On Block Triangular Preconditioners for the Interior Point Solution of PDE-Constrained Optimization Problems

John W. Pearson and Jacek Gondzio

Abstract We consider the numerical solution of saddle point systems of equations resulting from the discretization of PDE-constrained optimization problems, with additional bound constraints on the state and control variables, using an interior point method. In particular, we derive a Bramble–Pasciak Conjugate Gradient method and a tailored block triangular preconditioner which may be applied within it. Crucial to the usage of the preconditioner are carefully chosen approximations of the $(1,1)$-block and Schur complement of the saddle point system. To apply the inverse of the Schur complement approximation, which is computationally the most expensive part of the preconditioner, one may then utilize methods such as multigrid or domain decomposition to handle individual sub-blocks of the matrix system.

1 Introduction

A key application of domain decomposition methods, alongside a range of other numerical techniques, is within preconditioned iterative methods for linear systems of equations. In this paper, we examine such systems arising from optimization problems constrained by PDEs—in particular we wish to consider the application of interior point methods to formulations with additional bound constraints. The crucial computational element of such solvers is the development of a fast and robust method for the Newton systems that arise at each interior point iteration. We refer to [1, 3, 8, 13], and the references therein, for previous research on such iterative methods, as well as to [5] for the development of a multigrid scheme.

The key component of the authors' previous work [13] was the consideration of saddle point solvers for these linear systems. It was found that iterative methods

John W. Pearson
School of Mathematics, The University of Edinburgh, James Clerk Maxwell Building, The King's Buildings, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, United Kingdom, e-mail: j.pearson@ed.ac.uk

Jacek Gondzio
School of Mathematics, The University of Edinburgh, James Clerk Maxwell Building, The King's Buildings, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, United Kingdom, e-mail: j.gondzio@ed.ac.uk

accelerated by block triangular preconditioners are highly effective for the solution of such systems, often more so than those incorporating analogous block diagonal matrices; however, in general it is difficult to robustly predict the convergence rate of the iterative scheme when using block triangular preconditioners. In this work, we present a new Bramble–Pasciak Conjugate Gradient method which allows one to employ an efficient block triangular approximation, for which the preconditioned system is self-adjoint and positive definite in some non-standard inner product. This also enables one to predict the convergence of the algorithm based on the eigenvalues of the preconditioned system. Such guarantees are not available if one uses more standard Krylov subspace methods for non-symmetric systems, for instance GMRES or BICG. This also provides a framework for domain decomposition techniques, multigrid methods, or other tailored schemes to tackle the individual portions of the block matrix systems at hand. The main contribution of this paper is therefore the presentation of a new solver with the shared advantages of both its faster computational performance, due to the favourable properties of block triangular preconditioners, and the theoretical guarantees of convergence which it provides.

This paper is structured as follows. In Section 2 we describe the PDE-constrained optimization problem of which we wish to consider the numerical solution. In Section 3 we outline the Bramble–Pasciak Conjugate Gradient method, as well as the block triangular preconditioner that we apply within it. In Section 4 we ascertain the effectiveness of our methodology when applied to a number of practical problems.

2 PDE-Constrained Optimization Problem

The problem of which we consider the numerical solution in this paper is given as follows:

$$\begin{aligned} \min_{y,u} \quad & \frac{1}{2}\|y-\widehat{y}\|^2_{L_2(\Omega)} + \frac{\beta}{2}\|u\|^2_{L_2(\Omega)} \\ \text{s.t.} \quad & \mathscr{D}y = u, \quad \text{in } \Omega, \\ & y = f, \quad \text{on } \partial\Omega, \\ & y_a \leq y \leq y_b, \quad \text{a.e. in } \Omega, \\ & u_a \leq u \leq u_b, \quad \text{a.e. in } \Omega. \end{aligned}$$

This problem is solved on a domain $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, with boundary $\partial\Omega$. Here, y, $\widehat{y}$ and u represent the *state*, *desired state* and *control variables*, with $\mathscr{D}$ some given PDE operator. Further, β is a (positive) *regularization parameter*, with f, y_a, y_b, u_a, u_b given functions. The key to this problem is that we wish to find functions y and u which solve the minimization problem constrained by a system of PDEs, while also placing upper and lower bounds on the values that these functions may take.

