Iterative Solution of Saddle Point Linear Systems

Chen Greif

Department of Computer Science
University of British Columbia
Vancouver, Canada

Chinese Academy of Sciences
Beijing
May 17, 2010

Outline

Saddle Point Linear Systems

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

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

- A and B are sparse and large; m is typically smaller (possibly much smaller) than n.
- A is symmetric positive semidefinite (occurs often in practice): $\langle Ax, x\rangle \geq 0$ for all $x \in \mathbb{R}^n$. Note: A could be singular.
- B has full row rank: rank$(B) = m$.
- In this setting the saddle point matrix is nonsingular if and only if $\mathrm{null}(A) \bigcap \mathrm{null}(B) = \{0\}$.

Many Applications

- fluid flow: Stokes equations, Navier-Stokes,...
- electromagnetics
- linear, quadratic, semidefinite programming in optimization

Survey of solution methods: Benzi, Golub & Liesen, Acta Numerica, 2005 (137 pages, over 500 references).

Example: the Steady State Stokes Equations

$$-\Delta u + \nabla p = f,$$
$$\nabla \cdot u = 0;$$

on a domain Ω with appropriate boundary conditions (Dirichlet, Neumann, mixed).
u is a velocity vector (2D or 3D) and p is a pressure scalar.
FE or FD discretization yields

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

where A is symmetric positive definite, but ill-conditioned as the mesh is refined.
See Elman, Silvester, Wathen [2005].

Example: Quadratic Programming (Equality Constraints)

$$\min_x \frac{1}{2}\langle Ax, x\rangle - \langle f, x\rangle$$
$$\text{subject to} \;\; Bx = g.$$

Lagrangian:

$$\mathcal{L}(x, \lambda) = \frac{1}{2}\langle Ax, x\rangle - \langle f, x\rangle + \langle \lambda, g - Bx\rangle,$$

differentiate and equate to zero

$$\nabla \mathcal{L}(x, \lambda) = 0,$$

and obtain the saddle point system.
See Nocedal and Wright [2006].

Challenges and Goals

- matrix is indefinite (tough for iterative and direct solvers)
- want to exploit structure
- want to take advantage of properties of underlying continuous operators (if applicable)
- want to know how to deal with a singular (possibly with a high nullity) or an ill-conditioned $(1,1)$ block A:
    - interior point methods in optimization
    - The Maxwell equations in mixed form
    - fluid flow problems
    - magnetohydrodynamics problems
    - ...

A Few Possible Re-Formulations

Re-Formulation 1: Nonsymmetric Semidefinite System

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

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

Matrix is nonsymmetric but positive semidefinite;

$$\begin{pmatrix} 0 & B^T \\ -B & 0 \end{pmatrix}$$

is skew-symmetric;
can use Hermitian-Skew-Hermitian Splitting (HSS)
(Bai, Ng and Golub [2003], Benzi & Golub [2004]);
or Accelerated HSS (AHSS): Bai & Golub [2007].

Re-formulation 2: Regularization

$$\begin{pmatrix} A & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix},$$

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

Common in fluid flow for some low order mixed finite element discretizations

This approach gives a different solution than for $C = 0$.

C is typically positive semidefinite, but not always; see Bai, Ng and Wang [2009] for general symmetric (2,2) block.

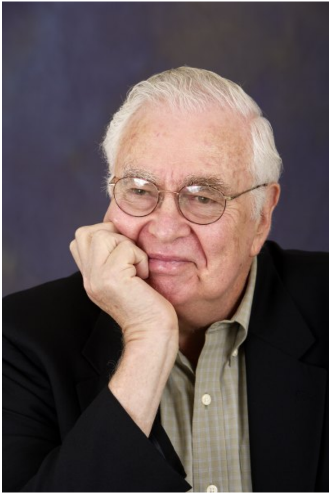
Re-Formulation 3: Augmented Lagrangian (Stabilization)

Augmented Saddle Point System:

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix} \to \begin{pmatrix} A + B^T W B & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}.$$

- A may be singular or highly ill-conditioned;
- may be interpreted as penalizing constraint violations ;
- explored mainly for optimization:
  Hestenes [1969], Fletcher[mid 1970s],... Fortin & Glowinski [mid 1980s], ...

