

Since we have the monotone operator $\Phi'_\mu(\tau)$, we achieve

$$(\frac{1}{\mu}(\sigma_\mu - \pi\sigma_\mu) - \frac{1}{\mu}(\sigma - \pi\sigma), \sigma_\mu - \sigma) \geq 0$$

and we can estimate

$$\|\sigma_\mu - \sigma\|^2 \leq (\sigma_\mu - \sigma, \sigma_\mu - \sigma) + (\frac{1}{\mu}(\sigma_\mu - \pi\sigma_\mu) - \frac{1}{\mu}(\sigma - \pi\sigma), \sigma_\mu - \sigma).$$

Furthermore, for the continuous solution σ holds

$$(\sigma - \pi\sigma) = 0$$

and

$$-(\operatorname{div}\sigma, \varphi) = -(\operatorname{div}\sigma_\mu, \varphi) = (f, \varphi).$$

Hence one receives:

$$\begin{aligned}\|\sigma_\mu - \sigma\|^2 &\leq (\sigma_\mu, \sigma_\mu - \sigma) + (\frac{1}{\mu}(\sigma_\mu - \pi\sigma_\mu), \sigma_\mu - \sigma) \\ &\quad - (\sigma, \sigma_\mu - \sigma) + (u_\mu, \operatorname{div}(\sigma_\mu - \sigma)) \\ &= -(\sigma, \sigma_\mu - \sigma) \to 0 \quad \text{for } \mu \to 0.\end{aligned}$$

So we received strong convergence of σ_μ out of the weak convergence proven by [Jo76/1]. In the following we concentrate on the estimation of $\|\sigma_\mu - \sigma_h\|$. Here, for case of simplicity, we redefine the notation $u_h := u_{\mu,h}$ and start with

$$
\begin{aligned}
(\sigma_\mu-&\sigma_h,\sigma_\mu-\sigma_h)+(\delta_1[u_\mu-u_h],[u_\mu-u_h])\\
&\leq |||(\sigma_\mu-\sigma_h,u_\mu-u_h)|||_\delta^2\\
&=(\sigma_\mu-\sigma_h,\sigma_\mu-\sigma_h)+(\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h,\delta_2(\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h))\\
&\quad+(\delta_1[u_\mu-u_h],[u_\mu-u_h])\\
&\leq \underbrace{(\sigma_\mu,\sigma_\mu-\sigma_h)+(u_\mu,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)}_{I}\underbrace{-(u_\mu-u_h,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)+(\delta_1[u_h],[u_h])}_{II}\\
&\quad\underbrace{+(\frac{1}{\mu}(\sigma_\mu-\pi\sigma_\mu)-\frac{1}{\mu}(\sigma_h-\pi\sigma_h),\sigma_\mu-\sigma_h)}_{\geq 0}\\
&\quad\underbrace{-(u_h,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)-(\sigma_h,\sigma_\mu-\sigma_h)}_{III}+(\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h,\delta_2(\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)).
\end{aligned}
$$

We consider that $(\sigma_h-\pi\sigma_h)=0$ and hence for term I we receive by testing (6.24) with $\sigma_\mu-\sigma_h$:

$$
(\sigma_\mu,\sigma_\mu-\sigma_h)+(\frac{1}{\mu}(\sigma_\mu-\pi\sigma_\mu),\sigma_\mu-\sigma_h)+(u_\mu,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)=0. \tag{6.25}
$$

The next part can be rewritten with the help of the following equation:

$$
\begin{aligned}
(\varphi_h,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)&=(\varphi_h,\operatorname{div}\sigma_\mu)-(\varphi_h,\operatorname{div}\sigma_h)\\
&\overset{(6.24),(6.20)}{=}-(\varphi_h,f)+(\varphi_h,f)-(\delta_1[\varphi_h],[u_h]).
\end{aligned}
$$

Let $u_i:=I_hu_\mu$ describe the interpolation of u_μ on the finite element space $I_h:V\to V_h$. Then for II we get

$$
\begin{aligned}
-(u_\mu-u_h,&\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)+(\delta_1[u_h],[u_h])\\
&=-(u_\mu-u_h+u_i-u_i,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)+(\delta_1[u_h],[u_h])\\
&=-(u_\mu-u_i,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)+(\delta_1[u_h],[u_h])-(\delta_1[u_h][u_h])+(\delta_1[u_h],[u_i])\\
&=-(u_\mu-u_i,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)+(\delta_1[u_h],[u_i]).
\end{aligned}
$$

Furthermore, for III there holds

$$
-(u_h,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)-(\sigma_h,\sigma_\mu-\sigma_h) \tag{6.26}
$$

$$
=-(u_h-Z,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)-(Z,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)-(\sigma_h,\sigma_\mu-\sigma_h) \tag{6.27}
$$

$$
=-(u_h-Z,\operatorname{div}\sigma_\mu-\operatorname{div}\sigma_h)+(\nabla Z-\sigma_h,\sigma_\mu-\sigma_h). \tag{6.28}
$$

Here, $Z \in W_h := \{w \in H_0^1(\Omega) |\, w \text{ bilinear on } T \in \mathbb{T}_h\}$ is determined by

$$(\frac{1}{p(Z)}\nabla Z, \nabla\Psi) = (\sigma_h, \nabla\Psi) \quad \forall\Psi \in W_h,$$

where $p(Z)$ denotes a projection of Z defined by

$$p(Z) = \begin{cases} 1, & \text{if } |\nabla Z| \leq 1, \\ |\nabla Z|, & \text{otherwise.} \end{cases} \tag{6.29}$$

Setting $\lambda_h = p(Z) - 1$, the last term of (6.28) can be estimated in the following way:

$$\begin{aligned}(\nabla Z - \sigma_h, \sigma_\mu - \sigma_h) &= \sum_T (\nabla Z - p(Z)\sigma_h, \sigma_\mu - \sigma_h)_T + (\lambda_h\sigma_h, \sigma_\mu - \sigma_h) \\ &\leq \sum_T \|\nabla Z - p(Z)\sigma_h\|_T \|\sigma_\mu - \sigma_h\|_T \\ &\quad + \int_\Omega \lambda_h\sigma_h\sigma_\mu - \int_\Omega \lambda_h|\sigma_h|^2 \\ &\leq \sum_T \frac{1}{2}\|\nabla Z - p(Z)\sigma_h\|_T^2 + \frac{1}{2}\|\sigma_\mu - \sigma_h\|^2 \\ &\quad + \int_\Omega \lambda_h|\sigma_h|\,|\sigma_\mu| - \int_\Omega \lambda_h|\sigma_h|^2 \\ &\leq \sum_T \frac{1}{2}\|\nabla Z - p(Z)\sigma_h\|_T^2 + \frac{1}{2}\|\sigma_\mu - \sigma_h\|^2 \\ &\quad + \int_\Omega \lambda_h|\sigma_h|(1 + c\sqrt{\mu}) - \int_\Omega \lambda_h|\sigma_h|^2.\end{aligned} \tag{6.30}$$

The last step includes the estimation $\frac{1}{2\mu}|\sigma_\mu - \pi\sigma_\mu|^2 \leq C$ (see Lemma 6.1). Collecting results (6.25)-(6.30) we achieve

$$\begin{aligned}&\|\sigma - \sigma_h\|^2 + \|\delta_1^{\frac{1}{2}}[u_h]\|^2 \\ &\qquad \leq (u_\mu - u_i, \text{div}\sigma_h - \underbrace{\text{div}\sigma}_{=-f}) + \underbrace{(\delta_1[u_h], [u_i])}_{VI} + (u_h - Z, \text{div}\sigma_h - \text{div}\sigma) \\ &\qquad + \frac{1}{2}\sum_T \|\nabla Z - p(Z)\sigma_h\|_T^2 + \frac{1}{2}\|\sigma - \sigma_h\|^2 \\ &\qquad + \int \lambda_h|\sigma_h|((1 + c\sqrt{\mu}) - |\sigma_h|) + \|\delta_2^{\frac{1}{2}}(\text{div}\,\sigma_h + f)\|^2.\end{aligned} \tag{6.31}$$

Taking the first term, we follow the interpolation estimate from [JH91]

$$|(f, v - v_i)| \leq C\|hf\|_{L_p(T)}\|\nabla v\|_{L_q(T)}, \quad \frac{1}{p} + \frac{1}{q} = 1,\ T \in \mathbb{T}_h,\ p, q \in [1, \infty] \tag{6.32}$$

where $v_i := I_h v$ denotes the interpolant of v on the finite element space, C independent of v, f, h, T and achieve:

$$\begin{aligned}(u_\mu - u_i, \operatorname{div}\sigma_h + f) &= \sum_T (u_\mu - u_i, f - f_i)_T \\ &\leq C \sum_T \|h_T(f - f_i)\|_{\infty,T} \|\nabla u_\mu\|_{L_1(T)}.\end{aligned}$$

Term VI on the right hand side of (6.31) can then be further estimated:

$$\begin{aligned}\delta_1([u_h],[u_i])_T &= \delta_1 \sum_{\Gamma\subset T} ([u_h], u_i|_T - u_i|_{T'_\Gamma})_\Gamma \\ &\leq \delta_1 \sum_{\Gamma\subset T} \Big[([u_h], u_i|_T - u_\mu)_\Gamma + ([u_h], u_\mu - u_i|_{T'_\Gamma})_\Gamma\Big],\end{aligned}$$

where $u_i|_{T'_\Gamma}$ describes the interpolation of u_μ on the neighbour cell of T with $T \cap T'_\Gamma = \Gamma$. Again we follow [JH91] by using interpolation estimate (6.32):

$$(\delta_1[u_h],[u_i])_T \leq C\delta_{1,T}\|h_T D_h u_h\|_{\infty,T}\|\nabla u_\mu\|_{L_1(T)}$$

with the maximal jump over an element edge

$$D_h u_h = \max_{\partial T} \frac{|[u_h]|}{h_T}.$$

Due to the considerations of [JH91] we can estimate $\|\nabla u_\mu\|_{L_1(T)} \leq C$ under the assumption that the safe load hypothesis if fulfilled (see inequality (6.18)). Alternatively, for a complete a posteriori estimator, we can replace $\|\nabla u_\mu\|_{L_1(T)}$ by $\|\nabla u_h\|_{L_1(T)}$. Since we have chosen piecewise constant elements this would be zero, so we rather choose the definition $\|D_h u_h\|_{L_1(T)}$. Collecting the results and passing to the limit $\mu \to 0$, we receive the following theorem:

Theorem 6.3.2. *For problem* (6.20) *there holds the a posteriori error bound*

$$\begin{aligned}\|\sigma - \sigma_h\|^2 + \|\delta_1^{\frac{1}{2}}[u_h]\|^2 \leq C \sum_T \Big(& h_T\|f - f_i\|_{\infty,T} + \delta_{1,T} h_T \|D_h u_h\|_{\infty,T} \\ & + \|\nabla Z - p(Z)\sigma_h\|_T^2 + (u_h - Z, \operatorname{div}\sigma_h + f)_T \\ & + \int_T \lambda_h|\sigma_h|(1 - |\sigma_h|) + \delta_{2,T}\|\operatorname{div}\sigma_h + f\|_T^2\Big),\end{aligned} \tag{6.33}$$

where for interior interelement boundaries $[u_h]$ *denotes the jump of displacement variable* u_h *over an element edge.*

6.4 Numerical results

We give some numerical results for the primal mixed system and for the dual one afterwards. As a test example we take the two dimensional domain $\Omega = [0,1]^2$ and a body force $f = 0.75\pi^2 \sin(\pi x)\sin(\pi y)$.

6.4.1 Primal problem

Testing estimator (6.15) of the primal system u is discretised by bilinear elements and κ_h are piecewise constant functions on the corresponding triangulation. In order to use (6.15), we take the L_2-projection of κ_h into the FE-space consisting of Q_1-Elements, that means $\mathcal{M} : Q_0 \to Q_1$, for calculating $\|(\mathcal{M}(\kappa_h) - \kappa_h)\nabla u_h\|_T^2$. If we refine the grid uniformly the estimator shows an optimal convergence order of $\mathcal{O}(h)$ as we can see in Table 6.2.

# cells	e_{res}		e_{jump}		e_κ		$\|e\|_1$	
16	1.309e+00	-	3.258e-01	-	0.000e+00	-	1.206e+00	-
64	6.543e-01	1.00	2.005e-01	0.70	5.467e-02	0.00	6.169e-01	0.97
256	3.271e-01	1.00	1.104e-01	0.86	3.438e-02	0.67	3.126e-01	0.98
1024	1.636e-01	1.00	5.689e-02	0.95	1.614e-02	1.09	1.566e-01	1.00
4096	8.178e-02	1.00	2.872e-02	0.99	7.804e-03	1.05	7.831e-02	1.00
16384	4.089e-02	1.00	1.440e-02	1.00	3.847e-03	1.02	3.916e-02	1.00
65536	2.045e-02	1.00	7.209e-03	1.00	1.917e-03	1.01	1.958e-02	1.00

Table 6.2: Setting $e_{\text{res}} = (\sum_T h_T^2 \|f + \kappa_h \Delta u_h + \nabla\kappa_h \nabla u_h\|_T^2)^{\frac{1}{2}}$, $e_{\text{jump}} = (\sum_E h_E \|n \cdot [\kappa_h \nabla u_h]\|_E^2)^{\frac{1}{2}}$ and $e_\kappa = (\sum_T \|(\mathcal{M}(\kappa_h) - \kappa_h)\nabla u_h\|_T^2)^{\frac{1}{2}}$, every component of the estimator converges optimal and so does the whole estimator (6.15) denoted by $|e|_1$. This can be observed in the right parts of every multicolumn, where there is always a value α determining the convergence order by $\mathcal{O}(h^\alpha)$ for every refinement step.

