Lecture Note #12 **MTH652: Advanced Numerical Analysis**
Winter 2025 Instructor: Will Pazner

Preconditioning Saddle-Point Systems

1 MINRES

Consider the saddle-point problem

$$\underbrace{\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}}_{\mathcal{A}} \underbrace{\begin{pmatrix} u \\ p \end{pmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{pmatrix} f \\ g \end{pmatrix}}_{\boldsymbol{b}}.$$

We assume that

1. A is a symmetric, positive-definite $n \times n$ matrix
2. B is a full-rank $m \times n$ matrix ($n \geq m$, B has independent rows)

These two conditions are sufficient to ensure that the matrix $\mathcal{A}$ is invertible. We saw previously that $\mathcal{A}$ has n positive eigenvalues and m negative eigenvalues. Since $\mathcal{A}$ is not SPD, we cannot use conjugate gradient. Instead, we use the "minimal residual method". This method generates the same iterates as GMRES (up to round-off differences), but uses a **short-term recurrence**. This is because the Arnoldi algorithm simplifies to the following *Lanczos* algorithm.

Algorithm 1 Lanczos Algorithm (simplification of Arnoldi for symmetric A)

```
1: q_0 ← 0
2: z ← b                              ▷ First Krylov vector
3: for k = 1, 2, ..., m do
4:     β ← ‖z‖
5:     q_k ← z/β                      ▷ Normalize
6:     z ← A q_k                      ▷ Next vector to orthogonalize
7:     z ← z − β q_{k−1}              ▷ Orthogonalize against q_{k−1}
8:     z ← z − (z^T q_k) q_k          ▷ Orthogonalize against q_k
9: end for
```

The convergence behavior of MINRES for indefinite problems is more complicated than for definite problems. Since at step k the residual $\|b - \mathcal{A}\boldsymbol{x}\|$ is minimized among all $\boldsymbol{x} \in \mathcal{K}_k(\mathcal{A}, \boldsymbol{b})$, it follows that

$$\|b - \mathcal{A}\boldsymbol{x}\| = \min_{p \in \mathcal{P}_k} \|p(\mathcal{A})\boldsymbol{b}\|.$$

In the above, $\mathcal{P}_k$ is the set of all polynomials of degree at most k (taking the value 1 at the origin). The right-hand side includes the quantity $p(\mathcal{A})$, which is the polynomial evaluated using the matrix $\mathcal{A}$ as the variable. Since $\mathcal{A}$ is symmetric (hence diagonalizable), we can write

$$\mathcal{A} = V \Lambda V^{-1}, \qquad \Lambda \text{ diagonal.}$$

From this, we can derive

$$\begin{aligned}
\|b - \mathcal{A}\boldsymbol{x}\| &= \min_{p \in \mathcal{P}_k} \|p(\mathcal{A})\boldsymbol{b}\| \\
&\leq \min_{p \in \mathcal{P}_k} \|p(\mathcal{A})\| \|\boldsymbol{b}\| \\
&= \min_{p \in \mathcal{P}_k} \|p(V\Lambda V^{-1})\| \|\boldsymbol{b}\| \\
&= \min_{p \in \mathcal{P}_k} \|V p(\Lambda) V^{-1}\| \|\boldsymbol{b}\| \\
&\leq \|V\| \|V^{-1}\| \|\boldsymbol{b}\| \min_{p \in \mathcal{P}_k} \|p(\Lambda)\| \\
&= \kappa(V) \|\boldsymbol{b}\| \min_{p \in \mathcal{P}_k} \max_{\lambda_j} \|p(\lambda_j)\| \\
&= \|\boldsymbol{b}\| \min_{p \in \mathcal{P}_k} \max_{\lambda_j} \|p(\lambda_j)\|,
\end{aligned}$$

since $\mathcal{A}$ has orthonormal eigenvectors, so $\|V\| = \|V^{-1}\| = 1$. In the above, λ_j is the jth eigenvalue of $\mathcal{A}$.