As illustrated in [13], we may solve this problem using a *discretize-then-optimize* strategy, where a Lagrangian is built on the discrete level and optimality conditions

are subsequently derived from it. The Lagrangian of which we wish to find the stationary point(s), when a finite element method is applied to tackle the barrier optimization problem, is given as follows:

$$\begin{aligned}\mathscr{L}(\mathbf{y},\mathbf{u},\boldsymbol{\lambda}) = \frac{1}{2}\mathbf{y}^T M\mathbf{y} - \mathbf{y}_d^T\mathbf{y} + \frac{\beta}{2}\mathbf{u}^T M\mathbf{u} + \boldsymbol{\lambda}^T(K\mathbf{y} - M\mathbf{u} - \mathbf{f}) \\ - \mu\sum_j \log\left(y_j - y_{a,j}\right) - \mu\sum_j \log\left(y_{b,j} - y_j\right) \\ - \mu\sum_j \log\left(u_j - u_{a,j}\right) - \mu\sum_j \log\left(u_{b,j} - u_j\right),\end{aligned}$$

where $\mathbf{y}$ and $\mathbf{u}$ are the discrete state and control variables, and y_j, $y_{a,j}$, $y_{b,j}$, u_j, $u_{a,j}$, $u_{b,j}$ denote the values of y, y_a, y_b, u, u_a, u_b at the j-th finite element node. The vector $\boldsymbol{\lambda}$ is the discrete *adjoint variable*, enforcing the PDE constraint (which in discretized form is given by $K\mathbf{y} - M\mathbf{u} = \mathbf{f}$). The matrix M is the well known finite element *mass matrix*, with entries defined by $[M]_{ij} = \int_\Omega \phi_i\phi_j \,\mathrm{d}\Omega$, where ϕ_i denote the finite element basis functions used. The matrix K relates to the weak form of the PDE operator $\mathscr{D}$. The vectors $\mathbf{y}_d$ and $\mathbf{f}$ correspond to the functions $\widehat{y}$ and f on the discrete level, and contain entries of the form $\int_\Omega \widehat{y}\phi_i \,\mathrm{d}\Omega$ and $\int_\Omega f\phi_i \,\mathrm{d}\Omega$ respectively. The (positive) *barrier parameter* μ precedes a sum of logarithmic terms which help to enforce the bound constraints on the state and control variables.

The essence of our interior point method is that at each step we wish to find the stationary point of the Lagrangian $\mathscr{L}$, with y_j and u_j updated to take account of the previous iterate, and with μ reduced at each iteration by a factor which is chosen in advance. The algorithm applied is stated in [13]—it is then shown that the main computational bottleneck is the solution of the Newton system

$$\begin{bmatrix} M+D_y & 0 & K^T \\ 0 & \beta M + D_u & -M \\ K & -M & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}\mathbf{y} \\ \boldsymbol{\delta}\mathbf{u} \\ \boldsymbol{\delta}\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mu(Y-Y_a)^{-1}\mathbf{e} - \mu(Y_b-Y)^{-1}\mathbf{e} + \mathbf{y}_d - M\mathbf{y}^* - K^T\boldsymbol{\lambda}^* \\ \mu(U-U_a)^{-1}\mathbf{e} - \mu(U_b-U)^{-1}\mathbf{e} - \beta M\mathbf{u}^* + M\boldsymbol{\lambda}^* \\ \mathbf{f} - K\mathbf{y}^* + M\mathbf{u}^* \end{bmatrix} \tag{1}$$

at each interior point step. The (diagonal) matrices D_y and D_u are given by

$$\begin{aligned} D_y &= (Y-Y_a)^{-1}Z_{y,a} + (Y_b-Y)^{-1}Z_{y,b}, \\ D_u &= (U-U_a)^{-1}Z_{u,a} + (U_b-U)^{-1}Z_{u,b}. \end{aligned}$$

