NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2011; **00**:1–15
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla

Preconditioners for state constrained optimal control problems with Moreau-Yosida penalty function

Martin Stoll[1,*] John Pearson[2], Andy Wathen[2]

[1] *Computational Methods in Systems and Control Theory, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg Germany, (martin.stoll80@gmail.com)*
[2] *Numerical Analysis Group, Mathematical Institute, 24 - 29 St Giles', Oxford, OX1 3LB, United Kingdom (wathen@maths.ox.ac.uk)*

Received ...

KEY WORDS: State-constrained problems, PDE-constrained optimization, Saddle point systems, Preconditioning, Newton method, Krylov subspace solver

1. INTRODUCTION

Optimization problems with constraints given by partial differential equations (PDEs) arise in a variety of applications (see [24]). Comprehensive introductions to this field can be found in [48, 24]. Throughout this paper, we consider the minimization of a functional $J(y,u)$ defined as

$$J(y,u) := \frac{1}{2}\left\|y-y_d\right\|^2_{L^2(\Omega_1)} + \frac{\beta}{2}\left\|u\right\|^2_{L^2(\Omega)}, \tag{1}$$

with $\Omega_1 \subset \Omega \in \mathbb{R}^d$. In (1) $\beta \in \mathbb{R}^+$ represents a regularization parameter and y_d is a given function that represents the desired state. In many practical applications this function exhibits design parameters representing the underlying application. The state y and the control u are linked via the Poisson equation

$$-\triangle y = u \text{ in } \Omega \tag{2}$$

with boundary conditions $y = g$ on $\partial\Omega$ or

$$-\triangle y + y = u \text{ in } \Omega \tag{3}$$

with boundary conditions $\frac{\partial y}{\partial n} = 0$ on $\partial\Omega$. We decide to consider both (2) and (3) as both play a significant role in the literature. The choice of g will typically be 0 or the projection of y_d onto the box defined by constraints. The introduction of box constraints on the control and the state, i.e.,

$$\underline{u} \leq u \leq \overline{u} \tag{4}$$

and

$$\underline{y} \leq y \leq \overline{y} \tag{5}$$

is of practical interest. In this paper we will focus on the numerical solution of the optimization problem given above when state constraints are present. Effective preconditioning strategies for the

*Correspondence to: Computational Methods in Systems and Control Theory, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg Germany, (martin.stoll80@gmail.com)

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls *[Version: 2010/05/13 v2.00]*

control constrained case can be found in [46, 23]. We will later show that a semi-smooth Newton method applied to the Moreau-Yosida regularization of (1) leads to a linear system in saddle point form. The saddle point matrix is symmetric and indefinite and a variety of methods exists to solve problems of this type efficiently (see [2] for a survey). In practice the linear system usually is of sufficiently high dimension that iterative solution methods are needed and it is never solved without the application of a preconditioner $\mathcal{P}$ which is chosen to enhance the convergence behavior of the iterative method. A variety of preconditioners exists to tackle saddle point problems. The aim of this paper is to present preconditioners that are tailored towards an efficient solution of the linear system arising from the discretization of an optimal control problem involving a partial differential equation and state constraints. In general, the state constrained problem is a considerably harder problem (see Section 2) than the control-constraint problem. In this paper,we will introduce preconditioning strategies that allow for a robust solution of the linear system with respect to both the regularization parameter β and the parameter coming from the Moreau-Yosida penalty term.

The paper is organized as follows. The problem we are interested in will be presented in detail in Section 2. Our focus in this paper is to derive efficient preconditioners for the optimal control problems and hence our focus is to introduce all methods from a linear algebra perspective. We show how for each method the saddle point system can be preconditioned and efficiently solved using a Krylov subspace technique. We successively introduce two preconditioners, where the first is derived from previous results for PDE-constrained optimization and the second follows a recent techniques focusing on robustness with respect to the regularization parameters. The numerical results presented in Section 4 illustrate the performance of the presented method.

2. THE MOREAU-YOSIDA FORMULATION

We consider the case when state constraints are introduced and assume that the functional $J(y,u)$ (1) has to be minimized for functions y and u defined over a domain $\Omega \in \mathbb{R}^d$. The case when minimization is sought over $\Omega_1 \subset \Omega$ is not discussed in this paper (see [44] for this case with no state constraints). The problem of minimizing (1) when bound constraints on the state are given is more complicated than the control constrained case [8, 24, 26] as in general the Lagrange multiplier is only a measure. Several remedies have been proposed for this problem. In [30] Meyer *et al.* consider regularized state-constraints , i.e.,

$$\underline{y} \leq \varepsilon u + y \leq \overline{y}. \tag{6}$$

An alternative approach is given by changing the objective function (1) using the Moreau-Yosida penalty function [25] to give

$$\begin{aligned} J(y,u) :=& \frac{1}{2}\left\|y-y_d\right\|^2_{L^2(\Omega)} + \frac{\beta}{2}\left\|u\right\|^2_{L^2(\Omega)} + \frac{1}{2\varepsilon}\left\|\max\left\{0, y-\overline{y}\right\}\right\|^2_{L^2(\Omega)} \\ &+ \frac{1}{2\varepsilon}\left\|\min\left\{0, y-\underline{y}\right\}\right\|^2_{L^2(\Omega)} \end{aligned} \tag{7}$$

subject to the above state equations with appropriate boundary conditions. For the remainder we will now assume that the state equations and hence $J(y,u)$ are considered in discretized form using an appropriate finite element discretization [23].

The discretized version of state equation (2) and (1) are given by

$$\begin{aligned} \text{Minimize } & \frac{1}{2}\left(y-y_d\right)^T M \left(y-y_d\right) + \frac{\beta}{2}u^T M u \\ &+ \frac{1}{2\varepsilon}\max\left\{0, y-\overline{y}\right\}^T M \max\left\{0, y-\overline{y}\right\} \\ &+ \frac{1}{2\varepsilon}\min\left\{0, y-\overline{y}\right\}^T M \min\left\{0, y-\overline{y}\right\} \\ \text{subject to } & Ky = Mu + d. \end{aligned} \tag{8}$$

Here K represents the stiffness matrix and M the mass matrix. We will only consider the lumped mass matrix here but comment later on how to precondition for the consistent mass matrix. Note

