An Optimal Solver for a KKT-System arising from an Interior-Point Formulation of a Topology Optimization Problem

Roman Stainko *

April 12, 2006

Abstract

There are basically two approaches to solve optimal design problems with a partial differential equation, usually called the state equation, as a constraint. The usual procedure is to eliminate the state variables and the state equation and only optimize in the design space. Another possibility is to keep the state equation and to treat it as a constraint throughout the optimization progress. This formulation is called simultaneous or one-shot optimization. Then, in order to satisfy the optimality conditions, large scale indefinite linear systems (KKT-systems) have to be solved. This is the drawback, or better the challenge, of this alternative approach. If it is possible to construct an optimal solver to this KKT-systems, we can benefit from the expected speed-up of the one-shot formulation.

In this work we consider a multigrid based solver to such a KKT-system, resulting from stress constrained topology optimization. As a proper smoothing procedure we use a multiplicative Schwarz-type smoother. Here, in each iteration step of the smoother, several small local saddle-point problems are solved. The numerical test examples show the typical multigrid convergence behaviour, i.e. asymptotic constant number of iterations and convergence rates.

Keywords: KKT-system, Multigrid methods, Schwarz-type smoother.

AMS Subject Classification: 49M15, 49K20, 65F10, 65M55, 65N55.

1 Introduction

In this work we will consider a topology optimization problem with local stress constraints as a starting point for a derivation of a KKT-system. Treating topology optimization problems with local stress constraints usually results in a large scale optimization problem with a large number of constraints, e.g. two local stress constraints per finite element after discretization. Moreover, the design domain (i.e. the feasible set defined by the constraints) may be nonconvex (even nonconnected) and contain degenerated appendices with lower measure. (Global) optima are very likely to be located in this lower dimensional regions (cf. Rozvany [14]), where constraint qualifications are lacking. In literature, this effect is often called the *singularity* problem. To overcome this difficulty we consider a reformulation of the given stress constraints, similar to Stolpe and Svanberg [17], which results in an even higher number

*Spezialforschungsbereich SFB F013 Numerical and Symbolic Scientific Computing, supported by the Austrian Science Fund "Fonds zur Förderung der wissenschaftlichen Forschung (FWF)". Altenbergerstr. 69, A 4040 Linz, Austria. e-mail: roman.stainko@sfb013.uni-linz.ac.at

of constraints and unknowns. Over the last two decades interior-point methods turned out to be efficient optimization methods for solving large-scale nonlinear optimization problems (cf. Subsection 2.1). Most of the computing time is actually spent in the solution of linear systems arising from the linearization of the primal-dual optimality equations. These optimality conditions lead to large scale KKT-systems, like

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \triangle\mathbf{x} \\ \triangle\mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} \tag{1}$$

For this system we have to ensure that $\mathbf{A}$ is positive definite on the null space of the matrix $\mathbf{B}$. It might also happen that $\mathbf{B}^T$ does not have full rank, as a consequence the system matrix in (1) is singular. Thus, it might be necessary to modify the matrix in the following way to sustain regularity:

$$\begin{pmatrix} \mathbf{A} + \delta_1\mathbf{I} & \mathbf{B}^T \\ \mathbf{B} & -\delta_2\mathbf{I} \end{pmatrix} \begin{pmatrix} \triangle\mathbf{x} \\ \triangle\mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix},$$

with some small $\delta_1, \delta_2 \geq 0$ (see also Section 5).

Multigrid methods certainly belong to the most efficient methods for solving large-scale systems, arising from discretized systems of partial differential equations of elliptic type. While the construction of such methods for symmetric and positive definite systems is quite standard, this is not the case for saddle point problems like (1). A successful construction of a solver with optimal complexity for linear systems like (1) would yield a significant speedup for an interior-point method, because these systems have to be solved in each iteration of an interior-point method. One of the most important ingredients of an efficient multigrid method is an appropriate smoother, i.e. a simple iterative smoothing procedure (cf. Subsection 2.2). In this chapter we consider a multiplicative Schwarz-type iteration method as a smoother in a multigrid method. Each iteration step of such a multiplicative Schwarz-type smoother consists of the solution of several small local saddle point problems, i.e. small local version of the problem (1).

In the next section we will give a brief introduction to interior-point methods, multigrid methods, and the stress constrained topology optimization problem, which optimality conditions we are going to solve. In the following section, we will deduce a saddle point problem from the primal-dual optimality conditions for the optimization problem presented in the previous chapter. More information about the used kind of smoother will be given in Section 4. Finally, in Section 5 we will present some numerical experiments from the application of a multigrid method with the mentioned smoother to the derived saddle point problem.

2 Basics

2.1 Interior–Point Methods

In the last two decades interior–point algorithms have evolved to efficient methods for large scale nonlinear programming since their revival in 1984. For a survey see e.g. the related chapters in Nocedal and Wright [13] and Wright [20] and the references therein.

The rediscovery of interior–point methods is rooted in the desire to find algorithms with a better complexity than the *simplex method* for linear programming by Dantzig in 1947. Since then, the simplex method dominated the field of linear programming, although its worst-case complexity is exponential in the size of the problem dimension. After Karmakar's announcement in 1984 of the *projective algorithm*, a polynomial-time method for linear programs, interior–point methods have been subject of intense research. In principle there are two ways to motivate these methods nowadays, namely minimizing a barrier function or perturbing the optimality conditions.

For a short introduction we consider the following general optimization problem:

$$\begin{array}{rrcll} & J(\mathbf{x}) & \to & \min\limits_{\mathbf{x}\in\mathbb{R}^n} & \\ \text{subject to} & c_i(\mathbf{x}) & = & 0, & i \in \mathcal{E}, \\ & c_i(\mathbf{x}) & \leq & 0, & i \in \mathcal{I}. \end{array} \tag{2}$$

where all appearing functions should be sufficiently differentiable. For the sake of simplified notation we will denote $\mathbf{c}_{\mathcal{I}}(\mathbf{x})$ as $(c_i(\mathbf{x}))_{i\in\mathcal{I}}$ and $\mathbf{c}_{\mathcal{E}}(\mathbf{x})$ as $(c_i(\mathbf{x}))_{i\in\mathcal{E}}$. This problem is then modified such that the restricting inequality constraints are treated implicitly by adding them to the objective functional using some barrier term. The predominant barrier function is the logarithmic barrier function and so the new barrier objective $J_\mu(\mathbf{x}) := J(\mathbf{x}) + B_\mu(\mathbf{x})$ is now the sum of the original one and a logarithmic interior part:

$$\begin{array}{rcl} J(\mathbf{x}) - \mu \sum\limits_{i\in\mathcal{I}} \ln\big(- c_i(\mathbf{x})\big) & \to & \min\limits_{\mathbf{x}\in\mathbb{R}^n} \\ \text{subject to} \qquad \mathbf{c}_{\mathcal{E}}(\mathbf{x}) & = & \mathbf{0}, \end{array} \tag{3}$$

where $\mu > 0$ is called the *barrier parameter.* A major characteristic of these methods is that all inequality constraints are (have to be) satisfied strictly, which leads to the nomenclature *interior-point* methods. Minimization of (3) for a decreasing sequence of the barrier parameter $\mu \to 0$ will result (under appropriate assumptions) in a sequence of minimizers $\overline{\mathbf{x}}_\mu \to \overline{\mathbf{x}}_0 = \overline{\mathbf{x}}$ converging to the minimizer $\overline{\mathbf{x}}$ of the original problem (2). The sequence $\overline{\mathbf{x}}_\mu$ also defines a path to $\overline{\mathbf{x}}$, which is either called the *central path* or the *barrier trajectory.* The central path is a path of strictly feasible points that satisfy the perturbed complementarity conditions, see below. It is the essential idea of most interior–point methods to follow this path numerically more or less exactly.

Using the following notation we state the first order necessary optimality conditions for (3): $\mathbf{C}_{\mathcal{I}}(\mathbf{x}) = \operatorname{diag}(c_i(\mathbf{x}), i \in \mathcal{I})$, $\boldsymbol{\lambda}_{\mathcal{E}}$ the vector of Lagrange multipliers for the equality constraints and $\mathbf{e}$ a vector of ones in the appropriate dimension:

$$\begin{array}{rl} \nabla J(\mathbf{x}) + \mu \nabla \mathbf{c}_{\mathcal{I}}(\mathbf{x})^T \mathbf{C}_{\mathcal{I}}(\mathbf{x})^{-1}\mathbf{e} + \nabla \mathbf{c}_{\mathcal{E}}(\mathbf{x})^T \boldsymbol{\lambda}_{\mathcal{E}} = & \mathbf{0}, \\ \mathbf{c}_{\mathcal{E}}(\mathbf{x}) = & \mathbf{0}. \end{array} \tag{4}$$

Usually Newton's method is used to solve (4) and to find minimizers $\overline{\mathbf{x}}$. Unfortunately, the poor-scaling of the objective $J_\mu(\mathbf{x})$ becomes worse as $\mu \to 0$. The extreme behavior of the barrier function close to the boundary of the feasible set translates to ill-conditioning in the barrier Hessian $\nabla^2_{\mathbf{x}} B_\mu(\mathbf{x})$. As a consequence the quadratic Taylor series approximation, on which Newton-like methods are based, does not reflect the behavior of the original function except in a small neighbourhood of $\overline{\mathbf{x}}$. This fact was one of the major motivations for the downfall of barrier methods before 1984, since it may cause poor numerical performance of unconstrained optimization methods ($\mathcal{E} = \emptyset$). Fortunately Newton's method (in a carefully implemented algorithm, see FORSGREN, GILL, AND WRIGHT [7]) can be made insensitive to this poor scaling.

The true reason for the inefficiency of classical barrier methods is another. Unfortunately, it is often not possible to take a full Newton step, because this step would move the current iterate out of the feasible region, especially when the current iterate is very close to a minimizer of $B_\mu(\mathbf{x})$ with a fixed μ. Suppose the current iterate is the minimizer $\overline{\mathbf{x}}_\mu$ of (3) with a fixed μ and the barrier parameter μ is now reduced to $\hat{\mu}$ with $\mu > \hat{\mu}$. If the ratio $\mu/\hat{\mu}$ exceeds a certain factor and the next Newton step is computed with respect to the new barrier parameter $\hat{\mu}$, a full Newton step will move the iterate to a significant infeasible point.

There are several remedies to overcome these poor steps that occur after a reduction of the barrier parameter, but the best one is to use *primal-dual* interior methods. In primal-dual methods we treat the primal variables and the dual variables (the Lagrangian multipliers of the problem) independently. In this spirit we now create an independent variable $\boldsymbol{\lambda}_\mathcal{I}$ of multipliers for the inequality constraints from the relation $\boldsymbol{\lambda}_\mathcal{I} = -\mu \mathbf{C}_\mathcal{I}(\mathbf{x})^{-1}\mathbf{e}$. Furthermore, if we consider $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_\mathcal{I}, \boldsymbol{\lambda}_\mathcal{E})$ and $\mathbf{c}(\mathbf{x}) = (\mathbf{c}_\mathcal{I}(\mathbf{x}), \mathbf{c}_\mathcal{E}(\mathbf{x}))$, we can rewrite (4) as a system in the primal variables $\mathbf{x}$ and the dual variables $\boldsymbol{\lambda}$:

$$\nabla J(\mathbf{x}) + \nabla \mathbf{c}(\mathbf{x})^T \boldsymbol{\lambda} = \mathbf{0}, \tag{5a}$$
$$\mathbf{C}_\mathcal{I}(\mathbf{x})\boldsymbol{\lambda}_\mathcal{I} + \mu \mathbf{e} = \mathbf{0}, \tag{5b}$$
$$\mathbf{c}_\mathcal{E}(\mathbf{x}) = \mathbf{0}. \tag{5c}$$

The second equation (5b) can be interpreted as the *perturbed complementarity condition* for the inequality constraints in the KKT conditions for (2). The success of primal–dual methods is now partly due to their effectiveness at following the central path, especially in steps where the barrier parameter is reduced.

The left-hand-side of (5) defines a function $F_\mu(\mathbf{x}, \boldsymbol{\lambda})$. Instead of minimizing (3) for $\mu \to 0$, we look for solutions of $F_\mu(\mathbf{x}, \boldsymbol{\lambda}) = 0$ for $\mu \to 0$. For a fixed μ (5) can be solved, e.g., using a modified Newton-type method such that $\mathbf{x}$ and $\boldsymbol{\lambda}_\mathcal{I}$ fulfill the inequality constraints $\mathbf{c}_\mathcal{I}(\mathbf{x}) \le 0$ and $\boldsymbol{\lambda}_\mathcal{I} \ge 0$ strictly. The Newton direction $(\triangle \mathbf{x}, \triangle \boldsymbol{\lambda})$ of such a method is defined as the solution of $\nabla F_\mu(\mathbf{x}, \boldsymbol{\lambda})(\triangle \mathbf{x}, \triangle \boldsymbol{\lambda}) = -F_\mu(\mathbf{x}, \boldsymbol{\lambda})$:

$$\begin{pmatrix} \nabla^2 \mathbf{H} & -\nabla \mathbf{c}_\mathcal{I}^T & \nabla \mathbf{c}_\mathcal{E}^T \\ \boldsymbol{\Lambda}_\mathcal{I} \nabla \mathbf{c}_\mathcal{I} & \mathbf{C}_\mathcal{I} & \mathbf{0} \\ \nabla \mathbf{c}_\mathcal{E} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \triangle \mathbf{x} \\ \triangle \boldsymbol{\lambda}_\mathcal{I} \\ \triangle \boldsymbol{\lambda}_\mathcal{E} \end{pmatrix} = - \begin{pmatrix} \nabla J + \nabla \mathbf{c}^T \boldsymbol{\lambda} \\ \mathbf{C}_\mathcal{I} \boldsymbol{\lambda}_\mathcal{I} + \mu \mathbf{e} \\ \mathbf{c}_\mathcal{E} \end{pmatrix}, \tag{6}$$

where $\boldsymbol{\Lambda}_\mathcal{I} = \mathrm{diag}(\boldsymbol{\lambda}_i, i \in \mathcal{I})$, $\mathbf{H}(\mathbf{x}, \boldsymbol{\lambda})$ denotes the Hessian of the Lagrangian of (2) and all arguments in (6) are omitted.

An efficient solver for systems like (6) are important for an efficient performance of an interior-point method, since they have to be solved in each iteration step of the optimization method.

2.2 The Multigrid Method

The multigrid method provides an optimal order algorithm for solving linear systems arising from finite element discretizations, as well as other discretization techniques applied to certain classes of PDEs. The number of iterations of standard iteration methods are increasing as $h \to 0$ if no proper preconditioning is applied. When using multigrid methods we get numbers of iterations that are asymptotically independent from the mesh parameter h. In other words the convergence speed does not deteriorate when the discretization is refined, whereas classical iterative methods slow down for decreasing mesh size. A fundamental attribute of the multigrid method is that it is working on a hierarchy of meshes and related discretizations of a boundary value problem. We recommend e.g. the books BRAMBLE [3] and HACKBUSCH [9] for detailed introduction.