Comparing the estimator to (6.5) the improvement gets obvious since estimator (6.15) has a convergence order of $\mathcal{O}(h)$ in every region whereas estimator (6.5) has a convergence order of $\mathcal{O}(h^{\frac{1}{2}})$ in plastic areas (see Figure 6.3).

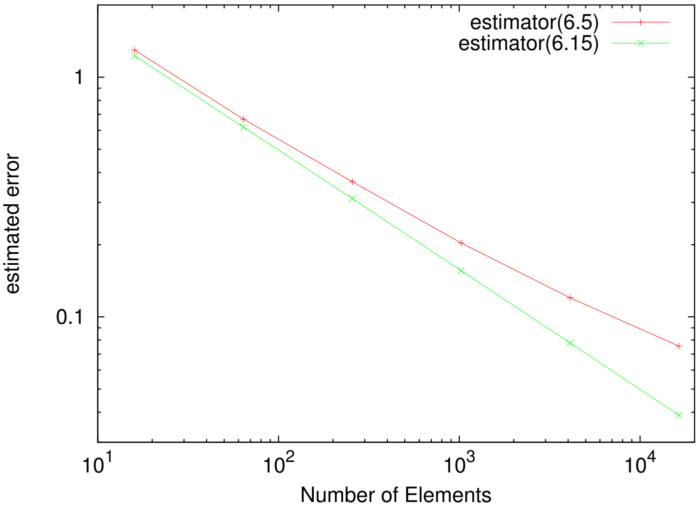


Figure 6.3: Comparison of estimator (6.5) and the improved estimator (6.15). The latter has a better convergence rate because the convergence order is still optimal in regions of plasticity which is not fulfilled in (6.5).

Adaptive refinement leads to a good mesh structure which outlines the critical zones.

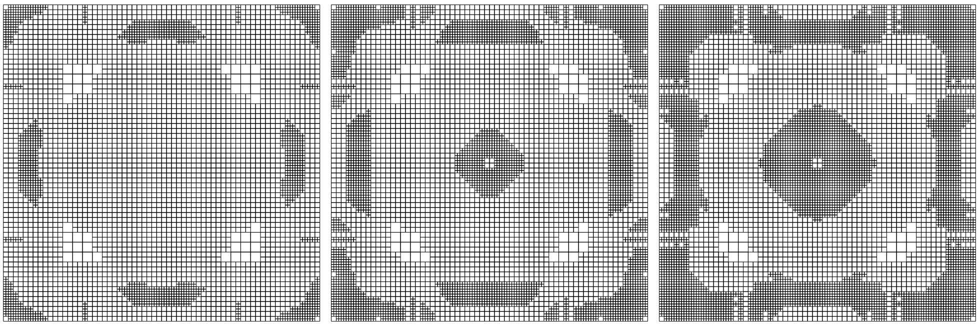
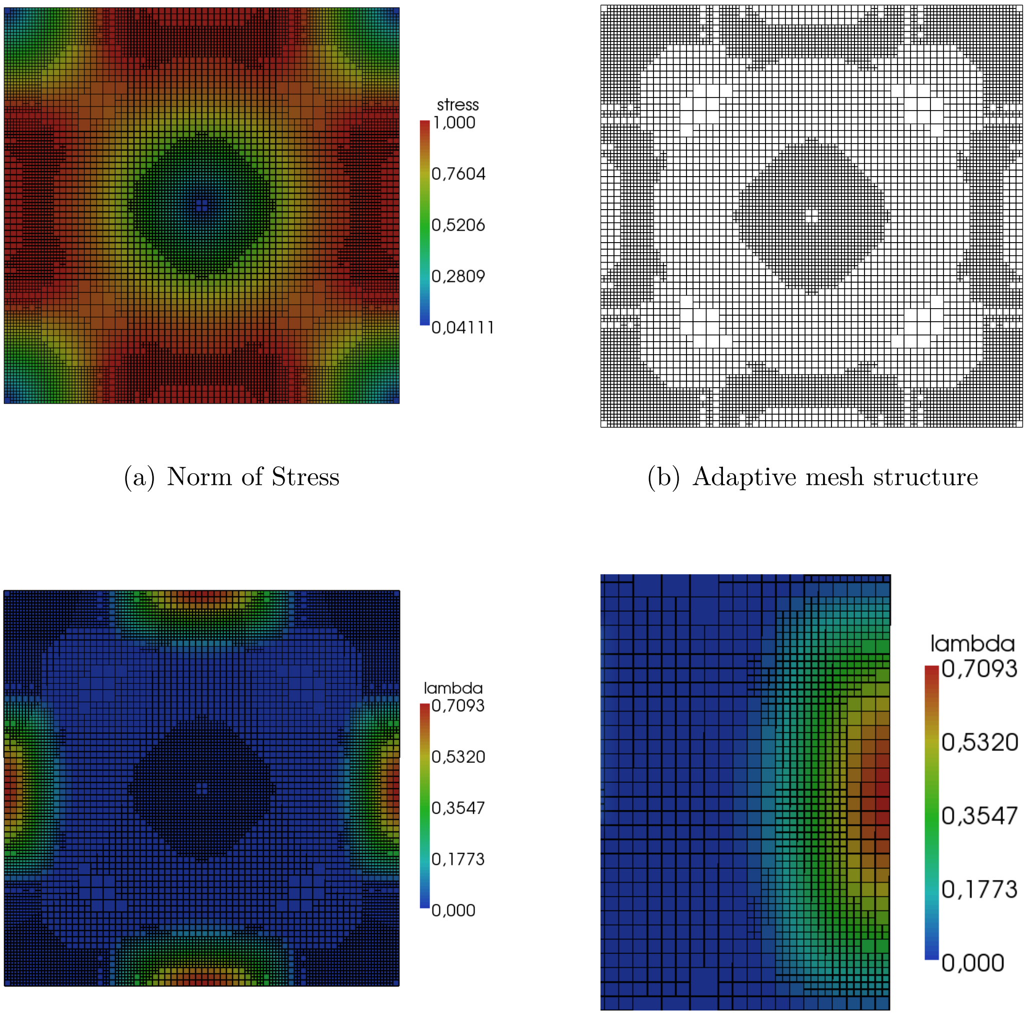


Figure 6.4: A sequence of grids created by estimator (6.15) within adaptive refinement.

Critical zones in this example are those where the material begins to plastify. Like in other examples before, the Lagrangian multiplier implies a good refinement of these areas which can be seen in Figures 6.4 and 6.5.



(a) Norm of Stress

(b) Adaptive mesh structure

(c) Lagrangian multiplier λ_h

(d) Zoom to λ_h

Figure 6.5: (a) shows the solution of the stress $\|\sigma_h\|$, (b) gives the corresponding mesh structure for the appropriate refinement step. In (c) and (d) we can observe that the most refined areas are those critical parts where the material begins to plastify.

Working on shapes which have singularities formular (6.15) turns out to produce almost optimal convergence order refining adaptively. So again we achieved an economical error estimator producing optimal mesh structures with the help of the Lagrangian parameter.

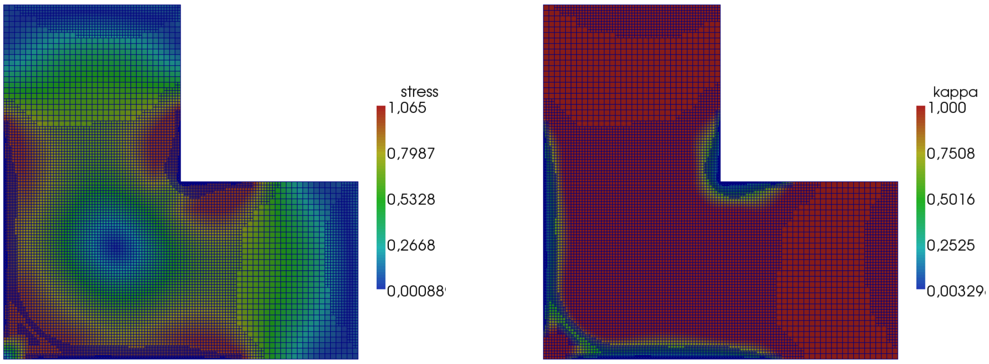


Figure 6.6: Norm of Stress $\|\sigma_h\|$ and κ_h calculated on an L-shape $\Omega = [-1,1]^2 \backslash (0,1]^2$. The singularity as well as zones where plasticity arise are well refined.

Figures 6.6 and 6.7 as well as Table 6.3 confirm these results.

# cells	$\|e\|_{1,\text{global}}$		# cells	$\|e\|_{1,\text{adaptive}}$	
192	9.986e-02	-	192	9.986e-02	-
768	1.105e-01	-0.15	540	8.843e-02	0.24
3072	8.073e-02	0.45	4053	3.791e-02	0.84
12288	5.134e-02	0.65	11889	2.372e-02	0.87
49152	3.111e-02	0.72	111150	8.106e-03	0.97
196608	1.857e-02	0.74	341571	4.708e-03	0.97

Table 6.3: Global and adaptive refinement on a shape having a singularity. The adaptive refinement using estimator (6.15) shows a better convergence rate than global refinement since the value α on the right side of the multicolumns determines the convergence order by $\mathcal{O}(h^\alpha)$ for every refinement step.

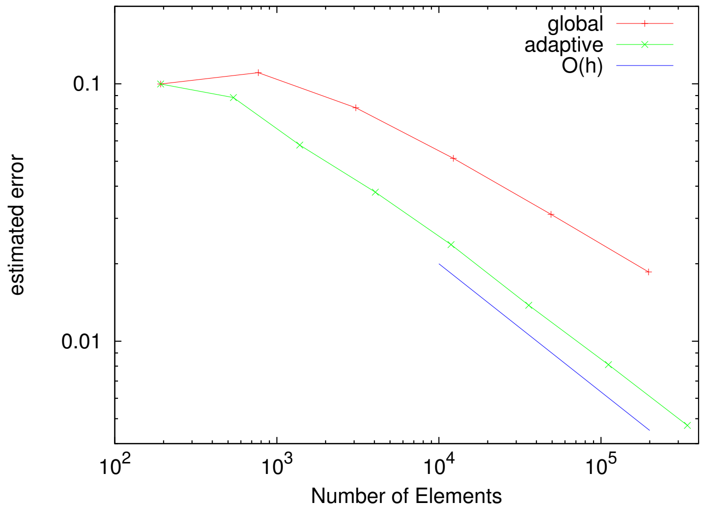


Figure 6.7: Adaptive refinement on a shape including singularities using estimator (6.15) offers a better convergence order than global refinements. The calculation was made on the L-shape of Figure 6.6

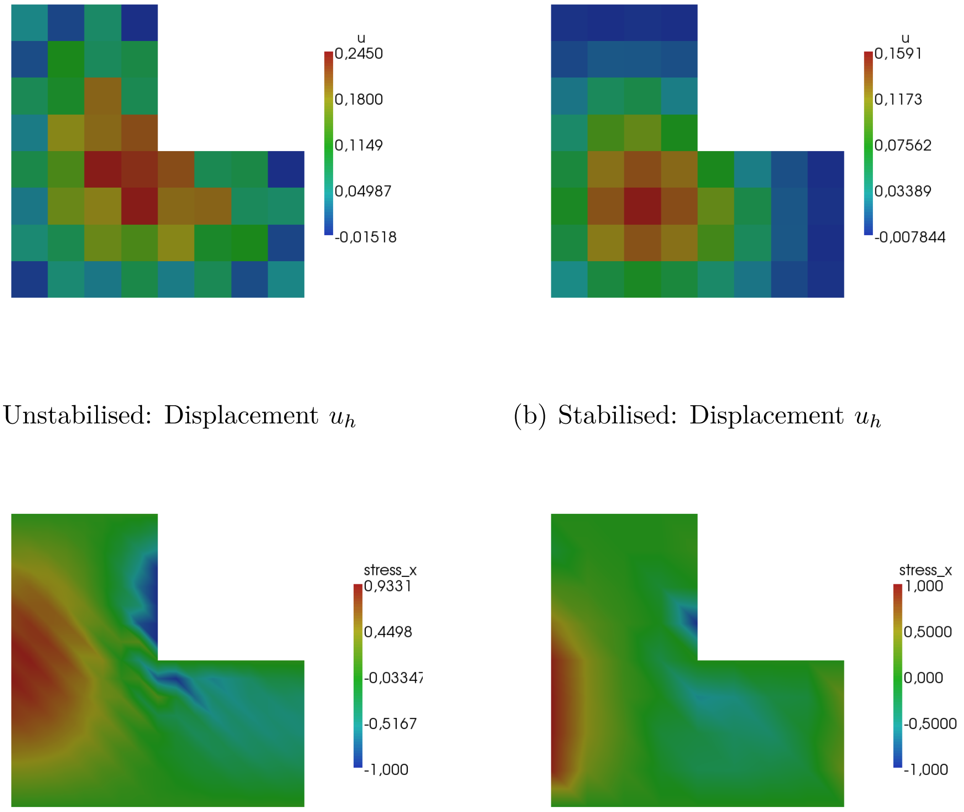
6.4.2 Dual problem

Taking the same domain and right hand side as above we first give a result about stability.

	unstable system				stabilised system			
# cells	$\|u - u_h\|_0$		$\|\sigma - \sigma_h\|_0$		$\|u - u_h\|_0$		$\|\sigma - \sigma_h\|_0$	
16	7.224e-02	-	1.031e-01	-	8.117e-02	-	6.195e-02	-
64	3.422e-02	1.08	5.275e-02	0.97	3.564e-02	1.19	1.641e-02	1.92
256	1.732e-02	0.98	1.821e-02	1.53	1.739e-02	1.04	3.969e-03	2.05
1024	8.685e-03	0.99	6.499e-03	1.49	8.688e-03	1.00	9.687e-04	2.03
4096	4.346e-03	0.99	2.340e-03	1.47	4.346e-03	0.99	2.442e-04	1.99

Table 6.4: Convergence rate of system (6.19) and (6.20) without projection on a wraped grid (see Figure 3.9 (a)). The unstable system does not show the expected L_2-convergence rate of the stress variable in contrast to the stabilised one.