A Tribute to Gene Golub (1932-2007)



Spectral Analysis ($W = \gamma I$)

[Golub and Greif, 2003]

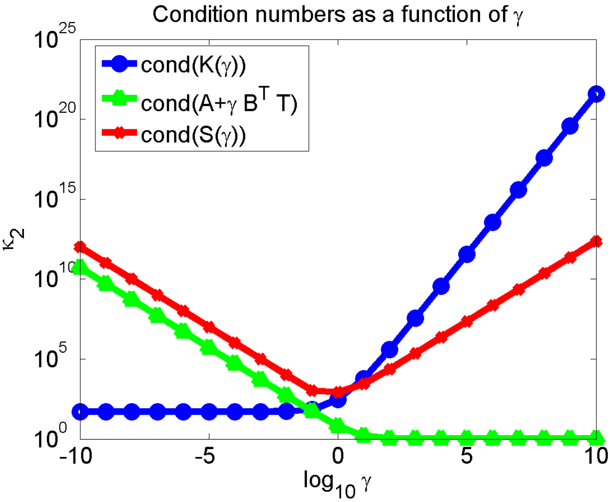
Let $\mathcal{K}(\gamma) = \begin{pmatrix} A + \gamma B^T B & B^T \\ B & 0 \end{pmatrix}$.

- Clustering: As γ grows large, the m negative eigenvalues cluster near to $-\frac{1}{\gamma}$.
- Dependence of $\kappa_2(\mathcal{K})$ on γ:

$$\kappa_2(\mathcal{K}(\gamma)) \rightarrow \gamma^2 \|B\|_2^2 \neq 0 \quad \text{as} \quad \gamma \rightarrow \infty.$$

So, need to choose a moderate value of γ to balance fast convergence with conditioning deterioration.

Condition Numbers: Singular Leading Block



The Inverse

Fletcher [1974]; Golub & Greif [2003]

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

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

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

Block Preconditioning

Basic Block Decompositions

Dual Schur Complement (A nonsingular):

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

Primal Schur Complement (regularizing (2,2) block added):

$$\begin{pmatrix} A & B^T \\ B & -W \end{pmatrix} = \begin{pmatrix} I & -B^TW^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A + B^TW^{-1}B & 0 \\ 0 & -W \end{pmatrix} \begin{pmatrix} I & 0 \\ -W^{-1}B & I \end{pmatrix}$$

Block Preconditioning

Three basic approaches with many possible variants:

- (Dual) Schur complement-based preconditioners:

$$\mathcal{M}_S = \begin{pmatrix} A & 0 \\ 0 & BA^{-1}B^T \end{pmatrix}.$$

- Constraint preconditioners:

$$\mathcal{M}_C = \begin{pmatrix} G & B^T \\ B & 0 \end{pmatrix}.$$

- **Augmentation (primal Schur complement-based) preconditioners:**

$$\mathcal{M}_A = \begin{pmatrix} A + B^T W^{-1} B & 0 \\ 0 & W \end{pmatrix}.$$

Augmentation Preconditioner

Greif and Schötzau [2006-2007]

Consider the preconditioner giving the preconditioned matrix

$$\underbrace{\begin{pmatrix} A + B^T W^{-1} B & 0 \\ 0 & W \end{pmatrix}^{-1}}_{\mathcal{M}} \underbrace{\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}}_{\mathcal{K}}.$$

- If A has nullity $r \leq m$, then the preconditioned matrix has the eigenvalue 1 of multiplicity n, the eigenvalue -1 of multiplicity r, and the remaining $m - r$ eigenvalues of $\mathcal{M}^{-1}\mathcal{K}$ satisfy

$$\lambda A v = B^T W^{-1} B v.$$

- Eigenvectors are known and for $\lambda = 1$ are readily available.

Blessing of Nullity

If A has nullity m, then two distinct eigenvalues:

$$1 \text{ (multiplicity } n)$$
$$-1 \text{ (multiplicity } m)$$

Also, the eigenvectors are explicitly known. In particular, $n - m$ of the eigenvectors are of the form $(z_i, 0)$, where z_i form a basis for the null space of B.