Here, Y, U, Y_a, Y_b, U_a, U_b are diagonal matrices containing the entries of y, u (at the previous Newton step), y_a, y_b, u_a, u_b; further, $Z_{y,a}$, $Z_{y,b}$, $Z_{u,a}$, $Z_{u,b}$ denote diagonal matrices with entries defined by Lagrange multipliers associated with bounds y_a, y_b, u_a, u_b, respectively. At each iteration, an interior point algorithm attempts to approximately satisfy the following centrality condition:

$$(Z_{y,a})_{jj} = \frac{\mu}{y_j - y_{a,j}}, \quad (Z_{y,b})_{jj} = \frac{\mu}{y_{b,j} - y_j},$$
$$(Z_{u,a})_{jj} = \frac{\mu}{u_j - u_{a,j}}, \quad (Z_{u,b})_{jj} = \frac{\mu}{u_{b,j} - u_j}.$$

The vector $\mathbf{e}$ contains a one at each entry, and the vectors $\mathbf{y}^*$, $\mathbf{u}^*$, $\boldsymbol{\lambda}^*$ contain the previous iterates for $\mathbf{y}$, $\mathbf{u}$, $\boldsymbol{\lambda}$. We wish to solve the matrix system (1) for $\boldsymbol{\delta}\mathbf{y}$, $\boldsymbol{\delta}\mathbf{u}$, $\boldsymbol{\delta}\boldsymbol{\lambda}$, the Newton updates of $\mathbf{y}$, $\mathbf{u}$, $\boldsymbol{\lambda}$, at each interior point iteration.

3 Bramble–Pasciak Conjugate Gradients and Preconditioning

We now wish to approach the main computational challenge within the interior point algorithm, namely the fast and efficient solution of the matrix system (1). This is an example of a *saddle point system*, which is defined in general as a system of equations of the form

$$\underbrace{\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}}_{\mathscr{A}} \underbrace{\begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \mathbf{b}^{(1)} \\ \mathbf{b}^{(2)} \end{bmatrix}}_{\mathbf{b}}.$$

There has been a great deal of research on the subject of the numerical solution of such systems, and we refer to [2] for a comprehensive survey. However, in the setting of interior point methods, we face the additional challenge that the $(1,1)$-block A is severely ill-conditioned, due to the presence of diagonal scaling matrices (defined as D_y and D_u in Section 2 for our problem).

In [13], a block diagonal preconditioner was presented, involving approximations $\widehat{A}$ and $\widehat{S}$ for the $(1,1)$-block and the (negative) Schur complement $S := BA^{-1}B^T$, respectively. These approximations were carefully chosen such that the preconditioned system $\mathscr{P}^{-1}\mathscr{A}$ had clustered eigenvalues, and also such that $\widehat{A}^{-1}$ and $\widehat{S}^{-1}$ could be applied cheaply. In this work, we wish to apply a suitable block triangular preconditioner

$$\mathscr{P} = \begin{bmatrix} \widehat{A} & 0 \\ B & -\widehat{S} \end{bmatrix}$$

within a non-standard Conjugate Gradient method. By doing so, we are able to exploit the often superior convergence properties of block triangular preconditioners, alongside the theoretical guarantees of convergence that Conjugate Gradient type methods provide. In particular, we may predict a certain rate of convergence of the iterative method by examining the eigenvalues of the preconditioned system.

The idea of the Bramble–Pasciak Conjugate Gradient method [4] is that we apply this method using an inner product within which the preconditioned system is self-adjoint and positive definite. A suitable inner product is given by $\langle \cdot, \cdot \rangle_{\mathscr{H}}$, with

$$\mathscr{H} = \begin{bmatrix} A - \widehat{A} & 0 \\ 0 & \widehat{S} \end{bmatrix}.$$

The structure of the algorithm is presented below, and we refer to [4, 17, 18] for further details.

Algorithm: Bramble–Pasciak Method for $\mathscr{A}\mathbf{x} = \mathbf{b}$ with Preconditioner $\mathscr{P}$

Initial vectors

Given $\mathbf{x}_0$, set $\mathbf{r}_0 = \mathscr{P}^{-1}(\mathbf{b} - \mathscr{A}\mathbf{x}_0)$, $\mathbf{p}_0 = \mathbf{r}_0$