The multigrid method has two main features: smoothing on a fine grid and error correction on a coarser grid. The starting point for this idea is the observation that classical iteration methods have smoothing properties, i.e. they remove the high oscillating parts of the error very quick. The smooth part of the error can be represented and corrected well on coarser grids. Hence, combining these two approaches makes the multigrid method are of the most efficient solvers. For a short introduction let us consider a hierarchy of l meshes (e.g. like in

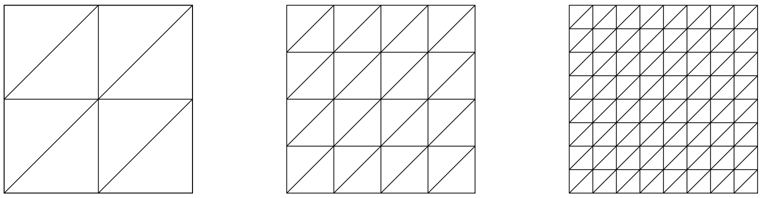

Figure 1: A hierarchy of 3 meshes: $\mathcal{T}_0 \subset \mathcal{T}_1 \subset \mathcal{T}_2$.

Figure 1)

$$\mathcal{T}_0 \subset \mathcal{T}_1 \subset \ldots \subset \mathcal{T}_l,$$

with corresponding finite element spaces $V_0 \subset \ldots \subset V_l$, mesh sizes $h_0 \geq \ldots \geq h_l$, and number of unknowns $n_0 \leq \ldots \leq n_l$. One of the ingredients of a successful multigrid method are the *intergrid transfer* operators:

Definition 1 (Intergrid Operators)**.** *The coarse-to-fine operator*

$$\mathcal{I}_{l-1}^{l} \ : \ V_{l-1} \rightarrow V_l$$

is called the (prolongation) *operator and the the fine-to-coarse operator*

$$\mathcal{I}_{l}^{l-1} \ : \ V_l \rightarrow V_{l-1}$$

is called the (restriction) *operator.*

Remark 1. *If we have a sequence $V_0 \subset \ldots \subset V_l$ of spaces, the prolongation operator $\mathcal{I}_{l-1}^{l}$ can be taken to be the natural injection. In other words, $\mathcal{I}_{l-1}^{l} v = v, \quad \forall\ v \in V_{l-1}$. Then the restriction operator is defined to be the adjoint of $\mathcal{I}_{l-1}^{l}$ with respect to $(\cdot,\cdot)_{l-1}$ and $(\cdot,\cdot)_l$ inner products. In other words, $(\mathcal{I}_{l}^{l-1} w, v)_{l-1} = (w, \mathcal{I}_{l-1}^{l} v)_l = (w, v)_l, \quad \forall v \in V_{l-1}, w \in V_l$.*

A proper choice of the intergrid operators influences the convergence speed considerably, and may even be necessary for convergence.

In addition to the intergrid operators we need an iteration method *(smoother)* for the smoothing iterations on the fine grids. For instance we choose the smoothing operator $\mathcal{S}$ to realize the Jacobi-relaxation with a damping parameter $\tau > 0$ (cf. Algorithm 3):

$$\mathbf{u} \longmapsto \mathcal{S}\mathbf{u} \ = \ \mathbf{u} - \tau(\mathbf{K}\mathbf{u} - \mathbf{f}).$$

Since it reduces the high frequency error components the smoothing operator $\mathcal{S}$ is an essential part in multigrid methods. Typically, a proper smoother for a problem takes the special

Algorithm 1 Two grid method

Choose a relative error bound $\varepsilon > 0$.
Choose a number ν_1 of pre-smoothing and a number ν_2 of post-smoothing steps.
Initialize start value $\mathbf{u}_h^{(0,0)}$.
$k = 0$;

while not converged **do**
  /* Pre-smoothing: */
  $\mathbf{u}_h^{(k,1)} = \mathcal{S}^{\nu_1}\mathbf{u}_h^{(k,0)}$;
  /* Coarse grid correction: */
  /* Defect calculation: */
  $\mathbf{d}_h^k = \mathbf{f}_h - \mathbf{K}_h\mathbf{u}_h^{(k,1)}$;
  /* Restriction onto coarse grid: */
  $\mathbf{d}_H^k = \mathbf{I}_h^H\mathbf{d}_h^k$;
  /* Solve coarse grid system: */
  $\mathbf{K}_H\mathbf{w}_H^k = \mathbf{d}_H^k$;
  /* Prolongation onto the fine grid: */
  $\mathbf{w}_h^k = \mathbf{I}_H^h\mathbf{w}_H^k$;
  /* Add coarse grid correction: */
  $\mathbf{u}_h^{(k,2)} = \mathbf{u}_h^{(k,1)} + \mathbf{w}_h^k$;
  /* Post-smoothing: */
  $\mathbf{u}_h^{(k+1,0)} = \mathcal{S}^{\nu_2}\mathbf{u}_h^{(k,2)}$;
  $k = k + 1$;
end while

structure of the system matrix into account. Beside the point Jacobi smoother also the point Gauß-Seidel (cf. Algorithm 4), the block Jacobi and block Gauß-Seidel smoother are suitable for a large class of finite element discretized problems. In Section 4 we will discuss a local patch smoother for an specific application case. Beside the smoothing operation we also need a coarse grid correction. Let us assume that we have the corresponding system matrices $\mathbf{K}_0, \ldots, \mathbf{K}_l$ and load vectors $\mathbf{f}_0, \ldots, \mathbf{f}_l$ for each level at hand. These can either be generated by assembling on each level or can be constructed by Galerkin's method, i.e.

$$\mathbf{K}_{l-1} = \mathbf{I}_l^{l-1}\mathbf{K}_l\mathbf{I}_{l-1}^l.$$

After these preliminaries we are ready to state a two level method, as in Algorithm 1, where we assume that $l = 1$ and we use the following notation $h_0 = 2h_1 = H$, $\mathbf{I}_0^1 = \mathbf{I}_H^h$, and $\mathbf{I}_1^0 = \mathbf{I}_h^H$ for better readability. The parameters ν_1 and ν_2 control the number of smoothing iterations. For benign problems like the Poisson equation it usually does not pay off to use more than two smoothing steps. In the case of more complex problems, such as saddle point problems, it can be necessary to use more smoothing iterations.

The restricted system on the coarse grid is by far easier to solve than the one on the finer grid. When switching to a mesh from mesh size h to $2h$ by uniform refinement, the number of unknowns decreases about to a quarter. But still, the complexity of solving the coarse grid system may be regarded to high. The idea to advance from a two grid method to a multigrid method is now to repeat this procedure recursively. That is to coarsen the grid until the coarsest grid yields a sufficiently small system, that is easy to solve. The linear system on the coarsest grid is usually solved directly, e.g. by some Cholesky factorization. So, instead of solving the coarse grid system, one or two multigrid steps are performed, resulting in a V-cycle or a W-cycle. The patterns in Fig. 2 will explain the naming, where $\circ$ denotes smoothing, $\bullet$ marks the solution of the system on the coarsest grid, $\searrow$ and $\nearrow$ stand for restriction and interpolation between the grids, respectively. In the early days the common choice was a

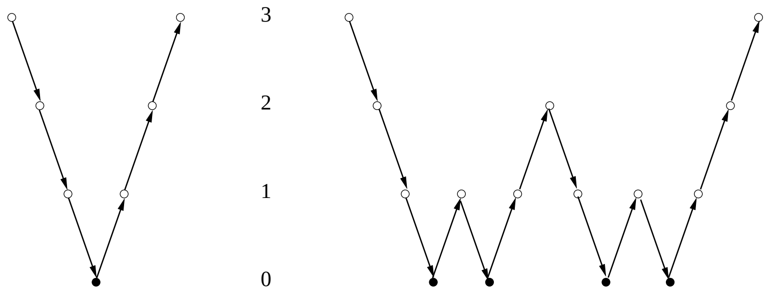

Figure 2: V-cycle and W-cycle on a hierarchy of 4 grids.

W-cycle to ensure that the error is not increasing too much when cycling between several grids. But most of the problems are so benign, that a V-cycle is more efficient. The following Algorithm 2 sketches the operations of the k^{th} multigrid iteration on level i with $1 \leq i \leq l$. For sake of readability we drop the iteration index k.

2.3 A Topology Optimization Problem with Local Stress constraints

In structural optimization there are two design - constraint combinations of particular importance, namely the maximization of material stiffness (minimizing the compliance) at given mass and the minimization of mass while keeping a certain stiffness. The first combination, also known as the minimal compliance problem, seems to be mathematically well understood and various successful numerical techniques to solve the problem have been proposed (see e.g. Bendsøe and Sigmund [1] and Stainko [16]). The treatment of the second problem is by far less understood and until now there seems to be no approach that is capable of computing reliable (global) optimal designs within reasonable computational effort. The main source of difficulties in this problem is the lack of constraint qualifications for the set of feasible designs, defined by the local stress constraints.

The approach we briefly introduce here is described in more details in Burger and Stainko [5]. Let $\Omega_{\text{mat}} = \{\mathbf{x} \in \Omega \mid \rho(\mathbf{x}) = 1\} \subset \Omega \subset \mathbb{R}^d$ $(d = 2, 3)$, denote the optimal design, which is of course initially unknown. Furthermore, let $\Gamma_{t_0} \subset \Gamma_t$ describe the part of the boundary Γ_t where the traction forces are zero, i.e. $\mathbf{t} = \mathbf{0}$. Then, the stress constrained

Algorithm 2 One multigrid method iteration MGM($\mathbf{K}_i, \mathbf{u}_i, \mathbf{f}_i, i$)

Let μ describe the number of MGM calls per level i.
Let ν_1 and ν_2 denote the number of pre- and post-smoothing steps.
Initialize start value $\mathbf{u}_i^0 = \mathbf{u}_i$;

if $i == 0$ **then**
  Solve the coarsest grid system, i.e. $\mathbf{u}_0 = \mathbf{K}_0^{-1}\mathbf{f}_0$;
  return;
else
  /* Pre-smoothing: */
  $\mathbf{u}_i^1 = \mathcal{S}^{\nu_1}\mathbf{u}_i^0$;
  /* Coarse grid correction: */
  /* Defect calculation: */
  $\mathbf{d}_i = \mathbf{f}_i - \mathbf{K}_i\mathbf{u}_i^1$;
  /* Restriction onto coarse grid: */
  $\mathbf{d}_{i-1} = \mathbf{I}_i^{i-1}\mathbf{d}_i$;
  /* Recursively call MGM for coarse grid approximation: */
  $\mathbf{w}_{i-1} = \mathbf{0}$;
  for $j = 1, \ldots, \mu$ **do**
    MGM($\mathbf{K}_{i-1}, \mathbf{w}_{i-1}, \mathbf{d}_{i-1}, i-1$);
  end for
  /* Prolongation onto the fine grid: */
  $\mathbf{w}_i = \mathbf{I}_{i-1}^i\mathbf{w}_{i-1}$;
  /* Add coarse grid correction: */
  $\mathbf{u}_i^2 = \mathbf{u}_i^1 + \mathbf{w}_i$;
  /* Post-smoothing: */
  $\mathbf{u}_i^3 = \mathcal{S}^{\nu_2}\mathbf{u}_i^2$;
end if

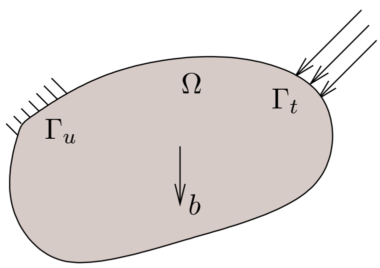

Figure 3: The reference domain and applied forces in a minimal mass problem.

topology optimization problem that we are going to investigate states as follows:

$$J(\rho) = \int_\Omega \rho(\mathbf{x})\; d\mathbf{x} \;\rightarrow\; \min_{\rho,\mathbf{u}} \tag{7a}$$

$$\text{subject to} \qquad \operatorname{div}\boldsymbol{\sigma} = 0, \qquad \text{in } \Omega_{\text{mat}}, \tag{7b}$$

$$\boldsymbol{\sigma} - \mathbf{C}\boldsymbol{\varepsilon}(u) = \mathbf{0}, \qquad \text{in } \Omega, \tag{7c}$$

$$\mathbf{u} = \mathbf{0}, \qquad \text{on } \Gamma_u, \tag{7d}$$

$$\boldsymbol{\sigma}\cdot\mathbf{n} = \mathbf{t}, \qquad \text{on } \Gamma_t, \tag{7e}$$

$$\boldsymbol{\sigma}\cdot\mathbf{n} = \mathbf{0}, \qquad \text{on } \big(\partial\Omega_{\text{mat}} \setminus \Gamma_t\big) \cup \Gamma_{t_0}, \tag{7f}$$

$$\rho(\mathbf{x}) \in \{0,1\}, \qquad \text{a.e. in } \Omega, \tag{7g}$$

$$\Phi^{\min} \le \Phi\big(\boldsymbol{\sigma}(\mathbf{x})\big) \le \Phi^{\max}, \qquad \text{a.e. in } \Omega_{\text{mat}}, \tag{7h}$$

$$\mathbf{u}^{\min} \le \mathbf{u}(\mathbf{x}) \le \mathbf{u}^{\max}, \qquad \text{a.e. in } \Omega. \tag{7i}$$

Thus, in a first formulation, the objective functional (7a) only consists of a mass term. Note that we only optimize with respect to the design ρ and the displacements $\mathbf{u}$, because the stresses $\boldsymbol{\sigma}$ can be eliminated using the stress-strain relation (7c). But for sake of better readability we will keep the stresses in the formulation. The constraints (7b) - (7f) describe the elasticity equations with corresponding boundary conditions (all to be interpreted in a weak sense), where we again neglect body forces for the sake of simplicity. In an ideal case, the material density ρ only attains two values, 1 for material and 0 for void, see the 0-1 constraint (7g). Moreover, the vectors $\mathbf{u}^{\min}$ and $\mathbf{u}^{\max}$ in the bound constraint (7i) are lower and upper bounds for the displacements $\mathbf{u}$. In the bound constraints (7h), Φ denotes a proper stress criterion. For $\Phi(\boldsymbol{\sigma}) = \boldsymbol{\sigma}$ we have that $\boldsymbol{\sigma}^{\min} \leq \boldsymbol{\sigma} \leq \boldsymbol{\sigma}^{\max}$ and we shall call this criterion total stress. For sake of simplicity we will only treat here the case of total stresses, but, e.g., von Mises stresses can be handled in the same way.