Applications

Semidefinite Programming (SDP)
Joint With Michael Overton [2010]

The Problem

$$\begin{array}{ll} \min_{X\in\mathcal{S}^n} & C \bullet X \\ \text{such that} & A_k \bullet X = b_k, \; k = 1, \ldots, m \\ & X \succeq 0. \end{array}$$

- $b \in \mathbf{R}^m$ and $C \in \mathcal{S}^n$, the space of $n \times n$ real symmetric matrices
- the A_k are linearly independent in $\mathcal{S}^n$
- the inner product $C \bullet X$ is $\mathrm{tr}\, CX = \sum_{i,j} C_{ij} X_{ij}$
- $X \succeq 0$ means X is positive semidefinite

The Dual

The dual standard form is:

$$\begin{array}{ll} \max_{y \in \mathbf{R}^m,\ Z \in \mathcal{S}^n} & b^T y \\ \text{such that} & \sum_{k=1}^m y_k A_k + Z = C \\ & Z \succeq 0. \end{array}$$

In practice: all matrices are block-diagonal.

Primal-Dual Interior Point Methods

Illustrate on LP:

$$\min_x c^T x, \qquad \text{subject to}: \quad Ax = b, \; x \geq 0.$$

$$\max_\lambda b^T \lambda, \qquad \text{subject to}: \quad A^T\lambda + s = c, \; s \geq 0.$$

Progress along a path in the interior of the domain.

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta s \end{pmatrix} = - \begin{pmatrix} r_c \\ r_b \\ XSe \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} D & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = \begin{pmatrix} -r_c + s \\ -r_b \end{pmatrix}$$

Matrix becomes increasingly ill-conditioned as solution is approached.

SDP: Many Important Applications

- Minimize the maximum eigenvalue (related to stabilizing PDEs)
- Minimize the L_2 norm of a matrix
- Generalize other optimization problems (linear programming, convex quadratically constrained programming, etc.)
- Control theory
- Minimization of nuclear norm
- ... and more

See Todd, Acta Numerica [2001]

The Augmented System

The linear algebra bottleneck is

$$H \begin{bmatrix} \text{vec } \Delta X \\ \Delta y \end{bmatrix} = f,$$

where

$$H = \begin{bmatrix} X^{-1} \otimes Z & A^T \\ A & 0 \end{bmatrix}, \quad A = \begin{bmatrix} (\text{vec } A_1)^T \\ \vdots \\ (\text{vec } A_m)^T \end{bmatrix}.$$

Here, "vec" maps $n \times n$ matrices to vectors in $\mathbf{R}^{n^2}$.

The Central Path

Assuming that the primal and dual SDPs have feasible points $X \succ 0$ and (y, Z) with $Z \succ 0$, the central path

$$\{(X^\mu, y^\mu, Z^\mu) \text{ feasible with } X^\mu Z^\mu = \mu I, \mu > 0\}$$

exists and converges to solutions of the primal and dual SDPs (which have the same optimal value) as $\mu \downarrow 0$.

At the solution, $XZ = 0$. Suppose the primal solution X has rank r and the dual slack solution Z has rank $n - r$, that is strict complementarity holds. Then X^μ and Z^μ respectively have r and $n - r$ eigenvalues that are $O(1)$ as $\mu \downarrow 0$; the others are $O(\mu)$.

Primal-dual interior-point path-following methods generate iterates that approximately follow the central path.

Iterative Methods for SDP

- Iterative methods for semidefinite programming problems have not been extensively used, perhaps because the problems have a strong dense component and have not been overly large.
- Toh [2003]: apply preconditioned SQMR to a newly introduced dense “reduced augmented system”, using diagonal preconditioners
- Zhao, Sun, Toh [2009]: a Newton-CG augmented Lagrangian method

Augmented System

The saddle point matrix is

$$H = \begin{bmatrix} X^{-1} \otimes Z & A^T \\ A & 0 \end{bmatrix}.$$

Since on the central path, $Z = \mu X^{-1}$, and $n - r$ eigenvalues of X are $O(\mu)$, H has order $n^2 + m$ and has

- $(n - r)^2$ eigenvalues that are $O(1/\mu)$
- the rest are $O(1)$.

Preconditioning the Augmented System

We consider preconditioning the augmented system matrix

$$H = \begin{bmatrix} X^{-1} \otimes Z & A^T \\ A & 0 \end{bmatrix}$$

by

$$K = \begin{bmatrix} X^{-1} \otimes Z + \beta B^T B & 0 \\ 0 & \beta^{-1} I \end{bmatrix}$$

in two cases:

1. $B = A$
2. B consists of s rows of A: $\beta B^T B = A^T V A$ where V is sparse, diagonal

Then we could iterate with MINRES, with an inner CG iteration to "invert" the K_{11} block.

