Commun. Comput. Phys.
doi: 10.4208/cicp.190914.080415a

Vol. **19**, No. 1, pp. 143-167
January 2016

A Fast Solver for an $\mathcal{H}_1$ Regularized PDE-Constrained Optimization Problem

Andrew T. Barker[1], Tyrone Rees[2,*] and Martin Stoll[3]

[1] *Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Mail Stop L-561, Livermore, CA 94551, USA.*
[2] *Numerical Analysis Group, Scientific Computing Department, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, United Kingdom.*
[3] *Computational Methods in Systems and Control Theory, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg, Germany.*

Received 19 September 2014; Accepted (in revised version) 8 April 2015

Abstract. In this paper we consider PDE-constrained optimization problems which incorporate an $\mathcal{H}_1$ regularization control term. We focus on a time-dependent PDE, and consider both distributed and boundary control. The problems we consider include bound constraints on the state, and we use a Moreau-Yosida penalty function to handle this. We propose Krylov solvers and Schur complement preconditioning strategies for the different problems and illustrate their performance with numerical examples.

AMS subject classifications: 49M25, 49K20, 65F10, 65N22, 65F50, 65N55

Key words: Preconditioning, Krylov methods, PDE-constrained optimization, optimal control of PDEs.

1 Introduction

As methods for numerically solving partial differential equations (PDEs) become more accurate and well-understood, some focus has shifted to the development of numerical methods for optimization problems with PDE constraints: see, e.g., [41,44,69] and the references mentioned therein. The canonical PDE-constrained optimization problem takes a given *desired state*, $\bar{y}$, and finds a *state*, y, and a *control*, u, to minimize the functional

$$\|y-\bar{y}\|_{\mathcal{Y}}^2+\frac{\beta}{2}R(u) \tag{1.1}$$

*Corresponding author. *Email addresses:* barker29@llnl.gov (A. T. Barker), tyrone.rees@stfc.ac.uk (T. Rees), stollm@mpi-magdeburg.mpg.de (M. Stoll)

 ©2016 Global-Science Press

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

subject to the constraints

$$\begin{aligned} &\mathcal{A}y = u, \\ &u_a \leq u \leq u_b, \\ &y_a \leq y \leq y_b, \end{aligned}$$

where $\|\cdot\|_{\mathcal{Y}}$ is some norm and $R(u)$ is a regularization functional. We are free to choose both the norm and the regularization functional here; appropriate choices often depend on the properties of the underlying application. In the description above $\mathcal{A}$ denotes a PDE with appropriate boundary conditions and β denotes a scalar regularization parameter. The focus of this manuscript is regularization based on the H_1 norm of the control, which we motivate below.

The simplest choice of $R(u)$ is $\|u\|^2_{L_2(\Omega)}$, where Ω denotes the domain on which the PDE is posed. This case has been well-studied in the literature, both from a theoretical and algorithmic perspective. However, the requirements of real-world problems has necessitated the application of alternative regularization terms.

One area where there has been much interest is in regularization using L_1 norms, see, e.g., the recent articles [12, 73]. A related norm is the total variation norm $R(u) = \|\nabla u\|_{L_1(\Omega)}$, has also aroused excitement recently – see e.g. [14, 59] and the references therein. These L_1 norms have the benefit that they allow discontinuous controls, which can be important in certain applications.

For certain applications it is desirable to have a smooth control – for this reason the $\mathcal{H}_1$ semi-norm, $R(u) = \|\nabla u\|^2_{L_2(\Omega)}$, has long been studied in the context of parameter-estimation problems [10, 46, 76], image-deblurring [13, 17, 48], image reconstruction [49], and flow control [18, 34], for example. Recently van den Doel, Ascher and Haber [19] argued that this norm can be a superior choice to its L_1-based cousin, total variation, for problems with particularly noisy data due to the smooth nature of controls which arise. The test problems in PDE constrained optimization by Haber and Hanson [31], which were designed to get academics solving problems more in-line with the needs of the real-world, suggest a regularization functional of the form $R(u) = \|u\|^2_{L_2(\Omega)} + \alpha\|\nabla u\|^2_{L_2(\Omega)}$ for a given α. Indeed, this form of regularization is commonly used in the ill-posed and inverse problem communities. Another example of a field where the standard L_2 regularization may not be appropriate is flow control – see, e.g., Gunzburger [28, Chapter 4].

At the heart of many techniques for solving the optimization problem, whether it is a linear problem or the linearization of a non-linear problem, lies the solution of a linear system [35, 41, 44, 70]. These systems are very often so-called saddle point matrices [4, 23], which have the form

$$\mathcal{A} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}, \tag{1.2}$$

where A represents the misfit and regularization terms in (1.1) and B represents the PDE constraint. In the systems we consider in this paper, A is symmetric positive semi-definite. Such saddle point matrices are invertible if B has full rank and $\ker(A) \cap \ker(B) =$

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

$\{\mathbf{0}\}$: this condition holds for most of the examples we consider here, and in the cases where it doesn't – e.g. (2.3-2.4) – there is a well-understood one dimensional null-space that can be straightforwardly dealt with [4, Section 3.2]. We are then left with the challenge of efficiently solving linear systems of the form (1.2).

Direct solvers based on factorizations [21] can be effective, but for large and, in particular, three-dimensional problems these are no longer sufficient. In such cases we turn to iterative Krylov subspace methods, which can deal with these large and sparse systems efficiently provided that they employ a preconditioner which enhances the convergence behaviour, ideally independent of problem-dependent parameters such as the mesh-size or the regularization parameter. For a general overview of preconditioners we refer to [29, 61], and in the particular case of saddle point problems see [4, 23, 77].

A number of preconditioners which are robust with respect to regularization parameters and mesh-parameters have recently been developed for PDE-constrained optimization [1, 15, 20, 36, 47, 52, 53, 65]. However, these methods are tailored for an optimization problem with $R(u) = \|u\|^2_{L_2(\Omega)}$ and heavily rely on the corresponding presence of a mass matrix in the A block of (1.2). Benzi, Haber and Taralli [5] consider a block preconditioner with of $R(u)$ given by (a variant of) the $\mathcal{H}_1-$norm, but their approach is general enough to work with most regularization and the form of this term is not exploited in the method. To the authors' knowledge there have been no other attempts to apply block preconditioners – which have proved so successful with L_2 regularization – in the case of other choices of $R(u)$. We address this issue here by considering a cost-functional where

$$R(u) = \|u\|^2_{L_2} + \|\nabla u\|^2_{L_2},$$

and we present preconditioners that show robustness with respect to the regularization parameter for this problem, which is more challenging from a linear algebra perspective.

In the following we use the heat equations as an example PDE. In principle the approaches described here can be extended to other PDEs, as for the L_2 regularization case. We deliberately choose to focus on the simplest PDE example to highlight the issues corresponding directly to the regularization, not the difficulties involved in using a more complicated model, which is discussed elsewhere.

The structure of the paper is as follows. We begin in Section 2 by stating the optimal control problem in the time-dependent and time-independent cases with both distributed and boundary control. We illustrate how to obtain discretized first order conditions from a so-called discretize-then-optimize approach. In Section 3 we describe how – following a method first proposed by Ito and Kunisch [43] – the state constraints can be handled using a Moreau-Yosida penalty approach and show how to incorporate this into possible preconditioning strategies. Sections 2 and 3, which describe the application of well known techniques for solving such optimal control problems, show how the bottleneck for such codes is the solution of a very large linear system.

In Section 4 we discuss the choice of possible Krylov solvers and introduce preconditioning strategies for both the time-dependent and time-independent control problem, with an emphasis on how to handle the $\mathcal{H}_1$ regularization term. This builds on the work

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

in the literature that has been used to efficiently solve L_2 regularized problems, but the use of the $\mathcal{H}_1$ norm in the cost functional causes difficulties which require novel techniques to overcome. The development of such techniques is the main contribution of the paper. Our numerical results shown in Section 6 illustrate the efficiency of our approach.

2 Problem setup and discretization

2.1 A stationary control problem

Before describing the time-dependent control problem we fix ideas by considering a stationary optimal control problem. We wish to minimize the functional

$$\begin{aligned}\mathcal{J}_1(y,u) &= \frac{1}{2}\|y-\bar{y}\|^2_{L_2(\Omega_1)} + \frac{\beta}{2}\|u\|^2_{\mathcal{H}_1(\Omega_2)} \\ &= \frac{1}{2}\|y-\bar{y}\|^2_{L_2(\Omega_1)} + \frac{\beta}{2}\|u\|^2_{L_2(\Omega_2)} + \frac{\beta}{2}\|\nabla u\|^2_{L_2(\Omega_2)},\end{aligned} \tag{2.1}$$

where both Ω_1 and Ω_2 are subdomains of $\Omega \in \mathbb{R}^d$ with $d = 2,3$. The constraint is given by the following elliptic PDE

$$-\triangle y = \begin{cases} u & \text{in } \Omega_2, \\ 0 & \text{in } \Omega \backslash \Omega_2, \end{cases} \tag{2.2}$$

together with Dirichlet boundary conditions, $y = g$ on $\partial\Omega$. We refer to y as the state and u as the corresponding control, which is used to drive the state variable as close as possible to the desired state (or observations) $\bar{y}$. The above problem is the distributed control problem, as u defines the forcing of the PDE over the interior subdomain Ω_2. Another important case is given by the Neumann boundary control problem, where $\Omega_2 = \partial\Omega$ together with the PDE constraint

$$-\triangle y = f \quad \text{in } \Omega, \tag{2.3}$$

$$\frac{\partial y}{\partial n} = u \quad \text{on } \partial\Omega, \tag{2.4}$$

where f represents a fixed forcing term.