Starting point of our analysis is a reformulation of the equality constraints describing the elastic equilibrium and the local inequality constraints for the stresses into a system of linear inequality constraints as recently proposed by Stolpe and Svanberg [17]. A remaining difficulty is that the arising problem also involves 0-1 constraints in addition to the linear inequalities. Instead of solving mixed linear programming problems we propose to use a phase–field relaxation of the reformulated problem. Due to the well-known ill-posedness of topology optimization problems we might add a perimeter penalization to the objective functional. The phase–field relaxation consists in using a material interpolation function $\eta(\rho) = \rho$, and additionally, a Cahn-Hilliard type penalization functional (see Cahn and Hilliard [6]) is used to approximate the perimeter. Up to the knowledge of the author, the phase–field method was first introduced by Bourdin and Chambolle [2] to the field of topology optimization.

For the reformulation of the set of constraints we introduce a $\beta > 0$, such that

$$\beta\left|\sigma_{ij}(\mathbf{x})\right| \leq 1, \qquad \text{a.e. in } \Omega,\ i,j = 1,\dots,d,$$

and an additional variable $\mathbf{s}$, such that $\mathbf{s}(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x})$ if $\rho(\mathbf{x}) = 1$ and $\mathbf{s}(\mathbf{x}) = \mathbf{0}$ if $\rho(\mathbf{x}) = 0$, i.e. $\mathbf{s} = \rho\boldsymbol{\sigma}$. Then the equivalent reformulation of the set of constraints looks like:

$$\begin{aligned}
\operatorname{div}\mathbf{s} &= 0, && \text{in } \Omega, && \text{(8a)}\\
\boldsymbol{\sigma} - \mathbf{C}\varepsilon(\mathbf{u}) &= \mathbf{0}, && \text{in } \Omega, && \text{(8b)}\\
\mathbf{u} &= \mathbf{0}, && \text{on } \Gamma_u, && \text{(8c)}\\
\mathbf{s}\cdot\mathbf{n} &= \mathbf{t}, && \text{on } \Gamma_t, && \text{(8d)}\\
\mathbf{s}\cdot\mathbf{n} &= \mathbf{0}, && \text{on } \Gamma_{t_0}, && \text{(8e)}\\
-(1-\rho)\mathbf{1} \leq \beta(\boldsymbol{\sigma} - \mathbf{s}) &\leq (1-\rho)\mathbf{1}, && \text{in } \Omega, && \text{(8f)}\\
\rho(\mathbf{x}) &\in \{0,1\}, && \text{a.e. in } \Omega, && \text{(8g)}\\
\rho(\mathbf{x})\boldsymbol{\sigma}^{\min} \leq \mathbf{s}(\mathbf{x}) &\leq \rho(\mathbf{x})\boldsymbol{\sigma}^{\max}, && \text{a.e. in } \Omega, && \text{(8h)}\\
\mathbf{u}^{\min} \leq \mathbf{u}(\mathbf{x}) &\leq \mathbf{u}^{\max}, && \text{a.e. in } \Omega. && \text{(8i)}
\end{aligned}$$

All the constraints in (8) are linear with respect to the vector of unknowns $(\rho, \mathbf{u}, \boldsymbol{\sigma}, \mathbf{s})$, except for $\rho(\mathbf{x}) \in \{0,1\}$ almost everywhere in Ω. We now replace the 0-1 constraint $\rho(\mathbf{x}) \in \{0,1\}$ by the following continuous version $\rho(\mathbf{x}) \in [0,1]$. Moreover, we approximate a perimeter term by the Cahn-Hilliard term and add it to the objective:

$$J_\epsilon(\rho) = \gamma \int_\Omega \rho(\mathbf{x})\ d\mathbf{x} + \frac{\epsilon}{2}\int_\Omega \left|\nabla\rho(\mathbf{x})\right|^2\ d\mathbf{x} + \frac{1}{\epsilon}\int_\Omega W\big(\rho(\mathbf{x})\big)\ d\mathbf{x}. \tag{9}$$

The term $\int_\Omega W(\rho(\mathbf{x})\ d\mathbf{x}$ favorites those designs which take values close to 0 or 1 (*phase separation*), while the term $\int_\Omega |\nabla\rho(\mathbf{x})|^2\ d\mathbf{x}$ penalizes the spatial inhomogeneity of ρ. The theorem of Modica and Mortola tells that the minimizers of (9) converge to the minimizers of $\int_\Omega \rho(\mathbf{x})\ d\mathbf{x}$ in the sense of Γ-convergence (see Modica and Mortola [12]). The resulting relaxed parameter dependent problem is now given by the objective functional (9) and by the constraints (8), where (8g) is replaced by $0 \leq \rho(\mathbf{x}) \leq 1$. The problem is now solved for a decreasing sequence of the parameter $\varepsilon \to 0$. For the relaxed problem it is now possible to show the existence of solutions in the corresponding set of feasible designs.

After a standard finite element discretization we end up with a large scale optimization problem, that now fulfills constraint qualifications, cf. Burger and Stainko [5]. We solved the discrete optimization problems using *Ipopt*, which is a free available optimization code realizing a primal-dual interior-point optimization method (see Wächter et al [19]).

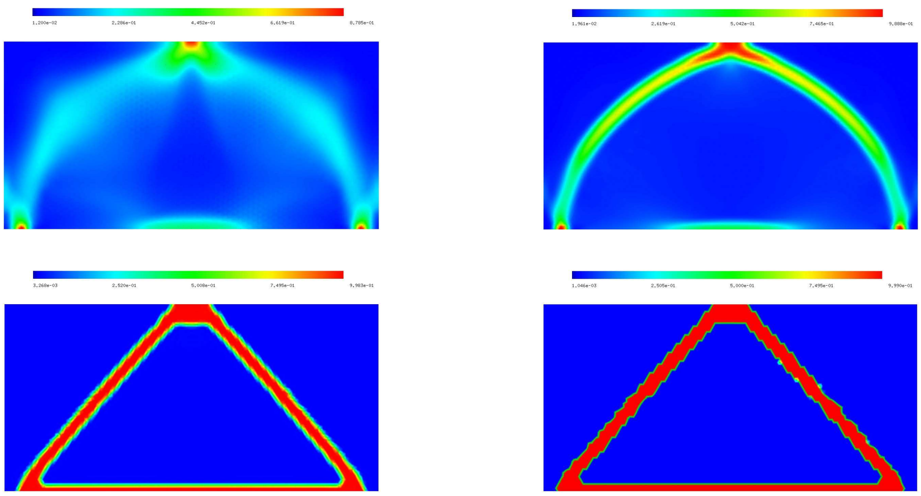

Figure 4: Optimal designs of the 4 level ϵ-continuation with $\epsilon_0 = 0.1$, $\epsilon_1 = 0.05$, $\epsilon_2 = 0.025$, and $\epsilon_3 = 0.0125$, respectively.

3 The Optimality System

In this section we will derive an optimality system for the optimization of the objective (9) subject to the constraints (8) from the previous Subsection 2.3. The derivation can be performed for both stress criteria, total stress and conservative von Mises stress, but we restrict ourselves to the case of total stress for the sake of simplicity. As a starting point we reconsider the following optimization problem:

$$\begin{aligned}
J_\epsilon(\rho) \;&\rightarrow\; \min_{\rho,\mathbf{u},\mathbf{s}} && && \text{(10a)}\\
\operatorname{div}\mathbf{s} \;&=\; 0, && \text{in } \Omega, && \text{(10b)}\\
\boldsymbol{\sigma} - \mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}) \;&=\; \mathbf{0}, && \text{in } \Omega, && \text{(10c)}\\
\mathbf{u} \;&=\; \mathbf{0}, && \text{on } \Gamma_u, && \text{(10d)}\\
\mathbf{s}\cdot\mathbf{n} \;&=\; \mathbf{t}, && \text{on } \Gamma_t, && \text{(10e)}\\
\mathbf{s}\cdot\mathbf{n} \;&=\; \mathbf{0}, && \text{on } \Gamma_{t_0}, && \text{(10f)}\\
\rho \;&=\; 1, && \text{on } \Gamma_t, && \text{(10g)}\\
-(1-\rho)\mathbf{1} \;\le\; \beta(\boldsymbol{\sigma}-\mathbf{s}) \;&\le\; (1-\rho)\mathbf{1}, && \text{in } \Omega, && \text{(10h)}\\
0 \;\le\; \rho(\mathbf{x}) \;&\le\; 1, && \text{a.e. in } \Omega, && \text{(10i)}\\
\rho(\mathbf{x})\boldsymbol{\sigma}^{\min} \;\le\; \mathbf{s}(\mathbf{x}) \;&\le\; \rho(\mathbf{x})\boldsymbol{\sigma}^{\max}, && \text{a.e. in } \Omega, && \text{(10j)}\\
\mathbf{u}^{\min} \;\le\; \mathbf{u}(\mathbf{x}) \;&\le\; \mathbf{u}^{\max}, && \text{a.e. in } \Omega. && \text{(10k)}
\end{aligned}$$

with

$$J_\epsilon(\rho) = \gamma \int_\Omega \rho(\mathbf{x})\; d\mathbf{x} + \frac{\epsilon}{2}\int_\Omega \left|\nabla\rho(\mathbf{x})\right|^2\; d\mathbf{x} + \frac{1}{\epsilon}\int_\Omega \rho(\mathbf{x})\big(1-\rho(\mathbf{x})\big)\; d\mathbf{x},$$

and the function space setting

$$(\rho,\mathbf{u},\mathbf{s}) \in \big(H^1(\Omega)\cap L_\infty(\Omega)\big) \times \big(H^1\big(\Omega;\mathbb{R}^2\big)\cap L_\infty\big(\Omega;\mathbb{R}^2\big)\big) \times L_\infty\big(\Omega;\mathbb{R}^{2\times 2}\big).$$

We mention that the conditions (10b), (10e), and (10f) have to be understood in a week sense, namely as

$$\int_\Omega \mathbf{s} : \boldsymbol{\varepsilon}(\mathbf{v})\; d\mathbf{x} = \int_{\Gamma_t} \mathbf{v}\cdot\mathbf{t}\; d\mathbf{a}, \qquad \forall\ \mathbf{v}\in H^1_{\Gamma_t},$$

with $H^1_{\Gamma_t} := \{\mathbf{v}\in H^1(\Omega;\mathbb{R}^2) \mid \mathbf{v}=\mathbf{0} \text{ on } \Gamma_t\}$. Because we aim at an interior-point optimization method to solve this problems, we will perform the derivation in a primal-dual interior-point framework. Especially, we consider a primal-dual barrier method to solve nonlinear optimization problems of the form

$$\begin{aligned}
f(\mathbf{x}) \;&\rightarrow\; \min_{\mathbf{x}\in\mathbb{R}^n}\\
\text{subject to} \qquad c(\mathbf{x}) \;&=\; \mathbf{0},\\
\mathbf{x}^{\min} \;\le\; \mathbf{x} \;&\le\; \mathbf{x}^{\max},
\end{aligned}$$

see, e.g. Wächter and Biegler [19]. Problems with inequality constraints, like (10), can be reformulated in the above form by introducing slack variables. Thus, we rewrite the inequality constraints (10h) and (10i) as equalities with the additional functions $\mathbf{z}_i \in L_2(\Omega;\mathbb{R}^{2\times 2})$, $i=1,\ldots,4$. Furthermore, to ease the representation of the problem, we **omit** the boundary conditions (10d) - (10g), and rely on their proper incorporation in the FE-discretization. As an additional simplification we eliminate $\boldsymbol{\sigma}$ using the identity (10c). Consequently, we end up with the following optimization problem:

$$
\begin{aligned}
J_\epsilon(\rho) &\;\to\; \min_{\rho,\mathbf{u},\mathbf{s},\mathbf{z}} && && \text{(11a)}\\
\operatorname{div}\mathbf{s}(\mathbf{x}) &= \mathbf{0}, && \text{in } \Omega, && \text{(11b)}\\
-\big(1-\rho(\mathbf{x})\big)\mathbf{1} - \beta\big(\mathbf{C}\boldsymbol{\varepsilon}\big(\mathbf{u}(\mathbf{x})\big) - \mathbf{s}(\mathbf{x})\big) + \mathbf{z}_1(\mathbf{x}) &= \mathbf{0}, && \text{in } \Omega, && \text{(11c)}\\
-\big(1-\rho(\mathbf{x})\big)\mathbf{1} + \beta\big(\mathbf{C}\boldsymbol{\varepsilon}\big(\mathbf{u}(\mathbf{x})\big) - \mathbf{s}(\mathbf{x})\big) + \mathbf{z}_2(\mathbf{x}) &= \mathbf{0}, && \text{in } \Omega, && \text{(11d)}\\
\rho(\mathbf{x})\boldsymbol{\sigma}^{\min} - \mathbf{s}(\mathbf{x}) + \mathbf{z}_3(\mathbf{x}) &= \mathbf{0}, && \text{in } \Omega, && \text{(11e)}\\
\mathbf{s}(\mathbf{x}) - \rho(\mathbf{x})\boldsymbol{\sigma}^{\max} + \mathbf{z}_4(\mathbf{x}) &= \mathbf{0}, && \text{in } \Omega, && \text{(11f)}\\
0 \;\le\; \rho(\mathbf{x}) &\le 1, && \text{a.e. in } \Omega, && \text{(11g)}\\
\mathbf{u}^{\min} \;\le\; \mathbf{u}(\mathbf{x}) &\le \mathbf{u}^{\max}, && \text{a.e. in } \Omega, && \text{(11h)}\\
\mathbf{z}_i(\mathbf{x}) &\ge \mathbf{0}, && \text{a.e. in } \Omega,\ i=1,\dots,4. && \text{(11i)}
\end{aligned}
$$

Interior-point methods propose to add the bound constraints to the objective functional and treat them implicitly by using a barrier function. With a barrier parameter $\mu > 0$ this leads to the barrier objective functional

$$
\begin{aligned}
J_{\epsilon,\mu}(\rho) = J_\epsilon(\rho) - \mu \Bigg(&\int_\Omega \ln\big(\rho(\mathbf{x})\big) + \ln\big(1-\rho(\mathbf{x})\big)\, d\mathbf{x} + \\
&+ \int_\Omega \ln\big(\mathbf{u}(\mathbf{x}) - \mathbf{u}^{\min}\big) + \ln\big(\mathbf{u}^{\max} - \mathbf{u}(\mathbf{x})\big)\, d\mathbf{x} + \int_\Omega \ln\big(\mathbf{z}(\mathbf{x})\big)\, d\mathbf{x}\Bigg),
\end{aligned}
$$

where we write $\ln\mathbf{u}$ instead of $\ln u_1 + \ln u_2$ for $\mathbf{u}\in\mathbb{R}^2$ and $\ln\mathbf{z}$ instead of $\sum_{i,j=1}^2 \ln z_{ij}$ for $\mathbf{z}\in\mathbb{R}^{2\times 2}$ for simplicity. The corresponding barrier problem is given by

$$
\begin{aligned}
J_{\epsilon,\mu}(\rho) &\;\to\; \min_{\rho,\mathbf{u},\mathbf{s},\mathbf{z}} \\
\operatorname{div}\mathbf{s}(\mathbf{x}) &= \mathbf{0}, && \text{in } \Omega,\\
-\big(1-\rho(\mathbf{x})\big)\mathbf{1} - \beta\big(\mathbf{C}\boldsymbol{\varepsilon}\big(\mathbf{u}(\mathbf{x})\big) - \mathbf{s}(\mathbf{x})\big) + \mathbf{z}_1(\mathbf{x}) &= \mathbf{0}, && \text{in } \Omega,\\
-\big(1-\rho(\mathbf{x})\big)\mathbf{1} + \beta\big(\mathbf{C}\boldsymbol{\varepsilon}\big(\mathbf{u}(\mathbf{x})\big) - \mathbf{s}(\mathbf{x})\big) + \mathbf{z}_2(\mathbf{x}) &= \mathbf{0}, && \text{in } \Omega,\\
\rho(\mathbf{x})\boldsymbol{\sigma}^{\min} - \mathbf{s}(\mathbf{x}) + \mathbf{z}_3(\mathbf{x}) &= \mathbf{0}, && \text{in } \Omega,\\
\mathbf{s}(\mathbf{x}) - \rho(\mathbf{x})\boldsymbol{\sigma}^{\min} + \mathbf{z}_4(\mathbf{x}) &= \mathbf{0}, && \text{in } \Omega,
\end{aligned}
\tag{12}
$$