Therefore, we calculate a fully elastic problem, that means we disregard the projection $|\sigma_h| \leq 1$. That does not change anything about stability but makes it easier to find an exact solution and hence to get the true error. Calculating the system on a wrapped grid reveals the instability of system (6.19). In contrast the stabilised system (6.20) offers the expected convergence order of the true error (see Table 6.4).



(a) Unstabilised: Displacement u_h

(b) Stabilised: Displacement u_h

(c) Unstabilised: Stress $\|\sigma_h\|$

(d) Stabilised: Stress $\|\sigma_h\|$

Figure 6.8: Calculation on an L-shape $\Omega = [0,1]^2 \backslash (0.5,1]^2$ with right hand side $f = 0.73\pi^2 \sin(\pi(x+0.25))\sin(\pi(y+0.25))$. Left: $\delta_1 = \delta_2 = 0$: Instabilities get obvious by oscillations in the displacement as well as in the stress variable. Right: $\delta_1 = \mathcal{O}(h)$, $\delta_2 = \mathcal{O}(h^2)$: We get smooth solutions u_h and σ_h.

For the further tests we calculate on a conform mesh and take account of the projection again. Global refinements of a square domain offer an optimal convergence order of estimator (6.33), too, which can be observed in Table 6.5.

# cells	ρ_1		ρ_2		ρ_3	
16	1.615e-01	-	1.744e-01	-	1.649e-01	-
64	9.741e-02	0.73	9.335e-02	0.90	9.034e-02	0.87
256	5.203e-02	0.90	4.703e-02	0.99	4.693e-02	0.94
1024	2.700e-02	0.95	2.343e-02	1.01	2.439e-02	0.94
4096	1.357e-02	0.99	1.174e-02	1.00	1.243e-02	0.97
16384	6.789e-03	1.00	5.864e-03	1.00	6.272e-03	0.99

# cells	ρ_4		ρ_5		$\|e\|_1$	
16	0.000e+00	-	6.624e-01	-	2.986e-01	-
64	2.140e-02	0.00	3.364e-01	0.98	1.649e-01	0.86
256	8.037e-03	0.98	1.761e-01	0.93	8.491e-02	0.96
1024	4.075e-03	0.98	9.385e-02	0.91	4.349e-02	0.96
4096	1.974e-03	1.04	5.245e-02	0.84	2.192e-02	0.99
16384	9.446e-04	1.06	3.004e-02	0.81	1.099e-02	0.99

Table 6.5: We set $\rho_1^2 = \delta_1 \sum_T([u_h],[u_h])_T$, $\rho_2^2 = \sum_T \|\nabla Z - (1+\lambda_h)\sigma_h\|_T^2$, $\rho_3^2 = \sum_T(u_h - Z, \operatorname{div}\sigma_h + f)_T$, $\rho_4^2 = \sum_T \int_T \lambda_h|\sigma_h|(1-|\sigma_h|)$ and $\rho_5^2 = \sum_T \|\operatorname{div}\sigma_h + f\|_T^2$. The estimator has a convergence rate of $\mathcal{O}(h)$.

Looking at Figure 6.9 and 6.10 we can see that the achieved estimator (6.33) fulfills our expectation to outline critical zones. As described above, these are the areas where plastification starts and we find them well refined.

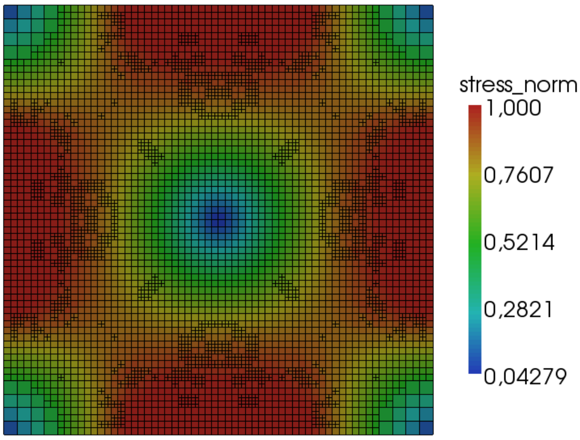


Figure 6.9: Refining adaptively the Lagrangian parameter provides the well refined transition zone between elastic and plastic areas.

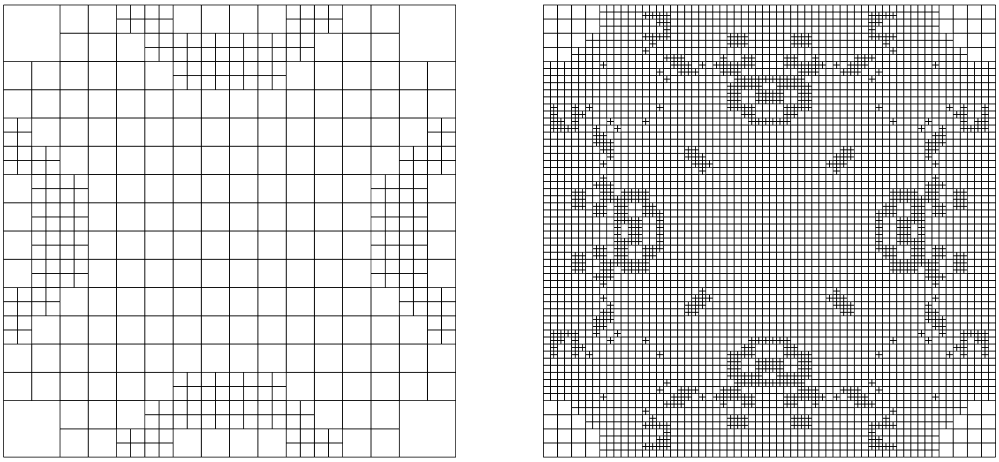


Figure 6.10: Sequence of grids created by estimator (6.33).

Furthermore, the estimator turns out to behave proper when refining shapes including singularities like in Figure 6.6. The convergence order is almost optimal whereas the estimated error within global refinement on such areas converges much slower (see Table 6.6 and Figure 6.11).

# cells	$\lvert e\rvert_{1,\text{global}}$		# cells	$\lvert e\rvert_{1,\text{adaptive}}$	
48	5.775e-01	-	48	5.77e-01	-
192	3.674e-01	0.65	162	3.33e-01	0.91
768	2.698e-01	0.45	975	1.47e-01	0.91
3072	2.055e-01	0.39	3156	8.07e-02	1.01
12288	1.583e-01	0.38	16944	4.03e-02	0.83

Table 6.6: Comparison of the estimated values within global and adaptive refinement on an L-shape. As expected, using global refinement steps the convergence order is low whereas it is almost optimal calculating on adaptive grids.

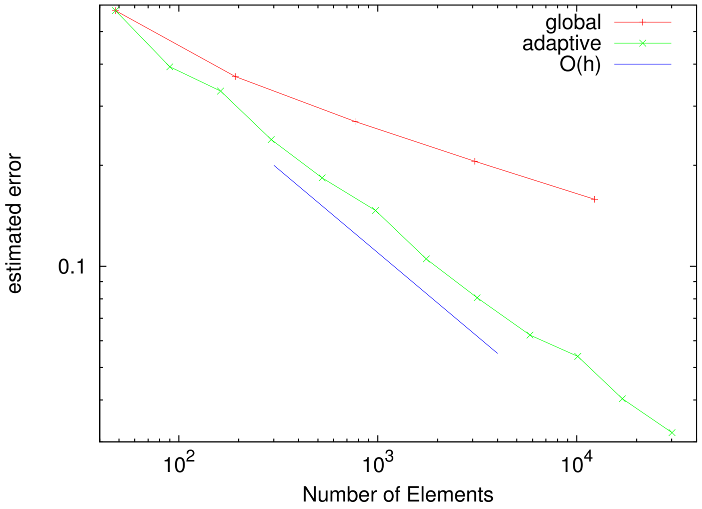


Figure 6.11: Using adaptive refinement on shapes with singularities, estimator (6.33) offers almost linear convergence.

7 Simplified Signorini problem

The technique of the Lagrangian formalism is transfered to a problem where the restriction lives on the boundary of the domain. A standard example in this case is the simplified Signorini problem which is again a contact situation but related to the boundary. After ensuring the well-posedness of the problem we introduce the saddle point formulation where the discrete version is not stable in contrast to the continuous case. To avoid instabilities we present two ways of stabilisation based on the least squares method and compare this technique to the one described in [Sc05]. It turns out by numerical tests that the consistent error estimator, taking account on the Lagrangian multiplier and the additional stabilisation term, works efficiently just as the utilised stabilisation does, which can be observed in figures and tables at the end of the chapter.

We want to extend the method of duality based error estimators by choosing the Lagrangian multiplier on the boundary of a domain. Therefore, we consider the so called simplified Signorini problem. There we have, similar to the obstacle problem, a membrane that is loaded by a body force f:

$$-\Delta u = f \text{ on } \Omega. \tag{7.1}$$

In addition, there are two different boundary conditions:

$$u = 0 \text{ on } \Gamma_1 \tag{7.2}$$

$$\partial_n u = g \text{ on } \Gamma_2 \tag{7.3}$$

with $\Gamma = \Gamma_1 \cup \Gamma_2$ and $\Gamma_1 \cap \Gamma_2 = \emptyset$. The variational formulation is written as

$$u \in H^1(\Omega, \Gamma_1): \quad (\nabla u, \nabla v) = (f, v) + \int_{\Gamma_2} g \cdot \gamma(v)\, d\Gamma_2 \quad \forall v \in H^1(\Omega, \Gamma_1),$$

with the trace operator $\gamma(\cdot)$ and

$$H^1(\Omega,\Gamma_1) := \{v \in H^1(\Omega) |\, v = 0 \text{ a.e. on } \Gamma_1\}.$$

Now we introduce an obstacle Ψ_{Γ_2} that lives on Γ_2 and the classical formulation of the restricted problem changes as follows:

$$\begin{aligned} -\Delta u &= f \quad \text{on } \Omega, \\ u &= 0 \quad \text{on } \Gamma_1, \\ g - \partial_n u &\leq 0 \quad \text{on } \Gamma_2, \\ (g - \partial_n u)(\gamma(u) - \Psi_{\Gamma_2}) &= 0 \end{aligned}$$

for $u \in C^2(\Omega) \cap C(\bar{\Omega})$.

7.1 Variational formulation

As we consider an obstacle that only applies to the Neumann boundary, v is an element of the restricted subspace

$$K := \{v \in H^1(\Omega,\Gamma_1) |\, \gamma(v) \geq \Psi_{\Gamma_2} \text{ a.e.}\}$$

with $\Psi_{\Gamma_2} \in L_2(\Gamma_2)$. So our primal problem is of the form

$$u \in K: \quad (\nabla u, \nabla(v-u)) \geq (f, v-u) + \int_{\Gamma_2} g \cdot \gamma(v-u)\, d\Gamma_2 \quad \forall v \in K. \tag{7.4}$$

K is convex because we can choose $\varepsilon \in [0,1]$, $v, w \in K$ and there holds

$$\begin{aligned} \gamma(\varepsilon v + (1-\varepsilon)w) - \Psi_{\Gamma_2} &= \varepsilon\gamma(v) + (1-\varepsilon)\gamma(w) - \Psi_{\Gamma_2} \qquad (7.5) \\ &= \varepsilon(\gamma(v) - \Psi_{\Gamma_2}) + (1-\varepsilon)(\gamma(w) - \Psi_{\Gamma_2}) \geq 0. \qquad (7.6) \end{aligned}$$

If $(v_n)_n \subset K$ and $v_n \to v$ in $H^1(\Omega,\Gamma_1)$, then $\gamma(v_n) \to \gamma(v)$, because $\gamma : H^1(\Omega,\Gamma_1) \to H^{\frac{1}{2}}(\Gamma_2)$ is continuous. Since $v_n \in K$, $\gamma(v_n) \geq \Psi_{\Gamma_2}$ a.e. on Γ_2. Therefore, $\gamma(v) \geq \Psi_{\Gamma_2}$ a.e. on Γ_2. Hence, $v \in K$ which shows that K is also closed.

Since K is a convex and closed cone the problem has a unique solution ensured by Theorem 2.3. For an approximation of the problem we consider the subspace $V_h \subset V$. The discrete version reads

$$u_h \in K_h : \quad (\nabla u_h, \nabla(v - u_h)) \geq (f, v - u_h) + \int_{\Gamma_2} g \cdot \gamma(v - u_h)\, d\Gamma_2 \quad \forall v \in K_h. \tag{7.7}$$

$\Psi_{\Gamma_2,h}$ is the bilinear interpolant of Ψ_{Γ_2} and K_h is given by

$$K_h = \{v \in V_h |\, \gamma(v) \geq \Psi_{\Gamma_2,h} \text{ a.e. on } \Gamma_2\}.$$

Like in Chapter 3, for sake of simplicity we assume $\Psi_{\Gamma_2} = \Psi_{\Gamma_2,h}$. By Theorem 2.1.1 this problem is unique solvable, too. Following the same line of estimations as it is shown in Section 3.1.2, we receive an a posteriori estimator for (7.7) of the following form:

Theorem 7.1. *For the simplified Signorini problem* (7.7) *there holds the error bound*

$$|e|_1^2 \leq C \sum_{T \in \mathbb{T}_h} \varrho_T^2 \tag{7.8}$$

with local residuals

$$\varrho_T = \begin{cases} h_T \|\Delta u_h + f\|_T + \frac{1}{2} h_T^{\frac{1}{2}} \|n \cdot [\nabla u_h]\|_{\partial T} & \text{if } \partial T \in \varepsilon_0 \\ h_T \|\Delta u_h + f\|_T + h_T^{\frac{1}{2}} \|g - \partial_n u_h\|_{\partial T} & \text{if } \partial T \subset \Gamma_2, \end{cases}$$

where $\varepsilon_0 \subset \mathbb{E}_h$ denotes the interior faces of a cell.