In practice, physical characteristics of the application will require *box constraints* on the control and/or the state. Typical bounds would be

$$u_a \leq u \leq u_b$$

for the control and

$$y_a \leq y \leq y_b$$

for the state. The numerical treatment of these constraints is by now well established [6,7,38] but nevertheless represents a computational challenge, in particular for the state constraints [11].

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

We follow the discretize-then-optimize paradigm and discretize the PDE and the objective function using Q1 finite elements [23, 67]; we employ the deal.II [2] finite element package for our numerical experiments.

We derive the discrete optimality system for the cost functional (2.1) with the PDE constraint (2.2), together with homogeneous Dirichlet boundary conditions for ease of exposition – the extension to other boundary conditions proceeds similarly. Let $\phi_1,\cdots\phi_n$ be a finite element basis for the interior of Ω, and suppose we extend this by $\phi_{n+1},\cdots,\phi_{n+\partial n}$ to include the boundary. Let $Y_0^h=\langle\phi_1\cdots\phi_n\rangle$, $U^h=\langle\phi_1\cdots\phi_n,\phi_{n+1},\phi_{n+\partial n}\rangle$. Furthermore, let $Y_{\Omega_1}:=\langle\hat{\phi}_1\cdots\hat{\phi}_{\hat{m}}\rangle$ and $U_{\Omega_2}:=\langle\bar{\phi}_1\cdots\bar{\phi}_{\bar{m}}\rangle$ denote the subsets of U^h with support on Ω_1 and Ω_2 respectively.

The finite dimensional analogue to (2.1), (2.2) is to find $y_h\in Y_0^h\subset\mathcal{H}_0^1(\Omega)$ and $u_h\in U^h\subset\mathcal{H}_1(\Omega)$ which satisfy

$$\min_{y_h\in Y_{\Omega_1},u_h\in U_{\Omega_2}}\frac{1}{2}||y_h-\bar{y}||^2_{L_2(\Omega_1)}+\frac{\beta}{2}||u_h||^2_{\mathcal{H}_1(\Omega_2)},$$
$$\text{s.t.}\quad\int_\Omega\nabla y_h\cdot\nabla v_h=\int_{\Omega_2}u_hv_h,\ \forall v_h\in Y_0^h.$$

We can write the optimization problem in terms of matrices as

$$\min_{\mathbf{y},\mathbf{u}}\frac{1}{2}\mathbf{y}^TM_y\mathbf{y}-\mathbf{y}^T\mathbf{b}+\frac{\beta}{2}\mathbf{u}^TM_u\mathbf{u}+\frac{\beta}{2}\mathbf{u}^TK_u\mathbf{u},\tag{2.5}$$
$$\text{s.t.}\quad K\mathbf{y}=M\mathbf{u},\tag{2.6}$$

where

$$(M_y)_{i,j}=\int_\Omega\hat{\phi}_i\hat{\phi}_j,\quad i,j=1,\cdots,\hat{m},\qquad (K_u)_{i,j}=\int_\Omega\nabla\bar{\phi}_i\cdot\nabla\bar{\phi}_j,\quad i,j=1,\cdots,\bar{m},$$
$$(M_u)_{i,j}=\int_\Omega\bar{\phi}_i\bar{\phi}_j,\quad i,j=1,\cdots,\bar{m},\qquad (K)_{i,j}=\int_\Omega\nabla\phi_i\cdot\nabla\phi_j,\ i,j=1,\cdots,n,$$
$$(M)_{i,j}=\int_\Omega\phi_i\bar{\phi}_j,\ i=1,\cdots,n,\ j=1,\cdots,\bar{m},\qquad \mathbf{b}_i=\int_\Omega\bar{y}\phi_i,\ i=1,\cdots,\hat{m}.$$

Note that in this paper we only discuss the case where $\Omega_2=\partial\Omega$ or $\Omega_2=\Omega$ and $\Omega_1=\Omega$. Other choices influence the matrix properties of M_y,M_u,K_u,M for which the techniques presented here are still applicable.

In the distributed control case the first order optimality conditions lead to the following saddle point system:

$$\begin{bmatrix}M_y & 0 & -K^T\\ 0 & \beta M_u+\beta K_u & M^T\\ -K & M & 0\end{bmatrix}\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\\ \mathbf{p}\end{bmatrix}=\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}.\tag{2.7}$$

Note that the addition of an $\mathcal{H}_1$ norm in the regularization leads to an optimality system with substantially different properties compared to the L_2 case; in particular, if $\mathbf{p}=\mathbf{0}$

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

on the boundary, we do not necessarily have that $\mathbf{u}=\mathbf{0}$ on the boundary here, which is known to be true if we use L_2 regularization (see [58, 69, Section 2.8]). If we were to use non-homogeneous boundary conditions the 3rd entry of the right hand side would hold the boundary data, as the state equation (2.6) would become $K\mathbf{y}=M\mathbf{u}-\mathbf{d}$ for some non-zero vector $\mathbf{d}$.

We treat the boundary control problem similarly. Here we get

$$\mathcal{J}_1(\mathbf{y},\mathbf{u})=\frac{1}{2}\mathbf{y}^T M_y\mathbf{y}-\mathbf{b}^T\mathbf{y}+\frac{\beta}{2}\mathbf{u}^T M_{u,b}\mathbf{u}+\frac{\beta}{2}\mathbf{u}^T K_{u,b}\mathbf{u} \tag{2.8}$$

together with

$$\widehat{K}\mathbf{y}=\widehat{N}\mathbf{u}+\mathbf{f}. \tag{2.9}$$

Here $M_{u,b}$ and $K_{u,b}$ are the boundary mass matrix and Laplacian, respectively, i.e.

$$(K_{u,b})_{i,j}=\int_{\partial\Omega}\nabla\mathrm{tr}(\phi_i)\cdot\nabla\mathrm{tr}(\phi_j),\ (M_{u,b})_{i,j}=\int_{\partial\Omega}\mathrm{tr}(\phi_i)\mathrm{tr}(\phi_j),\quad i,j=n+1,\cdots,n+\partial n,$$

where $\mathrm{tr}(\cdot)$ is the trace operator, which we use here to give us a finite element discretization of the boundary. The vector $\mathbf{f}$ represents the discretized forcing term, which for simplicity we take to be zero for the remainder of the paper. The matrix $\widehat{K}$ is the stiffness matrix, including the boundary nodes, and $\widehat{N}$ connects interior and boundary basis functions, in particular

$$(\widehat{N})_{ij}=\int_{\partial\Omega}\phi_i\mathrm{tr}(\phi_j),\quad i=1,\cdots,n+\partial n,\quad j=1,\cdots\partial n.$$

We obtain the following first order optimality system

$$\begin{bmatrix} M_y & 0 & -\widehat{K}^T \\ 0 & \beta M_{u,b}+\beta K_{u,b} & \widehat{N}^T \\ -\widehat{K} & \widehat{N} & 0 \end{bmatrix}\begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \end{bmatrix}=\begin{bmatrix} \mathbf{b} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{2.10}$$

2.2 Time-dependent problem

We now present a time-dependent version, which is of wide practical interest and will be the focus of our numerical tests. The objective function is now given by

$$\mathcal{J}_2(y,u)=\frac{1}{2}\int_0^T\int_{\Omega_1}(y-\bar{y})^2dxdt+\frac{\beta}{2}\int_0^T\int_{\Omega_2}u^2dxdt+\frac{\beta}{2}\int_0^T\int_{\Omega_2}(\nabla u)^2dxdt, \tag{2.11}$$

where all functions are simply time-dependent versions of their steady counterparts presented above. For the distributed control problem we apply the time-dependent parabolic constraint

$$\begin{aligned} &y_t-\triangle y=\begin{cases} u, & \text{for } (\mathbf{x},t)\in\Omega_2\times[0,T], \\ 0, & \text{for } (\mathbf{x},t)\in\Omega\backslash\Omega_2\times[0,T], \end{cases} \\ &y=g,\quad \text{on } \partial\Omega, \\ &y=y_0,\quad \text{at } t=0, \end{aligned}$$