In order to formulate the first order necessary conditions for (12), we consider the Lagrangian for the above problem. For this sake we introduce Lagrange multipliers $\boldsymbol{\lambda}_0 \in H_0^1(\Omega;\mathbb{R}^2) := \{\mathbf{v}\in H^1(\Omega;\mathbb{R}^2) \mid \mathbf{v} = \mathbf{0} \text{ on } \partial\Omega\}$, $\boldsymbol{\lambda}_1,\dots,\boldsymbol{\lambda}_4 \in L_2(\Omega;\mathbb{R}^{2\times 2})$ and state

$$
\begin{aligned}
\mathcal{L}(\rho,\mathbf{u},\mathbf{s},\mathbf{z},\boldsymbol{\lambda}) \;=\; & J_{\epsilon,\mu}(\rho) - \langle \mathbf{s}, \varepsilon(\boldsymbol{\lambda}_0)\rangle \;+\; \big\langle -(1-\rho)\mathbf{1} - \beta\big(\mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}) - \mathbf{s}\big) + \mathbf{z}_1, \boldsymbol{\lambda}_1\big\rangle \;+\\
&+\; \big\langle -(1-\rho)\mathbf{1} + \beta\big(\mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}) - \mathbf{s}\big) + \mathbf{z}_2, \boldsymbol{\lambda}_2\big\rangle \;+\\
&+\; \big\langle \rho\boldsymbol{\sigma}^{\min} - \mathbf{s} + \mathbf{z}_3, \boldsymbol{\lambda}_3\big\rangle \;+\; \big\langle \mathbf{s} - \rho\boldsymbol{\sigma}^{\max} + \mathbf{z}_4, \boldsymbol{\lambda}_4\big\rangle,
\end{aligned}
\tag{13}
$$

with the notation $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_0, \dots, \boldsymbol{\lambda}_4)$ and $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_4)$. The optimality conditions then look as:

$$
\begin{aligned}
\nabla_\rho \mathcal{L} \ &= \gamma + \epsilon\Delta\rho + \frac{1}{\epsilon}(1-2\rho) - \frac{\mu}{\rho} + \frac{\mu}{1-\rho} + \mathbf{1} : \boldsymbol{\lambda}_1 + \\
&\qquad\qquad\qquad + \ \mathbf{1} : \boldsymbol{\lambda}_2 + \boldsymbol{\sigma}^{\min} : \boldsymbol{\lambda}_3 - \boldsymbol{\sigma}^{\max} : \boldsymbol{\lambda}_4 \ = \ 0, && \text{(14a)}\\
\nabla_{\mathbf{u}} \mathcal{L} \ &= \ - \frac{\mu}{\mathbf{u} - \mathbf{u}^{\min}} + \frac{\mu}{\mathbf{u}^{\max} - \mathbf{u}} + \beta \mathbf{C} \operatorname{div} \boldsymbol{\lambda}_1 - \beta \mathbf{C} \operatorname{div} \boldsymbol{\lambda}_2 \ = \ \mathbf{0}, && \text{(14b)}\\
\nabla_{\mathbf{s}} \mathcal{L} \ &= \ - \boldsymbol{\varepsilon}(\boldsymbol{\lambda}_0) + \beta\boldsymbol{\lambda}_1 - \beta\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_3 + \boldsymbol{\lambda}_4 \ = \ \mathbf{0}, && \text{(14c)}\\
\nabla_{\mathbf{z}_i} \mathcal{L} \ &= \ - \frac{\mu}{\mathbf{z}_i} + \boldsymbol{\lambda}_i \ = \ \mathbf{0}, && \text{(14d)}\\
\nabla_{\boldsymbol{\lambda}_0} \mathcal{L} \ &= \ \operatorname{div} \mathbf{s} \ = \ \mathbf{0}, && \text{(14e)}\\
\nabla_{\boldsymbol{\lambda}_1} \mathcal{L} \ &= \ - (1-\rho)\mathbf{1} - \beta\big(\mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}) - \mathbf{s}\big) + \mathbf{z}_1 \ = \ \mathbf{0}, && \text{(14f)}\\
\nabla_{\boldsymbol{\lambda}_2} \mathcal{L} \ &= \ - (1-\rho)\mathbf{1} + \beta\big(\mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}) - \mathbf{s}\big) + \mathbf{z}_2 \ = \ \mathbf{0}, && \text{(14g)}\\
\nabla_{\boldsymbol{\lambda}_3} \mathcal{L} \ &= \rho\boldsymbol{\sigma}^{\min} - \mathbf{s} + \mathbf{z}_3 \ = \ \mathbf{0}, && \text{(14h)}\\
\nabla_{\boldsymbol{\lambda}_4} \mathcal{L} \ &= \mathbf{s} - \rho\boldsymbol{\sigma}^{\max} + \mathbf{z}_4 \ = \ \mathbf{0}. && \text{(14i)}
\end{aligned}
$$

In (14b) and (14d) (and further on) the fractions are meant by components. Moreover, we use the identity

$$\int_\Omega \operatorname{div} \boldsymbol{\sigma} \cdot \mathbf{v} \ d\mathbf{x} = - \int_\Omega \boldsymbol{\sigma} : \boldsymbol{\varepsilon}(\mathbf{v}) \ d\mathbf{x}$$

for $\mathbf{v} \in H_0^1(\Omega)$ for the derivation of (14) and for the statement of the Lagrangian (13). Moreover, the equality (14b) again has to be understood in a weak sense. In the spirit of primal-dual interior point methods we now introduce new independent variables ν_i that act as multipliers for the bound constraints (11g) - (11i). In particular we choose $\nu_1, \nu_2 \in H^1(\Omega)$, $\boldsymbol{\nu}_3, \boldsymbol{\nu}_4 \in H^1(\Omega; \mathbb{R}^2)$, and $\boldsymbol{\nu}_i \in L_2(\Omega; \mathbb{R}^{2\times 2})$ for $i = 5, \dots, 8$, such that

$$
\begin{gathered}
\nu_1 = \frac{\mu}{\rho}, \quad \nu_2 = \frac{\mu}{1-\rho}, \quad \boldsymbol{\nu}_3 = \frac{\mu}{\mathbf{u} - \mathbf{u}^{\min}}, \quad \boldsymbol{\nu}_4 = \frac{\mu}{\mathbf{u}^{\max} - \mathbf{u}}, \\
\boldsymbol{\nu}_5 = \frac{\mu}{\mathbf{z}_1}, \quad \boldsymbol{\nu}_6 = \frac{\mu}{\mathbf{z}_2}, \quad \boldsymbol{\nu}_7 = \frac{\mu}{\mathbf{z}_3}, \quad \boldsymbol{\nu}_8 = \frac{\mu}{\mathbf{z}_4}.
\end{gathered}
\qquad (15)
$$

Using the definition (15) of the dual variables, the optimality conditions (14) turn into the following system in the primal variables ρ, $\mathbf{u}$, $\mathbf{s}$, $\mathbf{z}$, and the dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$:

$$
\begin{aligned}
\gamma + \epsilon\Delta\rho + \frac{1}{\epsilon}(1-2\rho) - \nu_1 + \nu_2 + \mathbf{1} : \boldsymbol{\lambda}_1 + \mathbf{1} : \boldsymbol{\lambda}_2 + \boldsymbol{\sigma}^{\min} : \boldsymbol{\lambda}_3 - \boldsymbol{\sigma}^{\max} : \boldsymbol{\lambda}_4 \;&=\; 0, && \text{(16a)}\\
-\boldsymbol{\nu}_3 + \boldsymbol{\nu}_4 + \beta\mathbf{C}\operatorname{div}\boldsymbol{\lambda}_1 - \beta\mathbf{C}\operatorname{div}\boldsymbol{\lambda}_2 \;&=\; \mathbf{0}, && \text{(16b)}\\
-\boldsymbol{\varepsilon}(\boldsymbol{\lambda}_0) + \beta\boldsymbol{\lambda}_1 - \beta\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_3 + \boldsymbol{\lambda}_4 \;&=\; \mathbf{0}, && \text{(16c)}\\
-\boldsymbol{\nu}_5 + \boldsymbol{\lambda}_1 \;=\; -\boldsymbol{\nu}_6 + \boldsymbol{\lambda}_2 \;=\; -\boldsymbol{\nu}_7 + \boldsymbol{\lambda}_3 \;=\; -\boldsymbol{\nu}_8 + \boldsymbol{\lambda}_4 \;&=\; \mathbf{0}, && \text{(16d)}\\
\operatorname{div}\mathbf{s} \;&=\; \mathbf{0}, && \text{(16e)}\\
-(1-\rho)\mathbf{1} - \beta\big(\mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}) - \mathbf{s}\big) + \mathbf{z}_1 \;&=\; \mathbf{0}, && \text{(16f)}\\
-(1-\rho)\mathbf{1} + \beta\big(\mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}) - \mathbf{s}\big) + \mathbf{z}_2 \;&=\; \mathbf{0}, && \text{(16g)}\\
\rho\boldsymbol{\sigma}^{\min} - \mathbf{s} + \mathbf{z}_3 \;&=\; \mathbf{0}, && \text{(16h)}\\
\mathbf{s} - \rho\boldsymbol{\sigma}^{\max} + \mathbf{z}_4 \;&=\; \mathbf{0}, && \text{(16i)}\\
-\rho + \frac{\mu}{\nu_1} \;&=\; 0, && \text{(16j)}\\
\rho - 1 + \frac{\mu}{\nu_2} \;&=\; 0, && \text{(16k)}\\
-\big(\mathbf{u} - \mathbf{u}^{\min}\big) + \frac{\mu}{\boldsymbol{\nu}_3} \;&=\; \mathbf{0}, && \text{(16l)}\\
-\big(\mathbf{u}^{\max} - \mathbf{u}\big) + \frac{\mu}{\boldsymbol{\nu}_4} \;&=\; \mathbf{0}, && \text{(16m)}\\
-\mathbf{z}_1 + \frac{\mu}{\boldsymbol{\nu}_5} \;=\; -\mathbf{z}_2 + \frac{\mu}{\boldsymbol{\nu}_6} \;=\; -\mathbf{z}_3 + \frac{\mu}{\boldsymbol{\nu}_7} \;=\; -\mathbf{z}_4 + \frac{\mu}{\boldsymbol{\nu}_8} \;&=\; \mathbf{0}, && \text{(16n)}
\end{aligned}
$$

where the form of the equalities (16j) - (16n) is motivated to get a symmetric system matrix after discretization.

In the following we consider the discretization of the primal-dual equations (16). In order to construct a finite element approximation we assume that $\overline{\Omega} = \bigcup_{i=1}^{n} \overline{\tau}_i$ is partitioned into a proper triangulation $\mathcal{T} = \{\tau_i \mid i = 1, \ldots, n\}$ with n triangles τ_i. We shall use two different finite elements for the primal and dual variables. For the density ρ, the components of the displacements $\mathbf{u}$, the dual variables ν_1, ν_2, and the components of the dual variables $\boldsymbol{\nu}_3$, $\boldsymbol{\nu}_4$, we use the discrete H^1-subspace of linear elements

$$V^h := \big\{\tilde{v} \in C(\Omega);\ \big|\ \tilde{v}|_{\tau_i} \in \mathcal{P}_1(\tau_i),\ i = 1, \ldots, n\big\}.$$

For the components of the Lagrangian multiplier $\boldsymbol{\lambda}_0$ we use the discrete H_0^1-subspace V_0^h of linear elements with zero boundary conditions. The components of the stress $\mathbf{s}$, the slack variables $\mathbf{z}$, the dual variables $\boldsymbol{\nu}_5, \ldots \boldsymbol{\nu}_8$, and of the Lagrange multipliers $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_4$ are approximated by the L_∞-subspace of constant elements

$$Q^h := \big\{\tilde{q} \in L_\infty(\Omega) \ \big|\ \tilde{q}|_{\tau_i} \in \mathcal{P}_0(\tau_i),\ i = 1, \ldots, n\big\}.$$