We can therefore bound the relative residual $\|\boldsymbol{r}_k\|/\|\boldsymbol{b}\|$ by estimating the min-max problem for polynomials. For indefinite matrices, the min-max problem becomes complicated. Suppose the eigenvalues λ_j lie in the intervals

$$\lambda_j \in [\mu_1, \mu_2] \cup [\nu_1, \nu_2], \qquad \mu_1, \mu_2 < 0, quad \nu_1, \nu_2 > 0.$$

If $\mu_2 - \mu_1 = \nu_2 - \nu_1$, then it is possible to prove the following bound

$$\min_{p \in \mathcal{P}_k} \max_{\lambda_j} \|p(\lambda_j)\| \leq 2 \left(\frac{\sqrt{|\mu_1 \nu_2|} - \sqrt{|\mu_2 \nu_1|}}{\sqrt{|\mu_1 \nu_2|} + \sqrt{|\mu_2 \nu_1}} \right)^{\lceil k/2 \rceil} \tag{1}$$

For example, suppose $\mu_1 = -1, \nu_2 = 1$ and $-\mu_2 = \nu_1 = \kappa^{-1}$. Then, (1) reduces to

$$\min_{p \in \mathcal{P}_k} \max_{\lambda_j} \|p(\lambda_j)\| \leq 2 \left(\frac{\kappa - 1}{\kappa + 1} \right)^{\lceil k/2 \rceil}.$$

Compare this to the bound for CG applied to a positive-definite matrix:

$$\text{(CG bound)} \qquad 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k.$$

This means that the bound for MINRES on a matrix with condition number κ at step $2k$ corresponds to the bound for CG on a matrix with condition number κ^2 at step k; we could expect that MINRES with condition number κ will take twice the number of steps as CG with condition number κ^2. This bound is not sharp, and in practice we may do better. For indefinite matrices where the intervals are of different lengths, the bounds are much more complicated.

2 Preconditioning

Since convergence bounds depend on the eigenvalue distribution of the matrix $\mathcal{A}$, it makes sense to look for a **preconditioner** $\mathcal{M}$ such that $\mathcal{M}\mathcal{A}$ has a more favorable distribution of eigenvalues. Although MINRES works for any symmetric matrix, even if $\mathcal{M}$ is symmetric, $\mathcal{M}\mathcal{A}$ will generally not be symmetric, and so we cannot apply MINRES to $\mathcal{M}\mathcal{A}$. (We could apply GMRES; this may increase computational cost and memory requirements). Instead, we look for **SPD** preconditioners $\mathcal{M}$ that admit factorization $\mathcal{M} = EE^T$. Then, we apply MINRES to $E\mathcal{A}E^T$; this matrix has the same eigenvalues as $\mathcal{M}\mathcal{A}$. **This is equivalent to minimizing the residual in the norm induced by** $\mathcal{M}$ (rather than minimizing in the Euclidean norm). It is important to note: the factorization E is never needed in the actual algorithm. Instead, only the **action** of the matrix $\mathcal{M}$ is needed.

As before, define the **Schur complement** S by

$$S = -BA^{-1}B^T.$$

Since A is SPD and B is full row rank, the Schur complement is negative definite. Consider the **block-diagonal** preconditioner

$$\mathcal{D} = \begin{pmatrix} A^{-1} & 0 \\ 0 & -S^{-1} \end{pmatrix}.$$

It is immediate that $\mathcal{D}$ is symmetric and positive-definite. We consider the spectrum of $\mathcal{D}\mathcal{A}$.

Theorem 1. *Let $\mathcal{D}$ be the block-diagonal preconditioner defined above. Then the only distinct eigenvalues of $\mathcal{D}\mathcal{A}$ are*

$$1, \frac{1}{2}(1+\sqrt{5}), \frac{1}{2}(1-\sqrt{5}).$$

Proof. Let λ be an eigenvalue of $\mathcal{D}\mathcal{A}$, i.e.

$$\begin{aligned} A\boldsymbol{x} + B^T\boldsymbol{y} &= \lambda A\boldsymbol{x}, \\ B\boldsymbol{x} &= \lambda S\boldsymbol{y}, \end{aligned}$$

for some $(\boldsymbol{x}, \boldsymbol{y}) \neq 0$. Multiplying the first equation by BA^{-1} and eliminating $\boldsymbol{x}$, we obtain

$$(\lambda^2 - \lambda - 1)S\boldsymbol{y} = 0. \tag{2}$$

For $\boldsymbol{y} \neq 0$, this implies that

$$\lambda^2 - \lambda - 1 = 0,$$

from which we obtain the second two eigenvalues. If $\boldsymbol{y} = 0$, then the first equation implies $\lambda = 1$. □

From this, we can conclude that MINRES with $\mathcal{D}$ as a preconditioner will converge in at most three iterations. (Why?)

This preconditioner can also be applied to the slightly more general saddle-point system

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

where C is symmetric and positive-semidefinite.