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

that y, u, y_d, $\overline{y}$, and $\underline{y}$ now represent vectors. The optimality system of (8) looks as follows

$$-K^T\lambda = -M(y-y_d) - \varepsilon^{-1}\chi_{\mathcal{A}_+}M\max\{0, y-\overline{y}\} - \varepsilon^{-1}\chi_{\mathcal{A}_-}M\min\{0, y-\underline{y}\} \tag{9}$$

$$\beta Mu - M\lambda = 0 \tag{10}$$

$$-Ky + Mu = d \tag{11}$$

with $\chi_{\mathcal{A}_+}$ being the characteristic function for the indices where $y-\overline{y} > 0$ and $\chi_{\mathcal{A}_-}$ the characteristic function for the region where $y - \underline{y} < 0$. Note that $\mathcal{A}_+ = \{i : y_i > \overline{y}_i\}$ and $\mathcal{A}_- = \left\{i : y_i < \underline{y}_i\right\}$ are the active sets associated with the bound constraints on the state y at step k. If one now wants to apply a semi-smooth Newton method to (9) to (11) we have to solve the following system at every step

$$\begin{bmatrix} M + \varepsilon^{-1}G_{\mathcal{A}}MG_{\mathcal{A}} & 0 & -K^T \\ 0 & \beta M & M \\ -K & M & 0 \end{bmatrix} \begin{bmatrix} y^{(k+1)} \\ u^{(k+1)} \\ \lambda^{(k+1)} \end{bmatrix} = \begin{bmatrix} c_{\mathcal{A}} \\ 0 \\ d \end{bmatrix} \tag{12}$$

where $c_{\mathcal{A}} = My_d + \varepsilon^{-1}\left(G_{\mathcal{A}_+}MG_{\mathcal{A}_+}\overline{y} + G_{\mathcal{A}_-}MG_{\mathcal{A}_-}\underline{y}\right)$ defines part of the right hand side, $\mathcal{A} = \mathcal{A}_- \cup \mathcal{A}_+$ and the G matrices are projections onto the active sets defined by $\mathcal{A}$. The application of the semi-smooth Newton method to these problems has been studied (see [23, 25, 5]). Our task is the efficient solution of the linear system in (12) which is of saddle point type. Note that we do not focus on the discussion of the the inexact semi-smooth Newton's method here but rather refer to [27] where it was observed that with suitable preconditioning this method did just as well as the exact semi-smooth Newton method.

In the case of the state equation being defined by (3) we define $K := K_N + M$, where K_N is the stiffness matrix for a pure Neumann problem, and get the same formulation as shown above.

The Moreau-Yosida regularization has also recently been analyzed for semi-linear elliptic problems (see [28]).

3. SOLUTION OF THE LINEAR SYSTEM AND EIGENVALUE ANALYSIS

The system matrix

$$\mathcal{K} := \begin{bmatrix} L & 0 & -K^T \\ 0 & \beta M & M \\ -K & M & 0 \end{bmatrix} \tag{13}$$

is symmetric and indefinite with $L = M + \varepsilon^{-1}G_{\mathcal{A}}MG_{\mathcal{A}}$ for the remainder of the paper. Note, that the block $blkdiag(L, \beta M)$ is symmetric and positive definite as we have a mass matrix on the one hand and a mass matrix plus a submatrix of a mass matrix on the other. The matrix K is the stiffness matrix associated with the weak formulation of (2) or (3) and it is symmetric and positive definite. Benzi *et al.* [2] discuss properties and numerical methods to solve matrices in saddle point form. As $\mathcal{K}$ is a large and sparse, symmetric and indefinite matrix, a Krylov subspace solver [39, 20, 13] will be our method of choice. For smaller (and typically) $2D$ examples direct methods [9, 11] will prove very efficient but for large and/or $3D$ problems these methods are likely to run out of memory.

The choice of preconditioners that we mention in this Section is motivated by an observation about the eigenvalues of the preconditioned system $\mathcal{P}^{-1}\mathcal{K}$ for certain preconditioners $\mathcal{P}$. Murphy *et al.* show in [31] that for some idealized preconditioners the matrix $\mathcal{P}^{-1}\mathcal{K}$ has only a small number of eigenvalues (3 for block-diagonal and 2 for block-triangular).

One method that is a standard choice for symmetric and indefinite systems is the minimal residual method (MINRES) introduced in [32] as a method for minimizing the residual $\|r_k\|_2 = \|\mathcal{K}x_k - b\|_2$ over the current Krylov subspace

$$\text{span}\left\{r_0, \mathcal{K}r_0, \mathcal{K}^2r_0, \ldots, \mathcal{K}^{k-1}r_0\right\}.$$

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

In order to be able to use MINRES, we need the preconditioner to be symmetric and positive definite and hence block-diagonal preconditioners would present a natural choice [13, 15]. A preconditioner for MINRES and the above problem could look like the following

$$\mathcal{P} = \begin{bmatrix} A_0 & 0 & 0 \\ 0 & A_1 & 0 \\ 0 & 0 & S_0 \end{bmatrix}, \tag{14}$$

with A_0, A_1, and S_0 being approximations to the $(1,1)$, the $(2,2)$ block and the Schur-complement, respectively. The use of MINRES for optimal control problems has been recently investigated in [36, 35, 6, 45, 42]. Also note that MINRES is also applicable in the case of a semi-definite $(1,1)$-block, which is the case if we consider $\Omega_1 \subset \Omega$ as was done in [44]. We believe that all results given here can be extended to the subdomain case when MINRES with a block-diagonal preconditioner is used.

Another type of methods that has proven to be of interest is based on the fact that for some preconditioners the preconditioned saddle point matrix $\mathcal{P}^{-1}\mathcal{K}$ is symmetric and positive definite in an inner product defined by a matrix $\mathcal{H}$, i.e., $\langle \mathcal{P}^{-1}Ax, y\rangle_{\mathcal{H}} = \langle x, \mathcal{P}^{-1}Ay\rangle_{\mathcal{H}}$ where $\langle x, y\rangle_{\mathcal{H}} = x^T\mathcal{H}y$. There exists a variety of these methods [7, 40, 16, 3, 29, 10], which can also be combined to give rise to new methods [43, 41]. Herzog and Sachs [23] analyzed the method by Schöberl and Zulehner [40] for state and control constrained optimal control problems.

We want to focus our attention on the so-called Bramble-Pasciak CG introduced in [7], a method that uses a block-triangular preconditioner

$$\mathcal{P} = \begin{bmatrix} A_0 & 0 & 0 \\ 0 & A_1 & 0 \\ -K & M & -S_0 \end{bmatrix}, \tag{15}$$

with A_0, A_1, and S_0 being approximations just as above. Once, the preconditioner is applied to $\mathcal{K}$ the resulting preconditioned matrix $\widehat{\mathcal{K}} = \mathcal{P}^{-1}\mathcal{K}$ is not symmetric anymore but self adjoint in a non-standard inner product defined by

$$\mathcal{H} = \begin{bmatrix} L - A_0 & 0 & 0 \\ 0 & \beta M - A_1 & 0 \\ 0 & 0 & S_0 \end{bmatrix}. \tag{16}$$

It is clear that for $\mathcal{H}$ to define an inner product the diagonal blocks have to be symmetric and positive definite. While this is in general a rather tricky issue requiring an eigenvalue estimation problem, in the case of (lumped) mass matrices scaling is straightforward [46]. For the case of a consistent mass matrix Rees and Stoll showed that the scaling issues can be easily removed [36]. For more details on the implementation and properties of the non-standard inner product solver we refer to [7, 12, 23, 41, 37, 36, 40] and see Algorithm 1.