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

for some prescribed functions g, y_0. In case of a boundary control problem, where $\Omega_2 = \partial\Omega$ and again take the heat equation as our PDE constraint:

$$y_t - \triangle y = f \quad \text{for } (\mathbf{x},t) \in \Omega \times [0,T], \tag{2.12}$$

$$\frac{\partial y}{\partial n} = u \quad \text{on } \partial\Omega, \tag{2.13}$$

$$y = y_0, \quad \text{at } t = 0. \tag{2.14}$$

For the discretization of the time-dependent objective function we use the trapezoidal rule for the time integral and finite elements in space to give

$$\mathcal{J}_2(\mathbf{y},\mathbf{u}) = \frac{1}{2}\mathbf{y}^T\mathcal{M}_y\mathbf{y} + \widehat{\mathbf{b}}^T\mathbf{y} + \frac{\beta}{2}\mathbf{u}^T\mathcal{M}_u\mathbf{u} + \frac{\beta}{2}\mathbf{u}^T\mathcal{K}_u\mathbf{u}, \tag{2.15}$$

where $\widehat{\mathbf{b}} = [1/2\mathbf{b}^T, \mathbf{b}^T, \cdots, \mathbf{b}^T, 1/2\mathbf{b}^T]^T$,

$$\begin{aligned}
&\mathcal{M} = \text{blkdiag}(M_y, \cdots, M_y),\\
&\mathcal{M}_y = \text{blkdiag}(1/2M_y, M_y, \cdots, M_y, 1/2M_y),\\
&\mathcal{M}_u = \text{blkdiag}(1/2M_u, M_u, \cdots, M_u, 1/2M_u), \quad \text{and}\\
&\mathcal{K}_u = \text{blkdiag}(1/2K_u, K_u, \cdots, K_u, 1/2K_u),
\end{aligned}$$

which are simply block-variants of the previously defined matrices over the domains Ω_1 and Ω_2. Note that in the time-dependent case we abuse the notation $\mathbf{y}$, $\mathbf{u}$ defined earlier, i.e., $\mathbf{y} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T, \cdots, \mathbf{y}_{N_T}^T\right]^T$, etc.; we believe it will be clear from the context which of the two we are currently considering. Using this notation and a backward Euler scheme, we can write down a one-shot discretization of the time-dependent PDE as follows

$$-\underbrace{\begin{bmatrix} L & & & \\ -M & L & & \\ & \ddots & \ddots & \\ & & -M & L \end{bmatrix}}_{\mathcal{K}} \mathbf{y} + \tau\mathcal{M}\mathbf{u} = \mathbf{d} \tag{2.16}$$

with $L = M + \tau K$ and $\mathbf{d}$ holding the initial conditions for the heat equation. For more details see [5,20,66].

We form the Lagrangian and write down the first order conditions in a linear system,

$$\begin{bmatrix} \tau\mathcal{M}_y & 0 & -\mathcal{K}^T \\ 0 & \tau\beta(\mathcal{M}_u + \mathcal{K}_u) & \tau\mathcal{M} \\ -\mathcal{K} & \tau\mathcal{M} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \tau\widehat{\mathbf{b}} \\ 0 \\ \mathbf{d} \end{bmatrix}, \tag{2.17}$$

in the case of the distributed control problem, and

$$\begin{bmatrix} \tau\mathcal{M}_y & 0 & -\mathcal{K}^T \\ 0 & \tau\beta(\mathcal{M}_{u,b} + \mathcal{K}_{u,b}) & \tau\mathcal{N}^T \\ -\mathcal{K} & \tau\mathcal{N} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} M_y\bar{\mathbf{y}} \\ 0 \\ \mathbf{d} \end{bmatrix} \tag{2.18}$$

for boundary control, where $\mathcal{N} = \text{blkdiag}(N, \cdots, N)$.

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

3 Handling the state constraints

Box constraints for the state $\mathbf{y}$ can be dealt with efficiently using a penalty term. The Moreau-Yosida penalty function has proven to be a viable tool: see [36, 43, 52] and the references mentioned therein. One can also use the Moreau-Yosida technique for box constraints on the control but the primal-dual active set method [38] is mostly the method of choice. We briefly describe the Moreau-Yosida technique for the distributed control problem. A more thorough discussion can be found in the references mentioned earlier. The modified objective function becomes

$$\mathcal{J}_{MY}(y,u) = \mathcal{J}_2(y,u) + \frac{1}{2\varepsilon}\|\max\{0, y-y_b\}\|_Q^2 + \frac{1}{2\varepsilon}\|\min\{0, y-y_a\}\|_Q^2 \tag{3.1}$$

for the state constrained case. Here $Q = \Omega_1 \times [0,T]$ is the space-time cylinder. In accordance with [36], we can employ a semi-smooth Newton scheme that leads to the following linear system

$$\begin{aligned}
&\begin{bmatrix} \tau\mathcal{M}_y + \varepsilon^{-1} G_{\mathcal{A}} \mathcal{M}_y G_{\mathcal{A}} & 0 & -\mathcal{K}^T \\ 0 & \tau\beta(\mathcal{M}_{u,b} + \mathcal{K}_{u,b}) & \tau\mathcal{N}^T \\ -\mathcal{K} & \tau\mathcal{N} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{M}_y \bar{\mathbf{y}} + \varepsilon^{-1} \left(G_{\mathcal{A}_+} \mathcal{M}_y G_{\mathcal{A}_+} y_b + G_{\mathcal{A}_-} \mathcal{M}_y G_{\mathcal{A}_-} y_a \right) \\ 0 \\ \mathbf{d} \end{bmatrix},
\end{aligned} \tag{3.2}$$

where the block-diagonal matrix

$$G_{\mathcal{A}} \mathcal{M}_y G_{\mathcal{A}} = \text{blkdiag}(G_{\mathcal{A}^1} M_y G_{\mathcal{A}^1}, \cdots, G_{\mathcal{A}^{N_T}} M_y G_{\mathcal{A}^{N_T}})$$

defines the contribution of the penalty term with the active set $\mathcal{A}^k$ for time-step k defined as follows. We set where we define the active sets as $\mathcal{A}_+^k = \{i : \mathbf{y}_i^k > (y_b)_i^k\}$, and $\mathcal{A}_-^k \{i : \mathbf{y}_i^k < (y_a)_i^k\}$, and $\mathcal{A}^k = \mathcal{A}_+^k \cup \mathcal{A}_-^k$; the matrices G are diagonal matrix variants of the characteristic function for the corresponding sets, i.e.,

$$(G_{\mathcal{A}^k})_{ii} = \begin{cases} 1 & \text{for } i \in \mathcal{A}^k, \\ 0 & \text{otherwise.} \end{cases}$$

Our focus is on the efficient solution of the linear systems (3.2), which are of saddle point type. Note that the active sets defined above within an iterative process such as the semi-smooth Newton scheme are computed based on the state at the previous iteration, but for simplicity we neglect the iteration index. For more details of semi-smooth Newton methods we refer to [41, 44, 70]; there is also recent theory introducing path-following approaches for the penalty parameter ε [39].

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

4 Preconditioning

4.1 Choice of Krylov solver and Schur complement preconditioning

As mentioned in the introduction, the linear systems that arise from PDE-constrained optimization are very often too large for direct solvers to be effective, and for scalable and efficient solution of these linear systems the combination of a state-of-the-art solver with an efficient preconditioning technique is crucial. In this section we derive preconditioners for each of the problems presented earlier, but first mention the choice of the iterative scheme. Krylov solvers are for many applications the method of choice [64], as they are cheap to apply; at each step they only require a matrix vector product, the evaluation of the preconditioners, and the evaluation of inner products. These methods build up a low-dimensional subspace that can be used to approximate the solution to the linear system.

There are a variety of Krylov subspace methods, and the most effective to use depends on the properties of the linear system. Here we focus on the development of effective preconditioners and we will focus less on the choice of linear solver.

Schur-complement based preconditioners, based on approximations to $S := BA^{-1}B^T$, have proved to be effective. Popular choices are a block diagonal preconditioner $\mathcal{P}_1 = \text{blkdiag}(A,S)$, or a nonsymmetric preconditioner,

$$\mathcal{P}_2 = \begin{pmatrix} A & 0 \\ B & -S \end{pmatrix}.$$

Naturally, these are too expensive for any realistic problem, but if we can approximate both the $(1,1)$-block and the Schur-complement of $\mathcal{A}$, then the underlying Krylov method will converge in a small number of steps. In the following sections we describe how to find good approximations to these blocks for the application considered here.

4.2 The $(1,1)$-block

Our first goal is to efficiently approximate the $(1,1)$-block of the saddle point matrix. Parts of the $(1,1)$-block here consist of lumped mass matrices, which are diagonal and can simply be inverted. If, on the other hand, the user prefers to use consistent mass matrices they can use the Chebyshev semi-iteration [72]. If the $(1,1)$-block part corresponding to the discretization of the state misfit part of the objective function is only semi-definite, e.g., via a partial observation operator, we can add a small perturbation to the zero blocks within the preconditioned and hence make this part positive definite so the above applies. In more detail, we replace the zero blocks in A by blocks of the form ηI with η a small parameter greater than zero. Note that this technique can also be used for an approximation of the Schur-complement in case the $(1,1)$-block is semi-definite [5,66].

The matrix part corresponding to the discretization of the $\mathcal{H}_1$ term in the objective function is more complicated as it is not diagonal. The good news in this case is that

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

the operator and the corresponding matrix representation are not only symmetric but also positive definite. This allows the use of either geometric [32,75] or algebraic [24,60] multigrid techniques.

4.3 Schur-complement approximation

The methods described in Section 4.2 efficiently approximate the $(1,1)$-block, A, of the saddle point system; we use $\widehat{A}$ to represent such an approximation to A for the remainder of this paper. Our goal now is to introduce efficient approximations $\hat{S}$ to the Schur-complement S.

The Schur complement of the system matrix (2.17) is

$$S=\tau^{-1}\mathcal{K}\mathcal{M}_y^{-1}\mathcal{K}+\tau\beta^{-1}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1}\mathcal{M}^T. \tag{4.1}$$

There are various ways to approximate S; one of the simplest is

$$S\approx\tau^{-1}\mathcal{K}\mathcal{M}_y^{-1}\mathcal{K},$$

which for larger β often performs well but is not robust with respect to this parameter.