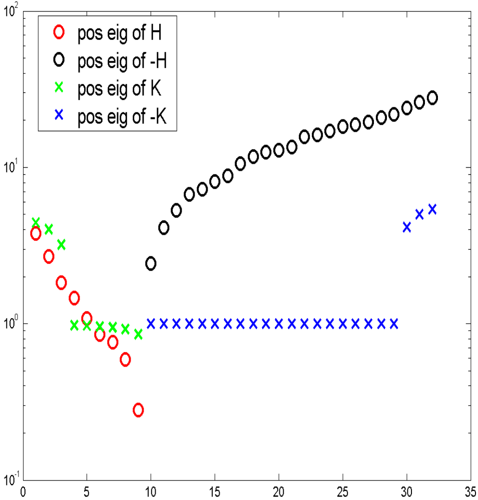
In second case require $s \geq r(r+1)2$, number of eigenvalues of $X^{-1} \otimes Z$ that are $O(\mu)$.

Low Rank correction ($m - s$ small)

Suppose $W^{-1} = \beta I_m$. Let V be a diagonal matrix with s of its diagonal values equal to β and the rest zero. Denote the eigenvalues of $K^{-1}H$ in this case by ν_j, ordered in descending order. Then, for β sufficiently large, the eigenvalues of the preconditioned matrix are given by

$$
\begin{aligned}
&\nu_j > 1, && j = 1, \ldots, m-s;\\
&\nu_j = 1, && j = m-s+1, \ldots, N;\\
&-1 < \frac{-\beta\gamma_{j+m-s}}{\beta\gamma_{j+m-s}+1} \le \nu_{N+j} \le \frac{-\beta\gamma_j}{\beta\gamma_j+1} < 0, && j = 1, \ldots, s;\\
&\nu_j < -1, && j = N+s+1, \ldots, N+m
\end{aligned}
$$

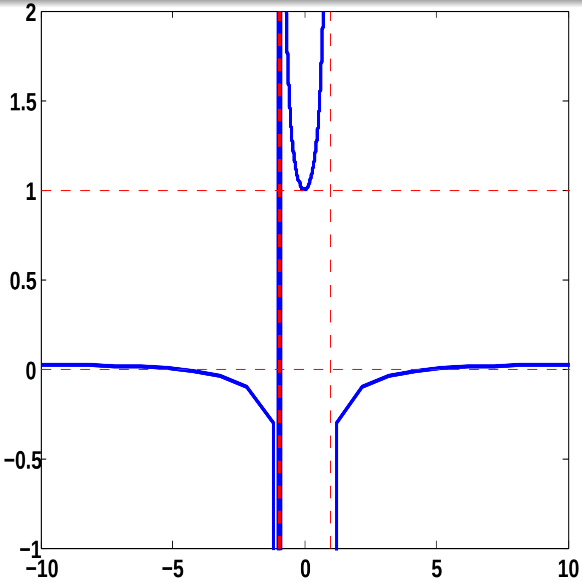
Eigenvalues

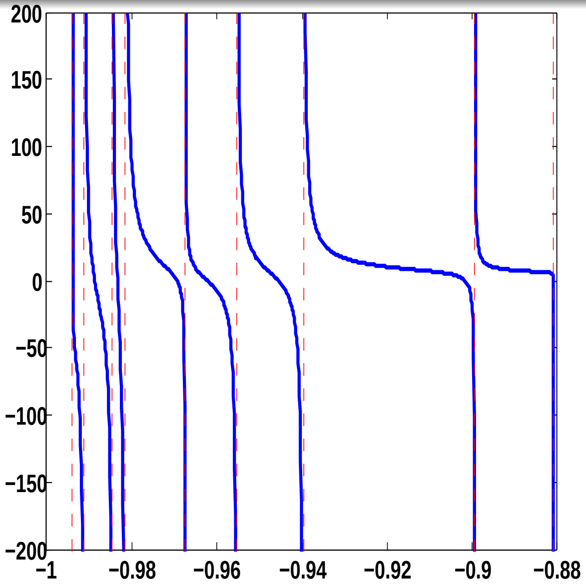


Interlacing

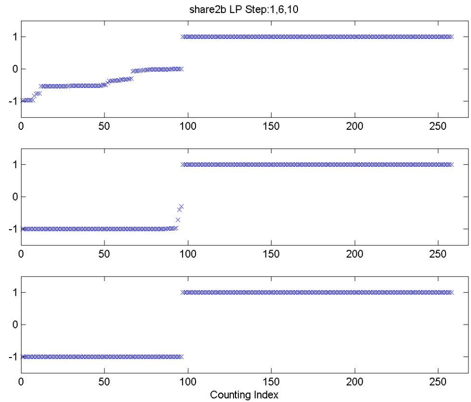
V is Low rank change of W^{-1}: developed an **interlacing theory** for a quadratic eigenvalue problem associated with the Schur complement matrix M

