Null-Space Based Preconditioners for Saddle-Point Systems

Chen Greif
The University of British Columbia
Vancouver
Canada

SIAM Applied Linear Algebra
October 26–30, 2015

Joint Work With Ron Estrin

Saddle-Point Systems

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

$$A \in \mathbb{R}^{n\times n},\ B \in \mathbb{R}^{m\times n},\ x, f \in \mathbb{R}^n,\ y, g \in \mathbb{R}^m.$$

- A and B are sparse and large; $m < n$ (may have $m \ll n$).
- Central assumption: A has nullity m (i.e., $\text{rank}(A) = n - m$), and the saddle-point matrix is nonsingular. We say that A is *maximally rank deficient.*
- A corollary: B has full row rank, i.e., $\text{rank}(B) = m$.
- We will assume symmetry of A throughout but this is not a strict requirement.

Highly Rank Deficient Leading Block

- Time-harmonic Maxwell
- Norm Minimization with Equality Constraints
- Interior-point methods (depends on the number of active constraints at the solution)
- Vorticity/Curl formulations of the Stokes equations
- Geophysical inverse problems
- Insert your problem here...

Formulas for the Inverse

Benzi, Golub, Liesen (2005), Section 3:

A formula based on the Schur complement:

$$\mathcal{K}^{-1} = \begin{pmatrix} A^{-1} - A^{-1}B^T S^{-1} B A^{-1} & S^{-1} B^T A^{-1} \\ A^{-1} B S^{-1} & -S^{-1} \end{pmatrix},$$

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

A formula based on null spaces (invertibility of A not required):

$$\mathcal{K}^{-1} = \begin{pmatrix} V & (I - VA)\, B^T \left(BB^T\right)^{-1} \\ \left(BB^T\right)^{-1} B\,(I - AV) & \left(BB^T\right)^{-1} B\,(A - AVA)\, B^T \left(BB^T\right)^{-1} \end{pmatrix}$$

where

$$V = Z\left(Z^T A Z\right)^{-1} Z^T,$$

and Z is a basis for the null space of B: $BZ = 0$.

The Surprising Nonzero Structure of the Inverse When null$(A) = m$

Not quite fully dense...

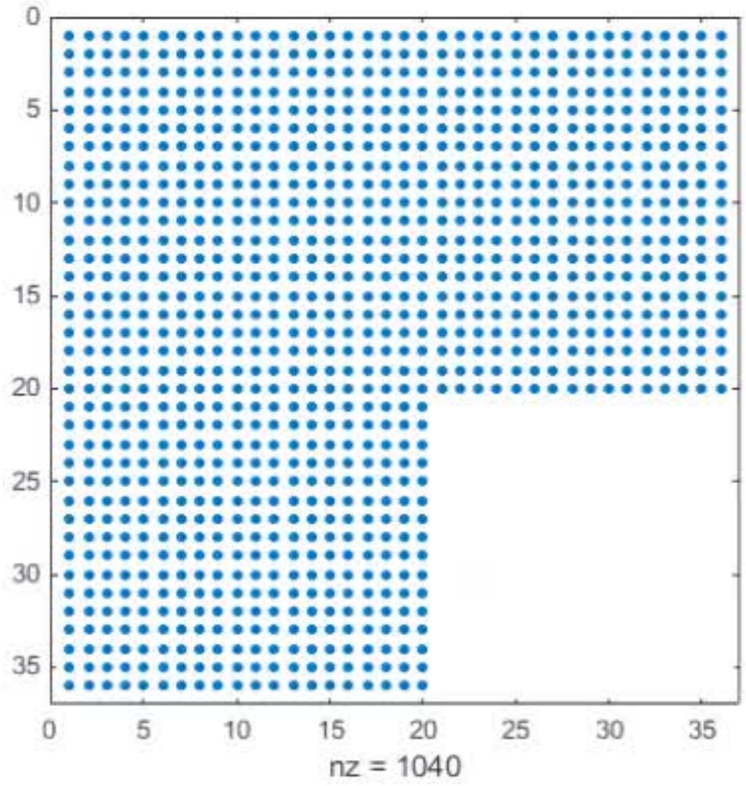

Figure: The inverse has a saddle-point form.