Conjugate Gradient loop

for $k = 0, 1, \ldots$

$$\alpha_k = \frac{\langle \mathbf{r}_k, \mathbf{r}_k \rangle_{\mathscr{H}}}{\langle \mathscr{P}^{-1}\mathscr{A}\mathbf{p}_k, \mathbf{p}_k \rangle_{\mathscr{H}}}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$$

$$\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathscr{P}^{-1}\mathscr{A}\mathbf{p}_k$$

$$\beta_k = \frac{\langle \mathbf{r}_{k+1}, \mathbf{r}_{k+1} \rangle_{\mathscr{H}}}{\langle \mathbf{r}_k, \mathbf{r}_k \rangle_{\mathscr{H}}}$$

$$\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_k \mathbf{p}_k$$

end

The key components within the algorithm involve computing terms of the form $\mathscr{P}^{-1}\mathbf{v}$ and $\mathscr{H}\mathscr{P}^{-1}\mathbf{v}$, where we write $\mathbf{v} = \left[\mathbf{v}_1^T,\ \mathbf{v}_2^T\right]^T$. The first of these tasks may be accomplished by applying $\widehat{A}^{-1}$ and $\widehat{S}^{-1}$ efficiently, whenever the inverse of the preconditioner is required. For the application of $\mathscr{H}\mathscr{P}^{-1}\mathbf{v}$, which is needed to compute terms of the form $\langle \mathscr{P}^{-1}\mathscr{A}\mathbf{p}_k, \mathbf{p}_k \rangle_{\mathscr{H}}$ and $\langle \mathbf{r}_k, \mathbf{r}_k \rangle_{\mathscr{H}}$ within the Bramble–Pasciak algorithm, we observe that

$$\mathscr{H}\mathscr{P}^{-1}\mathbf{v} = \begin{bmatrix} A - \widehat{A} & 0 \\ 0 & \widehat{S} \end{bmatrix} \begin{bmatrix} \widehat{A}^{-1}\mathbf{v}_1 \\ \widehat{S}^{-1}B\widehat{A}^{-1}\mathbf{v}_1 - \widehat{S}^{-1}\mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} A\widehat{A}^{-1}\mathbf{v}_1 - \mathbf{v}_1 \\ B\widehat{A}^{-1}\mathbf{v}_1 - \mathbf{v}_2 \end{bmatrix}.$$

Therefore, we are only required to apply $\widehat{A}^{-1}$ once in order to compute this term.

We therefore require efficient approximations for the $(1,1)$-block and Schur complement of the matrix system (1) under consideration. For this matrix,

$$A = \begin{bmatrix} M + D_y & 0 \\ 0 & \beta M + D_u \end{bmatrix}, \qquad B = \begin{bmatrix} K & -M \end{bmatrix},$$
$$S = BA^{-1}B^T = K(M + D_y)^{-1}K^T + M(\beta M + D_u)^{-1}M.$$

To approximate the $(1,1)$-block, we apply a Chebyshev semi-iteration method [6, 7] to the diagonally dominant matrices $M + D_y$ and $\beta M + D_u$. As it is necessary to ensure that $A - \widehat{A}$ is positive definite, in turn to guarantee that the inner product

matrix $\mathscr{H}$ is positive definite, we pre-multiply this approximation by a constant $0 \ll \gamma < 1$, which is chosen a priori such that this property holds (see [17]).

In order to approximate the Schur complement, we employ a 'matching strategy', which was derived in [14, 15, 16], and was demonstrated to be highly effective in the context of interior point methods in [13]. We write

$$\widehat{S} = \big(K + \widehat{M}\big)(M + D_y)^{-1}\big(K + \widehat{M}\big)^T,$$

where $\widehat{M} = M\big[\mathrm{diag}(\beta M + D_u)\big]^{-1/2}\big[\mathrm{diag}(M + D_y)\big]^{1/2}$, with the aim of capturing both terms of the exact Schur complement S within our approximation. The inverses of $K + \widehat{M}$ and its transpose may be efficiently approximated using multigrid, domain decomposition, or other methods.

Making use of our approximations of A and S, we may then compile our preconditioner

$$\mathscr{P} = \begin{bmatrix} \gamma(M + D_y)_{\mathrm{Cheb}} & 0 & 0 \\ 0 & \gamma(\beta M + D_u)_{\mathrm{Cheb}} & 0 \\ K & -M & -\widehat{S} \end{bmatrix},$$

which may be readily inverted, giving rise to a computationally efficient algorithm within the inner product $\langle \cdot, \cdot \rangle_{\mathscr{H}}$. Eigenvalue estimates for $\widehat{A}^{-1}A$ and $\widehat{S}^{-1}S$ are discussed in detail in [13]; applying these estimates within the Bramble–Pasciak method leads to robust estimates of convergence rates for the iterative solver, using previous research on this method for PDE-constrained optimization problems without additional bound constraints [17].