Given $\mathbf{x}_0 = 0$, set $\mathbf{r}_0 = \mathcal{P}^{-1}(\mathbf{b} - \mathcal{K}\mathbf{x}_0)$ and $\mathbf{p}_0 = \mathbf{r}_0$
for $k = 0, 1, \ldots$ **do**
  $\alpha = \frac{\langle \mathbf{r}_k, \mathbf{r}_k\rangle_{\mathcal{H}}}{\langle \mathcal{P}^{-1}\mathcal{K}\mathbf{p}_k, \mathbf{p}_k\rangle_{\mathcal{H}}}$
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha\mathbf{p}_k$
  $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha\mathcal{P}^{-1}\mathcal{K}\mathbf{p}_k$
  $\beta = \frac{\langle \mathbf{r}_{k+1}, \mathbf{r}_{k+1}\rangle_{\mathcal{H}}}{\langle \mathbf{r}_k, \mathbf{r}_k\rangle_{\mathcal{H}}}$
  $\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta\mathbf{p}_k$
end for

Algorithm 1: Non-standard inner-product CG

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

3.1. First preconditioner

The Schur complement of $\mathcal{K}$ is given by

$$S = KL^{-1}K^T + \beta^{-1}M. \tag{17}$$

For the case of $L = M$ it was proposed [35] to neglect the term $\beta^{-1}M$, which would in our case result in an approximation $S_0 = KL^{-1}K^T$ to S. For a symmetric system the clustering of the eigenvalues will govern the convergence of the iterative scheme and we want to analyze the eigenvalue distribution of $\widehat{\mathcal{K}} = \mathcal{P}^{-1}\mathcal{K}$ for an idealized case. We consider now the block-triangular preconditioner with the choice $A_0 = L$, $A_1 = \beta M$ and $S_0 = KM^{-1}K^T$, then the eigenvalues of the preconditioned matrix $\mathcal{P}^{-1}\mathcal{K}$ can be read off the diagonal blocks, i.e.,

$$\mathcal{P}^{-1}\mathcal{K} = \begin{bmatrix} I & 0 & -L^{-1}K^T \\ 0 & I & \beta^{-1}I \\ 0 & 0 & I + \beta^{-1}K^{-T}LK^{-1}M \end{bmatrix} \tag{18}$$

which shows that we have $2n$ eigenvalues at 1 and n eigenvalues are given by the eigenvalues of $I + \beta^{-1}K^{-T}LK^{-1}M$. Thus, we want to find eigenvalue bounds for $I + \beta^{-1}K^{-T}LK^{-1}M$. The eigenvalue bounds can be obtained from a field of value analysis[†]. Note that the matrix $I + \beta^{-1}K^{-T}LK^{-1}M$ is similar to the symmetric matrix $M^{1/2}\left(I + \beta^{-1}K^{-T}LK^{-1}M\right)M^{-1/2} = I + \beta^{-1}M^{1/2}K^{-T}LK^{-1}M^{1/2}$ and

$$\frac{x^Tx + \beta^{-1}x^TM^{1/2}K^{-T}LK^{-1}M^{1/2}x}{x^Tx} = 1 + \frac{\beta^{-1}\left(z^TLz\right)\left(x^TMx\right)\left(y^TK^{-T}K^{-1}y\right)}{\left(x^Tx\right)\left(y^Ty\right)\left(z^Tz\right)} \tag{19}$$

with $y = M^{1/2}x$ and $z = K^{-1}y$. The second term on the right hand side of (19) can be bounded using the results of Proposition 1.29 and Theorem 1.32 in [13], which provide bounds for the eigenvalues of the consistent mass matrix and the stiffness matrix. Namely, with h being the mesh-size of our finite element we get that

$$ch^2 \leq \frac{x^TMx}{x^Tx} \leq Ch^2 \text{ and } dh^2 \leq \frac{x^TKx}{x^Tx} \leq D$$

with c, C, d, and D being mesh-independent constants. Note that these are the bounds for a two-dimensional problem. For three-dimensional bounds we also refer to [13] but do not discuss them here. This directly gives bounds for almost all the terms in (19) and the only term that we need to analyze further is z^TLz/z^Tz. Using the definition of L we get $\left(z^TMz + \varepsilon^{-1}z^TG_{\mathcal{A}}MG_{\mathcal{A}}z\right)/z^Tz$, which obviously can be bounded above by $(1 + \varepsilon^{-1})Ch^2$. Hence, the overall bound is given by

$$\lambda_{\max}^{(I+\beta^{-1}K^{-T}LK^{-1}M)} \leq 1 + \frac{C^2}{\beta d^2} + \frac{C^2}{\beta\varepsilon d^2}. \tag{20}$$

Similarly the minimum eigenvalues are given by

$$1 \leq 1 + \frac{c^2h^4}{\beta D^2} + \frac{c^2h^4}{\beta\varepsilon D^2} \leq \lambda_{\min}^{(I+\beta^{-1}K^{-T}LK^{-1}M)} \tag{21}$$

Theorem 3.1. *For the consistent mass matrix M and the stiffness matrix K of a $Q1$ finite element space, the eigenvalues of the matrix*

$$I + \beta^{-1}K^{-T}LK^{-1}M$$

lie in the interval $\left[1, 1 + \frac{C^2}{\beta d^2} + \frac{C^2}{\beta\varepsilon d^2}\right]$.

[†]The field of values of a matrix $A \in \mathbb{R}^{n,n}$ is a set given by $\frac{x^TAx}{x^Tx}\ \forall x \neq 0, x \in \mathbb{R}^n$

Copyright © 2011 John Wiley & Sons, Ltd.

Prepared using nlaauth.cls
DOI: 10.1002/nla

We remark that the eigenvalue distribution depends on the regularization parameter β as was previously observed for the other cases (see [35, 46]). It also depends on the values of the penalty parameter ε : with decreasing value of ε the upper bound for the eigenvalues in Theorem 20 will increase.

We used the block $S_0 = KM^{-1}K$ as an approximation for the Schur-complement of the system matrix $\mathcal{K}$. This choice results in a good clustering of the eigenvalues but is too expensive for practical purposes as S_0^{-1} involves the term K^{-1} (the discretized PDE) twice. One now has to approximate the matrix K as best as possible. For this it is very important to take the structure of the infinite-dimensional problem into account. For both PDEs (2) and (3) the underlying operators are elliptic PDEs and hence multigrid provides a suitable and optimal preconditioner. The most efficient method would certainly be a geometric multigrid method as described in [21, 50]. It is well known that algebraic multigrid method provide very good approximations to the above operators while allowing greater flexibility than their geometric counterparts [38, 14]. As we implemented our method within the deal.II framework [1] we use the available interface to Trilinos [22] and the smoothed aggregation AMG implemented there [17]. Our choice will be to approximate K by a small number of V-cycles and a fixed number of steps of a Chebyshev smoother. The mass matrix M can be efficiently approximated with a variety of methods. In our case as we only work with lumped mass matrices, we can effortlessly solve for M. For consistent mass matrices the Chebyshev semi-iteration [18, 19] provides a powerful preconditioner [49, 36].