As in the previous chapters, $\mathcal{P}_k(\tau_i)$ represents the space of polynomials of maximal degree k over the triangle τ_i. Using these finite element approximations we discretize the system (16) by a standard finite element approach, i.e. we consider the weak formulations of the equations (16a) - (16n) and perform a partial integration for the divergence terms in the equations (16b) and (16e). Taking into account the Hilbert spaces $V = H^1(\Omega)$, $V_0 = H_0^1(\Omega)$, $V_{\Gamma_u} = H^1_{\Gamma_u}(\Omega)$, and $Q = L_2(\Omega)$ and having the application of a Newton type method in mind, we can write the weak linearized formulation of (16) in the following way: Find updates $\rho, \nu_1, \nu_2 \in V$, $\mathbf{u}, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4 \in V^2_{\Gamma_u}$, $\boldsymbol{\lambda}_0 \in V^2_0$, and $\mathbf{s}, \mathbf{z}_1, \ldots, \mathbf{z}_4, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_4, \boldsymbol{\nu}_5, \ldots, \boldsymbol{\nu}_8 \in Q^{2\times 2}$ for the current values of $\overline{\rho}, \overline{\mathbf{u}}, \overline{\mathbf{s}}, \overline{\mathbf{z}}, \overline{\boldsymbol{\lambda}}, \overline{\nu}_1, \overline{\nu}_2, \overline{\boldsymbol{\nu}}_3, \ldots, \overline{\boldsymbol{\nu}}_8$, respectively, such that

$$
\begin{aligned}
-\epsilon(\nabla\rho, \nabla v)_0 - \frac{2}{\epsilon}(\rho, v)_0 - (\nu_1, v)_0 + (\nu_2, v)_0 + (\mathbf{1} : \boldsymbol{\lambda}_1, v)_0 + \qquad & \\
+ (\mathbf{1} : \boldsymbol{\lambda}_2, v)_0 + (\boldsymbol{\sigma}^{\min} : \boldsymbol{\lambda}_3, v)_0 + (\boldsymbol{\sigma}^{\max} : \boldsymbol{\lambda}_4, v)_0 \;&=\; -\big(\gamma + \frac{1}{\epsilon}\big)(1, v)_0, \\
-(\boldsymbol{\nu}_3, \boldsymbol{\phi})_0 + (\boldsymbol{\nu}_4, \boldsymbol{\phi})_0 - \beta\big(\mathbf{C}\boldsymbol{\lambda}_1, \boldsymbol{\varepsilon}(\boldsymbol{\phi})\big)_0 + \beta\big(\mathbf{C}\boldsymbol{\lambda}_2, \boldsymbol{\varepsilon}(\boldsymbol{\phi})\big)_0 \;&=\; 0, \\
-\big(\boldsymbol{\varepsilon}(\boldsymbol{\lambda}_0), \mathbf{q}\big)_0 + \beta(\boldsymbol{\lambda}_1, \mathbf{q})_0 - \beta(\boldsymbol{\lambda}_2, \mathbf{q})_0 - (\boldsymbol{\lambda}_3, \mathbf{q})_0 + (\boldsymbol{\lambda}_4, \mathbf{q})_0 \;&=\; 0, \\
-(\boldsymbol{\nu}_5, \mathbf{q})_0 + (\boldsymbol{\lambda}_1, \mathbf{q})_0 \;=\; -(\boldsymbol{\nu}_6, \mathbf{q})_0 + (\boldsymbol{\lambda}_2, \mathbf{q})_0 \;=\; & \\
-(\boldsymbol{\nu}_7, \mathbf{q})_0 + (\boldsymbol{\lambda}_3, \mathbf{q})_0 \;=\; -(\boldsymbol{\nu}_8, \mathbf{q})_0 + (\boldsymbol{\lambda}_4, \mathbf{q})_0 \;&=\; 0, \\
-\big(\mathbf{s}, \boldsymbol{\varepsilon}(\boldsymbol{\psi})\big)_0 \;&=\; \int_{\Gamma_t} \mathbf{t} \cdot \boldsymbol{\psi} \, d\mathbf{x}, \\
(\rho\mathbf{1}, \mathbf{q})_0 - \beta\big(\mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}), \mathbf{q}\big)_0 + \beta(\mathbf{s}, \mathbf{q})_0 + (\mathbf{z}_1, \mathbf{q})_0 \;&=\; (\mathbf{1}, \mathbf{q})_0, \\
(\rho\mathbf{1}, \mathbf{q})_0 + \beta\big(\mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}), \mathbf{q}\big)_0 - \beta(\mathbf{s}, \mathbf{q})_0 + (\mathbf{z}_2, \mathbf{q})_0 \;&=\; (\mathbf{1}, \mathbf{q})_0, \\
(\rho\boldsymbol{\sigma}^{\min}, \mathbf{q})_0 - (\mathbf{s}, \mathbf{q})_0 + (\mathbf{z}_3, \mathbf{q})_0 \;&=\; 0, \\
(\mathbf{s}, \mathbf{q})_0 - (\rho\boldsymbol{\sigma}^{\max}, \mathbf{q})_0 + (\mathbf{z}_4, \mathbf{q})_0 \;&=\; 0, \\
-(\rho, v)_0 + \frac{\mu}{\overline{\nu}_1}(\nu_1, v)_0 \;&=\; 0, \\
(\rho, v)_0 + \frac{\mu}{\overline{\nu}_2}(\nu_2, v)_0 \;&=\; (1, v)_0, \\
-(\mathbf{u}, \boldsymbol{\phi})_0 + \frac{\mu}{\overline{\boldsymbol{\nu}}_3}(\boldsymbol{\nu}_3, \boldsymbol{\phi})_0 \;&=\; -(\mathbf{u}^{\min}, \boldsymbol{\phi})_0, \\
(\mathbf{u}, \boldsymbol{\phi})_0 + \frac{\mu}{\overline{\boldsymbol{\nu}}_4}(\boldsymbol{\nu}_4, \boldsymbol{\phi})_0 \;&=\; (\mathbf{u}^{\max}, \boldsymbol{\phi})_0, \\
-(\mathbf{z}_1, \mathbf{q})_0 + \frac{\mu}{\overline{\boldsymbol{\nu}}_5}(\boldsymbol{\nu}_5, \mathbf{q})_0 \;=\; -(\mathbf{z}_2, \mathbf{q})_0 + \frac{\mu}{\overline{\boldsymbol{\nu}}_6}(\boldsymbol{\nu}_6, \mathbf{q})_0 \;=\; & \\
=\; -(\mathbf{z}_3, \mathbf{q})_0 + \frac{\mu}{\overline{\boldsymbol{\nu}}_7}(\boldsymbol{\nu}_7, \mathbf{q})_0 \;=\; -(\mathbf{z}_4, \mathbf{q})_0 + \frac{\mu}{\overline{\boldsymbol{\nu}}_8}(\boldsymbol{\nu}_8, \mathbf{q})_0 \;&=\; 0,
\end{aligned}
$$

where the above equalities shall hold for all test functions $v \in V$, $\boldsymbol{\phi} \in V^2_0$, $\boldsymbol{\psi} \in V^2_{\Gamma_u}$, and $\mathbf{q} \in Q^{2\times 2}$. Let the vectors $\triangle\boldsymbol{\rho}^h$, $\triangle\mathbf{u}^h$, and so on, contain the coefficients of the finite element functions $\tilde{\rho} \in V^h$ and $\tilde{\mathbf{u}} \in (V^h_{\Gamma_u})^2$, and so on, respectively. We add the symbol $\triangle$ to emphasis that we consider the updates of current iterates in a Newton type iteration method. Moreover, we use the symmetries in the occurring variables $\mathbf{s}$, $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_4$, $\mathbf{z}_1, \ldots, \mathbf{z}_4$, and $\boldsymbol{\nu}_5, \ldots, \boldsymbol{\nu}_8$ (e.g. $s_{ij} = s_{ji}$) to reduce the number of unknowns. The discretized problem can now be written as:

$$
\begin{aligned}
-\epsilon \mathbf{K} \triangle \boldsymbol{\rho}^h - \frac{2}{\epsilon} \mathbf{M} \triangle \boldsymbol{\rho}^h - \mathbf{M} \triangle \boldsymbol{\nu}_1^h + \mathbf{M} \triangle \boldsymbol{\nu}_2^h + \tilde{\mathbf{N}}^T \triangle \boldsymbol{\lambda}_1^h + & \\
+ \tilde{\mathbf{N}}^T \triangle \boldsymbol{\lambda}_2^h + \tilde{\mathbf{N}}^T \boldsymbol{\Sigma}^{\min} \triangle \boldsymbol{\lambda}_3^h - \tilde{\mathbf{N}}^T \boldsymbol{\Sigma}^{\max} \triangle \boldsymbol{\lambda}_4^h &= -\big(\gamma + \frac{1}{\epsilon}\big) \mathbf{e}_{V^h}^h, & \text{(17a)} \\
-\mathbf{M}_2 \triangle \boldsymbol{\nu}_3^h + \mathbf{M}_2 \triangle \boldsymbol{\nu}_4^h - \beta \mathbf{D}^T \mathbf{C}^h \triangle \boldsymbol{\lambda}_1^h + \beta \mathbf{D}^T \mathbf{C}^h \triangle \boldsymbol{\lambda}_2^h &= \mathbf{0}, & \text{(17b)} \\
-\mathbf{D} \triangle \boldsymbol{\lambda}_0^h + \beta \mathbf{N} \triangle \boldsymbol{\lambda}_1^h - \beta \mathbf{N} \triangle \boldsymbol{\lambda}_2^h - \mathbf{N} \triangle \boldsymbol{\lambda}_3^h + \mathbf{N} \triangle \boldsymbol{\lambda}_4^h &= \mathbf{0}, & \text{(17c)} \\
-\mathbf{N} \triangle \boldsymbol{\nu}_5^h + \mathbf{N} \triangle \boldsymbol{\lambda}_1^h = -\mathbf{N} \triangle \boldsymbol{\nu}_6^h + \mathbf{N} \triangle \boldsymbol{\lambda}_2^h = & \\
= -\mathbf{N} \triangle \boldsymbol{\nu}_7^h + \mathbf{N} \triangle \boldsymbol{\lambda}_3^h = -\mathbf{N} \triangle \boldsymbol{\nu}_8^h + \mathbf{N} \triangle \boldsymbol{\lambda}_4^h &= \mathbf{0}, & \text{(17d)} \\
-\mathbf{D}^T \triangle \mathbf{s}^h &= \mathbf{t}^h, & \text{(17e)} \\
\tilde{\mathbf{N}} \triangle \boldsymbol{\rho}^h - \beta \mathbf{C}^h \mathbf{D} \triangle \mathbf{u}^h + \beta \mathbf{N} \triangle \mathbf{s}^h + \mathbf{N} \triangle \mathbf{z}_1^h &= \mathbf{e}_{(Q^h)^3}^h, & \text{(17f)} \\
\tilde{\mathbf{N}} \triangle \boldsymbol{\rho}^h + \beta \mathbf{C}^h \mathbf{D} \triangle \mathbf{u}^h - \beta \mathbf{N} \triangle \mathbf{s}^h + \mathbf{N} \triangle \mathbf{z}_2^h &= \mathbf{e}_{(Q^h)^3}^h, & \text{(17g)} \\
\boldsymbol{\Sigma}^{\min} \tilde{\mathbf{N}} \triangle \boldsymbol{\rho}^h - \mathbf{N} \triangle s^h + \mathbf{N} \triangle \mathbf{z}_3^h &= \mathbf{0}, & \text{(17h)} \\
-\boldsymbol{\Sigma}^{\max} \tilde{\mathbf{N}} \triangle \boldsymbol{\rho}^h + \mathbf{N} \triangle \mathbf{s}^h + \mathbf{N} \triangle \mathbf{z}_4^h &= \mathbf{0}, & \text{(17i)} \\
-\mathbf{M} \triangle \boldsymbol{\rho}^h + \mu \mathbf{M}_{\nu_1} \triangle \boldsymbol{\nu}_1^h &= \mathbf{0}, & \text{(17j)} \\
\mathbf{M} \triangle \boldsymbol{\rho}^h + \mu \mathbf{M}_{\nu_2} \triangle \boldsymbol{\nu}_2^h &= \mathbf{e}_{V^h}^h, & \text{(17k)} \\
-\mathbf{M}_2 \triangle \mathbf{u}^h + \mu \mathbf{M}_{\nu_3} \triangle \boldsymbol{\nu}_3 &= -\mathbf{M}_2 \mathbf{u}^{\min h}, & \text{(17l)} \\
\mathbf{M}_2 \triangle \mathbf{u}^h + \mu \mathbf{M}_{\nu_4} \triangle \boldsymbol{\nu}_4 &= \mathbf{M}_2 \mathbf{u}^{\max h}, & \text{(17m)} \\
-\mathbf{N} \triangle \mathbf{z}_1^h + \mu \mathbf{N}_{\nu_5} \triangle \boldsymbol{\nu}_5^h = -\mathbf{N} \triangle \mathbf{z}_2^h + \mu \mathbf{N}_{\nu_6} \triangle \boldsymbol{\nu}_6^h = & \\
-\mathbf{N} \triangle \mathbf{z}_3^h + \mu \mathbf{N}_{\nu_7} \triangle \boldsymbol{\nu}_7^h = -\mathbf{N} \triangle \mathbf{z}_4^h + \mu \mathbf{N}_{\nu_8} \triangle \boldsymbol{\nu}_8^h &= \mathbf{0}. & \text{(17n)}
\end{aligned}
$$

In (17a), $\mathbf{K}$ is a stiffness matrix arising from the finite element discretization of the Laplacian in V^h and $\mathbf{M}$ is a mass matrix for the identity in V^h. Furthermore, $\tilde{\mathbf{N}}$ is mixed mass matrix between the spaces V^h and $(Q^h)^3$. $\mathbf{e}_{V^h}^h$ and $\mathbf{e}_{(Q^h)^3}^h$ are vectors representing the coefficients of the constant function 1 with respect to the spaces V^h and $(Q^h)^3$, respectively. $\boldsymbol{\Sigma}^{\min}$ and $\boldsymbol{\Sigma}^{\max}$ are diagonal matrices representing the corresponding entries of $\boldsymbol{\sigma}^{\min}$ and $\boldsymbol{\sigma}^{\max}$, respectively. In (17b), $\mathbf{M}_2$ is a mass matrix for the identity in $V_{\Gamma_u}^h \times V_{\Gamma_u}^h$ and $\mathbf{C}^h$ is the discrete analogon of elasticity tensor $\mathbf{C}$. Moreover, $\mathbf{D}^T$ is the representation of the divergence operator (restricted to symmetric stress tensors). The mass matrix $\mathbf{N}$ in (17c) represents the mass matrix for the identity in $(Q^h)^3$. In the discretized partial differential equation (17e) $\mathbf{t}^h$ is a discrete representation of the traction forces. Moreover, in the equations (17j) - (17m) the matrices $\mathbf{M}_{\nu_1}, \ldots, \mathbf{M}_{\nu_4}$, and the matrices $\mathbf{N}_{\nu_5}, \ldots, \mathbf{N}_{\nu_8}$, in (17n) are weighted mass matrices with the weights $\nu_1, \ldots, \boldsymbol{\nu}_8$, respectively.