In order to develop a more robust method we look for a more sophisticated approximation inspired by [53] that more accurately mirrors S by also including the second term in (4.1). We have two options here, either a symmetric version,

$$\hat{S}_1=\tau^{-1}\left(\mathcal{K}+\hat{\mathcal{M}}\right)\mathcal{M}_y^{-1}\left(\mathcal{K}+\hat{\mathcal{M}}\right)^T,$$

which can be used within MINRES [51], or a non-symmetric approximation

$$\hat{S}_2=\tau^{-1}\left(\mathcal{K}+\hat{\mathcal{M}}_1\right)\mathcal{M}_y^{-1}\left(\mathcal{K}+\hat{\mathcal{M}}_2\right)^T$$

to be employed with a non-symmetric solver, e.g. GMRES [63] or BICG [25]. The goal is now to find $\hat{\mathcal{M}}_1$, $\hat{\mathcal{M}}_2$, and $\hat{\mathcal{M}}$ such that

$$\tau^{-1}\hat{\mathcal{M}}_1\mathcal{M}_y^{-1}\hat{\mathcal{M}}_2^T=\tau\beta^{-1}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1}\mathcal{M}^T$$

and

$$\tau^{-1}\hat{\mathcal{M}}\mathcal{M}_y^{-1}\hat{\mathcal{M}}^T=\tau\beta^{-1}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1}\mathcal{M}^T.$$

We start by deriving the symmetric approximation to S using

$$\hat{\mathcal{M}}:=\frac{\tau}{\sqrt{\beta}}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1/2}\mathcal{M}_y^{1/2}.$$

We then obtain the following Schur-complement approximation

$$\hat{S}_1=\tau^{-1}\left(\mathcal{K}+\frac{\tau}{\sqrt{\beta}}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1/2}\mathcal{M}_y^{1/2}\right)\mathcal{M}_y^{-1}\left(\mathcal{K}+\frac{\tau}{\sqrt{\beta}}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1/2}\mathcal{M}_y^{1/2}\right)^T.$$

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

This has the advantage that the approximation is symmetric and positive definite, which would allow us to use MINRES. However, the drawback is that this expression involves the square root of large-scale non-diagonal matrices, $\mathcal{K}_u$.

We now turn our attention to the non-symmetric approximation. Using

$$\hat{\mathcal{M}}_1 := \frac{\tau}{\sqrt{\beta}}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1}\mathcal{M}_y, \tag{4.2}$$

$$\hat{\mathcal{M}}_2 := \frac{\tau}{\sqrt{\beta}}\mathcal{M}, \tag{4.3}$$

we introduce the non-symmetric approximation

$$\hat{S}_2 = \tau^{-1}\left(\mathcal{K}+\frac{\tau}{\sqrt{\beta}}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1}\mathcal{M}_y\right)\mathcal{M}_y^{-1}\left(\mathcal{K}+\frac{\tau}{\sqrt{\beta}}\mathcal{M}\right)^T.$$

This configuration does not require the square root of a potentially very large matrix.

For any preconditioner to be effective we must be able to evaluate the inverse of the Schur-complement approximation quickly. We now focus on the non-symmetric approximation but discuss the symmetric approximation in Section 5 when we analyze the approximation quality of both Schur-complement approximations.

The second part $\left(\mathcal{K}+\frac{\tau}{\sqrt{\beta}}\mathcal{M}\right)$ of $\hat{S}_2$ is easy to approximate as this is simply a block-triangular matrix with symmetric positive definite matrices along the diagonal. We therefore use an algebraic multigrid approximation for the diagonal blocks and then proceed backwards approximating the inverse of $\left(\mathcal{K}+\frac{\tau}{\sqrt{\beta}}\mathcal{M}\right)$ requiring the application of N_T algebraic multigrid operators.

The approximation of the inverse of $\left(\mathcal{K}+\frac{\tau}{\sqrt{\beta}}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1}\mathcal{M}_y\right)$ is more involved. We are interested in solving systems of the form

$$\left(\mathcal{K}+\frac{\tau}{\sqrt{\beta}}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1}\mathcal{M}_y\right)u = f$$

and interpret this as the Schur-complement of the auxiliary system

$$\begin{bmatrix} \mathcal{K} & \mathcal{M} \\ \mathcal{M}_y & -\frac{\sqrt{\beta}}{\tau}(\mathcal{M}_u+\mathcal{K}_u) \end{bmatrix}\begin{bmatrix} u \\ * \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}. \tag{4.4}$$

Recalling the block-structure of the involved matrices, it is easy to see that we can proceed with a forward substitution that requires the solution of diagonal blocks given by

$$\begin{bmatrix} M+\tau K & M \\ M_y & -\frac{\sqrt{\beta}}{\tau}(M_u+K_u) \end{bmatrix}. \tag{4.5}$$

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

Even this block is not suitable to be inverted directly and we use a stationary iteration to approximate the solution to this system. Such an iteration proceeds by computing

$$u^{k+1} = u^k + \omega W^{-1} r_k,$$

where r_k is the residual for the system matrix used in (4.5) and a right-hand-side used within the preconditioner application. The matrix

$$W = \begin{bmatrix} \widehat{M+\tau K} & \\ & \frac{\sqrt{\beta}}{\tau}\widehat{(M_u + K_u)} \end{bmatrix}$$

is the preconditioner for (4.5). Here $\widehat{(\cdots)}$ signifies the algebraic multigrid approximation to the corresponding matrix.

Boundary control

The matrix structure in the case of a boundary control problem is very similar to the distributed control problem but nevertheless there are significant differences in the properties of some of the blocks. Therefore, we now discuss a Schur complement approximation for the boundary control problem driven by the system matrix (2.18). The Schur complement is now given by

$$S = \tau^{-1}\mathcal{K}\mathcal{M}_y^{-1}\mathcal{K}^T + \frac{\tau}{\beta}\mathcal{N}(\mathcal{M}_{u,b} + \mathcal{K}_{u,b})^{-1}\mathcal{N}^T.$$

For the reasons described above, we again focus on the non-symmetric approximation

$$\hat{S} = \tau^{-1}\left(\mathcal{K} + \frac{\tau}{\sqrt{\beta}}\mathcal{N}(\mathcal{M}_{u,b} + \mathcal{K}_{u,b})^{-1}\mathcal{N}^T\right)\mathcal{M}_y^{-1}\left(\mathcal{K} + \frac{\tau}{\sqrt{\beta}}\mathcal{M}_y\right)^T.$$

Again, the evaluation of the preconditioner $\hat{S}^{-1}$ needs to be discussed. While the term $\left(\mathcal{K} + \frac{\tau}{\sqrt{\beta}}\mathcal{M}_y\right)^{-T}$ can easily be approximated using multigrid techniques in combination with backward substitution, the term $\left(\mathcal{K} + \frac{\tau}{\sqrt{\beta}}\mathcal{N}(\mathcal{M}_{u,b} + \mathcal{K}_{u,b})^{-1}\mathcal{N}^T\right)^{-1}$ is more complicated to approximate. Note again that we only need to focus on the diagonal blocks of this matrix which correspond to the system

$$\begin{bmatrix} M+\tau K & N \\ N^T & -\frac{\sqrt{\beta}}{\tau}(M_{u,b} + K_{u,b}) \end{bmatrix}. \tag{4.6}$$

We proposed earlier the use of a stationary iteration, but found choice of damping parameter to be much more critical here; the value needed to be tuned by hand, which is

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

not desirable. We therefore use the non-linear iterative method GMRES [63] to evaluate the system (4.6), together with a preconditioner

$$W = \begin{bmatrix} \widehat{M+\tau K} & 0 \\ N^T & -\widehat{(M_{u,b}+K_{u,b})} \end{bmatrix}. \tag{4.7}$$

Here $\widehat{\cdots}$ again represents that the action of the inverse of these blocks is given by a fixed number of steps of an algebraic multigrid method. Note that, because of the use of GMRES as an inner iteration, the preconditioner $\mathcal{P}_2$ is nonlinear, and theory dictates that we should use a flexible outer method such as FGMRES [62]. By using a rather small tolerance to stop GMRES we seem to avoid convergence difficulties, allowing us to use a standard Krylov method; see Section 6 for details. An alternative would be to use a sparse direct method [22, 42] to solve for the sub-problem (4.6), giving us a hybrid solution method.