Theorem 2. *Let $\mathcal{D}$ be the block-diagonal preconditioner, with $S = C + BA{-}1B^T$ is the (negative) Schur complement of $\mathcal{A}$ with respect to the $(1,1)$-block. This preconditioner is optimal in the sense that*

$$\sigma(\mathcal{D}\mathcal{A}) \subseteq \Big[-1, (1-\sqrt{5})/2\Big) \cup \Big[1, (1+\sqrt{5})/2\Big) \text{ and } \kappa(\mathcal{D}\mathcal{A}) \leq \frac{\sqrt{5}+1}{\sqrt{5}-1} \approx 2.618\ldots$$

Proof. Let λ be an eigenvalue of $\mathcal{D}^{-1}\mathcal{A}$, i.e.

$$\begin{aligned} A\boldsymbol{x} + B^T\boldsymbol{y} &= \lambda A\boldsymbol{x}, \\ B\boldsymbol{x} - C\boldsymbol{y} &= \lambda S\boldsymbol{y}, \end{aligned}$$

for some $(\boldsymbol{x}, \boldsymbol{y}) \neq 0$. Multiplying the first equation by $BA{-}1$ and eliminating $\boldsymbol{x}$, we obtain

$$\big(\lambda C + (\lambda^2 - \lambda - 1)S\big)\,\boldsymbol{y} = 0. \tag{3}$$

Considering the case $\lambda \neq 1$, we have $\boldsymbol{y} \neq 0$ and so $(S\boldsymbol{y}, \boldsymbol{y}) \geq (C\boldsymbol{y}, \boldsymbol{y}) > 0$. Equation (3) then implies that

$$\lambda(C\boldsymbol{y}, \boldsymbol{y}) + (\lambda^2 - \lambda - 1)(S\boldsymbol{y}, \boldsymbol{y}) = 0,$$

and so λ and $(\lambda^2 - \lambda - 1)$ must have opposite signs. We consider the two cases:

- If $\lambda > 0$ then this implies $\lambda^2 - \lambda - 1 < 0$ and so $\lambda < (\sqrt{5}+1)/2$. On the other hand, $(C\boldsymbol{y}, \boldsymbol{y}) \leq (S\boldsymbol{y}, \boldsymbol{y})$ implies $\lambda(S\boldsymbol{y}, \boldsymbol{y}) + (\lambda^2 - \lambda - 1)(S\boldsymbol{y}, \boldsymbol{y}) \geq 0$, and so $\lambda^2 \geq 1$, and $\lambda \in [1, (\sqrt{5}+1)/2)$.
- If $\lambda < 0$ then $\lambda^2 - \lambda - 1 > 0$ and so $\lambda < (1-\sqrt{5})/2$. By the same reasoning as above, $\lambda^2 \leq 1$, so $\lambda \geq -1$, and $\lambda \in [-1, (1-\sqrt{5})/2)$. $\square$

In general, we say a preconditioner $\mathcal{M}$ is **optimal** if there exist constants c and C such that

$$c\boldsymbol{x}^T\mathcal{M}\boldsymbol{x} \leq \boldsymbol{x}^T\mathcal{A}^{-1}\boldsymbol{x} \leq C\boldsymbol{x}^T\mathcal{M}\boldsymbol{x}. \tag{4}$$

The above results show that $\mathcal{D}$ is an optimal preconditioner for $\mathcal{A}$.

Using block-diagonal preconditioning, we can reduce the problem of finding an optimal preconditioner for the indefinite saddle-point system to two positive-definite problems, one for A and one for $-S$. However, each of these problems may be very large, and we cannot practically compute either A^{-1} or $-S^{-1}$. It suffices to find *spectrally equivalent* approximations $B_A \sim A^{-1}$ and $B_S \sim -S^{-1}$, and then to use the preconditioner

$$\mathcal{B} = \begin{pmatrix} B_A & 0 \\ 0 & B_S \end{pmatrix}.$$

This can be seen as follows. We say B_A is spectrally equivalent to A^{-1} if there exist constants c and C such that

$$c\boldsymbol{x}^T B_A\boldsymbol{x} \leq \boldsymbol{x}^T A^{-1}\boldsymbol{x} \leq C\boldsymbol{x}^T B_A\boldsymbol{x}.$$

This is equivalent to saying that the spectrum of $B_A A$ lies in the interval $[c, C]$. Spectral equivalence for the individual blocks B_A and B_S implies that there exist constants c and C such that

$$c\boldsymbol{x}^T \mathcal{B}\boldsymbol{x} \leq \boldsymbol{x}^T \mathcal{D}\boldsymbol{x} \leq C\boldsymbol{x}\mathcal{B}\boldsymbol{x}. \tag{5}$$

Combining (4) and (5), we see that $\mathcal{B}$ is **also an optimal preconditioner** for $\mathcal{A}$. This means that we can replace the blocks in $\mathcal{D}$ with spectrally equivalent blocks and we pay only a constant factor.