The linear system (17) can be written in a compact representation as

$$
\boldsymbol{\mathcal{K}} \triangle \mathbf{x} = \mathbf{f}^h, \tag{18}
$$

with

$$
\triangle \mathbf{x} = \big(\triangle \boldsymbol{\rho}^h, \triangle \mathbf{u}^h, \triangle \mathbf{s}^h, \triangle \mathbf{z}_1^h, \ldots, \triangle \mathbf{z}_4^h, \triangle \boldsymbol{\lambda}_0^h, \ldots, \triangle \boldsymbol{\lambda}_4^h, \triangle \boldsymbol{\nu}_1^h, \ldots, \triangle \boldsymbol{\nu}_8^h\big)
$$

and

$$
\begin{aligned}
\mathbf{f}^h \ = \Big(& -(\gamma + \frac{1}{\epsilon})\mathbf{e}^h_{V^h}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{t}^h, \mathbf{e}^h_{(Q^h)^3}, \mathbf{e}^h_{(Q^h)^3}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \\
& \mathbf{e}^h_{V^h}, -\mathbf{M}_2\mathbf{u}^{\min h}, \mathbf{M}_2\mathbf{u}^{\max h}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}\Big).
\end{aligned}
$$

The coefficient matrix $\boldsymbol{\mathcal{K}}$ in (18) contains the matrices in (17) as block matrices and turns out to be large (even too large to be printed on one page). In order to reduce the size of the system to a more reasonable one, we reduce the system (but we keep the notation $\boldsymbol{\mathcal{K}}$ for the system matrix and $\mathbf{f}^h$ for the right-hand side after each of the following eliminations) using the following eliminations of the dual variables:

$$
\begin{aligned}
\triangle\boldsymbol{\nu}_1^h &= \frac{1}{\mu}\mathbf{M}_{\nu_1}^{-1}\mathbf{M}\triangle\boldsymbol{\rho}^h, \qquad \triangle\boldsymbol{\nu}_2^h = -\frac{1}{\mu}\mathbf{M}_{\nu_2}^{-1}\mathbf{M}\triangle\boldsymbol{\rho}^h + \frac{1}{\mu}\mathbf{M}_{\nu_2}^{-1}\mathbf{e}^h_{V^h}, \\
\triangle\boldsymbol{\nu}_3^h &= \frac{1}{\mu}\mathbf{M}_{\nu_3}^{-1}\mathbf{M}_2\triangle\mathbf{u}^h - \frac{1}{\mu}\mathbf{M}_{\nu_3}^{-1}\mathbf{M}_2\mathbf{u}^{\min h}, \\
\triangle\boldsymbol{\nu}_4^h &= -\frac{1}{\mu}\mathbf{M}_{\nu_4}^{-1}\mathbf{M}_2\triangle\mathbf{u}^h + \frac{1}{\mu}\mathbf{M}_{\nu_4}^{-1}\mathbf{M}_2\mathbf{u}^{\max h}, \\
\triangle\boldsymbol{\nu}_5^h &= \frac{1}{\mu}\mathbf{N}_{\nu_5}^{-1}\mathbf{N}\triangle\mathbf{z}_1^h, \qquad \triangle\boldsymbol{\nu}_6^h = \frac{1}{\mu}\mathbf{N}_{\nu_6}^{-1}\mathbf{N}\triangle\mathbf{z}_2^h, \\
\triangle\boldsymbol{\nu}_7^h &= \frac{1}{\mu}\mathbf{N}_{\nu_7}^{-1}\mathbf{N}\triangle\mathbf{z}_3^h, \qquad \triangle\boldsymbol{\nu}_8^h = \frac{1}{\mu}\mathbf{N}_{\nu_8}^{-1}\mathbf{N}\triangle\mathbf{z}_4^h.
\end{aligned}
\tag{19}
$$

This first elimination yields a smaller linear system like

$$
\boldsymbol{\mathcal{K}}\triangle\mathbf{x} = \mathbf{f}^h, \tag{20}
$$

with

$$
\triangle\mathbf{x} = \big(\triangle\boldsymbol{\rho}^h, \triangle\mathbf{u}^h, \triangle\mathbf{s}^h, \triangle\mathbf{z}_1^h, \ldots, \triangle\mathbf{z}_4^h, \triangle\boldsymbol{\lambda}_0^h, \ldots, \triangle\boldsymbol{\lambda}_4^h\big)
$$

and

$$
\begin{aligned}
\mathbf{f}^h \ = \Big(& -(\gamma + \frac{1}{\epsilon})\mathbf{e}^h_{V^h} - \frac{1}{\mu}\mathbf{M}\mathbf{M}_{\nu_2}^{-1}\mathbf{e}^h_{V^h}, -\frac{1}{\mu}\mathbf{M}_2\big(\mathbf{M}_{\nu_3}^{-1}\mathbf{u}^{\min h} + \mathbf{M}_{\nu_4}^{-1}\mathbf{u}^{\max h}\big), \\
& \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{t}^h, \mathbf{e}^h_{(Q^h)^3}, \mathbf{e}^h_{(Q^h)^3}, \mathbf{0}, \mathbf{0}\Big).
\end{aligned}
$$

The system matrix $\boldsymbol{\mathcal{K}}$ of (20) is given by

$$
\boldsymbol{\mathcal{K}} = \begin{pmatrix}
\boldsymbol{\mathcal{K}}_{\rho\rho} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{\mathbf{N}}^T & \tilde{\mathbf{N}}^T & \tilde{\mathbf{N}}^T\boldsymbol{\Sigma}^{\min} & -\tilde{\mathbf{N}}^T\boldsymbol{\Sigma}^{\max} \\
\mathbf{0} & \boldsymbol{\mathcal{K}}_{uu} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\beta\mathbf{D}^T\mathbf{C}^h & \beta\mathbf{D}^T\mathbf{C}^h & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{D} & \beta\mathbf{N} & -\beta\mathbf{N} & -\mathbf{N} & \mathbf{N} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{K}}_{z_1z_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{N} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{K}}_{z_2z_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{N} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{K}}_{z_3z_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{N} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{K}}_{z_4z_4} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{N} \\
\mathbf{0} & \mathbf{0} & -\mathbf{D}^T & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\tilde{\mathbf{N}} & -\beta\mathbf{C}^h\mathbf{D} & \beta\mathbf{N} & \mathbf{N} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\tilde{\mathbf{N}} & \beta\mathbf{C}^h\mathbf{D} & -\beta\mathbf{N} & \mathbf{0} & \mathbf{N} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\boldsymbol{\Sigma}^{\min}\tilde{\mathbf{N}} & \mathbf{0} & -\mathbf{N} & \mathbf{0} & \mathbf{0} & \mathbf{N} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
-\boldsymbol{\Sigma}^{\max}\tilde{\mathbf{N}} & \mathbf{0} & \mathbf{N} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{N} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{pmatrix},
$$

with

$$
\begin{aligned}
\boldsymbol{\mathcal{K}}_{\rho\rho} \ &= \ -\epsilon\mathbf{K} - \frac{2}{\epsilon}\mathbf{M} - \frac{1}{\mu}\mathbf{M}\big(\mathbf{M}_{\nu_1}^{-1} + \mathbf{M}_{\nu_2}^{-1}\big)\mathbf{M}, \\
\boldsymbol{\mathcal{K}}_{uu} \ &= \ -\frac{1}{\mu}\mathbf{M}_2\big(\mathbf{M}_{\nu_3}^{-1} + \mathbf{M}_{\nu_4}^{-1}\big)\mathbf{M}_2, \\
\boldsymbol{\mathcal{K}}_{z_iz_i} \ &= \ -\frac{1}{\mu}\mathbf{N}\mathbf{N}_{\nu_i}^{-1}\mathbf{N}, \qquad i = 1, \ldots, 4.
\end{aligned}
$$

We further eliminate we eliminate the slack variables $\triangle \mathbf{z}_i^h$, for $i = 1, \ldots, 4$, using (20):

$$\triangle \mathbf{z}_i^h = \mu \mathbf{N}^{-1} \mathbf{N}_{\nu_i} \triangle \boldsymbol{\lambda}_i^h, \qquad i = 1, \ldots, 4. \tag{21}$$

to we come up with the following system

$$\mathcal{K} \begin{pmatrix} \triangle \boldsymbol{\rho}^h \\ \triangle \mathbf{u}^h \\ \triangle \mathbf{s}^h \\ \triangle \boldsymbol{\lambda}_0^h \\ \triangle \boldsymbol{\lambda}_1^h \\ \triangle \boldsymbol{\lambda}_2^h \\ \triangle \boldsymbol{\lambda}_3^h \\ \triangle \boldsymbol{\lambda}_4^h \end{pmatrix} = \begin{pmatrix} -\big(\gamma + \frac{1}{\epsilon}\big)\mathbf{e}_{V^h}^h - \frac{1}{\mu}\mathbf{M}\mathbf{M}_{\nu_2}^{-1}\mathbf{e}_{V^h}^h \\ -\frac{1}{\mu}\mathbf{M}_2\big(\mathbf{M}_{\nu_3}^{-1}\mathbf{u}^{\min h} + \mathbf{M}_{\nu_4}^{-1}\mathbf{u}^{\max h}\big) \\ \mathbf{0} \\ \mathbf{t}^h \\ \mathbf{e}_{(Q^h)^3}^h \\ \mathbf{e}_{(Q^h)^3}^h \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \tag{22}$$

with the coefficient matrix

$$\mathcal{K} = \begin{pmatrix} \mathcal{K}_{\rho\rho} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{\mathbf{N}}^T & \tilde{\mathbf{N}}^T & \tilde{\mathbf{N}}^T\boldsymbol{\Sigma}^{\min} & -\tilde{\mathbf{N}}^T\boldsymbol{\Sigma}^{\max} \\ \mathbf{0} & \mathcal{K}_{uu} & \mathbf{0} & \mathbf{0} & -\beta\mathbf{D}^T\mathbf{C}^h & \beta\mathbf{D}^T\mathbf{C}^h & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{D} & \beta\mathbf{N} & -\beta\mathbf{N} & -\mathbf{N} & \mathbf{N} \\ \mathbf{0} & \mathbf{0} & -\mathbf{D}^T & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \tilde{\mathbf{N}} & -\beta\mathbf{C}^h\mathbf{D} & \beta\mathbf{N} & \mathbf{0} & \mu\mathbf{N}_{\nu_5} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \tilde{\mathbf{N}} & \beta\mathbf{C}^h\mathbf{D} & -\beta\mathbf{N} & \mathbf{0} & \mathbf{0} & \mu\mathbf{N}_{\nu_6} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Sigma}^{\min}\tilde{\mathbf{N}} & \mathbf{0} & -\mathbf{N} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mu\mathbf{N}_{\nu_7} & \mathbf{0} \\ -\boldsymbol{\Sigma}^{\max}\tilde{\mathbf{N}} & \mathbf{0} & \mathbf{N} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mu\mathbf{N}_{\nu_8} \end{pmatrix}.$$

As a last reduction we eliminate the updates concerning the Lagrange multipliers $\boldsymbol{\lambda}_1, \ldots \boldsymbol{\lambda}_4$, from the linear system (22) using

$$\begin{aligned} \triangle \boldsymbol{\lambda}_1^h &= \frac{1}{\mu}\mathbf{N}_{\nu_5}^{-1}\mathbf{e}_{(Q^h)^3}^h - \frac{1}{\mu}\mathbf{N}_{\nu_5}^{-1}\tilde{\mathbf{N}}\triangle\boldsymbol{\rho}^h + \frac{\beta}{\mu}\mathbf{N}_{\nu_5}^{-1}\mathbf{C}^h\mathbf{D}\triangle\mathbf{u}^h - \frac{\beta}{\mu}\mathbf{N}_{\nu_5}^{-1}\mathbf{N}\triangle\mathbf{s}^h, \\ \triangle \boldsymbol{\lambda}_2^h &= \frac{1}{\mu}\mathbf{N}_{\nu_6}^{-1}\mathbf{e}_{(Q^h)^3}^h - \frac{1}{\mu}\mathbf{N}_{\nu_6}^{-1}\tilde{\mathbf{N}}\triangle\boldsymbol{\rho}^h - \frac{\beta}{\mu}\mathbf{N}_{\nu_6}^{-1}\mathbf{C}^h\mathbf{D}\triangle\mathbf{u}^h + \frac{\beta}{\mu}\mathbf{N}_{\nu_6}^{-1}\mathbf{N}\triangle\mathbf{s}^h, \\ \triangle \boldsymbol{\lambda}_3^h &= -\frac{1}{\mu}\boldsymbol{\Sigma}^{\min}\mathbf{N}_{\nu_7}^{-1}\tilde{\mathbf{N}}\triangle\boldsymbol{\rho}^h + \frac{1}{\mu}\mathbf{N}_{\nu_7}^{-1}\mathbf{N}\triangle\mathbf{s}^h, \\ \triangle \boldsymbol{\lambda}_4^h &= \frac{1}{\mu}\boldsymbol{\Sigma}^{\max}\mathbf{N}_{\nu_8}^{-1}\tilde{\mathbf{N}}\triangle\boldsymbol{\rho}^h - \frac{1}{\mu}\mathbf{N}_{\nu_8}^{-1}\mathbf{N}\triangle\mathbf{s}^h. \end{aligned} \tag{23}$$

This elimination finally results in the symmetric saddle point problem

$$\begin{pmatrix} \mathcal{K}_{\rho\rho} & \mathcal{K}_{\rho u} & \mathcal{K}_{\rho s} & \mathbf{0} \\ \mathcal{K}_{\rho u}^T & \mathcal{K}_{uu} & \mathcal{K}_{us} & \mathbf{0} \\ \mathcal{K}_{\rho s}^T & \mathcal{K}_{us}^T & \mathcal{K}_{ss} & \mathbf{D}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{D} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \triangle\boldsymbol{\rho}^h \\ \triangle\mathbf{u}^h \\ \triangle\mathbf{s}^h \\ \triangle\boldsymbol{\lambda}_0^h \end{pmatrix} = \mathbf{f}^h, \tag{24}$$

with

$$\mathbf{f}^h = \begin{pmatrix} -\big(\gamma + \frac{1}{\epsilon}\big)\mathbf{e}_{V^h}^h - \frac{1}{\mu}\mathbf{M}\mathbf{M}_{\nu_2}^{-1}\mathbf{e}_{V^h}^h - \frac{1}{\mu}\tilde{\mathbf{N}}^T\big(\mathbf{N}_{\nu_5}^{-1} + \mathbf{N}_{\nu_6}^{-1}\big)\mathbf{e}_{(Q^h)^3}^h \\ -\frac{1}{\mu}\mathbf{M}_2\big(\mathbf{M}_{\nu_3}^{-1}\mathbf{u}^{\min h} + \mathbf{M}_{\nu_4}^{-1}\mathbf{u}^{\max h}\big) - \frac{\beta}{\mu}\mathbf{C}^h\mathbf{D}\big(\mathbf{N}_{\nu_5}^{-1} - \mathbf{N}_{\nu_6}^{-1}\big)\mathbf{e}_{(Q^h)^3}^h \\ -\frac{\beta}{\mu}\mathbf{N}_3\big(\mathbf{N}_{\nu_5}^{-1} - \mathbf{N}_{\nu_6}^{-1}\big)\mathbf{e}_{(Q^h)^3}^h \\ \mathbf{0} \end{pmatrix}$$

and the final block matrices

$$
\begin{aligned}
\mathcal{K}_{\rho\rho} &= -\epsilon\mathbf{K} - \frac{2}{\epsilon}\mathbf{M} - \frac{1}{\mu}\mathbf{M}\big(\mathbf{M}_{\nu_1}^{-1} + \mathbf{M}_{\nu_2}^{-1}\big)\mathbf{M} - \\
&\quad - \frac{1}{\mu}\tilde{\mathbf{N}}^T\big(\mathbf{N}_{\nu_5}^{-1} + \mathbf{N}_{\nu_6}^{-1} + \mathbf{\Sigma}^{\min 2}\mathbf{N}_{\nu_7}^{-1} + \mathbf{\Sigma}^{\max 2}\mathbf{N}_{\nu_8}^{-1}\big)\tilde{\mathbf{N}}, \\
\mathcal{K}_{uu} &= -\frac{1}{\mu}\mathbf{M}_2\big(\mathbf{M}_{\nu_3}^{-1} + \mathbf{M}_{\nu_4}^{-1}\big)\mathbf{M}_2 - \frac{\beta^2}{\mu}\mathbf{D}^T\mathbf{C}^{h2}\big(\mathbf{N}_{\nu_5}^{-1} + \mathbf{N}_{\nu_6}^{-1}\big)\mathbf{D}, \\
\mathcal{K}_{ss} &= -\frac{1}{\mu}\mathbf{N}\big(\beta^2\mathbf{N}_{\nu_5}^{-1} + \beta^2\mathbf{N}_{\nu_6}^{-1} + \mathbf{N}_{\nu_7}^{-1} + \mathbf{N}_{\nu_8}^{-1}\big)\mathbf{N}, \\
\mathcal{K}_{\rho u} &= \frac{\beta}{\mu}\tilde{\mathbf{N}}^T\big(\mathbf{N}_{\nu_5}^{-1} - \mathbf{N}_{\nu_6}^{-1}\big)\mathbf{C}^h\mathbf{D}^T, \\
\mathcal{K}_{\rho s} &= \frac{1}{\mu}\tilde{\mathbf{N}}^T\big(\beta\mathbf{N}_{\nu_6}^{-1} - \beta\mathbf{N}_{\nu_5}^{-1} + \mathbf{\Sigma}^{\min}\mathbf{N}_{\nu_7}^{-1} + \mathbf{\Sigma}^{\max}\mathbf{N}_{\nu_8}^{-1}\big)\mathbf{N}, \\
\mathcal{K}_{us} &= \frac{\beta^2}{\mu}\mathbf{C}^h\mathbf{D}^T\big(\mathbf{N}_{\nu_5}^{-1} + \mathbf{N}_{\nu_6}^{-1}\big)\mathbf{N}.
\end{aligned} \tag{25}
$$