Rank Structure of the (2,2) Block

Theorem

If A be symmetric positive semidefinite with $\text{null}(A) = r \leq m$,

$$\mathcal{K}^{-1} = \begin{pmatrix} * & * \\ * & X \end{pmatrix},$$

where $X \in \mathbb{R}^{m \times m}$, and $\text{rank}(X) = m - r$.

Corollary

Let A be symmetric positive semidefinite with $null(A) = m$ and suppose $\mathcal{K}$ is invertible. Then

$$\mathcal{K}^{-1} = \begin{pmatrix} * & * \\ * & 0 \end{pmatrix}.$$

Inverse of Augmented Lagrangian

Fletcher [1974]:

$$\mathcal{K}(W) := \begin{pmatrix} A + B^T W^{-1} B & B^T \\ B & 0 \end{pmatrix}.$$

Then if $\mathcal{K}(W)$ is nonsingular:

$$\mathcal{K}^{-1}(W) = \mathcal{K}^{-1}(0) - \begin{pmatrix} 0 & 0 \\ 0 & W^{-1} \end{pmatrix}.$$

- Works for nonsymmetric matrices and singular W too, Golub & G. [2003]
- Form of (1,1) block successfully used for preconditioning fluid flow problems, Benzi and Olshanskii [2006]
- Preconditioners for time-harmonic Maxwell, G. & Schötzau [2006]

Side Note: Optimal Conditioning for the Maximally Rank Deficient Setting

Quasi-direct sums (Fiedler [1981]):
Let $M, N \in \mathbb{R}^{n \times n}$, and let $\text{rank}(M) = r$, $\text{rank}(N) = n - r$, such that $M + N$ is nonsingular. Then we can perform a non-unitary simultaneous diagonalization:

$$M = P \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} Q^T; \quad N = P \begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix} Q^T,$$

and use this to minimize the condition number of $\widehat{A}$ for $W^{-1} = \gamma I$, with $M = A$ and $N = \gamma B^T B$; requires knowledge (or estimates) of the extremal eigenvalues of A and the extremal singular values of B.

Schur Complement

The rank of the (2,2) block of the inverse and the inverse result just mentioned lead to:

Theorem

Suppose $\text{null}(A) = m$ *and let* $W \in \mathbb{R}^{m \times m}$ *be an invertible matrix. Then*

$$B\left(A + B^T W^{-1} B\right)^{-1} B^T = W.$$

A New Formula for the Inverse

Using the above results and the Helmholtz decomposition,

$$\ker(A) \oplus \ker(B) = \mathbb{R}^n,$$

we get

$$\mathcal{K}^{-1} = \begin{pmatrix} \left(A + B^T L^{-1} B\right)^{-1} \left(I - B^T L^{-1} C^T\right) & C L^{-1} \\ L^{-1} C^T & 0 \end{pmatrix},$$

where $AC = 0, \quad L = BC.$

Preconditioners Based on the Null-Space of A

Inverse, again:

$$\mathcal{K}^{-1} = \begin{pmatrix} \left(A + B^T L^{-1} B\right)^{-1} (I - B^T L^{-1} C^T) & C L^{-1} \\ L^{-1} C^T & 0 \end{pmatrix}.$$

Two natural choices for preconditioners:

$$\mathcal{P}_1^{-1} = \begin{pmatrix} (A + R)^{-1} (I - B^T L^{-1} C^T) & C L^{-1} \\ L^{-1} C^T & 0 \end{pmatrix}.$$

$$\mathcal{P}_2^{-1} = \begin{pmatrix} (A + R)^{-1} & C L^{-1} \\ L^{-1} C^T & 0 \end{pmatrix}.$$

$R \approx B^T (BC)^{-1} B$ and $L \approx BC$, both symmetric and sparse.
(Not necessarily easy to find such matrices.)