3.2. A better preconditioner

As we have seen in Theorem 3.1 the regularization parameters β and ε enter the eigenvalue bounds of the preconditioned matrix. Based on recent efforts [34, 33, 47] we want to remove the dependency on the regularization parameters within the preconditioned matrix. For this we motivate a new preconditioner. It is based on the observation that all mass matrices are lumped and hence the matrix L can be split up in the following way

$$L = \begin{bmatrix} M_{\mathcal{I}} & 0 \\ 0 & (1+\varepsilon^{-1})M_{\mathcal{A}} \end{bmatrix}$$

where $M_{\mathcal{I}}$ is the part of the mass matrix that corresponds to the free variable and $M_{\mathcal{A}}$ analogously to the active sets. Our aim is to propose a preconditioner of the following form

$$\hat{S} = (K + \hat{M})L^{-1}(K + \hat{M}) \tag{22}$$

where $\hat{S}$ now approximates the Schur-complement $S = KL^{-1}K + \beta^{-1}M$ as well as possible. Hence, we look at $\hat{S}$ in more detail

$$\hat{S} = KL^{-1}K + \hat{M}L^{-1}\hat{M} + KL^{-1}\hat{M} + \hat{M}L^{-1}K, \tag{23}$$

where our goal is for $\hat{M}L^{-1}\hat{M}$ to approximate the term $\beta^{-1}M$ in the best possible way. Structuring $\hat{M}$ in the following way

$$\hat{M} = \begin{bmatrix} \alpha M_{\mathcal{I}} & 0 \\ 0 & \gamma M_{\mathcal{A}} \end{bmatrix}$$

gives for $\hat{M}L^{-1}\hat{M} = \beta^{-1}M$

$$\hat{M}L^{-1}\hat{M} = \begin{bmatrix} \alpha^2 M_{\mathcal{I}} & 0 \\ 0 & \gamma^2(1+\varepsilon^{-1})^{-1}M_{\mathcal{A}} \end{bmatrix} = \begin{bmatrix} \beta^{-1}M_{\mathcal{I}} & 0 \\ 0 & \beta^{-1}M_{\mathcal{A}} \end{bmatrix} = \beta^{-1}M. \tag{24}$$

Finally, this yield for the parameters α and γ that

$$\alpha = \frac{1}{\sqrt{\beta}} \text{ and } \gamma = \frac{\sqrt{1+\varepsilon^{-1}}}{\sqrt{\beta}}, \tag{25}$$

which we then use for $\hat{M}$ used in $\hat{S}$. The questions now is whether this results in better eigenvalue bounds for eigenvalues of the preconditioned matrix. For this we are interested in the eigenvalues

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

of the matrix $\hat{S}^{-1}S$, which are given by the Rayleigh quotient

$$\frac{v^TSv}{v^T\hat{S}v} = \frac{v^TKL^{-1}Kv + \beta^{-1}v^TMv}{v^TKL^{-1}Kv + \beta^{-1}v^TMv + v^T\hat{M}L^{-1}Kv + v^TKL^{-1}\hat{M}v}. \tag{26}$$

Equation (26) can be rewritten as

$$\frac{1}{\frac{v^TKL^{-1}Kv+\beta^{-1}v^TMv}{v^TKL^{-1}Kv+\beta^{-1}v^TMv} + \frac{v^T\hat{M}L^{-1}Kv+v^TKL^{-1}\hat{M}v}{v^TKL^{-1}Kv+\beta^{-1}v^TMv}} \tag{27}$$

or equivalently

$$\frac{1}{1 + \frac{v^T\hat{M}L^{-1}Kv+v^TKL^{-1}\hat{M}v}{v^TKL^{-1}Kv+\beta^{-1}v^TMv}} =: R. \tag{28}$$

We are left with analyzing the term

$$\frac{v^T\hat{M}L^{-1}Kv + v^TKL^{-1}\hat{M}v}{v^TKL^{-1}Kv + \beta^{-1}v^TMv} = \frac{b^Ta + ab^T}{a^Ta + b^Tb},$$

with $a = L^{-1/2}Kv$ and $b = L^{-1/2}\hat{M}v$ also using the fact that $\beta^{-1}M = \hat{M}L^{-1}\hat{M}$. We now want to show that $b^Ta + ab^T > 0$. For this we look at $b^Ta + ab^T$ in more detail and get

$$b^Ta + ab^T = \frac{v^T\left(DK + KD\right)v}{v^Tv} \tag{29}$$

$$= \frac{v^T\left(D^{1/2}\left(D^{1/2}Kd^{-1/2} + D^{-1/2}KD^{1/2}\right)D^{1/2}\right)v}{v^Tv} \tag{30}$$

$$= \frac{y^T\left(D^{1/2}KD^{-1/2} + D^{-1/2}KD^{1/2}\right)yv^TDv}{v^Tvy^Ty} \tag{31}$$

where $D = \hat{M}L^{-1} = L^{-1}\hat{M}$ because of the diagonal structure of L, M, and $\hat{M}$. Note that as D and K are both symmetric and positive definite (31) is trivially positive. From this it follows that $R < 1$ and from the fact that

$$(a-b)^T(a-b) \geq 0 \Leftrightarrow \frac{a^Tb + b^Ta}{a^Ta + b^Tb} \leq 1$$

for any a, b and $a^Ta + b^Tb > 0$ we get that $R \geq \frac{1}{2}$. Hence, the eigenvalues of $\hat{S}^{-1}S$ are in the interval $\left[\frac{1}{2}, 1\right)$. We have now proven the following Theorem.

Theorem 3.2. *With the above assumptions the eigenvalues of the matrix pencil $(S, \hat{S})$ lie in the interval $\left[\frac{1}{2}, 1\right)$.*

Note that the proof presented above used the diagonal structure of the matrices involving mass matrix terms. We now quickly show that the above also holds for $\hat{M}$ a symmetric positive definite matrix. Again, we consider

$$\frac{v^T\hat{M}L^{-1}Kv + v^TKL^{-1}\hat{M}v}{v^TKL^{-1}Kv + \beta^{-1}v^TMv} = \frac{b^Ta + ab^T}{a^Ta + b^Tb}. \tag{32}$$

For the above analysis to hold, we need $v^T\hat{M}L^{-1}Kv + v^TKL^{-1}\hat{M}v$ to be positive. Note that this is equal to

$$v^T\hat{M}^{1/2}\left[\hat{M}^{1/2}L^{-1}\hat{M}^{1/2}\hat{M}^{-1/2}K\hat{M}^{-1/2} + \hat{M}^{-1/2}K\hat{M}^{-1/2}\hat{M}^{1/2}L^{-1}\hat{M}^{1/2}\right]\hat{M}^{1/2}v,$$

which we can write as

$$v^T\hat{M}^{1/2}\left[D\hat{K} + \hat{K}D\right]\hat{M}^{1/2}v$$

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

with $D = \hat{M}^{1/2}L^{-1}\hat{M}^{1/2}$ symmetric and positive definite and $\hat{K} = \hat{M}^{-1/2}K\hat{M}^{-1/2}$ symmetric and positive definite. Finally, using the properties of L, K, and M the above also holds for a non-diagonal approximation $\hat{M}$ and also non-diagonal L, i.e., the case when consistent mass matrices are used. One possible Schur-complement approximation for the consistent mass matrix case is given in the following Theorem.

Theorem 3.3. *If we approximate the Schur complement S as follows:*

$$\hat{S}_2 = \left[K + \frac{1}{\sqrt{\beta}}M\left(I + \frac{1}{\sqrt{\epsilon}}G_{\mathcal{A}}\right)\right]\left(I + \frac{1}{\sqrt{\epsilon}}G_{\mathcal{A}}\right)^{-1}M^{-1}\left(I + \frac{1}{\sqrt{\epsilon}}G_{\mathcal{A}}\right)^{-1}\left[K + \frac{1}{\sqrt{\beta}}M\left(I + \frac{1}{\sqrt{\epsilon}}G_{\mathcal{A}}\right)\right]^T,$$

then the eigenvalue bound $\lambda(\hat{S}_2^{-1}S) \in [\frac{1}{2}, 2]$ holds.