7.2 Saddle point problem

In what follows we set $\langle \cdot, \cdot \rangle_{\Gamma_2} : L_2(\Gamma_2) \times L_2(\Gamma_2) \to \mathbb{R}$ the dual pairing on Γ_2. The Lagrangian formulation looks as follows:
Find a pair $(u, \lambda) \in V \times \Lambda := \{q \in L_2(\Gamma_2) |\, q \geq 0 \text{ a.e.}\}$ and $V = H^1(\Omega, \Gamma_1)$ with

$$\mathcal{L}(u, \lambda) = \inf_{\varphi \in V} \sup_{\omega \in \Lambda} \mathcal{L}(\varphi, \omega) \tag{7.9}$$

$$= \inf_{\varphi \in V} \sup_{\omega \in \Lambda} \left\{ \frac{1}{2} a(\varphi, \varphi) - (f, \varphi) - \langle g, \varphi \rangle_{\Gamma_2} - \langle \omega, \gamma(\varphi) - \Psi_{\Gamma_2} \rangle_{\Gamma_2} \right\}. \tag{7.10}$$

Derivation with respect to φ and ω leads to the system

$$\begin{array}{llcll} u \in V: & a(u,\varphi) - \langle \lambda, \varphi\rangle_{\Gamma_2} & = & (f,\varphi) + \langle g, \varphi\rangle_{\Gamma_2} & \forall \varphi \in V \\ \lambda \in \Lambda: & \langle u, \omega - \lambda\rangle_{\Gamma_2} & \geq & \langle \Psi_{\Gamma_2}, \omega - \lambda\rangle_{\Gamma_2} & \forall \omega \in \Lambda. \end{array} \tag{7.11}$$

Since $rg(\gamma(H^1(\Omega,\Gamma_1))) = H^{\frac{1}{2}}(\Gamma_2)$ is closed in $H^{\frac{1}{2}}(\Gamma_2)$, system (7.11) has a unique solution ensured by Theorem 2.9. Introducing the discrete version by setting

$$u_h \in V_h = \{v \in H^1(\Omega,\Gamma_1) | v \text{ bilinear on } T \in \mathbb{T}_h\} \tag{7.12}$$

and

$$\lambda_h \in \Lambda_h = \{\omega \in L_2(\Gamma) | \, \omega \text{ constant on } \partial T \subset \Gamma_2\}, \tag{7.13}$$

we receive the system:
Find a pair $(u_h, \lambda_h) \in V_h \times \Lambda_h$ fulfilling

$$\begin{array}{lcll} a(u_h,\varphi) - \langle \lambda_h, \varphi\rangle_{\Gamma_2} & = & (f,\varphi) + \langle g, \varphi\rangle_{\Gamma_2} & \forall \varphi \in V_h \\ \langle u_h, \omega - \lambda_h\rangle_{\Gamma_2} & \geq & \langle \Psi_{\Gamma_2}, \omega - \lambda_h\rangle_{\Gamma_2} & \forall \omega \in \Lambda_h. \end{array} \tag{7.14}$$

7.2.1 Stabilisation

System (7.14) does not fulfill the discrete inf-sup-condition by choosing the finite element spaces above. This gets obvious in Section 7.3, where the Lagrangian multiplier is plotted and shows oscillations which makes a stabilitsation necessary. For $u \in H^2(\Omega) \cap H^1(\Omega)$, system (7.11) can be rewritten as follows:
Find a pair $(u, \lambda) \in V \times \Lambda$ such that

$$\begin{array}{lcll} -(\Delta u,\varphi) - \langle \lambda, \varphi\rangle_{\Gamma_2} & = & (f,\varphi) + \langle g, \varphi\rangle_{\Gamma_2} - \langle \partial_n u, \varphi\rangle_{\Gamma_2} & \forall \varphi \in V \\ \langle u, \omega - \lambda\rangle_{\Gamma_2} & \geq & \langle \Psi_{\Gamma_2}, \omega - \lambda\rangle_{\Gamma_2} & \forall \omega \in \Lambda. \end{array} \tag{7.15}$$

Since we know $-\Delta u = f$ from (7.1) there holds

$$\lambda = \partial_n u - g \text{ on } \Gamma_2.$$

Let now $E \in \mathbb{E}_h$ be the outer boundary elements of a cell with $E \subset \Gamma_2$ and $\delta > 0$. Then the stabilisation looks as follows:

Find a pair $(u_h, \lambda_h) \in V_h \times \Lambda_h$ fulfilling

$$\begin{aligned} a(u_h, \varphi) - \langle \lambda_h, \varphi \rangle_{\Gamma_2} + \langle u_h, \omega - \lambda_h \rangle_{\Gamma_2} + \delta \sum_{E \in \mathbb{E}_h} h_E \langle \lambda_h - \partial_n u_h, \omega - \partial_n \varphi \rangle_E \\ \geq (f, \varphi) + \langle g, \varphi \rangle_{\Gamma_2} + \langle \Psi_{\Gamma_2}, \omega - \lambda_h \rangle_{\Gamma_2} - \delta \sum_{E \in \mathbb{E}_h} h_E \langle g, \omega - \partial_n \varphi \rangle_E \end{aligned} \tag{7.16}$$

for all $\varphi \in V_h$ and $\omega \in \Lambda_h$, where h_E is the diameter of E. We set

$$B_h(v, \mu; z, \nu) = (f, z) + \langle g, z \rangle_{\Gamma_2} - \langle \Psi_{\Gamma_2}, \nu \rangle_{\Gamma_2} + \delta \sum_{E \in \mathbb{E}_h} h_E \langle g, \nu - \partial_n z \rangle_E$$

with

$$B_h(v, \mu; z, \nu) := a(v, z) - \langle \mu, z \rangle_{\Gamma_2} - \langle v, \nu \rangle_{\Gamma_2} - \delta \sum_{E \in \mathbb{E}_h} h_E \langle \mu - \partial_n v, \nu - \partial_n z \rangle_E$$

and introduce the following mesh dependent norms (for a detailed adoption see [Pi80]):

$$\|v\|^2_{\frac{1}{2},h} = \sum_{E \in \mathbb{E}_h} h_E^{-1} \|v\|^2_{0,E} \quad \forall v \in H^1(\Omega), \tag{7.17}$$

$$\|\mu\|^2_{-\frac{1}{2},h} = \sum_{E \in \mathbb{E}_h} h_E \|\mu\|^2_{0,E} \quad \forall \mu \in L_2(\Gamma). \tag{7.18}$$

Obviously there holds

$$\langle v, z \rangle \leq \|v\|_{\frac{1}{2},h} \|z\|_{-\frac{1}{2},h} \quad \forall (v, z) \in H^1(\Omega) \times L_2(\Gamma_2) \tag{7.19}$$

and we also define

$$\|v\|_{1,h} = \|v\|_1 + \|v\|_{\frac{1}{2},h} \quad \forall v \in H^1(\Omega).$$

Following [St94] there holds:

Lemma 7.1. *There exists a constant C_I such that*

$$C_I \|\frac{\partial v}{\partial n}\|_{-\frac{1}{2},h} \leq \|\nabla v\|_0 \quad \forall v \in V_h$$

where

$$V_h = \{v \in H^1(\Omega) |\, v|_T \in P_k(T) \,\forall T \in \mathbb{T}_h\}$$

with $P_k(T)$ denoting the polynomials of degree $k \geq 1$ on T.

We show stability by proving

Theorem 7.2. *Suppose that* $0 < \delta < C_I$ *with* C_I *taken from Lemma 7.1. Then there holds*

$$\sup_{(z,\nu)\in V_h\times U_h} \frac{B_h(v,\mu;z,\nu)}{\|z\|_{1,h}+\|\nu\|_{-\frac{1}{2},h}} \geq C(\|v\|_{1,h}+\|\mu\|_{-\frac{1}{2},h}). \tag{7.20}$$

The proof is analogue to the one in [St94] where Dirichlet boundary conditions were enforced with the help of a Lagrangian multiplier. U_h is the discretisation of $L_2(\Gamma)$.

Proof. Let $(v,\mu)\in V_h\times U_h$ be arbitrary. With (7.17), (7.18) and Lemma 7.1 there holds

$$\begin{aligned}
B_h(v,\mu,v,-\mu) &= \|\nabla v\|_0^2 - \delta \sum_{E\in\mathbb{E}_h} h_E\langle \mu - \frac{\partial v}{\partial n}, -\mu - \frac{\partial v}{\partial n}\rangle_E \\
&= \|\nabla v\|_0^2 + \delta \sum_{E\in\mathbb{E}_h} h_E(\|\mu\|_{0,E}^2 - \|\frac{\partial v}{\partial n}\|_{0,E}^2) \\
&= \|\nabla v\|_0^2 + \delta \sum_{E\in\mathbb{E}_h} h_E\|\mu\|_{0,E}^2 - \delta\|\frac{\partial v}{\partial n}\|_{-\frac{1}{2},h}^2 \\
&\geq \left(1-\frac{\delta}{C_I}\right)\|\nabla v\|_0^2 + \delta \sum_{E\in\mathbb{E}_h} h_E\|\mu\|_{0,E}^2 \\
&\geq C_1(\|\nabla v\|_0^2 + \|\mu\|_{-\frac{1}{2},h}^2)
\end{aligned}$$

with the assumption $0<\delta<C_I$. Let $\Pi_h : L_2(\Gamma_2)\to U_h$ be the L_2-projection. Since the functions of U_h are discontinuous, we can define $\tilde{\mu}\in U_h$ by $\tilde{\mu}|_E = -h_E^{-1}\Pi_h v|_E$ for all $E\in\mathbb{E}_h$. Then we have

$$\|\tilde{\mu}\|_{-\frac{1}{2},h} = \|\Pi_h v\|_{\frac{1}{2},h}. \tag{7.21}$$

Using (7.19) and Lemma 7.1 we receive

$$\begin{aligned}
B_h(v,\mu;0,\tilde{\mu}) &= -\langle v,\tilde{\mu}\rangle_{\Gamma_2} - \delta \sum_{E\in\mathbb{E}_h} h_E\langle \mu - \frac{\partial v}{\partial n}, \tilde{\mu}\rangle_E \\
&= \|\Pi_h v\|_{\frac{1}{2},h}^2 - \delta \sum_{E\in\mathbb{E}_h} h_E\langle \frac{\partial v}{\partial n} - \mu, -\tilde{\mu}\rangle_E \\
&\geq \|\Pi_h v\|_{\frac{1}{2},h}^2 - \delta(\|\frac{\partial v}{\partial n}\|_{-\frac{1}{2},h} + \|\mu\|_{-\frac{1}{2},h})\|\Pi_h v\|_{\frac{1}{2},h} \\
&\geq \|\Pi_h v\|_{\frac{1}{2},h}^2 - C_2(\|\nabla v\|_0 + \|\mu\|_{-\frac{1}{2},h})\|\Pi_h v\|_{\frac{1}{2},h}.
\end{aligned}$$

Now using Young's inequality we achieve

$$\begin{aligned} B_h(v,\mu;0,\tilde{\mu}) &\geq \|\Pi_h v\|^2_{\frac{1}{2},h} - \frac{1}{2}C_2^2(\|\nabla v\|_0 + \|\mu\|_{-\frac{1}{2},h})^2 - \frac{1}{2}\|\Pi_h v\|^2_{\frac{1}{2},h} \\ &\geq \frac{1}{2}\|\Pi_h v\|^2_{\frac{1}{2},h} - C_3(\|\nabla v\|_0^2 + \|\mu\|^2_{-\frac{1}{2},h}). \end{aligned}$$

Combining these results by setting

$$(z,\nu) = (v, -\mu + \alpha\tilde{\mu}), \quad \alpha > 0$$

we get

$$\begin{aligned} B_h(v,\mu;z,\nu) &= B_h(v,\mu;v,-\mu) + \alpha B_h(v,\mu;0,\tilde{\mu}) \\ &\geq (C_1 - \alpha C_3)\|\nabla v\|_0^2 + \frac{1}{2}\alpha\|\Pi_h v\|^2_{\frac{1}{2},h} + (C_1 - \alpha C_3)\|\mu\|^2_{-\frac{1}{2},h} \\ &\geq C(\|\nabla v\|_0^2 + \|\Pi_h v\|^2_{\frac{1}{2},h} + \|\mu\|^2_{-\frac{1}{2}}) \end{aligned}$$

when choosing $\alpha < C_1/C_3$. One can prove by scaling that

$$\|\nabla v\|_0^2 + \|\Pi_h v\|^2_{\frac{1}{2},h} \geq C\|v\|^2_{1,h},$$

and with (7.21) we have

$$\|z\|_{1,h} + \|\nu\|_{-\frac{1}{2},h} \leq C(\|v\|_{1,h} + \|\mu\|_{-\frac{1}{2},h})$$

and the estimation (7.20) holds.