4 Numerical Experiments

To test the practical effectiveness of our method we implement an interior point scheme, within which we apply the Bramble–Pasciak Conjugate Gradient method with the preconditioner stated in Section 3. For each problem, we discretize the state, control and adjoint variables using $Q1$ finite elements. The Bramble–Pasciak method is run to a tolerance of 10^{-8} at each interior point step, with the outer (interior point) solver run to a tolerance of 10^{-6}. We measure the average number of Bramble–Pasciak iterations required per outer iteration, until convergence of the interior point method is achieved. The $(1,1)$-block of the matrix system (1) is approximated using 20 steps of Chebyshev semi-iteration, with parameter $\gamma = 0.95$ chosen to ensure positive definiteness of $\mathscr{H}$; the matrices $K + \widehat{M}$ and its transpose, within the Schur complement approximation, are approximately inverted using the Aggregation-based Algebraic Multigrid (AGMG) software [9, 10, 11, 12]. All tests are carried out using MATLAB R2017b, on a quad-core 3.2 GHz processor.

For our first test problem, we consider the Poisson operator $\mathscr{D} = -\nabla^2$, take $\widehat{y} = \sin(\pi x_1)\sin(\pi x_2)$, where $\mathbf{x} = [x_1, x_2]^T \in \Omega = [0,1]^2$, and set $y = 0$ on the bound-

Table 1 Results for the Poisson control example with state constraints, for a range of values of h and β. Presented are the average number of Bramble–Pasciak Conjugate Gradient iterations required, per interior point step.

		$\beta = 1$ $0 \le y \le 0.002$	$\beta = 10^{-1}$ $0 \le y \le 0.02$	$\beta = 10^{-2}$ $0 \le y \le 0.15$	$\beta = 10^{-3}$ $0 \le y \le 0.5$	$\beta = 10^{-4}$ $0 \le y \le 0.8$	$\beta = 10^{-5}$ $0 \le y \le 0.9$
h	2^{-2}	8.5	8.4	7.7	7.4	7.9	8.1
	2^{-3}	12.4	12.6	11.3	13.1	14.0	18.3
	2^{-4}	14.6	14.5	14.2	16.2	18.1	19.9
	2^{-5}	15.8	15.9	16.2	18.3	20.3	22.7
	2^{-6}	16.6	17.1	17.4	20.7	30.0	25.9
	2^{-7}	17.3	17.8	18.5	30.2	26.2	27.8

Table 2 Results for the convection–diffusion control example with control constraints, for a range of values of h and β. Presented are the average number of Bramble–Pasciak Conjugate Gradient iterations required, per interior point step.

		$\beta = 1$ $0 \le u \le 0.1$	$\beta = 10^{-1}$ $0 \le u \le 0.5$	$\beta = 10^{-2}$ $0 \le u \le 2$	$\beta = 10^{-3}$ $0 \le u \le 5$	$\beta = 10^{-4}$ $0 \le u \le 6$	$\beta = 10^{-5}$ $0 \le u \le 6$
h	2^{-2}	8.3	9.8	11.8	14.3	15.4	16.0
	2^{-3}	8.4	10.9	14.8	16.9	20.6	24.4
	2^{-4}	8.2	10.4	13.6	26.6	33.8	35.2
	2^{-5}	8.1	10.1	12.4	16.9	29.9	33.5
	2^{-6}	8.1	9.9	12.2	15.3	25.9	24.6
	2^{-7}	8.3	9.9	12.1	15.3	18.3	18.9

ary $\partial\Omega$ of Ω. We prescribe bound constraints on the state variable y, based on the physical properties of the problem. We solve the matrix systems for a range of step-sizes h and values of β, and present the results obtained in Table 1. We observe very low iteration numbers, considering the complexity of the problem and the accuracy to which we solve the matrix systems, with only moderate increases as h is decreased (i.e. as the dimensions of the matrix systems are increased). We also observe a benign increase in Bramble–Pasciak iterations as β is decreased.