Proof. We seek to prove that

$$S \approx K\left[\left(I + \frac{1}{\sqrt{\epsilon}}G_{\mathcal{A}}\right)M\left(I + \frac{1}{\sqrt{\epsilon}}G_{\mathcal{A}}\right)\right]^{-1}K^T + \frac{1}{\beta}M := \tilde{S} \approx \hat{S}_2,$$

which we do by considering the eigenvalues of $\tilde{S}^{-1}S$ and $\hat{S}_2^{-1}\tilde{S}$, using Rayleigh quotients.

We first examine the Rayleigh quotient

$$R_1 := \frac{\mathbf{v}^T(M + \epsilon^{-1}G_{\mathcal{A}}MG_{\mathcal{A}})\mathbf{v}}{\mathbf{v}^T(I + \epsilon^{-1/2}G_{\mathcal{A}})M(I + \epsilon^{-1/2}G_{\mathcal{A}})\mathbf{v}} = \frac{\mathbf{a}_1^T\mathbf{a}_1 + \mathbf{b}_1^T\mathbf{b}_1}{(\mathbf{a}_1 + \mathbf{b}_1)^T(\mathbf{a}_1 + \mathbf{b}_1)}, \quad \text{where} \quad \begin{array}{l}\mathbf{a}_1 = M^{1/2}\mathbf{v},\\ \mathbf{b}_1 = \epsilon^{-1/2}M^{1/2}G_{\mathcal{A}}\mathbf{v}.\end{array}$$

We can see by straightforward algebra (using that $\mathbf{a}_1^T\mathbf{a}_1 > 0$ by positive definiteness of M) that $R_1 \geq \frac{1}{2}$. Further, using the fact that $\mathbf{a}_1^T\mathbf{b}_1 = \mathbf{b}_1^T\mathbf{a}_1 > 0$ (by virtue of $G_{\mathcal{A}}$ being symmetric and positive semi-definite, and the fact that the product of two symmetric and positive semi-definite matrices is positive semi-definite), we can show that $R_1 \leq 1$. From these bounds, it is a simple matter to show that the Rayleigh quotient $\frac{\mathbf{v}^T S\mathbf{v}}{\mathbf{v}^T\tilde{S}\mathbf{v}} \in [1, 2]$.

Looking now at the eigenvalues of $\hat{S}_2^{-1}\tilde{S}$, we examine the Rayleigh quotient $\frac{\mathbf{v}^T\tilde{S}\mathbf{v}}{\mathbf{v}^T\hat{S}_2\mathbf{v}}$, writing

$$R_2 := \frac{\mathbf{v}^T\tilde{S}\mathbf{v}}{\mathbf{v}^T\hat{S}_2\mathbf{v}} = \frac{\mathbf{a}_2^T\mathbf{a}_2 + \mathbf{b}_2^T\mathbf{b}_2}{(\mathbf{a}_2 + \mathbf{b}_2)^T(\mathbf{a}_2 + \mathbf{b}_2)}, \quad \text{where} \quad \begin{array}{l}\mathbf{a}_2 = M^{-1/2}(I + \epsilon^{-1/2}G_{\mathcal{A}})^{-1}K^T\mathbf{v},\\ \mathbf{b}_2 = \beta^{-1/2}M^{1/2}\mathbf{v}.\end{array}$$

By algebraic manipulation (as $\mathbf{b}_2^T\mathbf{b}_2 > 0$), we see that $R_2 \geq \frac{1}{2}$. Using the inequality $\mathbf{a}_2^T\mathbf{b}_2 = \mathbf{b}_2^T\mathbf{a}_2 > 0$ (as a product of two SPD matrices is positive definite), we have that $R_2 < 1$.

Finally, using the above results, we have that the Rayleigh quotient

$$\frac{\mathbf{v}^T S\mathbf{v}}{\mathbf{v}^T\hat{S}_2\mathbf{v}} = \frac{\mathbf{v}^T S\mathbf{v}}{\mathbf{v}^T\tilde{S}\mathbf{v}} \cdot \frac{\mathbf{v}^T\tilde{S}\mathbf{v}}{\mathbf{v}^T\hat{S}_2\mathbf{v}} \in [\frac{1}{2}, 2],$$

and so the result is proved. □

As we only focus on lumped mass matrices in this paper we refrain from showing results for the non-lumped case.

The previous results were obtained for an idealized case where we use

$$\hat{S} = (K + \hat{M})L^{-1}(K + \hat{M}) \text{ with } \hat{S}^{-1} = (K + \hat{M})^{-1}L(K + \hat{M})^{-1}.$$

In practice we always use

$$\hat{S}^{-1} = \widehat{(K + \hat{M})}^{-1}L\widehat{(K + \hat{M})}^{-1}$$

where

$$\widehat{(K + \hat{M})}^{-1}$$

denotes the application of an algebraic or geometric multigrid method to the matrix $(K + \hat{M})$. Note that as $\hat{M}$ is changing with every Newton iteration, we have to recompute it at the beginning of each Newton step. Nevertheless, the reduction in iteration numbers is so significant that this is clearly the preferred approach especially for small values of the regularization parameters.

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

3.3. *Nested approach*

A strategy that will prove useful in the context of solving state constrained problems is the so-called nested approach [23]. This technique starts by computing the solution to the state-constrained problem on a very coarse grid. In the next step a uniform refinement is performed for the mesh and the solution from the coarse level is prolonged onto the fine mesh. This solution is then used as an initial guess for the Newton method on the fine level. Once the solution is computed to a desired accuracy we can proceed in the same way onto the next finer grid. It is hoped and will be shown in the Numerical Results section that this strategy reduces the number of Newton steps significantly.

4. NUMERICAL EXPERIMENTS

All results shown in this Section were computed using the deal.II [1] framework with an implementation of the Bramble-Pasciak CG method that uses the 2-norm of the relative preconditioned residual (10^{-6}) as a stopping criterion. The Newton method is stopped whenever the active sets stay unchanged [4]. For the approximation via algebraic multigrid (AMG) we use 10 steps of a Chebyshev smoother and 4 V-cycles of the smoothed aggregation algebraic multigrid implemented in Trilinos [17]. As the domain Ω we consider the unit cube. All results are performed on a Centos Linux machine with Intel(R) Xeon(R) CPU X5650 @ 2.67GHz CPUs and 48GB of RAM.