□

An alternative way of stabilisation is possible by using the jumps of the elementwise constant functions λ_h:
Find $(u_h,\lambda_h) \in V_h \times \Lambda_h$:

$$\begin{array}{lll} a(u_h,\varphi) & -\langle\lambda_h,\varphi\rangle_{\Gamma_2} & = (f,\varphi) + \langle g,\varphi\rangle_{\Gamma_2} \\ -\langle u_h, \omega - \lambda_h\rangle_{\Gamma_2} & -\delta\sum_{E\in\mathbb{E}_h} h_E\langle[\lambda_h],[\omega]\rangle_E & \leq -\langle\Psi_{\Gamma_2}, \omega-\lambda_h\rangle_{\Gamma_2} \end{array} \qquad (7.22)$$

for all $(\varphi, \omega) \in V_h \times \Lambda_h$ and a new $\delta > 0$. That is still a consistent way of stabilisation because these jumps vanish in the continuous case. We define

$$\begin{aligned} A_\delta((u_h, \lambda_h), (\varphi, \omega)) &:= a(u_h, \varphi) - \langle \lambda_h, \varphi \rangle_{\Gamma_2} \\ &\qquad + \langle u_h, \omega \rangle_{\Gamma_2} + \sum_{E \in \mathbb{E}_h} \overline{\delta}_E \langle [\lambda_h], [\omega] \rangle_E \\ F_\delta(\varphi, \omega) &:= (f, \varphi) + \langle g, \varphi \rangle_{\Gamma_2} + \langle \Psi_{\Gamma_2}, \omega \rangle_{\Gamma_2} \end{aligned}$$

with

$$A_\delta((u_h, \lambda_h), (\varphi, \omega)) = F_\delta(\varphi, \omega) \quad \forall (\varphi, \omega) \in V_h \times U_h \tag{7.23}$$

and the natural (semi-)norm in order to deal with (7.23) is defined by

$$|||(u_h, \lambda_h)|||_\delta^2 = \|\nabla u_h\|_0^2 + \|\overline{\delta}^{\frac{1}{2}} [\lambda_h]\|_{\partial \Gamma_2}^2,$$

with $\overline{\delta}$ being a piecewise constant parameter function fulfilling $\overline{\delta}_E \sim h_E$. It is clear that A_δ is positive definite:

$$A_\delta(\{\varphi, \omega\}, \{\varphi, \omega\}) \geq c|||\{\varphi, \omega\}|||_\delta^2, \quad 0 < c \leq 1,$$

and the solvability of the mixed problem and uniqueness of u_h is ensured.

7.2.2 A posteriori error analysis

We construct a consistent error estimator in case of the least squares stabilisation in (7.16):

Find a pair $(u_h, \lambda_h) \in V_h \times \Lambda_h$ fulfilling

$$\begin{aligned} a(u_h, \varphi) + \delta \textstyle\sum_E h_E \langle \partial_n u_h, \partial_n \varphi \rangle_E &- \langle \lambda_h, \varphi \rangle_{\Gamma_2} - \delta \textstyle\sum_E h_E \langle \lambda_h, \partial_n \varphi \rangle_E \\ &= (f, \varphi) + \langle g, \varphi \rangle_{\Gamma_2} + \delta \textstyle\sum_E h_E \langle g, \partial_n \varphi \rangle_E \\ \langle u_h, \omega - \lambda_h \rangle_{\Gamma_2} - \delta \textstyle\sum_E h_E \langle \partial_n u_h, \omega \rangle_E &+ \delta \textstyle\sum_E h_E \langle \lambda_h, \omega \rangle_E \\ &\geq \langle \Psi_{\Gamma_2}, \omega - \lambda_h \rangle_{\Gamma_2} - \delta \textstyle\sum_E h_E \langle g, \omega \rangle_E. \end{aligned} \tag{7.24}$$

We start estimating $(\nabla e, \nabla e_i)$. By (7.4) we receive

$$\begin{aligned}(\nabla e, \nabla e_i) &= (f, e_i) - (\nabla u_h, \nabla e_i) + (\nabla u, \nabla(e_i - e)) - (f, e_i - e) + (\nabla u, \nabla e) - (f, e)\\ &\leq (f, e_i) - (\nabla u_h, \nabla e_i) + (\nabla u, \nabla(e_i - e)) - (f, e_i - e) + \langle g, e\rangle_{\Gamma_2}.\end{aligned}$$

Using the first equation of (7.24) there holds:

$$\begin{aligned}(\nabla e, \nabla e_i) &\leq -\langle g, e_i\rangle_{\Gamma_2} + (\nabla u, \nabla(e_i - e)) - (f, e_i - e) + \langle g, e\rangle_{\Gamma_2}\\ &\quad - \langle \lambda_h, e_i\rangle_{\Gamma_2} + \delta \sum_E h_E \langle \partial_n u_h - \lambda_h - g, \partial_n e_i\rangle_E\\ &= \langle g, e - e_i\rangle_{\Gamma_2} + (\nabla u, \nabla(e_i - e)) - (f, e_i - e)\\ &\quad + \langle \lambda_h, e - e_i\rangle_{\Gamma_2} - \langle \lambda_h, e\rangle_{\Gamma_2} + \delta \sum_E h_E \langle \partial_n u_h - \lambda_h - g, \partial_n e_i\rangle_E.\end{aligned}$$

The last terms can be further estimated:

$$\begin{aligned}&-\langle \lambda_h,\, e\rangle_{\Gamma_2} + \delta \sum_E h_E \langle \partial_n u_h - \lambda_h - g, \partial_n e_i\rangle_E\\ &\overset{\text{Young}}{\leq} \langle \lambda_h, u_h - u - \Psi_{\Gamma_2} + \Psi_{\Gamma_2}\rangle_{\Gamma_2}\\ &\qquad + \frac{1}{2\varepsilon}\delta \sum_E h_E \|\partial_n u_h - \lambda_h - g\|_E^2 + \frac{\varepsilon}{2}\delta \sum_E h_E \|\partial_n e_i\|_E^2\\ &\overset{(7.18)}{=} \underbrace{\langle \lambda_h, \Psi_{\Gamma_2} - u\rangle_{\Gamma_2}}_{\leq 0} + \langle \lambda_h, u_h - \Psi_{\Gamma_2}\rangle_{\Gamma_2}\\ &\qquad + \frac{1}{2\varepsilon}\delta \sum_E h_E \|\partial_n u_h - \lambda_h - g\|_E^2 + \frac{\varepsilon}{2}\delta \|\partial_n e_i\|_{-\frac{1}{2},h}^2\\ &\overset{\text{Lemma 7.1}}{\leq} \langle \lambda_h, u_h - \Psi_{\Gamma_2}\rangle_{\Gamma_2}\\ &\qquad + \frac{1}{2\varepsilon}\delta \sum_E h_E \|\partial_n u_h - \lambda_h - g\|_E^2 + \frac{\varepsilon}{2}\delta C_I \|\nabla e_i\|^2\\ &\leq \langle \lambda_h, u_h - \Psi_{\Gamma_2}\rangle_{\Gamma_2}\\ &\qquad + \frac{1}{2\varepsilon}\delta \sum_E h_E \|\partial_n u_h - \lambda_h - g\|_E^2 + C_{II} \|\nabla e\|^2.\end{aligned}$$

Now, this result can be used for estimating the error in the energy norm:

$$\begin{aligned}
(\nabla e, \nabla e) &= (\nabla e, \nabla(e-e_i)) + (\nabla e, \nabla e_i)\\
&\leq (\nabla u, \nabla(e-e_i)) - (\nabla u_h, \nabla(e-e_i))\\
&\quad + \langle g, e-e_i\rangle_{\Gamma_2} + (\nabla u, \nabla(e_i - e)) - (f, e_i - e)\\
&\quad + \langle \lambda_h, e-e_i\rangle_{\Gamma_2} + \langle \lambda_h, u_h - \Psi_{\Gamma_2}\rangle_{\Gamma_2}\\
&\quad + \frac{1}{2\varepsilon}\delta \sum_E h_E \|\partial_n u_h - \lambda_h - g\|_E^2 + C_{II}\|\nabla e\|^2\\
&= -(\nabla u_h, \nabla(e-e_i)) + \langle g, e-e_i\rangle_{\Gamma_2} - (f, e_i - e)\\
&\quad + \langle \lambda_h, e-e_i\rangle_{\Gamma_2} + \langle \lambda_h, u_h - \Psi_{\Gamma_2}\rangle_{\Gamma_2}\\
&\quad + \frac{1}{2\varepsilon}\delta \sum_E h_E \|\partial_n u_h - \lambda_h - g\|_E^2 + C_{II}\|\nabla e\|^2.
\end{aligned}$$

We sum over the cells and integrate by parts:

$$\begin{aligned}
(\nabla e, \nabla e) &\leq \sum_{T\in\mathbb{T}_h} \left[(\Delta u_h + f, e-e_i)_T - \frac{1}{2}\int_{\partial T\setminus\Gamma} n\cdot[\nabla u_h](e-e_i)\partial T \right]\\
&\quad - \langle \partial_n u_h, e-e_i\rangle_{\Gamma_2} + \langle g, e-e_i\rangle_{\Gamma_2} + \langle \lambda_h, e-e_i\rangle_{\Gamma_2}\\
&\quad + \langle \lambda_h, u_h - \Psi_{\Gamma_2}\rangle_{\Gamma_2} + \frac{1}{2\varepsilon}\delta\sum_E h_E\|\partial_n u_h - \lambda_h - g\|_E^2 + C_{II}\|\nabla e\|^2\\
&\leq \sum_{T\in\mathbb{T}_h}\left[\|\Delta u_h + f\|_T \|e-e_i\|_T + \frac{1}{2}\|n\cdot[\nabla u_h]\|_{\partial T}\|e-e_i\|_{\partial T}\right]\\
&\quad + \|g - \partial_n u_h + \lambda_h\|_{\Gamma_2}\|e-e_i\|_{\Gamma_2}\\
&\quad + \langle \lambda_h, u_h - \Psi_{\Gamma_2}\rangle_{\Gamma_2} + \frac{1}{2\varepsilon}\delta\sum_E h_E\|\partial_n u_h - \lambda_h - g\|_E^2 + C_{II}\|\nabla e\|^2\\
&\leq \sum_{T\in\mathbb{T}_h} \varrho_T\omega_T + \|g - \partial_n u_h + \lambda_h\|_{\Gamma_2}\|e-e_i\|_{\Gamma_2}\\
&\quad + \langle \lambda_h, u_h - \Psi_{\Gamma_2}\rangle_{\Gamma_2} + \frac{1}{2\varepsilon}\delta\sum_E h_E\|\partial_n u_h - \lambda_h - g\|_E^2 + C_{II}\|\nabla e\|^2
\end{aligned}$$

with

$$\begin{aligned}
\varrho_T &:= h_T||f + \Delta u_h||_T + \frac{1}{2}h_T^{\frac{1}{2}}||n\cdot[\nabla u_h]\,||_{\partial T},\\
\omega_T &:= \max\{h_T^{-1}||e-e_i||_T, h_T^{-\frac{1}{2}}||e-e_i||_{\partial T}\}.
\end{aligned}$$

Estimation (3.5) and Young's inequality lead to

$$
\begin{aligned}
(\nabla e, \nabla e) \leq \sum_{T\in\mathbb{T}_h} & \Big[\varrho_T \|\nabla e\|_{\tilde{\omega}(T)} + C h_T^{\frac{1}{2}} \|g - \partial_n u_h + \lambda_h\|_E \|\nabla e\|_{\tilde{\omega}(T|_{\Gamma_2})} \Big] \\
& + \langle \lambda_h, u_h - \Psi_{\Gamma_2} \rangle_{\Gamma_2} + \frac{1}{2\varepsilon}\delta \sum_E h_E \|\partial_n u_h - \lambda_h - g\|_E^2 + C_{II}\|\nabla e\|^2 \\
\leq \sum_{T\in\mathbb{T}_h} & \Big[\frac{1}{2\varepsilon_1}\varrho_T^2 + \frac{\varepsilon_1}{2}\|\nabla e\|_{\tilde{\omega}(T)}^2 \\
& \quad + C\frac{1}{2\varepsilon_2} h_T \; \|g - \partial_n u_h + \lambda_h\|_E^2 + \frac{\varepsilon_2}{2}\|\nabla e\|_{\tilde{\omega}(T|_{\Gamma_2})}^2 \Big] \\
& + \langle \lambda_h, u_h - \Psi_{\Gamma_2} \rangle_{\Gamma_2} + \frac{1}{2\varepsilon}\delta \sum_E h_E \|\partial_n u_h - \lambda_h - g\|_E^2 + C_{II}\|\nabla e\|^2.
\end{aligned}
$$

So all in all we get the error estimation of the stabilised simplified Signorini problem

$$
\begin{aligned}
(\nabla e, \nabla e) \leq & C \sum_{T\in\mathbb{T}_h} \Big[\varrho_T^2 + h_T(1+\delta)\|g - \partial_n u_h + \lambda_h\|_E^2 \Big] \\
& + \langle \lambda_h, u_h - \Psi_{\Gamma_2} \rangle_{\Gamma_2}.
\end{aligned}
$$

As a result we preserve the following theorem:

Theorem 7.3. *For the simplified Signorini problem with a least squares stabilisation there holds the error bound*

$$
|e|_1^2 \leq C \left(\sum_{T\in\mathbb{T}_h} \eta_T^2 + \langle \lambda_h, u_h - \Psi_{\Gamma_2} \rangle_{\Gamma_2} \right) \tag{7.25}
$$

with local residuals

$$
\eta_T = \begin{cases} h_T\|\Delta u_h + f\|_T + \frac{1}{2} h_T^{\frac{1}{2}} \|n \cdot [\nabla u_h]\|_{\partial T} & \text{if } \partial T \in \varepsilon_0 \\ h_T\|\Delta u_h + f\|_T + h_T^{\frac{1}{2}}(1+\delta)\|g - \partial_n u_h + \lambda_h\|_E & \text{if } \partial T \subset \Gamma_2, \end{cases}
$$

where $\varepsilon_0 \subset \mathbb{E}_h$ denotes the interior boundaries of a cell and $E \subset \Gamma_2$ is the outer face of a boundary cell with $\partial T \subset \Gamma_2$.