Our second test problem involves a convection–diffusion operator $\mathscr{D} = -0.01\nabla^2 + \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^T \cdot \nabla$, a desired state $\widehat{y} = e^{-64\left((x_1-0.5)^2+(x_2-0.5)^2\right)}$, and the boundary condition $y = \widehat{y}$. On this occasion we provide bound constraints for the control variable u, as stated in Table 2. Once again, strong robustness of the Bramble–Pasciak method is observed when either h or β is altered, illustrating that our strategy may be applied to more varied differential operators and types of bound constraints.

We thus establish that the new Bramble–Pasciak Conjugate Gradient algorithm presented for this class of problems provides both enjoyable theoretical properties, and the fast, robust numerical solution of a range of practical examples. It may be concluded that this is therefore a suitable and effective technique for the interior point solution of a number of PDE-constrained optimization problems.

Acknowledgements JWP gratefully acknowledges financial support from the Engineering and Physical Sciences Research Council (EPSRC) Fellowship EP/M018857/2. JG gratefully acknowledges support from the EPSRC Grant EP/N019652/1. The authors thank an anonymous referee for their helpful comments.

References

1. A. Battermann and M. Heinkenschloss, *Preconditioners for Karush–Kuhn–Tucker matrices arising in the optimal control of distributed systems*. In: W. Desch, F. Kappel, and K. Kunisch (eds), *Control and Estimation of Distributed Parameter Systems*, pp.15–32, 1998.
2. M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numerica, 14, pp.1–137, 2005.
3. M. Benzi, E. Haber, and L. Taralli, *Multilevel algorithms for large-scale interior point methods*, SIAM Journal on Scientific Computing, 31, pp.4152–4175, 2009.
4. J. H. Bramble and J. E. Pasciak, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Mathematics of Computation, 50(181), pp.1–17, 1988.
5. A. Drăgănescu and C. Petra, *Multigrid preconditioning of linear systems for interior point methods applied to a class of box-constrained optimal control problems*, SIAM Journal on Numerical Analysis, 50(1), pp.328–353, 2012.
6. 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, pp.147–156, 1961.
7. G. H. Golub and R. S. Varga, *Chebyshev semi-iterative methods, successive over-relaxation iterative methods, and second order Richardson iterative methods. II*, Numerische Mathematik, 3, pp.157–168, 1961.
8. M. J. Grotte, J. Huber, D. Kourounis, and O. Schenk, *Inexact interior-point method for PDE-constrained nonlinear optimization*, SIAM Journal on Scientific Computing, 36, pp.A1251–A1276, 2014.
9. A. Napov and Y. Notay, *An algebraic multigrid method with guaranteed convergence rate*, SIAM Journal on Scientific Computing, 34(2), pp.A1079–A1109, 2012.
10. Y. Notay, *Aggregation-based algebraic multigrid for convection–diffusion equations*, SIAM Journal on Scientific Computing, 34(4), pp.A2288–A2316, 2012.
11. Y. Notay, *AGMG software and documentation*; see `http://homepages.ulb.ac.be/~ynotay/AGMG`.
12. Y. Notay, *An aggregation-based algebraic multigrid method*, Electronic Transactions on Numerical Analysis, 37, pp.123–146, 2010.
13. J. W. Pearson and J. Gondzio, *Fast interior point solution of quadratic programming problems arising from PDE-constrained optimization*, Numerische Mathematik, 137(4), pp.959–999, 2017.
14. J. W. Pearson, M. Stoll, and A. J. Wathen, *Regularization-robust preconditioners for time-dependent PDE-constrained optimization problems*, SIAM Journal on Matrix Analysis and Applications, 33(4), pp.1126–1152, 2012.
15. J. W. Pearson and A. J. Wathen, *A new approximation of the Schur complement in preconditioners for PDE-constrained optimization*, Numerical Linear Algebra with Applications, 19, pp.816–829, 2012.
16. J. W. Pearson and A. J. Wathen, *Fast iterative solvers for convection–diffusion control problems*, Electronic Transactions on Numerical Analysis, 40, pp.294–310, 2013.
17. T. Rees and M. Stoll, *Block triangular preconditioners for PDE-constrained optimization*, Numerical Linear Algebra with Applications, 17(6), pp.977–996, 2010.
18. M. Stoll, *Solving Linear Systems Using the Adjoint*, DPhil Thesis, University of Oxford, 2008.