4.1. *Results for Dirichlet problems*

$2D$ results The first example we compute is a Dirichlet problem with boundary condition $y = P_{[\underline{y},\overline{y}]}(y_d)$ *on* $\partial\Omega$ defined as

$$P_{[\underline{y},\overline{y}]}y_i = \begin{cases} y_{di} & \text{if } \underline{y}_i < y_i < \overline{y}_i \\ \overline{y}_i & \text{if } y_i \geq \overline{y}_i \\ \underline{y}_i & \text{if } y_i \leq \underline{y}_i. \end{cases}$$

Figure 1 shows the desired state y_d, computed control u and state y for the case without bound constraints. In Figure 2 we show the computed state and control for a bound constrained problem. The problem is unconstrained from below and the upper bound is given by $\overline{y} = 0.1$. It can be seen that there is a small active set where y_d is attained which results in the 'hole' in the control (see Figure 2b and the active set (black contour) in 2a). Here the desired state is given by

$$y_d = \sin(2\pi x_1 x_2).$$

Table I shows results for $\beta = 1e-2, \varepsilon = 1e-6$ and the upper bound $\overline{y} = 0.1$. It can be seen that for this setup the preconditioner as well as the Newton method are in this case almost independent of the mesh parameter. In our experience the performance of the algebraic multigrid preconditioner deteriorated for meshes with smaller mesh size h. The increase in iteration numbers could not be observed if a factorization of $K + \hat{M}$ was used, which for large problems is not feasible. Hence, we chose a rather large number of V-cycles, namely 4, to approximate the matrix well. A parameter independent approximation of $K + \hat{M}$ should be investigated in future research. Note that the timings shown also include the setup of the preconditioner for each Newton step in the improved preconditioner. As we can see from Table II where we show results for the same setup but with the non-robust preconditioner presented in Section 3.1, the improvement is substantial as for the setup with $\beta = 1e-2, \varepsilon = 1e-6$ the Newton method did not converge within 30 iterations. For $\beta = 1e-2$, $\varepsilon = 1e-4$ we show the results with 4 multigrid cycles in Table III an observe good convergence for this setup of parameters. As for the non-robust preconditioner the algebraic multigrid only has to approximate K a smaller number of V-cycles produces the results shown in Table IV Note that as observed in [27] the quality of the preconditioner determines the convergence of the semi-smooth Newton method with inexact solves.

Copyright © 2011 John Wiley & Sons, Ltd.

Prepared using nlaauth.cls
DOI: 10.1002/nla

DoF	Newton steps	Total CG	CG per Newton	Time for Newton
1089	14	259	18	11.27
4225	8	138	17	12.63
16641	7	108	15	30.3
66049	7	118	16	123.54
263169	6	103	17	462.78
1050625	7	183	26	3267.28

Table I. $2D$-Results for non-zero Dirichlet boundary,$\beta = 1e-2$, $\varepsilon = 1e-6$ and $\overline{y} = 0.1$.

DoF	Newton steps	Total CG	CG per Newton	Time for Newton
1089	14	780	55	29.31
4225	2	14	7	1.25
16641	5	59	11	13.63
66049	10	170	17	131.98
263169	No convergence after 40 Newton steps			

Table II. Non-robust preconditioner: $2D$-Results for non-zero Dirichlet boundary, $\beta = 1e-2$, $\varepsilon = 1e-6$ and $\overline{y} = 0.1$.

DoF	Newton steps	Total CG	CG per Newton	Time for Newton
1089	9	145	16	5.55
4225	3	21	7	1.87
16641	4	34	8	7.75
66049	4	39	9	31.26
263169	9	172	19	530.38

Table III. Non-robust preconditioner: $2D$-Results for non-zero Dirichlet boundary,$\beta = 1e-2$, $\varepsilon = 1e-4$ and $\overline{y} = 0.1$ with 4 V-cycles.

DoF	Newton steps	Total CG	CG per Newton	Time for Newton
1089	9	150	16	5.64
4225	3	24	8	1.76
16641	7	124	17	22.63
66049	8	307	38	191.54
263169	10	1059	105	2507.48

Table IV. Non-robust preconditioner: $2D$-Results for non-zero Dirichlet boundary,$\beta = 1e-2$, $\varepsilon = 1e-4$ and $\overline{y} = 0.1$ with 2 V-cycles.

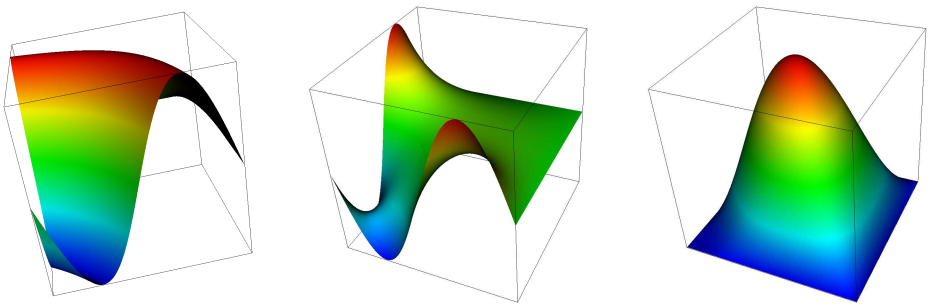
The next comparison we want to make is the quality of the preconditioner for different values of the parameters. As we mentioned earlier some dependence of the AMG on the parameters could be observed. Hence, our choice is on a factorization of $K + \hat{M}$ for a smaller mesh with 16641 degrees of freedom. The results shown in Table V show that having no deterioration in the approximation of $K + \hat{M}$ results in almost constant low iteration number s for the CG steps per iteration. We also show the behaviour of one instance of β and varying ε for the non-robust preconditioner when used with a factorization of K. In practice one should of course use approximations to $K + \hat{M}$.

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

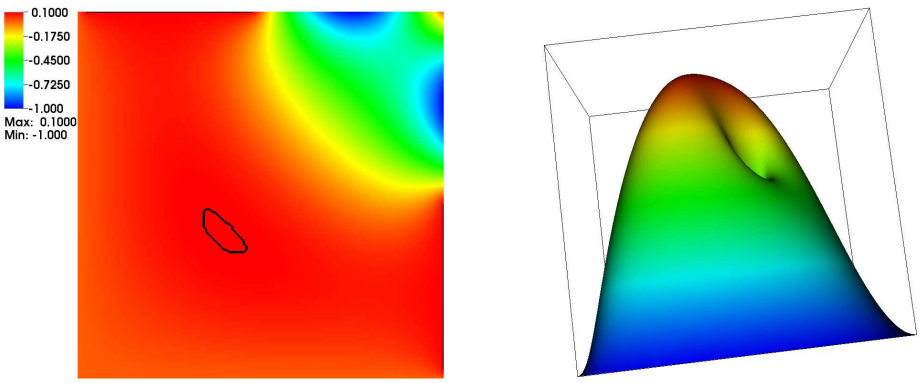
$\varepsilon \downarrow \beta \rightarrow$	1e-2	1e-4	1e-6	1e-2 (NR)
1e-4	19	27	26	30
1e-6	26	34	28	108
1e-8	32	36	28	430

Table V. Different values for β and ε using a direct factorization of $K + \hat{M}$ showing the number of CG iterations per Newton steps. We also show the non-robust (NR) preconditioner employing a factorization of K. The example was again the $2D$-results for non-zero Dirichlet boundary and $\overline{y} = 0.1$.