Preconditioning A is standard; for the Stokes problem, this is a Laplace operator. We will see approaches including domain decomposition and multigrid to address this problem. The Schur complement $S = -BA^{-1}B^T$ is more challenging, because the matrix S is defined in terms of the **inverse** of A, and so approximating $-S^{-1}$ can be difficult. This is usually problem-specific

In the case of Stokes, we have the following result.

Theorem 3. *Let S be the Schur complement of the Stokes system, $S = -BA^{-1}B^T$, where A is the (Laplacian) stiffness matrix, and B is the divergence matrix. Let M be the mass matrix defined on the pressure space P, i.e.*

$$\boldsymbol{q}^T M\boldsymbol{p} = (q, p)_{L^2(\Omega)}.$$

Then,

$$\beta^2 \boldsymbol{x}^T M\boldsymbol{x} \leq \boldsymbol{x}^T(-S^{-1})\boldsymbol{x} \leq \boldsymbol{x}^T M\boldsymbol{x}$$

where β is the inf–sup constant for the velocity and pressure spaces.

Proof. The inf–sup condition is

$$\inf_{q\in P} \sup_{v\in V} \frac{b(v,q)}{\|v\|_V \|q\|_P} \geq \beta.$$

In matrix form, this is

$$\inf_{\boldsymbol{q}} \sup_{\boldsymbol{v}} \frac{\boldsymbol{q}^T B\boldsymbol{v}}{(\boldsymbol{v}^T A\boldsymbol{v})^{1/2}(\boldsymbol{q}^T M\boldsymbol{q})^{1/2}} \geq \beta.$$

Setting $\boldsymbol{w} = A^{1/2}\boldsymbol{v}$, this gives

$$\begin{aligned}
\beta &\leq \inf_{\boldsymbol{q}} \sup_{\boldsymbol{w}=A^{1/2}\boldsymbol{v}} \frac{\boldsymbol{q}^T BA^{-1/2}\boldsymbol{w}}{(\boldsymbol{w}^T\boldsymbol{w})^{1/2}(\boldsymbol{q}^T M\boldsymbol{q})^{1/2}} \\
&= \inf_{\boldsymbol{q}} \sup_{\boldsymbol{w}=A^{1/2}\boldsymbol{v}} \frac{(A^{-1/2}B^T\boldsymbol{q})^T\boldsymbol{w}}{(\boldsymbol{w}^T\boldsymbol{w})^{1/2}(\boldsymbol{q}^T M\boldsymbol{q})^{1/2}} \\
&= \inf_{\boldsymbol{q}} \frac{((A^{-1/2}B^T\boldsymbol{q})^T A^{-1/2}B^T\boldsymbol{q})^{1/2}}{(\boldsymbol{q}^T M\boldsymbol{q})^{1/2}} \\
&= \inf_{\boldsymbol{q}} \frac{(\boldsymbol{q}^T BA^{-1}B^T\boldsymbol{q})^{1/2}}{(\boldsymbol{q}^T M\boldsymbol{q})^{1/2}}
\end{aligned}$$

which gives the lower bound.

For the upper bound,

$$(q, \nabla \cdot \boldsymbol{v}) \leq \|q\|_{L^2} \|\nabla \cdot \boldsymbol{v}\|_{L^2} \leq \|q\|_{L^2} \|\nabla \boldsymbol{v}\|_{L^2}.$$

In matrix form,

$$\boldsymbol{q}^T B \boldsymbol{v} \leq (\boldsymbol{q}^T M \boldsymbol{q})^{1/2} (\boldsymbol{v}^T A \boldsymbol{v})^{1/2}.$$

Repeating a similar argument as above, letting $\boldsymbol{w} = A^{1/2} \boldsymbol{v}$, we obtain

$$\frac{\boldsymbol{q}^T B A^{-1} B^T \boldsymbol{q}}{\boldsymbol{q}^T \boldsymbol{q}} \leq \frac{\boldsymbol{q}^T M \boldsymbol{q}}{\boldsymbol{q}^T \boldsymbol{q}}. \qquad \square$$

As a consequence of this result, we can choose $B_S = M^{-1}$, where M is the pressure mass matrix. We showed that $\kappa(M) = \mathcal{O}(1)$ (its maximum and minimum eigenvalues are both on the order of h^2), and so we can further replace M^{-1} with a simple approximation like its diagonal.