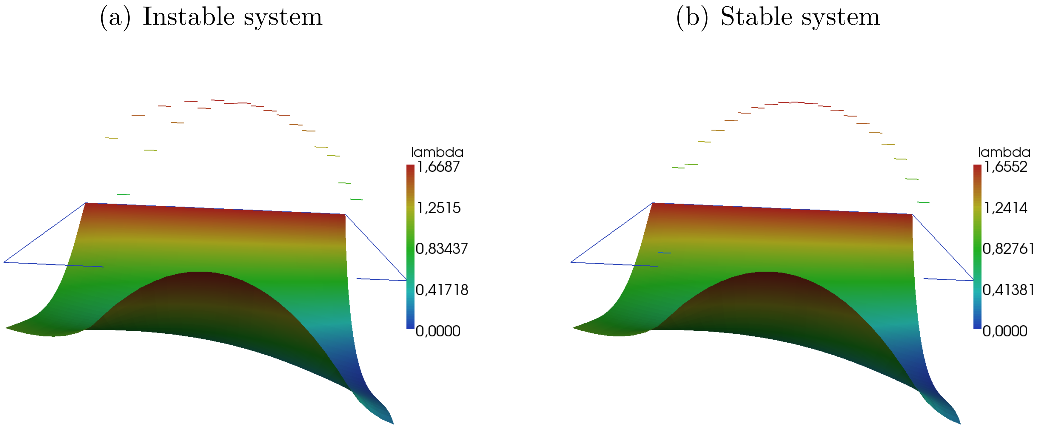
7.3 Numerical results

We give some numerical results considering stability and adaptivity by utilising different test examples. Conserning stability there are similar results given in [Sc05] who uses another stabilisation technique, so we discuss both strategies briefly.

7.3.1 Stability

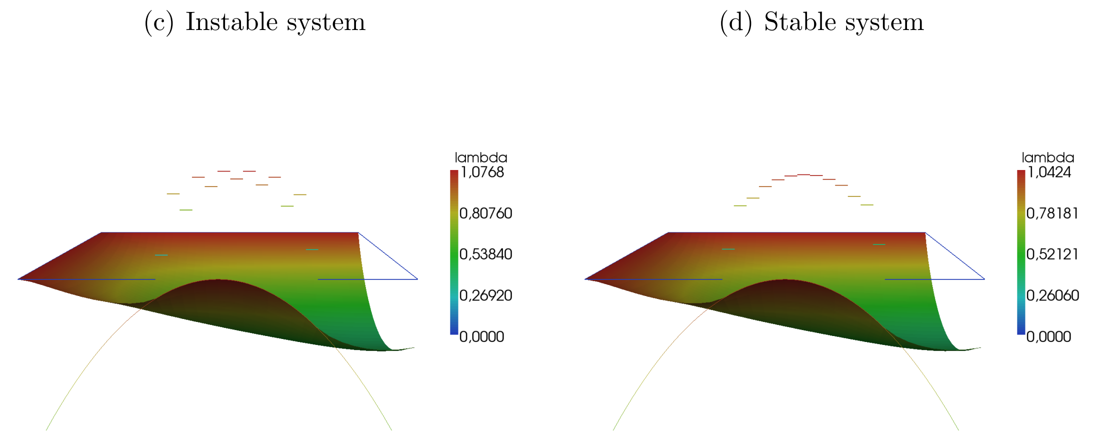
We choose a domain $\Omega = [-1, 1]^2$ and a constant body force $f = -1$. In Figure 7.1 the membrane is fixed at one side and the others are restricted to the Neumann condition $g = -0.25y^3$. The obstacle is set to $\Psi_{\Gamma_2} = -y^2$ if $x = 1$ and $\Psi_{\Gamma_2} = -10$ elsewhere.

In Figure 7.2 the membrane is fixed at two sides and the other sides are restricted to the same Neumann condition as above. As we can see in Firgure 7.1(a) and 7.2(c) the unstabilised systems show oscillations in the Lagrangian parameter. We receive stabilisation by using system (7.16). Figure 7.1(b) and Figure 7.2(d) offer smooth values.


(a) Instable system

(b) Stable system

Figure 7.1: (a): On one side of the area where $\lambda_h \neq 0$ there are oscillations of the Lagrangian multiplier. (b): Using the least squares stabilisation (7.16) we avoid oscillations.



(c) Instable system

(d) Stable system

Figure 7.2: (c): Instability of the Lagrangian multiplier. (d): There are no oscillations of the Lagrangian multiplier by using the least square stabilisation (7.16).

Using stabilisation (7.22) for the example of Figure 7.2 leads to the same results. The values of λ_h agree with the ones of Figure 7.2 (d).

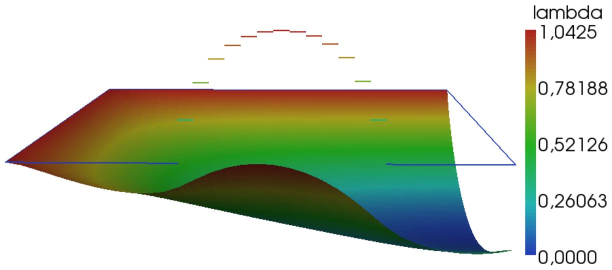


Figure 7.3: Stabilisation with the help of (7.22). Stabilisation by jump terms has the same effect as the one using the consistent terms in (7.16).

We compare these results to A. Schröder's work [Sc05], who used a different kind of stabilisation. There, the Lagrangian multiplier lives on a coarser mesh than the primal variable. That eliminates the oscillations, too, and the stabilised figures look the same. However, both variants have their advantages. Comparing the complexity of programming, the method that is presented here is the one demanding less effort. We only have to add another term in the assembling matrix. Using the method described in [Sc05], calculations have to be carried out on different meshes, especially a patch for the Lagrangian parameter's mesh must be generated which is difficult when using a mesh generator. So we have to handle different dimensions of finite elements as well as different meshes. The resolution of the contact situation is connected to the patch mesh and not as accurate as it could be in view of the mesh for the displacement. However, the creation of the patch is easy if hierarchic meshes are used which is mostly the case in adaptive finite elements. The stabilisation approach also can be used for higher order finite elements. For further information see [Sc05]. Using the least squares method the solving algorithm has to deal with an additional term and another matrix-vector-multiplication that costs time. Furthermore, we get a constant parameter δ which has to be determined. For convenience the method runs very robust even on irregular meshes.

7.3.2 Adaptivity

If we perform some global refinement steps we observe the estimator goes with the optimal convergence rate of $\mathcal{O}(h)$ which is proven by Table 7.1. As a further test case we choose the obstacle

$$\Psi_{\Gamma_2} = \begin{cases} -x^2, \text{ if } y = -1 \\ -y^2, \text{ if } x = 1 \\ -10, \text{ elsewhere} \end{cases}$$

with the constant body force $f = -1$ and no further boundary conditions.

# cells	e_{res}		e_{jump}		N		$\lvert e\rvert_1$	
64	5.000e-01	-	3.267e-01	-	1.200e-01	-	8.038e-01	-
256	2.500e-01	1.00	1.776e-01	0.88	5.225e-02	1.20	4.151e-01	0.96
1024	1.250e-01	1.00	9.069e-02	0.97	2.326e-02	1.17	2.097e-01	0.99
4096	6.250e-02	1.00	4.620e-02	0.97	1.101e-02	1.08	1.055e-01	0.99
16384	3.125e-02	1.00	2.337e-02	0.99	5.360e-03	1.04	5.298e-02	0.99
65536	1.562e-02	1.00	1.183e-02	0.99	2.647e-03	1.02	2.667e-02	0.99

Table 7.1: Convergence rate of the estimator and its components. We set $e_{\text{res}}^2 = \sum_T h_T^2 \|\Delta u_h + f\|_T^2$, $N = \langle \lambda_h, u_h - \Psi_{\Gamma_2} \rangle_{\Gamma_2}$, $e_{\text{jump}}^2 = \sum_T \frac{1}{4} h_T \|n \cdot [\nabla u_h]\|_{\partial T}^2$ if $\partial T \in \varepsilon_0$ and $e_{\text{jump}}^2 = \sum_T h_T \|g - \partial_n u_h + \lambda_h\|_{T|_{\Gamma_2}}^2$ if $\partial T \in \Gamma_2$. $|e|_1$ describes the whole estimator (7.25). In the right columns there is always a value α determining the convergence order by $\mathcal{O}(h^\alpha)$ for every refinement step.

When we refine the domain $\Omega = [-1, 1]^2$ adaptively we observe very fine mesh structures at the contact boundaries when using estimator (7.8) which is very uneconomical by the same argument as described in Section 3.4.2. The over-refinement can be seen in Figure 7.4, left column, or in Figure 7.5, first row, which is a zoom to the red rectangle in Figure 7.4 for every refinement step. In contrast, with the help of the counter force performed by the Lagrangian multiplier in estimator (7.25) we avoid such uneconomical mesh structures which is illustrated in the right column of Figure 7.4 or the second row of Figure 7.5. Furthermore, calculations happen faster here because the boundaries where λ_h exists are not so well-refined. That means less work for Uzawa's algorithm which runs slowly as we exhibited in Section 3.2.3.

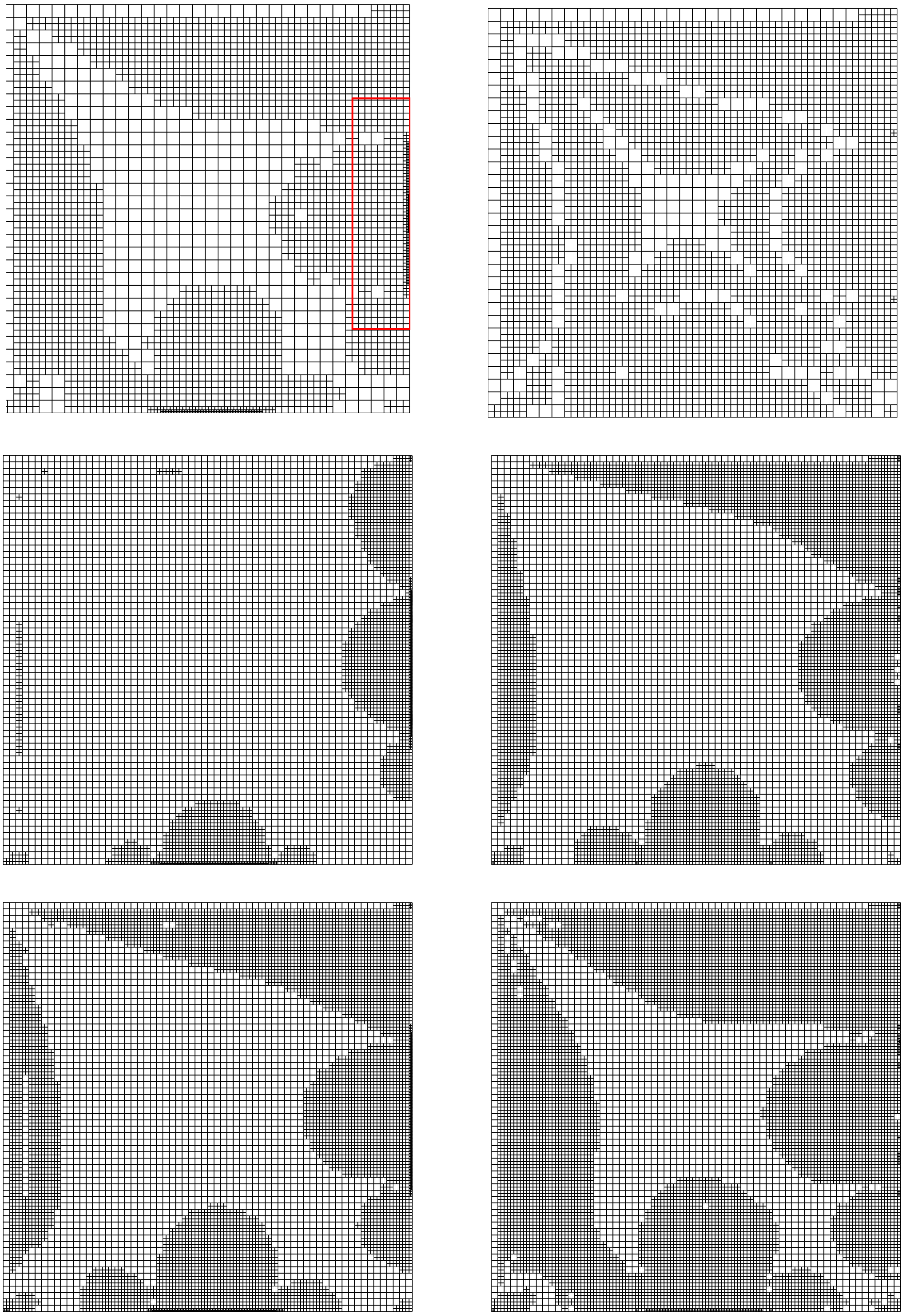


Figure 7.4: Left column: A sequence of grids created by estimator (7.8). The parts of the boundary where the obstacle exists are always very well refined. Right column: A sequence of grids created by estimator (7.25). The Lagrangian multiplier prevents an over-refinement at the boundaries.