(a) Desired state (b) State (c) Control

Figure 1. Desired state, state and control for unconstrained problem



(a) Computed state (b) Computed control

Figure 2. Computed state and control for constrained problem

$3D$ results We now want to show results for the $3D$ example with the desired state given by

$$y_d = \sin(2\pi x_1 x_2 x_3)$$

and a zero Dirichlet boundary condition. In this case we again consider the upper bound defined as $\overline{y} = 0.1$, and the parameters $\beta = 1e-2$ and $\varepsilon = 1e-4$. The results shown in Table VI show that the iteration numbers per Newton step as well as the number of Newton steps stays constant.

Copyright © 2011 John Wiley & Sons, Ltd.

Prepared using nlaauth.cls
DOI: 10.1002/nla

DoF	Newton steps	Total CG	CG per Newton	Time for Newton
729	4	48	12	1.05
4913	4	53	13	6.17
35937	4	53	13	37.78
274625	3	41	13	228.05

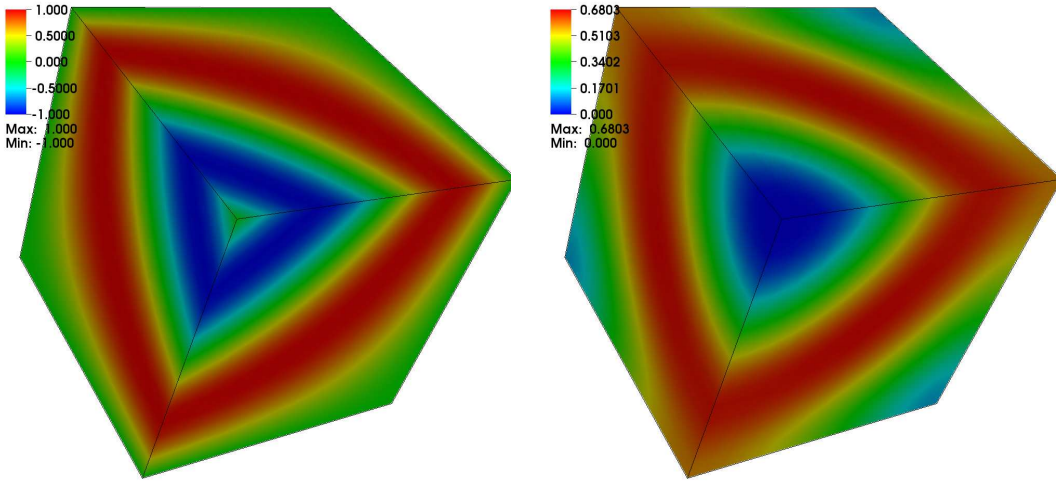
Table VI. $3D$-Results for zero Dirichlet boundary and $\overline{y} = 0.1$.

4.2. Results for Neumann boundary

In this section we are only considering three-dimensional results for the problem with the state equation given by (3). We start with the desired state given by

$$y_d = \sin(2\pi x_1 x_2 x_3)$$

and the lower bound $\underline{y} = 0$ the results are shown in Table VII. Here we define $\beta = 1e-3$ and $\varepsilon = 1e-5$. An illustration of the desired state and the constrained state is shown in Figure 3.



(a) Desired state (b) State

Figure 3. Desired state and state with lower bound $\underline{y} = 0$ for problem in $3D$

DoF	Newton steps	Total CG	CG per Newton	Time for Newton
729	6	78	13	1.95
4913	4	63	15	7.49
35937	4	72	18	51.61
274625	4	75	18	413.21
2146689	5	104	20	4458.58

Table VII. $3D$-Results for Neumann boundary and $\underline{y} = 0.2$.

We next compute an example presented in [23] where the desired state is given as

$$y_d = \begin{cases} 1 & \text{if } x_0 < 0.5 \\ -2 & \text{otherwise.} \end{cases} \tag{33}$$

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

The upper bound is given by $\overline{y} = 0$ and $\beta = 1e-2$ and $\varepsilon = 1e-4$. The results are shown in Table VIII where in comparison to [23] we observe parameter independent convergence.

DoF	Newton steps	Total CG	CG per Newton	Time for Newton
729	5	52	10	1.32
4913	1	13	13	1.67
35937	2	30	15	22.19
274625	2	32	16	181.56
2146689	1	16	16	758.11

Table VIII. $3D$-Results for Neumann boundary and $\overline{y}$ from Herzog and Sachs.

5. CONCLUSIONS AND OUTLOOK

In this paper we introduced preconditioners for the state-constrained problem when solved using the Moreau-Yosida penalization. The Krylov subspace solvers we used showed very promising performance as we could theoretically and practically obtain parameter independent convergence of the preconditioned Krylov solver.

In the future, the subdomain case in combination with MINRES should be investigated. Also, the choice of multilevel method for the parameter dependent matrix $K + \hat{M}$ should be reconsidered as we could observe dependency on the parameters for smaller meshes within the AMG preconditioner. Geometric multigrid and more advanced algebraic multigrid preconditioner have to be investigated. Incorporation of a more sophisticated scheme for the parameter ε is needed for future implementations. Also the extension of the results presented here to time-dependent problems and more difficult PDEs.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their comments. This publication is partially based on work done when the first author was supported by Award No. KUK-C1-013-04 , made by King Abdullah University of Science and Technology (KAUST)

REFERENCES

1. W. BANGERTH, R. HARTMANN, AND G. KANSCHAT, *deal.II—a general-purpose object-oriented finite element library*, ACM Transactions on Mathematical Software, 33 (2007), pp. Art. 24, 27.
2. M. BENZI, G. H. GOLUB, AND J. LIESEN, *Numerical solution of saddle point problems*, Acta Numerica, 14 (2005), pp. 1–137.
3. M. BENZI AND V. SIMONCINI, *On the eigenvalues of a class of saddle point matrices*, Numerische Mathematik, 103 (2006), pp. 173–196.
4. M. BERGOUNIOUX, K. ITO, AND K. KUNISCH, *Primal-dual strategy for constrained optimal control problems*, SIAM Journal on Control and Optimization , 37 (1999), pp. 1176–1194.
5. M. BERGOUNIOUX AND K. KUNISCH, *Primal-dual strategy for state-constrained optimal control problems*, Computational Optimization and Applications, 22 (2002), pp. 193–224.
6. L. BLANK, L. SARBU, AND M. STOLL, *Preconditioning for Allen-Cahn variational inequalities with non-local constraints*, Submitted.
7. J. H. BRAMBLE AND J. E. PASCIAK, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Mathematics of Computation, 50 (1988), pp. 1–17.
8. E. CASAS, *Control of an elliptic problem with pointwise state constraints*, SIAM Journal on Control and Optimization, 24 (1986), pp. 1309–1318.
9. T. DAVIS, *Umfpack version 4.4 user guide*, tech. rep., Dept. of Computer and Information Science and Engineering Univ. of Florida, Gainesville, FL, 2005.
10. C. R. DOHRMANN AND R. B. LEHOUCQ, *A primal-based penalty preconditioner for elliptic saddle point systems*, SIAM Journal on Numerical Analysis , 44 (2006), pp. 270–282.