State constraints

The situation is not much different in the case when state constraints are present. Here the system matrix is

$$\begin{bmatrix} \tau\mathcal{M}_\varepsilon & 0 & -\mathcal{K}^T \\ 0 & \tau\beta(\mathcal{M}_u+\mathcal{K}_u) & \tau\mathcal{M} \\ -\mathcal{K} & \tau\mathcal{M} & 0 \end{bmatrix}, \tag{4.8}$$

where each of the blocks of $\mathcal{M}_\varepsilon$ is now given by $\mathcal{M}_y+\varepsilon^{-1}G_{\mathcal{A}_i}M_yG_{\mathcal{A}_i}$. The Schur-complement is now

$$S = \tau^{-1}\mathcal{K}\mathcal{M}_\varepsilon^{-1}\mathcal{K}^T + \frac{\tau}{\beta}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1}\mathcal{M}.$$

We can now proceed as n the absence of state constraints. An approximation of S is chosen to be

$$\widehat{S} = \tau^{-1}\left(\mathcal{K}+\frac{\tau}{\sqrt{\beta}}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1}\mathcal{M}_\varepsilon^{1/2}\right)\mathcal{M}_\varepsilon^{-1}\left(\mathcal{K}^T+\frac{\tau}{\sqrt{\beta}}\mathcal{M}_\varepsilon^{1/2}\mathcal{M}\right).$$

We use the symmetric matrix $\mathcal{M}_\varepsilon^{1/2}$ because this makes all factors of the approximation $\hat{S}$ dependent on ε. A solve with $\widehat{S}$ is approximated as before, where the diagonal blocks of $\left(\mathcal{K}^T+\frac{\tau}{\sqrt{\beta}}\mathcal{M}_\varepsilon^{1/2}\mathcal{M}\right)^{-1}$ are approximated by an algebraic multigrid technique. Note that due to the matrix $\mathcal{M}_\varepsilon$ the diagonal blocks of this matrix are different at each outer iteration, and we need to recompute the algebraic multigrid approximation; update techniques to exploiting this structure should be investigated in the future to streamline the solver.

The term $\left(\mathcal{K}+\frac{\tau}{\sqrt{\beta}}\mathcal{M}(\mathcal{M}_u+\mathcal{K}_u)^{-1}\mathcal{M}_\varepsilon^{1/2}\right)$ is again harder to deal with and as previously we use an auxiliary system

$$\begin{bmatrix} \mathcal{K} & \mathcal{M} \\ \mathcal{M}_\varepsilon^{1/2} & \frac{-\sqrt{\beta}}{\tau}(\mathcal{M}_u+\mathcal{K}_u) \end{bmatrix},$$

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

which we can permute to be of block-triangular form. We are then left with approximately solving a system for

$$\begin{bmatrix} M+\tau K & M \\ M_{\varepsilon,i}^{1/2} & \frac{-\sqrt{\beta}}{\tau}(M_u+K_u) \end{bmatrix},$$

where i indicates the i-th block corresponding to the i-th point in time and its corresponding structure coming from the active set. Our strategy is again to use an accurate solution via a preconditioner GMRES method employing the preconditioner

$$W_i = \begin{bmatrix} \widehat{[M+\tau K]} & 0 \\ M_{\varepsilon,i}^{1/2} & -\widehat{[M_u+K_u]} \end{bmatrix}.$$

Here $\widehat{[\cdots]}$ indicates the use of an algebraic multigrid in the inversion of this matrix.

5 Eigenvalue analysis

Our goal here is to analyze the quality of the preconditioners proposed earlier. As described in the previous section, the approximation of the $(1,1)$-block is relatively straightforward using standard tools, such as multigrid, which are well understood. We therefore focus solely on the quality of the Schur-complement approximation.

We use the methodology introduced by Pearson and Wathen [54] for the stationary case that was later generalized for the time-dependent case (see [53]). There a symmetric Schur-complement approximation was chosen, and the quality of the approximation was measured by bounding the eigenvalues of $\hat{S}^{-1}S$ via the Rayleigh quotient

$$R := \frac{v^T S v}{v^T \hat{S}_1 v}.$$

We here briefly illustrate their argument in order to assess what parts carry over here. One can write

$$\frac{v^T S v}{v^T \hat{S}_1 v} = \frac{a^T a + b^T b}{a^T a + b^T b + b^T a + a^T b} \tag{5.1}$$

with suitably chosen vectors a and b. It is easy to see from

$$0 \leq (a-b)^T (a-b) = a^T a + b^T b - a^T b - b^T a \tag{5.2}$$

that $R \geq \frac{1}{2}$. Pearson and co-authors then proceeded by showing that $a^T b + b^T a$ is positive to conclude that the R is bounded by 1 from above. For the distributed control case this is both true in the steady [54] and transient [53] case.

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

Our goal is to carry this analysis over to our setup. We focus on the distributed control case here, where

$$a := \tau^{-1/2} \mathcal{M}_y^{-1/2} \mathcal{K}^T v, \tag{5.3}$$

$$b := \tau^{1/2} \beta^{-1/2} (\mathcal{M}_u + \mathcal{K}_u)^{-1/2} \mathcal{M}^T v. \tag{5.4}$$

Using

$$v^T S v = a^T a + b^T b$$

as well as (5.2) we can see that the lower bound for the approximation $\hat{S}_1$ is still valid, i.e., $R \geq \frac{1}{2}$ regardless of the mesh-parameter and the regularization parameter.

The interesting question from now on is whether the upper bound $R \leq 1$ is still valid. For this to be true we immediately see from (5.1) that $a^T b + b^T a$ needs to be positive. We proceed by considering the simpler time-independent case for which the matrix structure is very similar. In that case we obtain

$$a^T b + b^T a = \beta^{-1/2} v^T \left(K M_y^{-1/2} (M_u + K_u)^{-1/2} M^T + M (M_u + K_u)^{-1/2} M_y^{-1/2} K \right) v.$$

To see that in the L_2 norm case this was positive we set $K_u = 0$ and $M_u = M_y = M$ and obtain

$$\beta^{-1/2} v^T \left(K + K^T \right) v$$

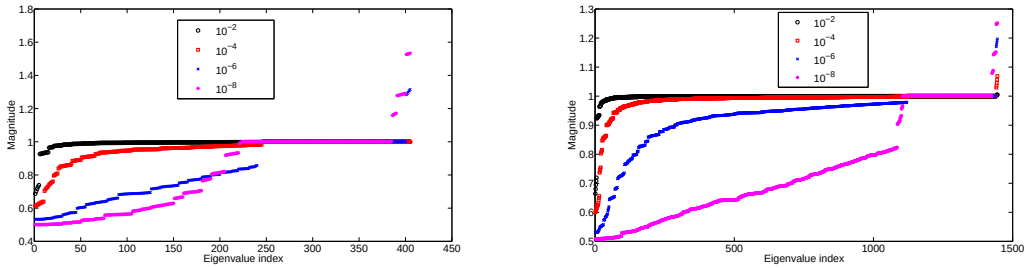
which is obviously positive. The same is true for the time-dependent problem without $\mathcal{H}_1$-norm (see [53] for a proof). Unfortunately, once the $\mathcal{H}_1$-norm is considered the positivity of $a^T b + b^T a$ is lost.

So why does the $\mathcal{H}_1$-norm cause a problem as its discretization only introduces a symmetric and positive definite matrix $M_u + K_u$? This is clear from the fact that in general $WV + VW \not> 0$ even when both V and W are symmetric and positive definite matrices. A simple example is given when W and V correspond to a Dirichlet and Neumann Laplacian, respectively.

Note that in our case $a^T b + b^T a$ is of precisely this form and the computation of the eigenvalues of $\left(K M_y^{-1/2} (M_u + K_u)^{-1/2} M^T + M (M_u + K_u)^{-1/2} M_y^{-1/2} K \right)$ reveals several negative eigenvalues.

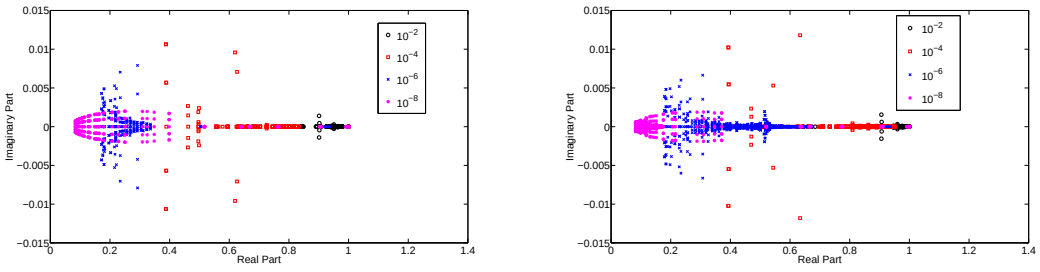
Nevertheless, the spread of the eigenvalues above the desired value of 1 with varying β and mesh-parameter is not severe as illustrated by the eigenvalues shown in Fig. 1. Fig. 1(a) shows the eigenvalues of $\hat{S}_1$ for a coarse mesh and four values of the regularization parameter β and Fig. 1(b) shows the eigenvalues for the same values of β but on a finer mesh. From these pictures we can see that the magnitude of the eigenvalues does not increase for the finer mesh and that most of the eigenvalues are contained in the interval $\left[\frac{1}{2}, 1\right]$ with some outliers that do not move much beyond 1 when the regularization parameter is decreased. The major disadvantage of the approximation $\hat{S}_1$ is the use of the matrix square roots, which is infeasible for large systems. We hence move to the

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press



(a) Coarse (b) Finer

Figure 1: Eigenvalues for two different meshes and a variety of regularization parameters. We show coarse mesh on the left and slightly finer mesh on the right. We use $N_T = 5$.



(a) Coarse (b) Finer

Figure 2: Eigenvalues for two different meshes and a variety of regularization parameters. We show coarse mesh on the left and slightly finer mesh on the right. We use $N_T = 5$.