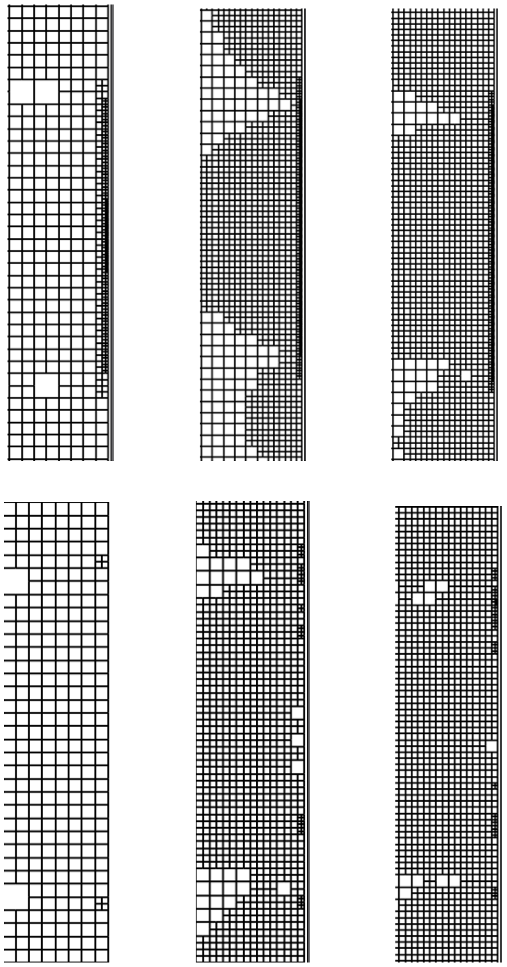


Figure 7.5: First row: Zoom to the contact boundary in the red box of Figure 7.4 for estimator (7.8) (left column in Figure 7.4) which refines the contact zone very well. Second row: Zoom to the contact zone of estimator (7.25) (right column in Figure 7.4). There is less refinement at the boundaries due to the consistency of the estimator.

The consistency of the improved estimator is outlined in Figure 7.6. We compare the terms $(\sum_E \|g - \partial_n u_h + \lambda_h\|_E^2)^{\frac{1}{2}}$ and $(\sum_E \|g - \partial_n u_h\|_E^2)^{\frac{1}{2}}$ in areas of contact. The term without λ_h stays almost constant with every refinement step whereas the other one falls continuously. Again, λ_h is to be understood as a counter force that fills the gap $|g - \partial_n u_h|$ where the restriction $\partial_n u_h = g$ cannot be fulfilled due to the obstacle.

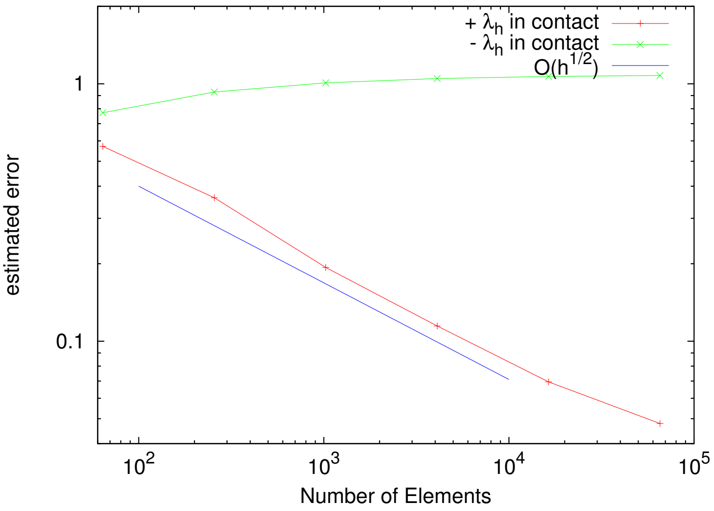


Figure 7.6: Comparison of the terms $(\sum_E \|g - \partial_n u_h + \lambda_h\|_E^2)^{\frac{1}{2}}$ and $(\sum_E \|g - \partial_n u_h\|_E^2)^{\frac{1}{2}}$ in areas of contact. The consistency of the improved estimator can be seen clearly.

Comparing the whole estimators (7.8) and (7.25), we get a better convergence rate for the improved one (see Figure 7.7).

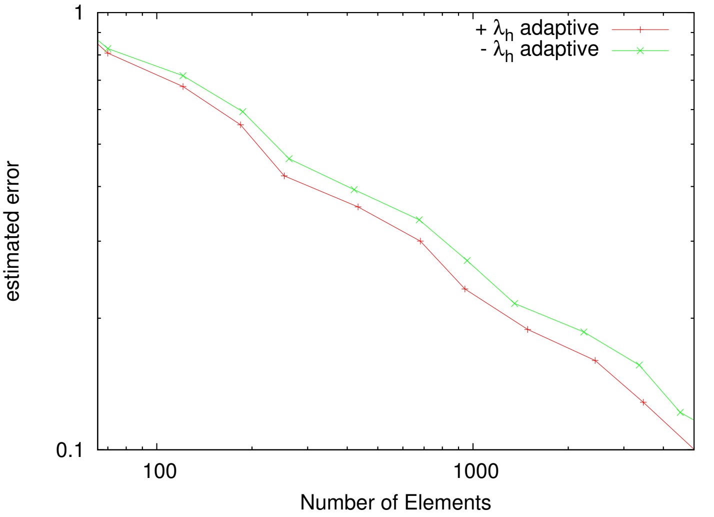


Figure 7.7: Comparison of the estimators (7.8) and (7.25). The one including the Lagrangian parameter offers a better convergence rate within adaptive refinement.

In order to show the effectivity of adaptive refinement using estimator (7.25) we choose a discontinuous obstacle at the free boundary $y = -1$ of Ω which is set by

$$\Psi = \begin{cases} -0.3 & \text{if } 0 < x < 0.3 \\ -0.2 & \text{if } 0.2 < x \leq 0 \;\vee\; 0.3 \leq x < 0.5 \\ -0.1 & \text{elsewhere on y=-1} \end{cases}$$

and a body force $f = 5y$ if $y < 0$ and $f = 0$ elsewhere.

#cells	$\lvert e\rvert_{1,\text{global}}$		#cells	$\lvert e\rvert_{1,\text{adaptive}}$	
64	1.571e+00	-	46	1.458e+00	-
256	8.394e-01	0.90	157	7.798e-01	1.02
1024	4.750e-01	0.82	565	4.228e-01	0.96
4096	2.937e-01	0.69	1888	2.411e-01	0.93
16384	1.720e-01	0.77	5923	1.274e-01	1.11
65536	1.065e-01	0.69	26153	6.301e-02	0.95

Table 7.2: Table of convergence of the error estimator using a discontinuous obstacle within global and adaptive refinement. Using global mesh refinement the error converges with a smaller convergence rate.

Table 7.2 and Figure 7.9 confirm the better convergence rate of adaptive calculations.

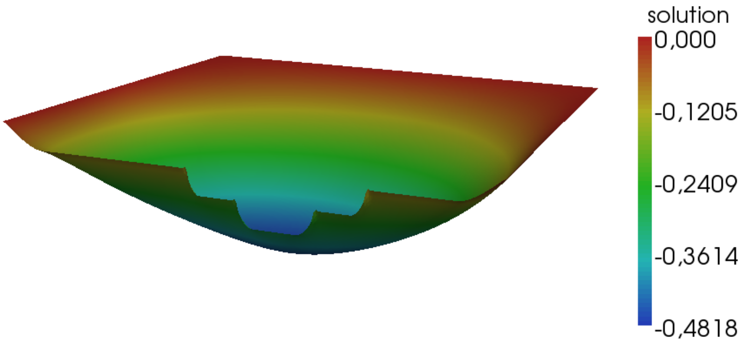


Figure 7.8: Solution of the test case for adaptive and global refinement.

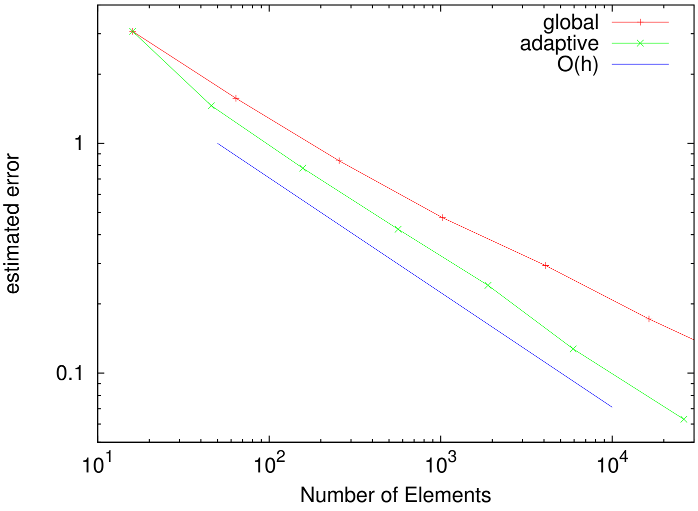
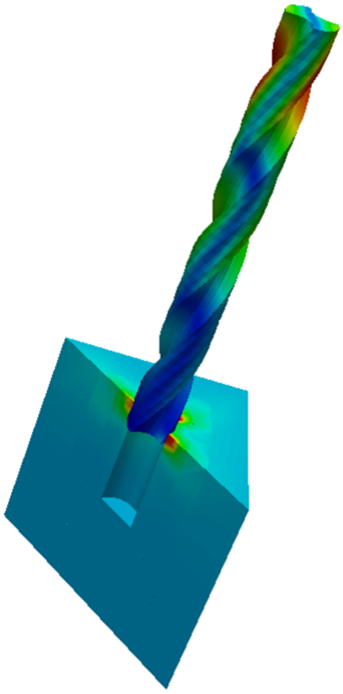


Figure 7.9: Adaptive mesh refinement by the help of estimator (7.25) leads to an almost optimal convergence order in contrast to global refinement steps when using a discontinuous obstacle.

8 Application: Deep drilling

The techniques introduced in this work have been used in context of a project supported by the Deutsche Förderungsgesellschaft within the program of emphasis "Modellierung, Simulation und Kompensation von thermischen Bearbeitungseinflüssen für komplexe Zerspanungsprozesse". In cooperation with the Institute of Mathematics (LSX) and the Institute of Chipping (ISF) in Dortmund we participated with the project "Numerische Analyse und effiziente Implementierung komplexer FE-Modelle maschineller Fertigungsprozesse am Beispiel des Tiefbohrens" (BL 256/11-1, SU 245/6-1). The aim of our work group is to develop a long-time process simulation of deep drilling with low lubrication using high performance computing methods. The main focus is put on the heat flux in the workpiece. The high cutting velocity and feed rate which can be achieved, result in a high heat generation in the working zone with enormous thermical load which can cause unwanted deformations. The achievement of this project should be a better understanding of the processes in the workpiece caused by heat transfer and to develop strategies to compensate unwanted effects. In a first phase of the project the ISF did some experimental efforts to offer input values which we used for modeling a short-time process simulation of the static tempera-



ture disposition in the workpiece. The figure on the right is taken from Dr. Heiko Kleemann, TU Dortmund, showing a first simulation of the drilling process. With the resulting data from this simulation the ISF will generate a long-time process simulation later on. For an efficient simulation there are many numerical aspects to be analysed and improved such as contact algorithms, nonlinear material laws at finite strains, thermoelastic and thermoplastic aspects, friction, heat generation and heat flux as well as mesh-generation. Furthermore, we have a complex underlying geometry which makes it necessary, in view of saving calculating time by keeping accuracy, to use adaptive meshes based on appropriate error indicators. Since the used contact algorithms are based on mixed formulations, the here developed techniques of consistent error indicators form a solid basis. Many principle aspects like contact, nonlinear material laws and torsion are already examined. In further phases of the project they will have to be coupled and further improved. In spite of this, instabilisations coming along with the discretisation of the mixed problems are eliminated by the least squares techniques presented here. So oscillations of the dual variable which affect the physical relevance are resolved and a calculation on robust systems is possible.

In the first phase, which is almost finished, a couple of benchmarks were made to improve the algorithms of the simulation code. As an example, in order to find effective methods to solve the contact algorithms, there was made one of several solvers on a linear elastic Signorini problem in two dimensions with E-module $E = 6.2 \cdot 10^5$ and $\nu = 0.22$. The obstacle that was pressed on the boundary had the geometry of the drill bit used in the experiments. The resulting cg-iterations of every solver are shown in Figure 8.2. The tests were made in cooperation with J. Frohne and A. Rademacher.

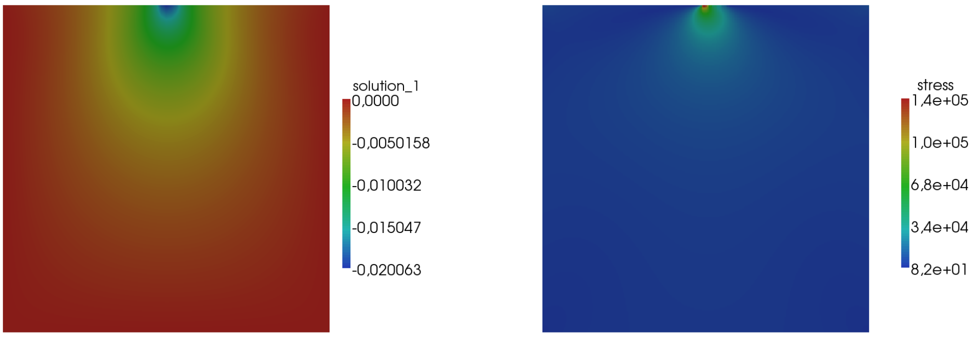


Figure 8.1: Signorini problem for a model problem based on plain strain and an obstacle which represents the drill bit. Left: displacement in y-direction. Right: Von Mises stress.