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

11. I. DUFF, *Sparse numerical linear algebra: direct methods and preconditioning*, Technical Report TR/PA/96/22, CERFACS, Toulouse, France, 1996. Also RAL Report RAL 96-047.
12. H. C. ELMAN, *Multigrid and Krylov subspace methods for the discrete Stokes equations*, in Seventh Copper Mountain Conference on Multigrid Methods, N. D. Melson, T. A. Manteuffel, S. F. McCormick, and C. C. Douglas, eds., vol. CP 3339, Hampton, VA, 1996, NASA, pp. 283–299.
13. H. C. ELMAN, D. J. SILVESTER, AND A. J. WATHEN, *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2005.
14. R. FALGOUT, *An Introduction to Algebraic Multigrid*, Computing in Science and Engineering, 8 (2006), pp. 24–33. Special Issue on Multigrid Computing.
15. B. FISCHER, *Polynomial based iteration methods for symmetric linear systems*, Wiley-Teubner Series Advances in Numerical Mathematics, John Wiley & Sons Ltd, Chichester, 1996.
16. B. FISCHER, A. RAMAGE, D. J. SILVESTER, AND A. J. WATHEN, *Minimum residual methods for augmented systems*, BIT, 38 (1998), pp. 527–543.
17. M. GEE, C. SIEFERT, J. HU, R. TUMINARO, AND M. SALA, *ML 5.0 smoothed aggregation user's guide*, Tech. Rep. SAND2006-2649, Sandia National Laboratories, 2006.
18. G. H. GOLUB AND R. S. VARGA, *Chebyshev semi-iterative methods, successive over-relaxation iterative methods, and second order Richardson iterative methods. I*, Numerische Mathematik, 3 (1961), pp. 147–156.
19. ———, *Chebyshev semi-iterative methods, successive over-relaxation iterative methods, and second order Richardson iterative methods. II*, Numerische Mathematik., 3 (1961), pp. 157–168.
20. A. GREENBAUM, *Iterative methods for solving linear systems*, vol. 17 of Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
21. W. HACKBUSCH, *Multigrid methods and applications*, vol. 4 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1985.
22. M. HEROUX, R. BARTLETT, V. H. R. HOEKSTRA, J. HU, T. KOLDA, R. LEHOUCQ, K. LONG, R. PAWLOWSKI, E. PHIPPS, A. SALINGER, H. THORNQUIST, R. TUMINARO, J. WILLENBRING, AND A. WILLIAMS, *An Overview of Trilinos*, Tech. Rep. SAND2003-2927, Sandia National Laboratories, 2003.
23. R. HERZOG AND E. W. SACHS, *Preconditioned conjugate gradient method for optimal control problems with control and state constraints*, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 2291–2317.
24. M. HINZE, R. PINNAU, M. ULBRICH, AND S. ULBRICH, *Optimization with PDE Constraints*, Mathematical Modelling: Theory and Applications, Springer-Verlag, New York, 2009.
25. K. ITO AND K. KUNISCH, *Semi-smooth Newton methods for state-constrained optimal control problems*, Systems & Control Letters, 50 (2003), pp. 221–228.
26. ———, *Lagrange multiplier approach to variational problems and applications*, vol. 15 of Advances in Design and Control, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.
27. C. KANZOW, *Inexact semismooth newton methods for large-scale complementarity problems*, Optimization Methods and Software, 19 (2004), pp. 309–325.
28. K. KRUMBIEGEL, I. NEITZEL, AND A. RÖSCH, *Sufficient optimality conditions for the Moreau-Yosida-type regularization concept applied to the semilinear elliptic optmimal control problems with pointwise state constraints*, Tech. Rep. 1503/2010, WIAS, 2010.
29. J. LIESEN AND B. N. PARLETT, *On nonsymmetric saddle point matrices that allow conjugate gradient iterations*, Numerische Mathematik, 108 (2008), pp. 605–624.
30. C. MEYER, U. PRÜFERT, AND F. TRÖLTZSCH, *On two numerical methods for state-constrained elliptic control problems*, Optimization Methods and Software, 22 (2007), pp. 871–899.
31. M. F. MURPHY, G. H. GOLUB, AND A. J. WATHEN, *A note on preconditioning for indefinite linear systems*, SIAM Journal on Scientific Computing, 21 (2000), pp. 1969–1972.
32. C. C. PAIGE AND M. A. SAUNDERS, *Solutions of sparse indefinite systems of linear equations*, SIAM Journal on Numerical Analysis , 12 (1975), pp. 617–629.
33. J. W. PEARSON, M. STOLL, AND A. J. WATHEN, *Regularization-robust preconditioners for time-dependent PDE-constrained optimization problems*, Submitted, (2011).
34. J. W. PEARSON AND A. J. WATHEN, *A New Approximation of the Schur Complement in Preconditioners for PDE Constrained Optimization*, Submitted, (2010).
35. T. REES, H. S. DOLLAR, AND A. J. WATHEN, *Optimal solvers for PDE-constrained optimization*, SIAM Journal on Scientific Computing, 32 (2010), pp. 271–298.
36. T. REES AND M. STOLL, *Block-triangular preconditioners for pde-constrained optimization*, Numerical Linear Algebra with Applications, 17 (2010), pp. 977–996.
37. T. REES, M. STOLL, AND A. WATHEN, *All-at-once preconditioners for PDE-constrained optimization*, Kybernetika, 46 (2010), pp. 341–360.
38. J. W. RUGE AND K. STÜBEN, *Algebraic multigrid*, in Multigrid methods, vol. 3 of Frontiers Appl. Math., SIAM, Philadelphia, PA, 1987, pp. 73–130.
39. Y. SAAD, *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003.
40. J. SCHÖBERL AND W. ZULEHNER, *Symmetric indefinite preconditioners for saddle point problems with applications to pde-constrained optimization problems*, SIAM Journal on Matrix Analysis and Applications , 29 (2007), pp. 752–773.
41. M. STOLL, *Solving Linear Systems using the Adjoint*, PhD thesis, University of Oxford, 2009.
42. ———, *All-at-once solution of a time-dependent time-periodic PDE-constrained optimization problems*, Submitted, (2011).
43. M. STOLL AND A. WATHEN, *Combination preconditioning and the Bramble–Pasciak$^+$ preconditioner*, SIAM Journal on Matrix Analysis and Applications , 30 (2008), pp. 582–608.

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla

44. M. STOLL AND A. WATHEN, *All-at-once solution of time-dependent PDE-constrained optimization problems*, Submitted, (2010).
45. ———, *All-at-once solution of time-dependent Stokes control*, Submitted, (2011).
46. ———, *Preconditioning for PDE-constrained optimization with control constraints*,Submitted, (2011).
47. S. TAKACS AND W. ZULEHNER, *Convergence analysis of multigrid methods with collective point smoothers for optimal control problems*, Submitted, (2011).
48. F. TRÖLTZSCH, *Optimale Steuerung partieller Differentialgleichungen: Theorie, Verfahren und Anwendungen*, Vieweg Verlag, Wiesbaden, 2005.
49. A. J. WATHEN AND T. REES, *Chebyshev semi-iteration in preconditioning for problems including the mass matrix*, Electronic Transactions in Numerical Analysis, 34 (2008), pp. 125–135.
50. P. WESSELING, *An introduction to multigrid methods*, Pure and Applied Mathematics (New York), John Wiley & Sons Ltd., Chichester, 1992.

Copyright © 2011 John Wiley & Sons, Ltd.
Prepared using nlaauth.cls

DOI: 10.1002/nla