The linear system (24) yields a solution in the variables $\triangle\boldsymbol{\rho}^h, \triangle\mathbf{u}^h, \triangle\mathbf{s}^h, \triangle\boldsymbol{\lambda}_0^h$. Using this solution, the other variables are determined by the substitutions (23), (21), and finally (23).

4 A Multigrid KKT Solver

In this section we consider (additive and multiplicative) Schwarz-type iteration methods as smoothers in a multigrid method for saddle point problems. Each iteration step of such a Schwarz-type smoother consists of the solution of several small local saddle point problems in a Jacobi- or Gauss-Seidel-type manner. The computational domain is therefore divided into overlapping cells, also called *patches*. One iteration step of a Schwarz-type smoother consists now of solving a local saddle point problem for each patch. This is done in a Jacobi- or Gauß-Seidel-type manner and thus, called additive or multiplicative Schwarz-type smoother.

To begin with, we state the two most basic iterative methods for a linear system

$$\mathbf{K}\mathbf{u} = \mathbf{f},$$

which are used as smoothing methods, namely the Jacobi- and the Gauss-Seidel iterations, being the origins of the additive and multiplicative Schwarz methods, respectively. In Algorithm 3 we present the *Jacobi* iteration. The algorithm is simple, but with the disadvantage of slow convergence (note the analogy to the Richardson iteration). We state the Jacobi iteration without any consideration about convergence, but refer e.g. to Jung and Langer [10]. One criterion for the convergence of the damped Jacobi iteration is the symmetry and positive definiteness of the system matrix $\mathbf{K}$. A similar method to the Jacobi iteration is the *Gauss-Seidel* method. In difference to the Jacobi iteration we use for the computation of the i-th component u_i^k the already updated components u_j^k, for $j = 1, \dots, i-1$, in iteration k. The Gauß-Seidel method is presented in Algorithm 4. Again, the Gauss-Seidel iteration converges if the system matrix $\mathbf{K}$ is symmetric and positive definite. For more information we refer again e.g. to Jung and Langer [10] or to Hackbusch [8]. Both iteration methods (Jacobi iterations after under-relaxation, damped Jacobi) have smoothing properties in the sense that they reduce the high frequency part of the error components and are cheap to apply.

Algorithm 3 Damped Jacobi iteration

Choose a damping parameter τ, $0 < \tau < \frac{2}{\lambda_{\max}(\mathrm{diag}(\mathbf{K})^{-1}\mathbf{K})}$.
Choose a relative error bound $\varepsilon > 0$.
Initialize start value $\mathbf{u}^0 \in \mathbb{R}^n$.
$k = 0$;

while not converged **do**
  for $i = 1, \ldots, n$ **do**
$$u_i^{k+1} = (1-\tau)u_i^k + \frac{\tau}{K_{ii}}\left(f_i - \sum_{\substack{j=1\\ j\neq i}}^{n} K_{ij}u_j^k\right) \;;$$
  end for
  $k = k + 1$;
end while

Algorithm 4 Gauß-Seidel iteration

Choose a relative error bound $\varepsilon > 0$.
Initialize start value $\mathbf{u}^0 \in \mathbb{R}^n$.
$k = 0$;

while not converged **do**
$$u_1^{k+1} = \frac{1}{K_{11}}\left(f_1 - \sum_{j=2}^{n} K_{1j}u_j^k\right) \;;$$
  for $i = 2, \ldots, n-1$ **do**
$$u_i^{k+1} = \frac{1}{K_{ii}}\left(f_i - \sum_{j=1}^{i-1} K_{ij}u_j^{k+1} - \sum_{j=i+1}^{n} K_{ij}u_j^k\right) \;;$$
  end for
$$u_n^{k+1} = \frac{1}{K_{nn}}\left(f_n - \sum_{j=1}^{n-1} K_{nj}u_j^k\right) \;;$$
  $k = k + 1$;
end while

Note, that the Gauß-Seidel iteration depends on the ordering of the unknowns and that the Jacobi iteration is independent of the ordering of the unknowns, see e.g. Hackbusch [9]. In order to get a symmetric multigrid operator the post smoothing has to be arranged in a backward fashion for the Gauß-Seidel-type iteration. Moreover, the same number of pre- and post-smoothing steps has to be used.

The above definitions of the iteration methods would lead to *pointwise* methods. Below we shall define smoothing operators in terms of subspace decompositions, which will lead to a *blockwise* iteration method. These procedures are related to overlapping domain decomposition algorithms and to the classical Schwarz method. They are generalizations of Jacobi and Gauß-Seidel iteration procedures.

We start to introduce the Schwarz-type smoothers in an abstract framework of mixed variational problems (cf. Brezzi and Fortin [4]). For this sake let V and Q be Hilbert spaces and let $a(\cdot,\cdot) : V \times V \to \mathbb{R}$, $b(\cdot,\cdot) : V \times Q \to \mathbb{R}$, and $c(\cdot,\cdot) : Q \times Q \to \mathbb{R}$ be continuous bilinear forms. Furthermore, let $f(\cdot) : V \to \mathbb{R}$ and $g(\cdot) : Q \to \mathbb{R}$ be continuous linear functionals. Then we can formulate the following mixed variational problem: Find $u \in V$ and $p \in Q$ such that

$$\begin{array}{rcrcll} a(u,v) & + & b(v,p) & = & f(v), & \forall\, v \in V, \\ b(u,q) & - & c(p,q) & = & g(q), & \forall\, q \in Q. \end{array} \tag{26}$$

Following on the framework of multigrid methods we would introduce now a hierarchy of finite element spaces $V_0 \subset \ldots \subset V_l \subset V$, $Q_0 \subset \ldots \subset Q_l \subset Q$ on a corresponding hierarchy of increasingly finer meshes and so on. But since the smoothing procedure involves only one level of the sequence of spaces, we will omit these notations and fix one level i, $0 < i < l$. For simplification of notation we will also drop the subindex k when denoting spaces, matrices and so on. Following a standard finite element discretization let the vectors $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{q} \in \mathbb{R}^n$ contain the coefficients of the corresponding finite element functions with respect to some bases of V and Q. Moreover, we introduce the matrix representation of the mixed variational problem (26):

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix}. \tag{27}$$

For our specific problem (24) it turns out that the the above block matrices are given by

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\mathcal{K}}_{\rho\rho} & \boldsymbol{\mathcal{K}}_{\rho u} & \boldsymbol{\mathcal{K}}_{\rho s} \\ \boldsymbol{\mathcal{K}}_{\rho u}^T & \boldsymbol{\mathcal{K}}_{uu} & \boldsymbol{\mathcal{K}}_{us} \\ \boldsymbol{\mathcal{K}}_{\rho s}^T & \boldsymbol{\mathcal{K}}_{us}^T & \boldsymbol{\mathcal{K}}_{ss} \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} \mathbf{0}, \mathbf{0}, \mathbf{D} \end{pmatrix}, \qquad \text{and} \qquad \mathbf{C} = \mathbf{0}.$$

In the sequel we will again abbreviate the system matrix with $\boldsymbol{\mathcal{K}}$, i.e.

$$\boldsymbol{\mathcal{K}} = \begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\mathbf{C} \end{pmatrix}.$$

As a consequence from the properties from the bilinear forms, we assume that $\mathbf{A}$ is a symmetric positive semi-definite $n \times n$ matrix, $\mathbf{C}$ is a symmetric positive semi-definite $m \times m$ matrix, that $\mathbf{B}$ is a $m \times n$ matrix, and that $\boldsymbol{\mathcal{K}}$ is regular.

We shall start with a decomposition of the spaces

$$V = \sum_{i=1}^{l} V^i \qquad \text{and} \qquad Q = \sum_{i=1}^{l} Q^i.$$

Before we define the additive and multiplicative smoother we have to introduce linear operators for each subspace to set up the local sub-problems:

$$\mathbf{P}_{V\,i} : \mathbb{R}^{n_i} \to \mathbb{R}^n \qquad \text{and} \qquad \mathbf{P}_{Q_i} : \mathbb{R}^{m_i} \to \mathbb{R}^m, \qquad \text{for } i = 1, \ldots, l, \tag{28}$$

with n_i, m_i denoting the dimensions of the local subspaces V_i and Q_i, respectively. The matrices $\mathbf{P}_{V\,i}$ and $\mathbf{P}_{Q_i}$ denote prolongation operators with the associated restriction operators $\mathbf{P}_{V_i}^T$ and $\mathbf{P}_{Q_i}^T$, respectively. Furthermore, let the operators (28) satisfy the conditions

$$\sum_{i=1}^{l} \mathbf{P}_{V\,i} \mathbf{P}_{V_i}^T = \mathbf{I} \qquad \text{and} \qquad \sum_{i=1}^{l} \mathbf{P}_{Q_i} \mathbf{P}_{Q_i}^T \text{ is regular.} \tag{29}$$

With these preliminaries we can now define two Schwarz-type smoothers assuming that $\mathbf{u}^k$ and $\mathbf{p}^k$ are some approximations for the exact solutions $\mathbf{u}$ and $\mathbf{p}$ of (27).

The first one will be called an *additive Schwarz smoother* and is defined by

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \sum_{i=1}^{l} \mathbf{P}_{V\,i}\mathbf{v}_i, \qquad \mathbf{p}^{k+1} = \mathbf{p}^k + \sum_{i=1}^{l} \mathbf{P}_{Q\,i}\mathbf{q}_i,$$

with $\mathbf{v}_i$ and $\mathbf{q}_i$, $i = 1, \ldots, l$, solving the local saddle point problem

$$\begin{pmatrix} \hat{\mathbf{A}}_i & \mathbf{B}_i^T \\ \mathbf{B}_i & \mathbf{B}_i\hat{\mathbf{A}}_i^{-1}\mathbf{B}_i^T - \hat{\mathbf{S}}_i \end{pmatrix} \begin{pmatrix} \mathbf{v}_i \\ \mathbf{q}_i \end{pmatrix} = \begin{pmatrix} \mathbf{P}_{V\,i}^T(\mathbf{f} - \mathbf{A}\mathbf{u}^k - \mathbf{B}^T\mathbf{p}^k) \\ \mathbf{P}_{Q\,i}^T(\mathbf{g} - \mathbf{B}\mathbf{u}^k + \mathbf{C}\mathbf{p}^k) \end{pmatrix},$$

where $\hat{\mathbf{S}}_i = \frac{1}{\tau}(\mathbf{C}_i + \mathbf{B}_i\hat{\mathbf{A}}_i^{-1}\mathbf{B}_i^T)$, with some damping parameter $\tau > 0$. Thus, the actual residuum is restricted to the smaller spaces. Then the local saddle point problems are solved for all patches, and the solutions are finally prolongated back onto the whole space. This Jacobi-type process can be seen as an additive Schwarz method and the corresponding smoothing operator $\boldsymbol{\mathcal{S}}_A$ can be written as

$$\boldsymbol{\mathcal{S}}_A \begin{pmatrix} \mathbf{u}^k \\ \mathbf{p}^k \end{pmatrix} = \begin{pmatrix} \mathbf{u}^k \\ \mathbf{p}^k \end{pmatrix} + \sum_{i=1}^{l} \mathbf{P}_i\hat{\boldsymbol{\mathcal{K}}}_i^{-1}\mathbf{P}_i^T \left(\begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} - \boldsymbol{\mathcal{K}} \begin{pmatrix} \mathbf{u}^k \\ \mathbf{p}^k \end{pmatrix} \right),$$

where we used the abbreviations

$$\hat{\boldsymbol{\mathcal{K}}}_i = \begin{pmatrix} \hat{\mathbf{A}}_i & \mathbf{B}_i^T \\ \mathbf{B}_i & \mathbf{B}_i\hat{\mathbf{A}}_i^{-1}\mathbf{B}_i^T - \hat{\mathbf{S}}_i \end{pmatrix} \qquad \text{and} \qquad \mathbf{P}_i = \begin{pmatrix} \mathbf{P}_{V\,i} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{Q\,i} \end{pmatrix}.$$

Moreover, we define the *multiplicative Schwarz smoother* based on the above subspace decomposition as the following procedure: Set $\mathbf{w}^0 = \mathbf{0}$ and $\mathbf{r}^0 = \mathbf{0}$ and compute

$$\begin{pmatrix} \mathbf{w}^i \\ \mathbf{r}^i \end{pmatrix} = \begin{pmatrix} \mathbf{w}^{i-1} \\ \mathbf{r}^{i-1} \end{pmatrix} + \mathbf{P}_i\hat{\boldsymbol{\mathcal{K}}}_i^{-1}\mathbf{P}_i^T \left(\begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} - \boldsymbol{\mathcal{K}} \begin{pmatrix} \mathbf{w}^{i-1} \\ \mathbf{r}^{i-1} \end{pmatrix} \right), \qquad \text{for } i = 1, \ldots, l, \tag{30}$$

where we set $\tau = 1$, i.e. the local saddle point problems resemble the global saddle point problem in shape. Finally we define the multiplicative smoother as

$$\boldsymbol{\mathcal{S}}_M \begin{pmatrix} \mathbf{u}^k \\ \mathbf{p}^k \end{pmatrix} = \begin{pmatrix} \mathbf{u}^k \\ \mathbf{p}^k \end{pmatrix} + \begin{pmatrix} \mathbf{w}^l \\ \mathbf{r}^l \end{pmatrix}. \tag{31}$$

So far we did not pose any conditions on the local matrices $\hat{\mathbf{A}}_i$, $\mathbf{B}_i$, and $\mathbf{C}_i$. For the additive case we can state the following theorem, under which assumptions it is possible to interpret the additive Schwarz iteration as a symmetric inexact Uzawa method. Then, the smoothing property, an important part of a convergence proof for multigrid methods, can be shown (cf. Schöberl and Zulehner [15]).