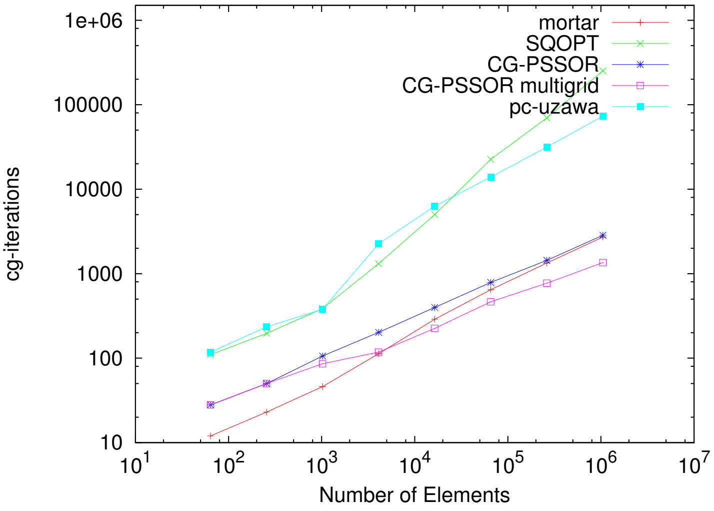


Figure 8.2: Benchmark of several solvers for the Signorini problem shown in Figure 8.1. "mortar" is a primal-dual active set method, developed by Wohlmuth [HW09], SQOPT solves the Schur complement, CG-PSSOR denotes the projected cg-solver including an SSOR-step (see Section 3.3), whereas CG-PSSOR multigrid is the same solver preconditioned by a multigrid method and the last one is the preconditioned Uzawa method (see Section 3.3).

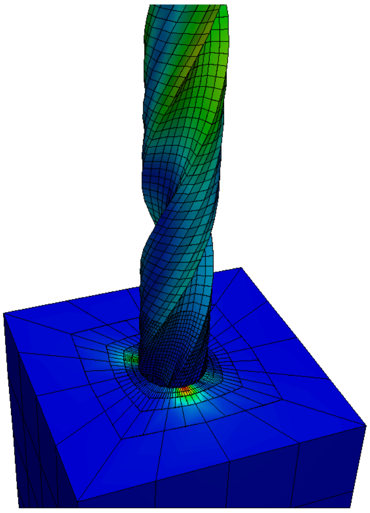


Figure 8.3: Primary calculations of the drilling process by Dr. Heiko Kleemann, TU Dortmund.

9 Conclusion and outlook

The subject matter deals with mixed formulations of restricted problems which are also called saddle point problems. They evolve from variational inequalities by using the Lagrangian formalism. Our intention was to develop consistent error estimators, giving economical mesh structures by adaptive refinements. The main component here is the Lagrangian parameter which eliminates inconsistent parts of the estimator and helps to generate meshes that outline critical zones. By discretising these formulations with the help of the Finite Element Method, instabilities may appear, which often have negative effects concerning the dual variable. Our studies are based on principle problems that often show such effects. So a second aim of the work was to stabilise these systems with the help of the least squares method to have optimal conditions for studying efficient error controls. We analysed existence and uniqueness of the variational formulation as well as the mixed system of several basic problems, stabilised the system if necessary and developed a consistent a posteriori error estimator. Numerical tests in each chapter, including different possibilities restrictions may occur, show that it was succeeded to stabilise systems in a consistent way and also generate economical meshes. More precisely, we eliminated unwanted oscillations in the Lagrangian multiplier by using additional stabilisation terms. The positive effect is that the dual variable retrieves its physical relevance and solvers work more reliable. The new error estimators, which take account of the Lagrangian variable as well as the stabilisation terms, are able to outline critical zones which typically are transition areas. In addition to this we examined different solving algorithms to get some (faster) alternatives to standard Uzawa's method and compared them in a benchmark. Furthermore, we gave a small insight in the possible development of DWR-estimators with respect to the Lagrangian multiplier, using the example of a torsion problem. At last, we presented a current project sponsored by the DFG dealing with development and spreading of heat flux in a workpiece in case of deep drilling. For a fast and efficient calculation the estimators at hand give useful criteria for mesh refinement in this application.

Since we have studied basic problems separately the next assignment is to combine the estimators. Taking for example the application of deep drilling we have several components, plastification, contact and torsion, which have to be coupled. Another open task is to find an a posteriori estimator concerning $\|\lambda - \lambda_h\|$ to have control of the dual variable, too. Furthermore, an idea for further studies is to make more efforts relating to the DWR-Method because it is very important to engineers who are often interested in stresses or other values in a single point.

Concerning stabilisation, it would be interesting to compare the method of including additional bubble-functions to the one of least squares stabilisation. The resulting finite element is called MINI-Element in literature. Similar to the least squares terms they produce an additional term in the C-block. The advantage is that they also deliver the appropriate order of the stabilisation parameter δ. So one relieves the fitting processes that we had to perform to find a suitable value for δ.

Bibliography

[AH09] K.Atkinson, W.Han: *Theoretical Numerical Analysis.*
Springer-Verlag, 2009.

[AO00] M.Ainsworth, J.T.Oden: *A posteriori error estimation in finite element analysis.*
Pure and Applied Mathematics. Wiley-Interscience, Chichester, 2000.

[BBS04] H.Blum, D.Braess, F.T.Suttmeier: *A Cascadic Multigrid Algorithm for Variational Inequalities.*
Computing and Visualization in Science, Volume 7, Issue 3-4 , pp 153-157, Springer-Verlag, 2004.

[Be95] R.Becker: *An adaptive Finite Element Method for the Incompressible Navier-Stokes Equations on Time-dependent Domains.*
Dissertation, 1995.

[BF91] F.Brezzi, M.Fortin: *Mixed and Hybrid Finite Element Methods.*
Springer-Verlag, 1991.

[BHR77] F.Brezzi, W.W.Hager, P.A.Raviart: *Error estimates for the finite element solution of variational inequalities.*
Num. Math., 28:431-443, 1977.

[Br07] D.Braess: *Finite Elemente.*
Springer-Verlag 2007.

[Br72] H.Brézis: *Problèmes unilatéraux.*
J.Math. Pures Appl., 1972.

[Ca03] Z.H.Cao: *Fast Uzawa algorithm for generalized saddle point problems.*
Applied Numerical Mathematics 46:157-171, 2003.

[Ce78] J.Cea: *Optimization-Theory and algorithms.*
Springer-Verlag, 1978.

[Cu02] M.Ciu: *A sufficient condition for the convergence of the inexact Uzawa algorithm for saddle point problems.*
Journal of Computal and Applied Mathematics 139:189-196, 2002.

[DL76] G.Duvaut, J.L.Lions: *Inequalities in Mechanics and Physics.*
Springer-Verlag, 1976.

[Dr11] N.Dröge: *Stabile FE-Diskretisierungen für eine dual-gemischte Formulierung des Laplace-Problems.*
Masterthesis, Univeristy of Siegen, 2011.

[ET99] I.Ekland, R.Temam: *Convex Analysis and Variational Problems.*
Siam, In Applied Mathematics 28, 1999.

[Fa74] R.S.Falk: *Error estimates for the approximation of a class of variational inequalities.*
Math. Comp., 28:963-971, 1974.

[FS91] L.P.Franca, R.Stenberg: *Error analysis of some Galerkin least squares methods for the elasticity equations.*
SIAM J. Numer. Anal. 28, 1991.

[GHS10] F.Gimbel, P.Hansbo, F.T.Suttmeier: *An adaptive low-order FE-scheme for Stokes flow with cavitation.*
J. Numer. Math. Vol. 18, No.3, 177-185, 2010.

[Gi12] F.Gimbel: *Modelling and Numerical Simulation of Contact and Lubrication.*
Dissertation, Univeristy of Siegen, 2012.

[GK02] C.Geiger, Ch.Kanzow: *Theorie und Numerik restringierter Optimierungsaufgaben.*
Springer-Verlag, Berlin-Heidelberg, 2002.

[Gl83] R.Glowinski: *Numerical Methods for Nonlinear Variational Problems.*
Springer-Verlag, 1983.

[GLT76] R.Glowinski, J.L.Lions, R.Trémoliéres: *Numerical Analysis of Variational Inequalities.*
North-Holland, 1976.

[GR92] Ch.Grossmann, G.Roos: *Numerik partieller Differentialgleichungen.*
Springer-Verlag, 1983.

[GT97] Ch.Grossmann, J.Ternos: *Numerik der Optimierung.*
Vieweg + Teubner, 1997.

[Ha77] S.P.Han: *A Globally Convergent Method for Nonlinear Programming.*
Journal of Optimization Theory and Applications, 22:297-309, 1977.

[HFB86] T.J.R.Hughes, L.P.Franca, M.Balestra: *A new finite element formulation for computational fluid dynamics: V. Circumventing the Babusca-Brezzi condition: A stable Petrov-Galerkin formulation for the stokes problem accommodating equal order interpolation.*
Comp. Meth. mech. Eng. 59, 85-99, 1986.

[HKS11] D.Hage, N.Klein, F.T.Suttmeier: *Adaptive finite elements for a certain class of variational inequalities of second kind.*
Calcolo, Journal ID 10092, Article ID 40, 2011.

[HW09] S.Hüeber, B.I.Wohlmuth: *Thermo-mechanical contact problems on non-matching meshes.*
Comput. Methods Appl. Mech. Engrg., 198:1338-1350, 2009.

[JH91] C.Johnson, P.Hansbo: *Adaptive finite element methods for small strain elasto-plasticity.*

Technical report, Chalmers University of Technology, University Göteborg, 1991.

[Jo76/1] C.Johnson: *Existence theorems for plasticity problems.*
J.Math. Pures Appl., 55:431-444, 1976.

[Jo76/2] C.Johnson: *A mixed finite element method for plasticity problems with hardening.*
SIAM J. Numer. Anal. 14, 1976.

[Jo78] C.Johnson: *On Plasticity with Hardening.*
Journal of math Analysis and Applications, 62:325-336, 1978.

[Kl09] N.Klein: *Modelladaptive FE-Techniken bei Elastizitätsproblemen mit grossen Verformungen.*
Masterthesis, Univeristy of Siegen, 2009.

[KR00] M.Kunze, J.F.Rodrigues: *An elliptic quasi-variational inequality with gradient constraint and some of its applications.*
Math. Meth.Appl.Sci., 23:897-908, 2000.

[LP66] E.S.Levitin, B.T.Polyak: *Minimization methods under constraints.*
Zhurn. Vychisl. Matem. i Matem. Fiz. 6, 787-823, 1966.

[LS67] J.L.Lions, G.Stampacchia: *Variational inequalities.*
Commun. Pure Appl. Math. 20, 493-519, 1967.

[LS71] H.Lewy, G.Stampacchia: *On Existence and Smoothness of Solutions of Some Non-Coercive Variational Inequalities.*
Arch. Rat. Mech. Analysis, 1971.

[LS92] J.L.Lions, G.Stampacchia: *Variational Inequalities.*
Comm. Pure Appl. Math., 20:493-519, 1967.

[LY00] W.Lin, N.Yan: *A Posteriori Error Estimates for a Class of Variational Inequalities.*
Journal of Scientific Computing, Vol.15, No.3, 2000.

[MJ05] M.Jacob: *Simulation des Temperaturfeldes und Eigenspannung von einer MIG-Schweissung an einem Werkstück unter Berücksichtigung der temperaturabhängigen Materialeigenschaft von Aluminiumlegierungen.*
Diplomarbeit, Bauhaus-University Weimar, 2005.

[Pi80] J.Pitkäranta: *Local stability condition for the Babŭska method of lagrange multipiers.*
Math. Comput., 35:1113-1129, 1980.

[Sa03] Y.Saad: *Iterative Methods for Sparse Linear Systems.*
Society for Industrial and Applied Mathematics, 2003.

[Sc05] A.Schröder: *Fehlerkontrollierte adaptive h- und hp-Finite-Elemente-Methoden für Kontaktprobleme mit Anwendungen in der Fertigungstechnik.*
Dissertation, University of Dortmund, 2005.

[Se91] G.A.Seregin: *On the regularity of weak solutions of vaiational problems in plasticity theory.*
Soviet Math. Dokl., 42(2), 1991.

[St79] G.Strang: *A minimax problem in plasticity theory.*
Functional analysis methods in numerical analysis, Spec. Sess., AMS, St. Louis 1977, Lect. Notes Math. 701.319-333. Springer, 1979.

[St94] R.Stenberg: *On some techniques for approximating boundary conditions in the finite element method.*
Journal of Computational and Applied Mathematics 63, 139-148, 1994.

[Su08] F.T.Suttmeier: *Numerical solution of Variational Inequalities by Adaptive Finite Elements.*
Vieweg and Teubner, 2008.

[Su96] F.T.Suttmeier: *Adaptive Finite Element Approximation of Problems in Elasto-Plasticity Theory.*
Dissertation, Institute for Applied Mathematics, University of Heidelberg, 1996.

[Su10] F.T.Suttmeier: *Localised FE-Analysis of Strang's problem based on Lagrange techniques.*
J. Numer. Math., Vol. 18, No. 2, pp 135-141, 2010.

[Tr87] Troianiello: *Elliptic Differential Equations and Obstacle Problems.*
Plenum Press, 1987.

[Ve96] R.Verfürth: *A review of a posteriori error estimation and adaptive mesh-refinement techniques.*
Wiley-Teubner Series Advances in Numerical Mathematics. John Wiley and Sons, B.G. Teubner, Chichester, Stuttgart, 1996.

[Ze88] E.Zeidler: *Nonlinear Functional Analysis and its Applications. IV: Applications to Mathematical Physics.*
Springer-Verlag, New York, 1988.

[ZZ87] O.C.Zienkiewicz and J.Z.Zhu: *A simple error estimator and adaptive procedure for practical engineering analysis.*
Int. J. Numer. Methods Engrg., 24:337-357, 1987.