Preconditioners for Saddle Point Problems by Interpolation

Markus Kollmann

11th January 2010

Outline

Main Task: Construction of efficient preconditioners for so-called saddle point problems (optimal control problem)

We consider problems, where some critical model parameter α and (due to discretization) a discretization parameter h are involved.

We want to construct preconditioners such that the convergence rate of some preconditioned Krylov subspace method (MINRES) is robust with respect to these parameters.

Optimal Control Problem

We consider the following optimal control problem (distributed elliptic optimal control):

Find the state $y \in V = H_0^1(\Omega)$ and the control $u \in Q = L^2(\Omega)$ that minimizes the cost functional

$$J(y,u) = \frac{1}{2}||y - y_d||_{L^2}^2 + \frac{\alpha}{2}||u||_{L^2}^2$$

subject to the state equation

$$\begin{aligned} -\Delta y &= u \quad \text{in} \quad \Omega \\ y &= 0 \quad \text{on} \quad \Gamma \end{aligned}$$

or, more precisely, subject to the state equation in its weak form, given by

$$(\nabla y, \nabla z)_{L^2} = (u, z)_{L^2} \quad \forall z \in V.$$

The Lagrangian functional associated to this problem is given by:

$$\mathcal{L}(y,u,p) = J(y,u) + (\nabla y, \nabla p)_{L^2} - (u,p)_{L^2},$$

which leads to the following optimality system

$$\begin{array}{rcl} (y,z)_{L^2} + (\nabla z, \nabla p)_{L^2} & = & (y_d, z)_{L^2} \quad \forall z \in V, \\ \alpha(u,v)_{L^2} - (v,p)_{L^2} & = & 0 \quad \forall v \in Q, \\ (\nabla y, \nabla q)_{L^2} - (u,q)_{L^2} & = & 0 \quad \forall q \in V, \end{array}$$

which characterizes the solution $(y,u) \in V \times Q$ of the optimal control problem with Lagrangian multiplier (co-state) $p \in V$.

From the second equation we have that $u = \alpha^{-1} p$, therefore we get the reduced optimality system

$$\begin{aligned} (y,z)_{L^2} + (\nabla z, \nabla p)_{L^2} &= (y_d, z)_{L^2} \quad \forall z \in V, \\ (\nabla y, \nabla q)_{L^2} - \alpha^{-1}(p,q)_{L^2} &= 0 \quad \forall q \in Q. \end{aligned} \tag{1}$$

Discretization

Discretization of (1) with the finite element method on a simplicial subdivision of Ω with continuous and piecewise linear functions for the state and co-state leads to the linear system

$$\mathcal{A}\begin{pmatrix} \underline{y} \\ \underline{p} \end{pmatrix} = \begin{pmatrix} M\underline{y_d} \\ 0 \end{pmatrix} \quad \text{with} \quad \mathcal{A} = \begin{pmatrix} M & K \\ K & -\alpha^{-1}M \end{pmatrix} \tag{2}$$

- M...mass matrix representing the L^2 inner product
- K...stiffness matrix representing the elliptic operator of the state equation

The system matrix in (2) is indefinite.
proof:
M and K are symmetric and positive definite $\Longrightarrow S = \alpha^{-1}M + KM^{-1}K$ is symmetric and positive definite.
We have:

$$\begin{pmatrix} M & K \\ K & -\alpha^{-1}M \end{pmatrix} = \begin{pmatrix} I & 0 \\ KM^{-1} & I \end{pmatrix}\begin{pmatrix} M & 0 \\ 0 & -S \end{pmatrix}\begin{pmatrix} I & M^{-1}K \\ 0 & I \end{pmatrix}.$$

Since $\begin{pmatrix} M & 0 \\ 0 & -S \end{pmatrix}$ is indefinite, the statement follows from Sylvester's law of inertia. $\square$.

Now:

Construct block diagonal preconditioner $\mathcal{P} = \begin{pmatrix} \mathcal{P}_V & 0 \\ 0 & \mathcal{P}_Q \end{pmatrix}$ such that

- Estimation of condition number of $\mathcal{P}^{-1}\mathcal{A}$ independent of α and h
- $\Longrightarrow$ robust convergence rates for Krylov subspace method

Introduction

For the solution of a linear system

$$Kx = b$$

Krylov subspace methods produce a chain of vectors

$$x_0 \longrightarrow x_1 \longrightarrow \ldots \longrightarrow x_m$$

which converges to the exact solution x as m increases. In some sense, the Krylov subspace methods can be seen as improvements of a simple fixed point iteration of the form

$$x_{m+1} = x_m + \alpha r_m$$

where r_m is the residual in the m-th step

$$r_m = b - Kx_m$$

but more robust and more efficient.