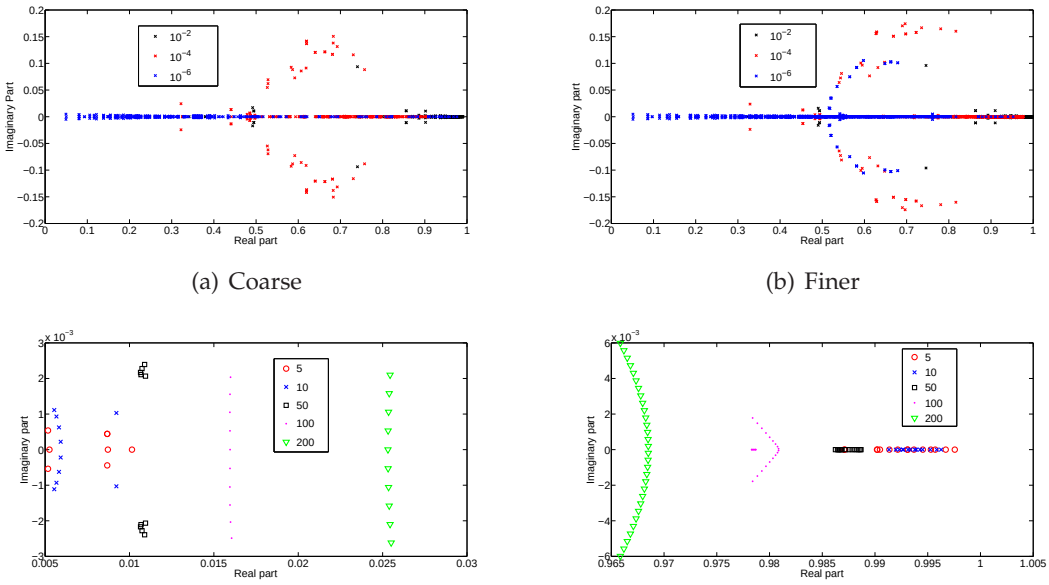
nonsymmetric approximation $\hat{S}_2$ for which the above used Rayleigh quotient analysis is unfortunately not applicable.

Nevertheless, we expect the eigenvalues of the pencil $(S, \hat{S}_2)$ to provide guidance on the speed of convergence of our iterative scheme. We here want to numerically study the eigenvalues of $\hat{S}_2^{-1}S$ to obtain information that can allow us to understand the convergence of a nonsymmetric solver using the nonsymmetric Schur-complement approximation $\hat{S}_2$.

Fig. 2 shows eigenvalue distributions of $\hat{S}_2^{-1}S$ for two different mesh-sizes and a variety of regularization parameters. The comparison of both plots 2(a) for the coarse mesh and 2(b) for the refined one indicates that for very small values of β the eigenvalues move closer towards the origin but stay sufficiently far away from zero. Additionally, this behaviour does not change when the mesh is refined so we expect robust iteration numbers with respect to a refinement in space. We computed approximations to the eigenvalues closest to the origin of $\hat{S}_2^{-1}S$ for one further mesh and found these to be in the same region as the smallest eigenvalues shown in Fig. 2.

Our numerical results given in Section 6 indicate that this choice of Schur complement approximation allows for good convergence with relatively robust iteration numbers.

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press



(a) Coarse (b) Finer

(c) Leftmost eigenvalues (d) Rightmost eigenvalues

Figure 3: Eigenvalues for two different meshes and a variety of regularization parameters. We show a coarse mesh on the left and slightly finer mesh on the right. The top row shows the eigenvalues for $\hat{S}_2^{-1}S$ for $N_T=5$. The lower figures illustrate the dependency of the smallest and largest eigenvalues of $\hat{S}_2^{-1}S$ on the number of time-steps and hence τ.

6 Numerical results

We now want to illustrate how the preconditioners presented above perform when applied to a variety of problems. As mentioned earlier we employ a finite element discretization, here done with the finite element package deal.II [2]. We discretize the state, control and adjoint state variables using **Q1** elements. For symmetric methods the stopping criterion is often inherent to the problem [74]. In the nonsymmetric context the debate is much more open an we decide to use the relative residual with $x_0=0$ based on the discussion in [3]. Hence, we present results for both a tolerance of 10^{-4} and a tighter tolerance of 10^{-6} for the relative residual within BICG using the preconditioner $\mathcal{P}_2$. For the algebraic multigrid preconditioner we use the Trilinos ML package [27] that implements a smoothed aggregation AMG. Within the algebraic multigrid we used 6 steps of a Chebyshev smoother in combination with the application of two V-cycles. For time-dependent problems we show the degrees of freedom only for one grid point in time (i.e. for a single time-step) and we are implicitly solving a linear system of dimension 3 times the number of time-steps (N_t) times the degrees of freedom of the spatial discretization (n). For example, a spatial discretization with 274625 spatial unknowns and 20 time-steps corresponds to an overall linear system of dimension 16 477 500.

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

6.1 Distributed control

No state constraints

In this section we show results for the time-dependent case. First, we consider the case when no state constraints are present. Here, we work with a fixed time-step $\tau = 0.05$, which results in 20 time-steps. In all tables we only show the degrees of freedom associated with the discretization of the spatial domain. The desired state is now given by

$$\bar{\mathbf{y}} = \exp\left(-64\left((x_0-0.5)^2+(x_1-0.5)^2\right)\right)$$

and $\mathbf{y} = \bar{\mathbf{y}}$ on $\partial\Omega$, where the domain is $[0,1]^2$. The results for this setup are shown in Table 1 for various mesh-parameters and values of the regularization parameter β.

Table 1: Results for the distributed control problem and varying mesh and regularization parameter. This table shows iteration numbers and timings for BICG with a nonsymmetric Schur complement approximation using 10 Uzawa steps and a damping parameter $\omega = 0.1$. The tolerance of the iterative solver is set to 10^{-6}.

DoF	$\beta = 10^{-2}$	$\beta = 10^{-4}$	$\beta = 10^{-6}$
	# it(t)	# it(t)	# it(t)
1089	15(40.1)	17(45.9)	28(72.9)
4225	15(129.2)	18(153.5)	29(242.1)
16641	18(554.2)	22(669.7)	31(932.2)
66049	19(1627.6)	27(2280.1)	36(2995.4)
263169	23(5922.8)	28(7203.9)	44(11389.2)

Table 2: Results for the distributed control problem and varying mesh and regularization parameter. This table shows iteration numbers and timings for BICG with a nonsymmetric Schur complement approximation using 10 Uzawa steps and a damping parameter $\omega = 0.1$. The tolerance of the iterative solver is set to 10^{-6}.

DoF	$\beta = 10^{-2}$	$\beta = 10^{-4}$	$\beta = 10^{-6}$
	# it(t)	# it(t)	# it(t)
1089	13(35.1)	13(35.2)	22(57.3)
4225	13(112.6)	15(128.8)	22(184.8)
16641	15(462.3)	15(462.2)	25(756.1)
66049	17(1442.6)	20(1691.4)	31(2578.7)
263169	19(4928.3)	22(5843.9)	34(8368.3)

State constraints

We now consider the problem with state constraints. The defining parameters are given by the desired state

$$\bar{\mathbf{y}} = -tx_0\exp\left(-\left((x_0-0.5)^2+(x_1-0.5)^2\right)\right)$$

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

Table 3: Results for the distributed control problem and varying mesh and regularization parameter. This table shows iteration numbers and timings for BICG with a nonsymmetric Schur complement approximation using 10 Uzawa steps and a damping parameter $\omega=0.1$. The tolerance of the iterative solver is set to 10^{-6}.

DoF	$\beta=10^{-2}$ # it(t)	$\beta=10^{-4}$ # it(t)	$\beta=10^{-6}$ # it(t)
1089	13(35.1)	13(35.2)	22(57.3)
4225	13(112.6)	15(128.8)	22(184.8)
16641	15(462.3)	15(462.2)	25(756.1)
66049	17(1442.6)	20(1691.4)	31(2578.7)
263169	19(4928.3)	22(5843.9)	34(8368.3)

Table 4: Results for the state-constrained problem. We here vary the penalization parameter ε. Shown are the Newton iteration numbers for a Newton tolerance of 10^{-3} for the first two columns and a tolerance for 10^{-2} for the case $\varepsilon=10^{-4}$. As the number of Newton iterations increased we here only show iteration numbers for a stopping tolerance of 10^{-2} for the outer iteration. Further we give the average number of BICG iterations and the maximal number of GMRESiterations needed for the evaluation of the preconditioner. The tolerance of the iterative solver is set to 10^{-6}.

DoF	$\varepsilon=10^{0}$ AS/BICG/GMRES	$\varepsilon=10^{-2}$ AS/BICG/GMRES	$\varepsilon=10^{-4}$ AS/BICG/GMRES
81	3/23.7/12	7/21.3/19	6/25.8/41
289	3/32.7/16	7/26.9/23	6/35.2/52
1089	3/51.3/19	6/37.0/27	2/45.5/75
4225	3/74.0/23	6/53.3/43	2/56.5/109

with zero initial and boundary condition. We then consider a fixed regularization parameter $\beta=10^{-4}$, which then allows us to consider the lower bound $-0.1\leq\mathbf{y}$ for all time-steps. The results are shown in Table 4, where we vary the penalization parameter from 1 to 10^{-4}. The iteration numbers obtained show a small increase with respect to the mesh-size. This might be due to the approximation quality of the diagonal blocks used within the evaluation of the preconditioner W. We observed that we needed to increase the number of V-cycles within the AMG method to 8 to obtain a robust performance. Future research should be devoted to obtaining preconditioners that allow updating to deal with the changing blocks involving components from the active sets and also show more robustness with respect to parameter-dependent matrices (here in particular β and ε).