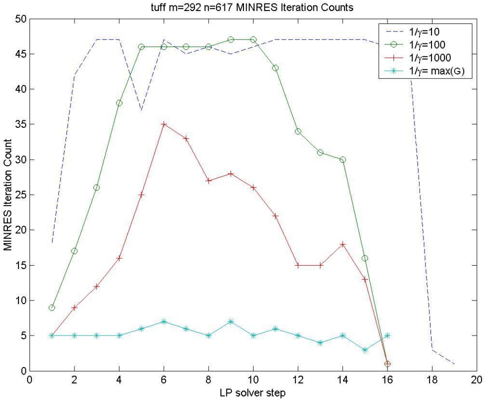
$$\left(\nu^2(M^{-1} + \beta I_m - uu^T) - \nu M^{-1} - \beta I_m\right)\tilde{z} = 0.$$


2
1.5
1
0.5
0
−0.5
−1
−10
−5
0
5
10


200
150
100
50
0
−50
−100
−150
−200
−1
−0.98
−0.96
−0.94
−0.92
−0.9
−0.88

Solution of LPs and QPs (with Rees)




tuff m=292 n=617 MINRES Iteration Counts
MINRES Iteration Count
LP solver step
1/γ=10
1/γ=100
1/γ=1000
1/γ= max(G)
0
5
10
15
20
25
30
35
40
45
50
0
2
4
6
8
10
12
14
16
18
20

Time-Harmonic Maxwell (with Li and Schötzau)

Dominik Schötzau

Dan Li

The Maxwell Problem in Mixed Form

Greif and Schötzau [2007], Greif, Li and Schötzau [2009-2010]
The time-harmonic Maxwell equations with constant coefficients in lossless media with perfectly conducting boundaries:

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

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: $k \ll 1$.

Mixed Finite Element Discretization: Saddle Point System

$$\begin{pmatrix} A - k^2 M & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} g \\ 0 \end{pmatrix}.$$

$A \in \mathbb{R}^{n\times n}$ is the discrete curl-curl; $B \in \mathbb{R}^{m\times n}$ is a discrete divergence operator with full row rank; $M \in \mathbb{R}^{n\times n}$ is a vector mass matrix.

Note: A is semidefinite with nullity m; (1,1) block is **indefinite** if $k \neq 0$.

Define also $L \in \mathbb{R}^{m\times m}$ as the scalar Laplacian.

Much Related Work

- Arnold and Falk [2000]
- Bochev, Hu, Siefert, Tuminaro [2007]
- Demkowicz and Vardapetyan [1998]
- Hiptmair [2002] (Acta Numerica survey)
- Hiptmair and Xu [2006]
- Hu, Tuminaro, Bochev, Garasi and Robinson [2005]
- Hu and Zou [2004]
- Reitzinger and Schöberl [2002]
- Römer, Witzigmann, Chinellato and Arbenz [2007]

and more...

A Few Key Properties

- Discrete Helmholtz Decomposition: $\mathbb{R}^n = \text{null}(A) \oplus \text{null}(B)$.
- There is a 'gradient matrix' $C \in \mathbb{R}^{n \times m}$ such that for any $u \in \text{null}(A)$ there is a unique $q \in \mathbb{R}^m$ such that $u = Cq$.
  - $AC = 0$
  - $BC = L$
  - $MC = B^T$
- Coercivity, Continuity and Inf-Sup,...
- And more...

Preconditioner

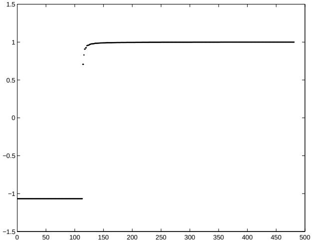
- MINRES preconditioned with

$$\mathcal{P}_{M,L} = \begin{pmatrix} A + \gamma M & 0 \\ 0 & L \end{pmatrix},$$

  where γ is a scalr, L is the scalar Laplace matrix and M is the vector mass matrix
- $\mathcal{P}_{M,L}^{-1}\mathcal{K}$ has eigenvalues $\mu_+ = 1$ and $\mu_- = -1/(1-k^2)$ of high multiplicities
- The rest of the eigenvalues are bounded; bound depends on the shape regularity of the mesh and the finite element approximation order.