The iterates x_m satisfy

$$x_m \in x_0 + \mathcal{K}_m(K, r_0)$$

where

$$\mathcal{K}_m(K, r_0) = span\{r_0, Kr_0, ..., K^{m-1} r_0\}$$

is the Krylov subspace of order m, x_0 is some initial guess and r_0 is the initial residual.
For each m, the Krylov subspace $\mathcal{K}_m(K, r_0)$ has dimension m. Because of the m degrees of freedom in the choice of the iterate x_m, m constraints are required to make x_m unique.

Minimal Residual Method

MINRES

The Minimal Residual Method is a Krylov subspace method for the solution of a symmetric and indefinite matrix system

$$Kx = b.$$

In this method, $x_m \in x_0 + \mathcal{K}_m(K, r_0)$ is characterized by

$$x_m = \underset{y \in x_0 + \mathcal{K}_m}{\arg\min} \, ||b - Ky||^2$$

i.e., x_m minimizes the residual (three-term recurrence relation).

In our purposes:

$$P^{-1}Kx = P^{-1}b$$

with P symmetric and positive definite.

For the preconditioned MINRES method we have the following convergence result:

Theorem (Greenbaum)

$$\|r_{2m}\|_{P^{-1}} \leq \frac{2q^m}{1+q^{2m}} \|r_0\|_{P^{-1}}$$

where

$$r_m = b - Kx_m \quad \textit{and} \quad q = \frac{\kappa(P^{-1}K)-1}{\kappa(P^{-1}K)+1},$$

κ denotes the condition number:

$$\kappa(P^{-1}K) = ||P^{-1}K||_P||(P^{-1}K)^{-1}||_P = ||P^{-1/2}KP^{-1/2}||_2||(P^{-1/2}KP^{-1/2})^{-1}||_2.$$

Note that $P^{-1}K$ and $P^{-1/2}KP^{-1/2}$ have the same spectrum.

The General Case

In general, saddle point problems are of the following form:

$$\begin{pmatrix} A & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \tag{3}$$

with A and C being symmetric and positive semidefinite. In the following, we will present some examples where preconditioners, which lead to robust convergence rates, are known.

Case 1

- $C = 0$
- A non-singular $\Longrightarrow$ A positive definite
- Schur complement $S = BA^{-1}B^T$ non-singular $\Longrightarrow$ B has full rank

Then problem (3) reads

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}$$

with indefinite system matrix.

Case 1 contd. I

Now we have the following theorem

Theorem

If

$$\mathcal{A} = \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$$

is preconditioned by

$$\mathcal{P} = \begin{pmatrix} A & 0 \\ 0 & S \end{pmatrix}$$

then the preconditioned matrix $\mathcal{T} = \mathcal{P}^{-1}\mathcal{A}$ *has exactly three distinct eigenvalues, namely* 1 *and* $\left(1 \pm \sqrt{5}\right)/2$*.*

proof:

$$T = \begin{pmatrix} I & A^{-1}B^T \\ S^{-1}B & 0 \end{pmatrix};$$

Case 1 contd. II

Eigenvalue problem:

$$\begin{pmatrix} I & A^{-1}B^T \\ S^{-1}B & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$$

$$\Longleftrightarrow \qquad \begin{aligned} &1) \quad x + A^{-1}B^T y = \lambda x \\ &2) \quad S^{-1}Bx = \lambda y \end{aligned}$$

From 1): $(\lambda - 1)x = A^{-1}B^T y$

- Case $\lambda = 1$: Since B^T has full rank $\Longrightarrow y = 0$. Plug into 2) $\Longrightarrow$ $x \in \mathrm{Ker} B \Longrightarrow 1$ is eigenvalue.
- Case $\lambda \neq 1$: $x = \frac{1}{\lambda - 1} A^{-1}B^T y$. Plug into 2) $\Longrightarrow (\lambda^2 - \lambda - 1)y = 0$. Therefore the statement follows. $\square$.

Case 1 contd.

Corollary

If

$$\mathcal{A} = \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$$

is preconditioned by

$$\mathcal{P} = \begin{pmatrix} A & 0 \\ 0 & S \end{pmatrix}$$

then the preconditioned matrix $\mathcal{T} = \mathcal{P}^{-1}\mathcal{A}$ *satisfies*

$$(\mathcal{T} - I)\left(\mathcal{T} - \frac{1}{2}\left(1 + \sqrt{5}\right) I\right)\left(\mathcal{T} - \frac{1}{2}\left(1 - \sqrt{5}\right) I\right) = 0.$$

Case 1 contd.