6.2 Boundary control

We now show results for the boundary control case where the desired state is given by

$$\bar{\mathbf{y}}=-\exp(t)\sin(2\pi x_0x_1x_2)\exp\left(-\left((x_0-0.5)^2+(x_1-0.5)^2+(x_2-0.5)^2\right)\right)$$

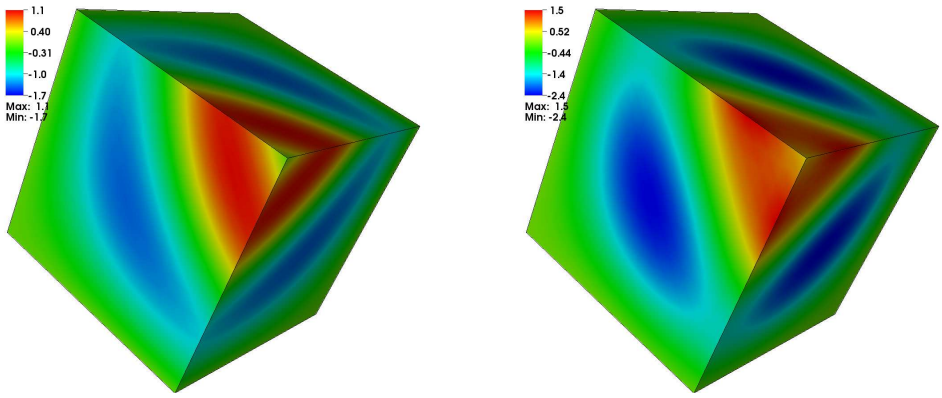
on the three-dimensional domain $\Omega=[0,1]^3$. The results with the Schur complement approximation $\hat{S}$ with varying mesh-size and regularization parameter β are shown in

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

Table 5: Results for the boundary control problem and varying mesh and regularization parameter. This table shows iteration numbers for BICG and the maximal number of GMRES iterations used for the preconditioner. The tolerance of the iterative solver is set to 10^{-6}.

DoF	$\beta=10^{-2}$ BICG/GMRES	$\beta=10^{-4}$ BICG/GMRES
729	18(21)	19(38)
4913	18(22)	17(40)
35937	19(24)	17(44)
274625	19(25)	19(47)

Table 5. We again want to emphasize that we use preconditioned GMRES to evaluate the diagonal-blocks of the Schur-complement approximation. The tolerance is set rather tight on the one hand to guarantee that as an outer iteration BICG is still suited and on the other hand to guarantee that we obtain robustness with respect to parameter changes. We additionally state for every problem the maximal number of iterations that was needed for GMRES. It can be seen that the number of BICGiterations are robust with respect to parameter changes. The number of iterations for the GMRES preconditioner increases slightly with a decrease of the regularization parameter.



(a) Desired state $\bar{\mathbf{y}}_{12}$ (b) Computed state $\mathbf{y}_{12}$

Figure 4: Desired state and computed state for boundary control problem. Here the regularization parameter was set to $\beta=10^{-6}$.

7 Conclusions and outlook

In this paper we presented optimal control problems subject to the Poisson equation or the heat equation in a distributed or boundary control setting. The control was added to the objective function as a regularization term in the $\mathcal{H}_1$ norm. We introduced the corre-

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

sponding discrete optimality system and introduced preconditioners for both the steady as well as the transient problem. Due to the Laplacian term coming from the $\mathcal{H}_1$ norm we were not able to introduce preconditioners that are fully independent of the regularization parameter but for the simple preconditioners we introduced the dependence on the regularization parameter seemed rather weak. We also showed that our approach works for state-constrained problems, which were treated using a Moreau-Yosida penalty approach. Numerical results showed that our preconditioners provided satisfactory results when applied to three-dimensional test problems.

The method presented here has not focused on the storage efficiency of our all-at-once approach. One might employ checkpointing [30] techniques when alternately solving forward and adjoint PDEs. Multiple shooting approaches are one way of splitting up the time-interval [33] and can lead to the same type of system. A possible way forward is to compute suboptimal solutions on a sequential splitting of the time-interval [33] or to use a parallel implementation of our approach. It is also possible to reduce the storage requirements by performing block-eliminations of some form, usually via a Schur-complement approach.

References

[1] Bai, Zhong-Zhi: Block preconditioners for elliptic PDE-constrained optimization problems. Computing **91**(4), 379–395 (2011)
[2] Bangerth, W., Hartmann, R., Kanschat, G.: deal.II—a general-purpose object-oriented finite element library. ACM Trans. Math. Software **33**(4), Art. 24, 27 (2007)
[3] Barrett, R., Berry, M., Chan,T. F., Demmel, J., Donato,J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., and der Vorst, H. V.: Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods. 2nd Edition, SIAM, Philadelphia, PA, 1994.
[4] Benzi, M., Golub, G.H., Liesen, J.: Numerical solution of saddle point problems. Acta Numer **14**, 1–137 (2005)
[5] Benzi, M., Haber, E., Taralli, L.: A preconditioning technique for a class of PDE-constrained optimization problems. Advances in Computational Mathematics **35**, 149–173 (2011)
[6] Bergounioux, M., Ito, K., Kunisch, K.: Primal-dual strategy for constrained optimal control problems. SIAM J. Control Optim. **37**(4), 1176–1194 (1999)
[7] Bergounioux, M., Kunisch, K.:Primal-dual strategy for state-constrained optimal control problems, Comput. Optim. Appl. **22**(2), 193–224 (2002) DOI 10.1023/A:1015489608037
[8] Bochev, P., Lehoucq, R.: On the finite element solution of the pure neumann problem. SIAM review **47**(1), 50–66 (2005)
[9] Bramble, J.H., Pasciak, J.E.: A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. Math. Comp **50**(181), 1–17 (1988)
[10] Cai, Xiao-Chuan and Liu, Si and Zou, Jun: An overlapping domain decomposition method for parameter identification problems. Domain Decomposition Methods in Science and Engineering XVII, **60**, 451–458 (2008)
[11] Casas, E.: Control of an elliptic problem with pointwise state constraints. SIAM J. Control Optim. **24**(6), 1309–1318 (1986). DOI 10.1137/0324078. URL `http://dx.doi.org/10.1137/0324078`

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

[12] Casas, E., Herzog, R. and Wachsmuth, G.: Approximation of sparse controls in semilinear equations by piecewise linear functions. Numerische Mathematik **122**, 645–669 (2012)
[13] Chan, R.H. and Chan, T.F. and Wan, WL and others: Multigrid for differential-convolution problems arising from image processing. Proc. Workshop on Scientific Computing, pp. 58–72 (1997)
[14] Chan, T.F. and Tai, X.C.: Identification of discontinuous coefficients in elliptic problems using total variation regularization. SIAM Journal on Scientific Computing **25**(3) 881–904 (2003)
[15] Choi, Y., Farhat, C., Murray, W. and Saunders, M.: A practical factorization of a Schur complement for PDE-constrained Distributed Optimal Control. arXiv preprint arXiv:1312.5653 (2013)
[16] Christofides, P.: Nonlinear and robust control of PDE systems: Methods and applications to transport-reaction processes. Birkhauser (2001)
[17] Cimrák, I. and Melicher, V.: Mixed Tikhonov regularization in Banach spaces based on domain decomposition submitted to Applied Mathematics and Computation (2012)
[18] Collis, S.S., Ghayour, K., Heinkenschloss, M., Ulbrich, M. and Ulbrich, S.: Numerical solution of optimal control problems governed by the compressible Navier-Stokes equations International series of numerical mathematics, 43–56 (2002)
[19] van den Doel, K. and Ascher, U. and Haber, E.: The lost honour of l2-based regularization Submitted (2012)
[20] Du, X., Sarkis, M., Schaerer, C.E., Szyld, D.B.: Inexact and truncated parareal-in-time Krylov subspace methods for parabolic optimal control problems. Tech. Rep. 12-02-06, Department of Mathematics, Temple University (2012)
[21] Duff, I.S., Erisman, A.M., Reid, J.K.: Direct methods for sparse matrices. Monographs on Numerical Analysis. The Clarendon Press Oxford University Press, New York (1989)
[22] Duff, Iain S.: MA57—a code for the solution of sparse symmetric definite and indefinite systems. ACM Transactions on Mathematical Software (TOMS) **30**(2) 118-144 (2004)
[23] Elman, H.C., Silvester, D.J., Wathen, A.J.: Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics. Numerical Mathematics and Scientific Computation. Oxford University Press, New York (2005)
[24] Falgout, R.: An Introduction to Algebraic Multigrid. Computing in Science and Engineering, 8 (2006), pp. 24–33. Special Issue on Multigrid Computing.
[25] Fletcher, R.: Conjugate gradient methods for indefinite systems, *Lecture Notes in Mathematics*, vol. 506. Springer–Verlag, Berlin (1976), 73–89.
[26] Freund, R.W., Nachtigal, N.M.: QMR: a quasi-minimal residual method for non-Hermitian linear systems. Num. Math. **60**(1) 315–339 (1991)
[27] Gee, M., Siefert, C., Hu, J., Tuminaro, R., Sala, M.: ML 5.0 smoothed aggregation user's guide. Tech. Rep. SAND2006-2649, Sandia National Laboratories (2006)
[28] Gunzburger, Max D.: Perspectives in flow control and optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2003)
[29] Greenbaum, A.: Iterative methods for solving linear systems, *Frontiers in Applied Mathematics*, vol. 17. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1997)
[30] Griewank, A., Walther, A.: Algorithm 799: revolve: an implementation of checkpointing for the reverse or adjoint mode of computational differentiation. ACM Transactions on Mathematical Software (TOMS) **26**(1), 19–45 (2000)
[31] Haber, E. and Hanson, L.: Model Problems in PDE-Constrained Optimization Emory University TR-2007-009 (2007)