Eigenvalues

Use Laplacian L for augmentation, then replace $B^T L^{-1} B$ by a simple mass matrix M which is spectrally equivalent.
For inner iteration, use Hiptmair & Xu's auxiliary nodal projection.



Grid	$n+m$	$k=0$	$k=\frac{1}{8}$	$k=\frac{1}{4}$	$k=\frac{1}{2}$
G1	113	5	5	5	5
G2	481	5	5	5	5
G3	1,985	5	5	5	5
G4	8,065	6	6	5	6
G5	32,513	6	6	6	6
G6	130,561	6	6	6	6
G7	523,265	6	6	6	6

Table: Iteration counts for a typical example with a divergence free right hand side and various values of k and various meshes, using MINRES for solving the saddle point system with the preconditioner $\mathcal{P}_{M,L}$. The iteration was stopped once the initial relative residual was reduced by a factor of 10^{-10}.

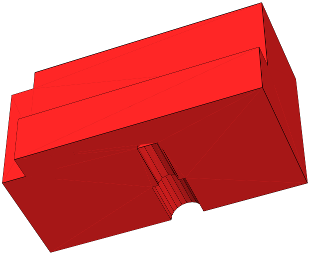
Large Scale Implementation

- $A + \gamma M$:
  - CG preconditioned with Hiptmair & Xu's solver:

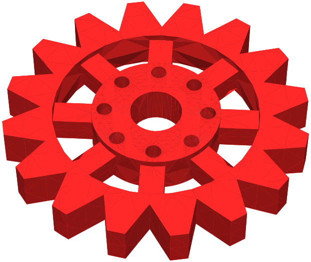
    $$\mathcal{P}_V^{-1} = \mathrm{diag}(A + \gamma M)^{-1} + P(\bar{L} + \gamma \bar{Q})^{-1} P^T + \gamma^{-1} C (L^{-1}) C^T$$

  - $\kappa_2(\mathcal{P}_V^{-1}(A + \gamma M))$ is independent of the mesh size
  - Matrix inversions are approximated with one AMG V-cycle
- L:
  - CG with algebraic multigrid preconditioner

3D Maxwell Test Problems



3D box-shaped object



3D gear

Programming Language and External Packages

- Programming language: C++
- Iterative solvers: PETSc and Hypre libraries
- Mesh generator: Some meshes were generated with TetGen
- Mesh partition: METIS
- Parallelization: MPICH2

Maxwell Numerical Results: 3D Box

np	Nel	DOFs	its	its_{i_1}	its_{i_2}	t_s	t_a
3	3,693,263	4,738,111	18	31	6	3716.27s	38.99s
6	7,380,288	9,509,347	17	32	6	4186.47s	40.01s
9	14,760,128	19,082,621	17	33	7	4796.96s	41.92s

MINRES with $rtol = 1e-6$, preconditioned with $\mathcal{P}_{I,L}$

- its, its_{i1}, its_{i2}: iteration counts for outer iterations, $(1,1)$ and $(2,2)$ blocks of the inner iterations
- $t_{\{s,a\}}$: times (sec) for {solution, assembly}

Maxwell Numerical Results: 3D Gear

np	Nel	DOFs	its	its_{i_1}	its_{i_2}	t_s	t_a
4	596,011	685,736	8	47	5	126.04s	6.43s
8	1,046,568	1,245,761	8	42	5	293.87s	5.55s
16	1,965,622	2,398,793	8	42	5	292.65s	5.16s
32	3,802,327	4,725,385	8	42	5	346.04s	5.13s

MINRES with $rtol = 1e - 6$, preconditioned with $\mathcal{P}_{I,L}$

- its, its_{i1}, its_{i2}: iteration counts for outer iterations, $(1,1)$ and $(2,2)$ blocks of the inner iterations
- $t_{\{s,a\}}$: times (sec) for {solution, assembly}

Summary

- Solution of saddle point systems: an important theme in numerical linear algebra, optimization, solution of PDEs.
- Iterative solution taking center-stage due to increasing size of problems. Preconditioning is a must when it comes to iterative solvers.
- Block preconditioning is largely based on finding approximations to primal and dual Schur complements; nullity of $(1,1)$ block plays a role in augmentation preconditioning
- Interesting and tough problems and applications

THANK YOU!