Remark

It directly follows from the Corollary that for any vector r the Krylov subspace

$$\mathcal{K}_m(\mathcal{T}, r) = span\{r, \mathcal{T}r, ..., \mathcal{T}^{m-1}r\}$$

is of dimension at most 3, therefore MINRES will terminate in at most 3 iterations with the solution.

Remark

Forming the preconditioned matrix $\mathcal{T}$ is essentially as expensive as computing the inverse of $\mathcal{A}$ directly. In practice, the exact preconditioner $\mathcal{P}$ needs to be replaced by an approximation,

$$\hat{\mathcal{P}} = \begin{pmatrix} \hat{A} & 0 \\ 0 & \hat{S} \end{pmatrix}$$

where both $\hat{A}$ and $\hat{S}$ are approximations of A and S, respectively.

Case 1 contd. I

Remark

For the condition number of $\mathcal{T}$ we therefore obtain

$$\kappa(\mathcal{T}) = \frac{|\lambda_{max}|}{|\lambda_{min}|} = \frac{\left(1+\sqrt{5}\right)/2}{\left(\sqrt{5}-1\right)/2} \approx 2.62.$$

It follows that:

$$\underline{c}||z||_{\mathcal{P}} \leq ||\mathcal{A}z||_{\mathcal{P}^{-1}} \leq \overline{c}||z||_{\mathcal{P}}$$

where

$$\underline{c} = \frac{\sqrt{5}-1}{2} \quad \text{and} \quad \overline{c} = \frac{1+\sqrt{5}}{2}$$

Case 1 contd. II

proof:

$$\begin{aligned}
\frac{||\mathcal{A}z||_{\mathcal{P}^{-1}}}{||z||_{\mathcal{P}}} &\leq \sup_{z\neq 0} \frac{||\mathcal{A}z||_{\mathcal{P}^{-1}}}{||z||_{\mathcal{P}}} \\
&= \sup_{z\neq 0} \frac{\sqrt{\langle \mathcal{P}^{-1}\mathcal{A}z, \mathcal{A}z\rangle}}{\sqrt{\langle \mathcal{P}z, z\rangle}} \\
&= \sup_{x\neq 0} \frac{\sqrt{\langle \mathcal{P}^{-1/2}\mathcal{A}\mathcal{P}^{-1/2}x, \mathcal{P}^{-1/2}\mathcal{A}\mathcal{P}^{-1/2}x\rangle}}{\sqrt{\langle x, x\rangle}} \\
&= ||\mathcal{P}^{-1/2}\mathcal{A}\mathcal{P}^{-1/2}||_2.
\end{aligned}$$

The lower estimate can be shown analoguously $\square$.

Case 2a

- $C \neq 0$
- A non-singular
- Schur complement $S = C + BA^{-1}B^T$ non-singular

Then problem (3) reads

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

with indefinite system matrix.

Case 2a contd. I

Theorem

If

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

is preconditioned by

$$\mathcal{P} = \begin{pmatrix} A & 0 \\ 0 & S \end{pmatrix}$$

then the eigenvalues of the preconditioned matrix $\mathcal{T} = \mathcal{P}^{-1}\mathcal{A}$ *are in* $\left(-1, \frac{1}{2}(1-\sqrt{5})\right] \cup \{1\} \cup \left(1, \frac{1}{2}(1+\sqrt{5})\right]$.

proof:

$$T = \begin{pmatrix} I & A^{-1}B^T \\ S^{-1}B & -S^{-1}C \end{pmatrix};$$

Case 2a contd. II

Eigenvalue problem:

$$\begin{aligned} 1) \quad & x + A^{-1}B^T y = \lambda x \\ 2) \quad & S^{-1}Bx - S^{-1}Cy = \lambda y \end{aligned}$$

From 1): $(\lambda - 1)x = A^{-1}B^T y$

- Case $\lambda = 1$: Since B^T has full rank $\Longrightarrow y = 0$. Plug into 2) $\Longrightarrow$ $x \in \mathrm{Ker} B \Longrightarrow 1$ is eigenvalue.
- Case $\lambda \neq 1$: $x = \frac{1}{\lambda - 1} A^{-1}B^T y$. Plug into 2):

$$\begin{aligned} & \frac{1}{\lambda - 1} S^{-1}BA^{-1}B^T y - S^{-1}Cy = \lambda y \\ \Longleftrightarrow \quad & -\lambda Cy = (\lambda^2 - \lambda - 1)Sy. \end{aligned}$$