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

[32] Hackbusch,W.:Multigrid methods and applications. vol. 4 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1985.
[33] Heinkenschloss, M.: A time-domain decomposition iterative method for the solution of distributed linear quadratic optimal control problems. J. Comput. Appl. Math. **173**(1), 169–198 (2005). DOI 10.1016/j.cam.2004.03.005
[34] Heinkenschloss, M.: Formulation and analysis of a sequential quadratic programming method for the optimal Dirichlet boundary control of Navier-Stokes flow Optimal Control, Theory, Algorithms, and Applications (1998)
[35] Heinkenschloss, Matthias, Denis Ridzal: A matrix-free trust-region SQP method for equality constrained optimization. Technical Report 11-17, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 2011.
[36] Herzog, R., Sachs, E.W.: Preconditioned conjugate gradient method for optimal control problems with control and state constraints. SIAM J. Matrix Anal. Appl. **31**(5), 2291–2317 (2010)
[37] Hestenes, M.R., Stiefel, E.: Methods of conjugate gradients for solving linear systems. J. Res. Nat. Bur. Stand **49**, 409–436 (1953) (1952)
[38] Hintermüller, M., Ito, K., Kunisch, K.: The primal-dual active set strategy as a semismooth Newton method. SIAM J. Optim. **13**(3), 865–888 (2002)
[39] Hintermüller, M., Kunisch, K.: Path-following methods for a class of constrained minimization problems in function space. SIAM J. Optim. **17**(1), 159–187 (2006)
[40] Hinze, M., Köster, M., Turek, S.: A Hierarchical Space-Time Solver for Distributed Control of the Stokes Equation. Tech. rep., SPP1253-16-01 (2008)
[41] Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE Constraints. Mathematical Modelling: Theory and Applications. Springer-Verlag, New York (2009)
[42] Hogg, Jonathan D., and Jennifer A. Scott. HSL_MA97: A bit-compatible multifrontal code for sparse symmetric systems. Science and Technology Facilities Council, 2011.
[43] Ito, K., Kunisch, K.: Semi-smooth Newton methods for state-constrained optimal control problems. Systems Control Lett. **50**(3), 221–228 (2003). DOI 10.1016/S0167-6911(03)00156-7
[44] Ito, K., Kunisch, K.: Lagrange multiplier approach to variational problems and applications, *Advances in Design and Control*, vol. 15. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2008)
[45] Kanzow, C.: Inexact semismooth Newton methods for large-scale complementarity problems. Optimization Methods and Software **19**(3-4), 309–325 (2004).
DOI 10.1080/10556780310001636369.
URL `http://www.tandfonline.com/doi/abs/10.1080/10556780310001636369`
[46] Keung, Y.L. and Zou, J.: Numerical identifications of parameters in parabolic systems Inverse Problems **14**(1), 83–100 (1999)
[47] Kollmann, M., Kolmbauer, M.: A Preconditioned MinRes Solver for Time-Periodic Parabolic Optimal Control Problems. Sumitted,Numa-Report 2011-06 (August 2011)
[48] Li, F. and Shen, C. and Li, C.: Multiphase Soft Segmentation with Total Variation and H_1 Regularization Journal of Mathematical Imaging and Vision **37**(2), 98–111 (2010)
[49] Ng, M.K. and Chan, R.H. and Chan, T.F. and Yip, A.M.: Cosine transform preconditioners for high resolution image reconstruction Linear Algebra and its Applications **316**(1) 89–104 (2000)
[50] Murphy, M.F., Golub, G.H., Wathen, A.J.: A note on preconditioning for indefinite linear systems. SIAM J. Sci. Comput **21**(6), 1969–1972 (2000)
[51] Paige, C.C., Saunders, M.A.: Solutions of sparse indefinite systems of linear equations. SIAM

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

J. Numer. Anal **12**(4), 617–629 (1975)
[52] Pearson, J.W., Stoll, M., Wathen, A.: Preconditioners for state constrained optimal control problems with Moreau-Yosida penalty function. Numerical Linear Algebra with Applications **21**(1), 81-97, (2014)
[53] Pearson, J.W., Stoll, M., Wathen, A.J.: Regularization-robust preconditioners for time-dependent PDE-constrained optimization problems. SIAM J. Matrix Anal.Appl **33**, 1126–1152 (2012)
[54] Pearson, J.W., Wathen, A.J.: A new approximation of the Schur complement in preconditioners for PDE-constrained optimization. Numerical Linear Algebra with Applications **19**, 816–829 (2012). DOI 10.1002/nla.814. URL `http://dx.doi.org/10.1002/nla.814`
[55] Peirce, A., Dahleh, M., Rabitz, H.: Optimal control of quantum-mechanical systems: Existence, numerical approximation, and applications. Physical Review A **37**(12), 4950 (1988)
[56] Pironneau, O.: Optimal shape design for elliptic systems. System Modeling and Optimization pp. 42–66 (1982)
[57] Rees, T., Dollar, H.S., Wathen, A.J.: Optimal solvers for PDE-constrained optimization. SIAM Journal on Scientific Computing **32**(1), 271–298 (2010). DOI http://dx.doi.org/10.1137/080727154
[58] Rees, T., Stoll, M., Wathen, A.: All-at-once preconditioners for PDE-constrained optimization. Kybernetika **46**, 341–360 (2010)
[59] De los Reyes, Juan-Carlos and Schönlieb, Carola-Bibiane: Image denoising: learning noise distribution via PDE-constrained optimization, http://arxiv.org/abs/1207.3425 (2012)
[60] Ruge, J. W. and Stüben, K.: Algebraic multigrid. in Multigrid methods, vol. 3 of Frontiers Appl. Math., SIAM, Philadelphia, PA, 1987, pp. 73–130.
[61] Saad, Y.: Iterative methods for sparse linear systems. Society for Industrial and Applied Mathematics, Philadelphia, PA (2003)
[62] Saad, Y.: A flexible inner-outer preconditioned GMRES algorithm. SIAM Journal on Scientific Computing, 14 (1993), pp. 461–461.
[63] Saad, Y., Schultz, M.H.: GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Stat. Comput., **7**(3), 856–869 (1986).
[64] Simoncini, V., Szyld, D.: Recent computational developments in Krylov subspace methods for linear systems. Numer. Linear Algebra Appl **14**(1), 1–61 (2007).
[65] Stoll, M.: All-at-once solution of a time-dependent time-periodic PDE-constrained optimization problems. IMA J Numer Anal (2013)
[66] Stoll, M., Wathen, A.: All-at-once solution of time-dependent PDE-constrained optimization problems. Technical Report, University of Oxford, (2010)
[67] Strang, G., Fix, G.: An Analysis of the Finite Element Method 2nd Edition, 2nd edn. Wellesley-Cambridge (2008)
[68] Takacs, S., Zulehner, W.: Convergence analysis of multigrid methods with collective point smoothers for optimal control problems. Computing and Visualization in Science **14**, 131–141 (2011)
[69] Tröltzsch, F.: Optimal Control of Partial Differential Equations: Theory, Methods and Applications. Amer Mathematical Society (2010)
[70] Ulbrich, M.: Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems. SIAM Philadelphia (2011)
[71] Van Der Vorst, H.A.: BI–CGSTAB: A fast and smoothly converging variant of BI–CG for the solution of nonsymmetric linear systems. SIAM J. Sci. Stat. Comput. **13**(2), 631–634 (1992).
[72] Wathen, A.J., Rees, T.: Chebyshev semi-iteration in preconditioning for problems including

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press

the mass matrix. Electronic Transactions in Numerical Analysis **34**, 125–135 (2008)
[73] Wachsmuth, D. and Wachsmuth, G.: Necessary conditions for convergence rates of regularizations of optimal control problems, RICAM Report **4** (2012)
[74] Wathen, A. J.: Preconditioning and convergence in the right norm. International Journal of Computer Mathematics, 84 (2007), pp. 1199–1209.
[75] Wesseling,P. : An introduction to multigrid methods. Pure and Applied Mathematics (New York), John Wiley & Sons Ltd., Chichester, 1992.
[76] Wilson, J. and Patwari, N. and Vasquez, F.G.: Regularization methods for radio tomographic imaging. 2009 Virginia Tech Symposium on Wireless Personal Communications (2009)
[77] Zulehner, W.: Analysis of iterative methods for saddle point problems: a unified approach. Math. Comp **71**(238), 479–505 (2002)

https://doi.org/10.4208/cicp.190914.080415a Published online by Cambridge University Press