Spectral Analysis

$$\mathcal{P}_1^{-1} = \begin{pmatrix} (A+R)^{-1}(I - B^T L^{-1} C^T) & CL^{-1} \\ L^{-1}C^T & 0 \end{pmatrix}.$$

$$\mathcal{P}_2^{-1} = \begin{pmatrix} (A+R)^{-1} & CL^{-1} \\ L^{-1}C^T & 0 \end{pmatrix}.$$

Theorem

The matrices $\mathcal{P}_1^{-1}\mathcal{K}$ and $\mathcal{P}_2^{-1}\mathcal{K}$ have eigenvalue 1 of algebraic multiplicity at least $2m$. In $\mathcal{P}_1^{-1}\mathcal{K}$, the eigenvalue 1 has geometric multiplicity at least $2m$ with eigenvectors of the form (Cy, u) where $u, y \in \mathbb{R}^m$. In $\mathcal{P}_2^{-1}\mathcal{K}$ the eigenvalue 1 has geometric multiplicity at least m with eigenvectors of the form $(Cy, 0)$.

Krylov Subspaces

- It is possible to show, for example, that if $RC = B^T$ then

$$\mathcal{P}_2^{-1}\mathcal{K} = \begin{pmatrix} (A+R)^{-1}(A+B^TL^{-1}B) & C \\ 0 & I \end{pmatrix}.$$

Krylov Subspaces

- It is possible to show, for example, that if $RC = B^T$ then

$$\mathcal{P}_2^{-1}\mathcal{K} = \begin{pmatrix} (A+R)^{-1}(A + B^T L^{-1} B) & C \\ 0 & I \end{pmatrix}.$$

- If $RC = B^T$ and f is divergence-free, can use $\mathcal{P}_2$ as a preconditioner for the Conjugate Gradient method. That is, can use CG for an indefinite preconditioner with an indefinite coefficient matrix.

Krylov Subspaces

- It is possible to show, for example, that if $RC = B^T$ then
$$\mathcal{P}_2^{-1}\mathcal{K} = \begin{pmatrix} (A+R)^{-1}(A+B^TL^{-1}B) & C \\ 0 & I \end{pmatrix}.$$
- If $RC = B^T$ and f is divergence-free, can use $\mathcal{P}_2$ as a preconditioner for the Conjugate Gradient method. That is, can use CG for an indefinite preconditioner with an indefinite coefficient matrix.
- Even better: if $RC = B^T$ and f is not divergence-free, can use $\mathcal{P}_1$ as a preconditioner for CG.
- Matrix-vector products with $\mathcal{P}_1$ are not much more expensive than with $\mathcal{P}_2$.
- Potential limitation: need to know null-space of leading block (at least approximately), want it ideally sparse.

In Search of a Recipe

Theorem

Suppose Z is a null matrix of B, namely $BZ = 0$, and C is a null matrix of A. Then $R = B^T L^{-1} B$ if and only if R satisfies the conditions (i) $RZ = 0$, (ii) $RC = B^T$, and (iii) $R = R^T$.

An inf-sup like condition:

Theorem

If R satisfies $\|(A+R)^{-1} R u_B\| \leq \alpha$ for $\alpha < 1$ and $u_B \in \ker(B)$, then the eigenvalues λ of $\mathcal{P}_i^{-1}\mathcal{K}, i = 1, 2$, satisfy $1 - \alpha \leq \lambda \leq 1 + \alpha$.

The condition $RZ = 0$ and its relaxed form primarily affect the distribution of the eigenvalues, whereas the conditions $RC = B^T$ and $R = R^T$ are needed in order to be able to potentially use CG.