Multiplication by y^T from the left and the fact that S is positive definite and C is positive semidefinite yields:

$$\frac{-\lambda^2 + \lambda + 1}{\lambda} = \rho$$

Case 2a contd. III

where $\rho \in [0, 1)$. Therefore

$$\lambda = -\frac{1}{2}\left((\rho - 1) \pm \sqrt{(1-\rho)^2 + 4}\right) \tag{4}$$

Due to ρ and the fact that (4) is monotonically decreasing in ρ, the statement follows $\square$.

Case 2a contd.

Now we have

$$\underline{c}||z||_{\mathcal{P}} \leq ||\mathcal{A}z||_{\mathcal{P}^{-1}} \leq \overline{c}||z||_{\mathcal{P}}$$

with $\underline{c}$ and $\overline{c}$ independent of parameters.

Case 2b

- $C \neq 0$
- C non-singular
- $R = A + B^T C^{-1} B$ non-singular

With

$$\mathcal{P} = \begin{pmatrix} R & 0 \\ 0 & C \end{pmatrix}$$

we have, analoguously as in Case 2a,

$$\underline{c}||z||_{\mathcal{P}} \leq ||\mathcal{A}z||_{\mathcal{P}^{-1}} \leq \overline{c}||z||_{\mathcal{P}}$$

with $\underline{c}$ and $\overline{c}$ independent of parameters.

Back to our problem

$$\mathcal{A}\begin{pmatrix} \underline{y} \\ \underline{p} \end{pmatrix} = \begin{pmatrix} \underline{y}_d \\ 0 \end{pmatrix} \quad \text{with} \quad \mathcal{A} = \begin{pmatrix} M & K \\ K & -\alpha^{-1}M \end{pmatrix}$$

This fits into the general Case 2 with $A = M$, $B = B^T = K$ and $C = \alpha^{-1}M$. Therefore we have the following two preconditioners

$$\begin{aligned} \mathcal{P}_0 &= \begin{pmatrix} M & 0 \\ 0 & S \end{pmatrix} \\ \mathcal{P}_1 &= \begin{pmatrix} R & 0 \\ 0 & \alpha^{-1}M \end{pmatrix} \end{aligned}$$

where

$$S = \alpha^{-1}M + KM^{-1}K \quad \text{and} \quad R = M + \alpha KM^{-1}K$$

which lead to robust estimates

$$\underline{c}||z||_{\mathcal{P}_0} \leq ||\mathcal{A}z||_{\mathcal{P}_0^{-1}} \leq \overline{c}||z||_{\mathcal{P}_0}$$
$$\underline{c}||z||_{\mathcal{P}_1} \leq ||\mathcal{A}z||_{\mathcal{P}_1^{-1}} \leq \overline{c}||z||_{\mathcal{P}_1}.$$

$S = \alpha^{-1}M + KM^{-1}K$ and $R = M + \alpha KM^{-1}K$:

matrices which result from the discretization of a differential operator of fourth order

$\Longrightarrow$ Possible but hard to work with

$\Longrightarrow$ Interpolation of the two preconditioners

Interpolation Theorem

Theorem

Let $\mathcal{A} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *with*

$$||\mathcal{A}z||_{Y_0} \leq c_0 ||z||_{X_0} \quad \text{and} \quad ||\mathcal{A}z||_{Y_1} \leq c_1 ||z||_{X_1},$$

where the norms $||.||_{X_i}$ *and* $||.||_{Y_i}$ *are the norms associated to the inner products:*

$$(x,y)_{X_i} = \langle M_i x, y\rangle \quad \text{and} \quad (x,y)_{Y_i} = \langle N_i x, y\rangle$$

Then, for $X_\Theta = [X_0, X_1]_\Theta$ *and* $Y_\Theta = [Y_0, Y_1]_\Theta$, *we have*

$$||\mathcal{A}z||_{Y_\Theta} \leq c_0^{1-\Theta} c_1^{\Theta} ||z||_{X_\Theta}.$$