Theorem 1. *Assume that* (29) *is satisfied, the matrices* $\hat{\mathbf{A}}_i$ *and* $\hat{\mathbf{S}}_i$ *are symmetric and positive definite, and there is a symmetric positive definite* $n \times n$ *matrix* $\hat{\mathbf{A}}$ *such that*

$$\mathbf{P}_{V\,i}\hat{\mathbf{A}} = \hat{\mathbf{A}}_i\mathbf{P}_{V\,i}^T, \qquad \textit{for } i = 1, \ldots, l.$$

Furthermore, assume that the matrices $\mathbf{B}_i$ *satisfy the condition*

$$\mathbf{P}_{Q_i}^T\mathbf{B} = \mathbf{B}_i\mathbf{P}_{Q_i}^T, \qquad \textit{for } i = 1,\ldots,l.$$

Then we have

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{v}^k \qquad \textit{and} \qquad \mathbf{p}^{k+1} = \mathbf{p}^k + \mathbf{q}^k, \tag{32}$$

where $\mathbf{v}^k$ *and* $\mathbf{q}^k$ *satisfy the equation*

$$\begin{pmatrix} \hat{\mathbf{A}} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{B}\hat{\mathbf{A}}^{-1}\mathbf{B}^T - \hat{\mathbf{S}} \end{pmatrix} \begin{pmatrix} \mathbf{v}^k \\ \mathbf{q}^k \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} - \mathcal{K} \begin{pmatrix} \mathbf{u}^k \\ \mathbf{p}^k \end{pmatrix} \tag{33}$$

and

$$\hat{\mathbf{S}} = \left(\sum_{i=1}^{l} \mathbf{P}_{Q_i} \hat{\mathbf{S}}_i^{-1} \mathbf{P}_{Q_i}^T \right)^{-1}.$$

Proof. See Schöberl and Zulehner [15]. □

Up to the knowledge of the author, there is no theory available for the multiplicative Schwarz-type smoother. We refer to Schöberl and Zulehner [15] for a theoretical analysis for the convergence and smoothing properties of the additive smoother. But in practice, the multiplicative version turns out to much more efficient than the additive iteration scheme. So, we realized our numerical test examples with the multiplicative Schwarz-type smoother, as presented in the next section. The verification of the assumptions of Theorem 1 for our particular case, as well as numerical experiments for the additive version, are still missing and will be part of future research.

5 Numerical experiments

In this chapter we only investigate the numerical behaviour of the multiplicative Schwarz smoother applied to our specific linear system (24). The behaviour of the overall interior-point method was already investigated in Burger and Stainko [5].

For the numerical results we choose $\Omega = (0,1) \times (0,1)$ and decompose it into a regular triangulation $\mathcal{T}_h^k = \{\tau_i \mid i = 1,\ldots,n_k\}$ for each level k of a hierarchy of l nested meshes with $3 \le k \le l$. That means that level $k = 3$ is the coarsest grid where the corresponding linear system is solved exactly. For each level k we assemble the block matrices that finally build up the saddle point system (24). For convenience we state the system matrix again:

$$\mathcal{K}_k = \begin{pmatrix} \mathcal{K}_{\rho\rho} & \mathcal{K}_{\rho u} & \mathcal{K}_{\rho s} & \mathbf{0} \\ \mathcal{K}_{\rho u}^T & \mathcal{K}_{uu} & \mathcal{K}_{us} & \mathbf{0} \\ \mathcal{K}_{\rho s}^T & \mathcal{K}_{us}^T & \mathcal{K}_{ss} & \mathbf{D}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{D} & \mathbf{0} \end{pmatrix}, \tag{34}$$

with the block matrices (25). In order to test the multiplicative patch smoother (30) - (31) we solved the saddle point system (24) on a hierarchy with an increasing number of meshes. We set $\mathbf{f}_k = \mathbf{0}$ and used randomly chosen starting values for $\triangle\mathbf{x}_k^0$ for the exact solutions $\triangle\mathbf{x}_k$. For constructing the local subproblems we decomposed the grid $\mathcal{T}_h^k$ into m_k overlapping patches, where m_k denotes the number of nodes on level k. Each patch consists of the at most 6 surrounding triangles for each node. As mentioned in the previous section, we approximated

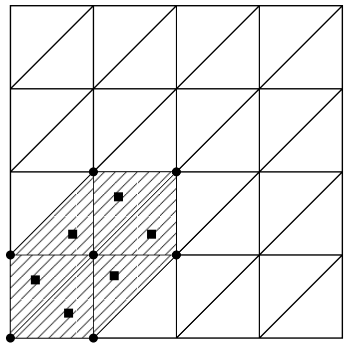

Figure 5: Patch of a local saddle point problem.

the density ρ, the displacements $\mathbf{u}$, and the Lagrangian multiplier $\boldsymbol{\lambda}_0$ with linear elements and the stresses $\mathbf{s}$ with constant elements. The corresponding subspaces V_i, for $i = 1, \ldots, m_k$, consist now of the degrees of freedom of the node i, related to the approximations of the density and the displacement components, and the degrees of freedom in the surrounding elements, related to the stress components. The subspaces Q_i, for $i = 1, \ldots, m_k$, consist of the unknowns at node i with respect to the approximation of the Lagrangian multiplier $\boldsymbol{\lambda}_0$. Figure 5 shows an example of a patch, where the places marked with a '■' indicate the unknowns of the constant elements and the places marked with a '●' indicate the unknowns of the linear elements. For the actual numerical tests we chose the local block matrix $\hat{\mathbf{A}}_i = \mathbf{A}_i$

Level	Unknowns	Smoothing steps			
		2		4	
		Iterations	Conv. Factor	Iterations	Conv. Factor
4	725	25	0.478	14	0.255
5	2853	27	0.510	15	0.269
6	11333	26	0.489	14	0.255
7	45189	25	0.479	13	0.230
8	180485	23	0.445	12	0.209

Table 1: Convergence rates for a W-cycle and an error reduction by a factor of 10^{-8} ($\epsilon = 0.1$, $\mu = 0.1$, $\boldsymbol{\nu}_i^h = \mathbf{1}$ for $i = 1, \ldots, 8$).

and used a W-cycle with 2 smoothing steps (one pre- and one post-smoothing step. We stopped the iteration process when the initial defect was reduced by a factor of 10^{-8}, measured by the Euclidean norm. In Table 1 we list the convergence data for the following choice of parameters: $\epsilon = 0.1$, $\mu = 0.1$, and $\boldsymbol{\nu}_i^h = \mathbf{1}$ for $i = 1, \ldots, 8$. The table shows the typical multigrid convergence behavior, i.e., convergence rates that are asymptotic independent of the grid level and an asymptotic constant number of iterations. For the next test example we set μ and ϵ so smaller values, as these parameters are supposed to tend to zero in actual computations. We chose $\mu = 10^{-6}$ and $\epsilon = 10^{-4}$, even smaller values as have actually been used to compute reasonable designs in Burger and Stainko [5]. All in all, Table 2 shows again the expected behavior. Both, Table 1 and Table 2, show the robust behaviour of the multigrid method with respect to the parameters μ and ϵ.

Level	Unknowns	Smoothing steps			
		2		4	
		Iterations	Conv. Factor	Iterations	Conv. Factor
4	725	39	0.621	19	0.376
5	2853	25	0.478	14	0.258
6	11333	24	0.460	13	0.226
7	45189	22	0.427	12	0.210
8	180485	22	0.425	12	0.211

Table 2: Convergence rates for a W-cycle and an error reduction by a factor of 10^{-8} ($\epsilon = 10^{-4}$, $\mu = 10^{-6}$, $\boldsymbol{\nu}_i^h = \mathbf{1}$ for $i = 1, \ldots, 8$).

Also arbitrary values of the dual variables $\boldsymbol{\nu}_1^h, \ldots, \boldsymbol{\nu}_4^h$, in $[10^{-6}, 10^1]$, possibly after suitable scaling, do not change this behaviour. However, the dual variables $\boldsymbol{\nu}_5^h, \ldots, \boldsymbol{\nu}_8^h$, that act as Lagrange multipliers for the slack variables $\mathbf{z}_1^h, \ldots, \mathbf{z}_4^h$ (see (15) in Subsection 3) may cause troubles. In the case that $\mathbf{v}_5^h \leq \mathbf{v}_6^h$ and $\mathbf{v}_7^h \leq \mathbf{v}_8^h$ the condition number of the system matrix is raising, but the multigrid iteration still achieves convergence with more then 4 smoothing steps. See Table 3 for the convergence data for extreme values $\boldsymbol{\nu}_5^h = \boldsymbol{\nu}_7^h = \mathbf{10}$ and $\boldsymbol{\nu}_6^h = \boldsymbol{\nu}_8^h = \mathbf{10^{-6}}$ with 5 and 9 pre- and post-smoothing steps, respectively. Unfortunately, for

Level	Unknowns	Smoothing steps			
		10		18	
		Iterations	Conv. Factor	Iterations	Conv. Factor
4	725	52	0.701	26	0.490
5	2853	45	0.664	23	0.446
6	11333	34	0.582	18	0.358
7	45189	29	0.529	17	0.321
8	180485	27	0.500	16	0.298

Table 3: Convergence rates for a W-cycle and an error reduction by a factor of 10^{-8}. ($\epsilon = 10^{-4}$, $\mu = 10^{-6}$, $\boldsymbol{\nu}_5^h = \boldsymbol{\nu}_7^h = \mathbf{10}$, and $\boldsymbol{\nu}_6^h = \boldsymbol{\nu}_8^h = \mathbf{10^{-6}}$).

some choices $\mathbf{v}_5^h > \mathbf{v}_6^h$ and $\mathbf{v}_7^h > \mathbf{v}_8^h$ the upper left block of the system matrix (34) becomes almost indefinite, e.g., $\lambda_{\min} \in [-9 \cdot 10^{-6}, 9 \cdot 10^{-6}]$. Similar behaviour is reported, e.g. in Maar and Schulz [11] and Wächter and Biegler [19]. If the upper left block looses its property to be positive definite, the smoother fails and the multigrid iteration diverges. For instance for the choices $\mathbf{v}_5^h = \mathbf{v}_7^h = 10^2$ and $\mathbf{v}_6^h = \mathbf{v}_8^h = 10^{-6}$ we get $\lambda_{\max} = 336$ and $\lambda_{\min} = -8 \cdot 10^6$.

A known remedy for the above situation is to add a small multiple of the identity matrix to the upper left block, which is called *inertia correction* in literature and is, e.g. used in the software packages Ipopt, see Wächter and Biegler [19], and LOQO, see Vanderbei and Shanno [18].

In our test examples a addition of $\delta\mathbf{I}$, with $\delta = 10^{-3}$, to the upper left part removed the mentioned difficulties.

6 Conclusions and Outlook

The optimal solver presented in this report shows the potential of Schwarz-type smoothers in the multigrid framework, applied to saddle-point problems. The linear complexity solver should be embedded in a dual-primal interior-point optimization method to show its true potential.

Moreover, a comparison between the computational behaviour of the additive and the multiplicative version of the smoother is still missing, as well as the validation of the assumptions of Theorem 1 for the additive case. It is expected that the additive version will not behave as good as the multiplicative version. But, since there is no theory for the multiplicative version available, the theoretical analysis of the additive case will be an interesting future task.

References

[1] M.P. Bendsøe and O. Sigmund. *Topology Optimization: Theory, Methods and Applications*. Springer, Berlin, 2003.

[2] B. Bourdin and A. Chambolle. Design-dependent loads in topology optimization. *ESAIM: Control, Optimisation and Calculus of Variations*, 9:19–48, 2003.

[3] J.H. Bramble. *Multigrid methods*, volume 294 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow, 1993.

[4] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer, New York, 1991.

[5] M. Burger and R. Stainko. Phase–field relaxation of topology optimization with local stress constraints. To appear in *SIAM Journal on Control and Optimization*, 2006.

[6] J.W. Cahn and J.E. Hilliard. Free energy of a nonuniform system I - Interfacial free energy. *Journal of Chemical Physics*, 28:258–267, 1958.

[7] A. Forsgren, P.E. Gill, and M.H. Wright. Interior methods for nonlinear optimization. *SIAM Review*, 55(4):525–597, 2002.

[8] W. Hackbusch. *Iterative Lösung großer schwachbesetzter Gleichungssysteme*. B.G. Teubner, Stuttgart, 1991.

[9] W. Hackbusch. *Multi-Grid Methods and Applications*. Springer, Berlin, 2003.

[10] M. Jung and U. Langer. *Methode der finiten Elemente für Ingenieure*. B.G. Teubner, Stuttgart - Leipzig - Wiesbaden, 2001.

[11] B. Maar and V. Schulz. Interior point multigrid methods for topology optimization. *Structural and Multidisciplinary Optimization*, 19:214–224, 2000.

[12] L. Modica and S. Mortola. Un esempio di Γ-convergenza. *Boll. Unione Mat. Ital.*, 14(B):285–299, 1977.

[13] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, New York, 1999.

[14] G.I.N. Rozvany. Aims, scope, methods, history and unified terminology of computer-aided topology optimization in structural mechanics. *Structural and Multidisciplinary Optimization*, 21:90–108, 2001.

[15] J. Schöberl and W. Zulehner. On Schwarz-type smoothers for saddle point problems. *Numerische Mathematik*, 95(2):377–399, 2003.

[16] R. Stainko. An adaptive multilevel approach to minimal compliance topology optimization. *Communications in Numerical Methods in Engineering*, 22:109–118, 2006.

[17] M. Stolpe and K. Svanberg. Modeling topology optimization problems as linear mixed 0–1 programs. *International Journal for Numerical Methods in Engineering*, 57(5):723–739, 2003.

[18] R.J. Vanderbei and D.F. Shanno. An interior-point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications*, 13:231–252.

[19] A. Waechter and L.T. Biegler. On the implementation of an interior–point filter line–search algorithm for large–scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

[20] S.J. Wright. *Primal-Dual Interior-Point Methods.* SIAM, Society for Industrial and Applied Mathematics, Philadelphia, 1997.