A Numerical Illustration: Time-Harmonic Maxwell

The time-harmonic Maxwell equations with constant coefficients in lossless media with perfectly conducting boundaries:

$$\begin{aligned} \nabla \times \nabla \times u - k^2 u + \nabla p = f \qquad & \text{in } \Omega, \\ \nabla \cdot u = 0 \qquad & \text{in } \Omega, \\ u \times n = 0 \qquad & \text{on } \partial\Omega, \\ p = 0 \qquad & \text{on } \partial\Omega. \end{aligned}$$

u is an electric vector field; p is a scalar multiplier.
$k^2 = \omega^2 \epsilon \mu$, where ω is the temporal frequency, and ϵ and μ are permittivity and permeability parameters.
Assume throughout small wave number: $k \ll 1$.
See Hiptmair, Monk,...
(Preconditioners based on augmented Lagrangian approach, Schötzau & G. [2006])

Eigenvalues of Preconditioned Matrix

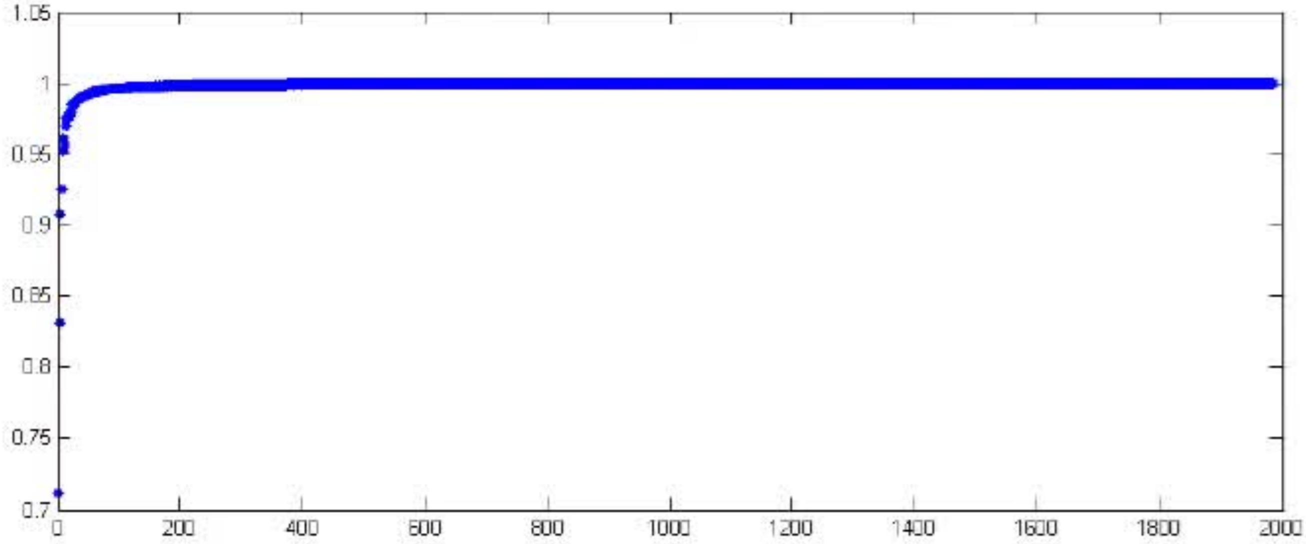

Figure: Eigenvalue distribution of the preconditioned matrix $\mathcal{P}_1^{-1}\mathcal{K}$ for the time-harmonic Maxwell problem.

A Numerical Illustration: Geophysical Inverse Problem

Thanks to Eldad Haber and Kristofer Davis.

$$
\begin{array}{ll}
\min & \frac{1}{2}\|P^T u - d\|_{W_d}^2 + \frac{\alpha}{2}\|Lm\|^2 \\
\text{s.t.} & F(m)u + Gm = f,
\end{array}
$$

where m is a model, L is a regularization operator (typically a discretized second order differential operator), and the forward problem $F(m)u + Gm = f$ is typically a second order differential equation; G is an averaging operator. P is an observation matrix, and W_d is the standard deviation.