$||.||_{X_\Theta}$, $||.||_{Y_\Theta}$ *are the norms associated to the inner products*

$$(x,y)_{X_\Theta} = \langle M_\Theta x, y\rangle \quad \text{with} \quad M_\Theta = M_0^{1/2}\left(M_0^{-1/2} M_1 M_0^{-1/2}\right)^{\Theta} M_0^{1/2}$$

$$(x,y)_{Y_\Theta} = \langle N_\Theta x, y\rangle \quad \text{with} \quad N_\Theta = N_0^{1/2}\left(N_0^{-1/2} N_1 N_0^{-1/2}\right)^{\Theta} N_0^{1/2}$$

In our problem:

$M_0 = \mathcal{P}_0$, $M_1 = \mathcal{P}_1$, $N_0 = \mathcal{P}_0^{-1}$ and $N_1 = \mathcal{P}_1^{-1}$

We have:

$$||\mathcal{A}z||_{\mathcal{P}_0^{-1}} \leq \overline{c}||z||_{\mathcal{P}_0} \quad \text{and} \quad ||\mathcal{A}z||_{\mathcal{P}_1^{-1}} \leq \overline{c}||z||_{\mathcal{P}_1}$$

$$||\mathcal{A}^{-1}y||_{\mathcal{P}_0} \leq \underline{c}||y||_{\mathcal{P}_0^{-1}} \quad \text{and} \quad ||\mathcal{A}^{-1}y||_{\mathcal{P}_1} \leq \underline{c}||y||_{\mathcal{P}_1^{-1}}.$$

Interpolation with $\Theta = \frac{1}{2}$ gives:

$$\mathcal{P}_{1/2} = \begin{pmatrix} M + \sqrt{\alpha} K & 0 \\ 0 & \alpha^{-1} M + \alpha^{-1/2} K \end{pmatrix}$$

proof:
Since we have block diagonal matrices:

$$\mathcal{P}_{1/2} = \begin{pmatrix} [M, M + \alpha K M^{-1} K]_{1/2} & 0 \\ 0 & [\alpha^{-1} M + K M^{-1} K, \alpha^{-1} M]_{1/2} \end{pmatrix}.$$

Observe that

$$\frac{1}{\sqrt{2}}(1 + \sqrt{x}) \leq \sqrt{1 + x} \leq 1 + \sqrt{x} \quad \forall x \geq 0,$$

therefore

$$(I + L)^{1/2} \sim I + L^{1/2} \quad \forall L \geq 0 \quad \text{(matrix function)}.$$

Now we have:

$$\begin{aligned}
[M, M+\alpha K M^{-1} K]_{1/2} &= M^{1/2}\left(M^{-1/2}(M+\alpha K M^{-1} K) M^{-1/2}\right)^{1/2} M^{1/2} \\
&= M^{1/2}\left(I+\alpha M^{-1/2} K M^{-1} K M^{-1/2}\right)^{1/2} M^{1/2} \\
&\sim M + M^{1/2}\sqrt{\alpha}\left(M^{-1/2} K M^{-1} K M^{-1/2}\right)^{1/2} M^{1/2} \\
&= M+\sqrt{\alpha} K,
\end{aligned}$$

and

$$\begin{aligned}
&[\alpha^{-1} M + K M^{-1} K, \alpha^{-1} M]_{1/2} = [\alpha^{-1} M, \alpha^{-1} M + K M^{-1} K]_{1/2} \\
&\quad = \alpha^{-1/2} M^{1/2}\left(\sqrt{\alpha} M^{-1/2}(\alpha^{-1} M + K M^{-1} K)\sqrt{\alpha} M^{-1/2}\right)^{1/2} \alpha^{-1/2} M^{1/2} \\
&\quad = \alpha^{-1/2} M^{1/2}\left(I+\alpha M^{-1/2} K M^{-1} K M^{-1/2}\right)^{1/2} \alpha^{-1/2} M^{1/2} \\
&\quad \sim \alpha^{-1} M + \alpha^{-1/2} K \quad \square.
\end{aligned}$$

$\Longrightarrow$ matrix which results from the discretization of a differential operator of second order

Interpolation Theorem gives:

$$\underline{c}||z||_{\mathcal{P}_{1/2}} \lesssim ||\mathcal{A}z||_{\mathcal{P}_{1/2}^{-1}} \lesssim \overline{c}||z||_{\mathcal{P}_{1/2}}$$

$\Longrightarrow$ robust estimate of the condition number of $\mathcal{P}_{1/2}^{-1}\mathcal{A}$

Conclusion

- Two preconditioners which lead to robust convergence rates
- But hard to realize (fourth order)
- Interpolation of the two preconditioners
- $\Longrightarrow$ Preconditioner which leads to robust convergence rates
- Preconditioner results from the discretization of a differential operator of second order