Preconditioned iterative linear solvers for PDE-constrained Optimization problems

Andy Wathen
Oxford University

joint work with
Tyrone Rees, Martin Stoll, Sue Thorne and John Pearson

Let $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{n}}$ be symmetric and positive definite

$$\mathbf{min}_{\boldsymbol{x}} \ \ \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{b}$$

equivalent to solving

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

Constrained Optimization and Saddle-point problems:

$$\min_x \ \frac{1}{2}x^TAx - x^Tb$$

subject to $Bx = c$

Constrained Optimization and Saddle-point problems:

$$\min_x \ \frac{1}{2}x^TAx - x^Tb$$

subject to $Bx = c$

Lagrangian: $\mathcal{L}(x,\lambda) = \frac{1}{2}x^TAx - x^Tb + (Bx-c)^T\lambda$

λ : Lagrange multipliers.

$$\begin{array}{lcll} \min_x & \Rightarrow & Ax + B^T\lambda & = b \\ \max_\lambda & \Rightarrow & Bx & = c \end{array}$$

$$\Rightarrow \underbrace{\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}}_{\mathcal{A}} \underbrace{\begin{bmatrix} x \\ \lambda \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} b \\ c \end{bmatrix}}_{f}$$

(*Benzi, Golub & Liesen (2005)*)

Classic problem of this type: the Stokes problem

'Minimise energy subject to conserving mass' $\Rightarrow$

$$\begin{aligned} -\nabla^2 \vec{\mathbf{y}} + \nabla \mathbf{p} &= \vec{\mathbf{u}} \\ \text{subject to} \qquad \nabla \cdot \vec{\mathbf{y}} &= \mathbf{0} \end{aligned}$$

$\vec{\mathbf{y}}$: velocity, $\mathbf{p}$: pressure, $\vec{\mathbf{u}}$: body forces

Classic problem of this type: the Stokes problem

'Minimise energy subject to conserving mass' $\Rightarrow$

$$\begin{aligned} -\nabla^2 \vec{\mathbf{y}} + \nabla \mathbf{p} &= \vec{\mathbf{u}} \\ \text{subject to} \qquad \nabla \cdot \vec{\mathbf{y}} &= \mathbf{0} \end{aligned}$$

$\vec{\mathbf{y}}$: velocity, $\mathbf{p}$: pressure, $\vec{\mathbf{u}}$: body forces

Mixed finite elements/MAC finite difference approximation:

$$\begin{bmatrix} \underline{\boldsymbol{K}} & \boldsymbol{B}^{\boldsymbol{T}} \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{g} \end{bmatrix}, \quad \underline{\boldsymbol{K}} = \begin{bmatrix} \boldsymbol{K} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{K} \end{bmatrix} \text{ in } \mathbb{R}^{\mathbf{2}}$$

$\boldsymbol{K}$: discrete Laplacian

$$\mathcal{A} = \left[\begin{array}{cc} K & B^T \\ \hline B & 0 \end{array} \right]$$

is large, sparse, symmetric and indefinite $\rightarrow$ iterative solver

Leading contender: Krylov subspace method

- require 1 matrix$\times$vector multiplication at each iteration
- convergence depends only on eigenvalues

e.g. for symmetric positive definite problems, Conjugate Gradient method *(Hestenes & Stiefel (1952))*:

$$\frac{\|x - x_k\|_A}{\|x - x_0\|_A} \leq \min_{p \in \Pi_k, p(0)=1} \max_{\lambda \in \sigma(A)} |p(\lambda)|$$

but for symmetric (including indefinite) matrices, MINRES

$$\frac{\|r_k\|}{\|r_0\|} \leq \min_{p \in \Pi_k, p(0)=1} \max_{\lambda \in \sigma(\mathcal{A})} |p(\lambda)|$$

MINRES *(Paige & Saunders (1975))*:

$$\frac{\|r_k\|}{\|r_0\|} \leq \min_{p \in \Pi_k, p(0)=1} \max_{\lambda \in \sigma(\mathcal{A})} |p(\lambda)|$$

$$r_k = f - \mathcal{A} y_k$$

For fast convergence:

MINRES *(Paige & Saunders (1975))*:

$$\frac{\|\boldsymbol{r_k}\|}{\|\boldsymbol{r_0}\|} \leq \min_{\boldsymbol{p}\in\boldsymbol{\Pi_k},\boldsymbol{p(0)=1}} \max_{\boldsymbol{\lambda}\in\boldsymbol{\sigma(\mathcal{A})}} |\boldsymbol{p(\lambda)}|$$

$$\boldsymbol{r_k = f - \mathcal{A}y_k}$$

For fast convergence:

- few distinct eigenvalues
- cluster eigenvalues

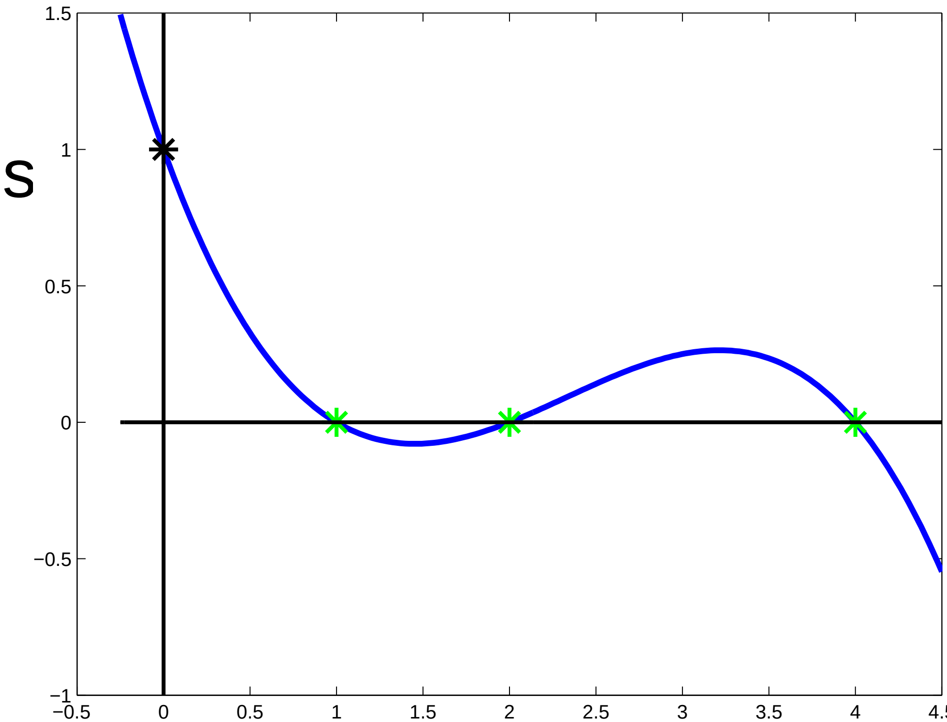
MINRES *(Paige & Saunders (1975))*:

$$\frac{\|r_k\|}{\|r_0\|} \leq \min_{p \in \Pi_k, p(0)=1} \max_{\lambda \in \sigma(\mathcal{A})} |p(\lambda)|$$

$$r_k = f - \mathcal{A} y_k$$

For fast convergence:

- few distinct eigenvalues
- cluster eigenvalues



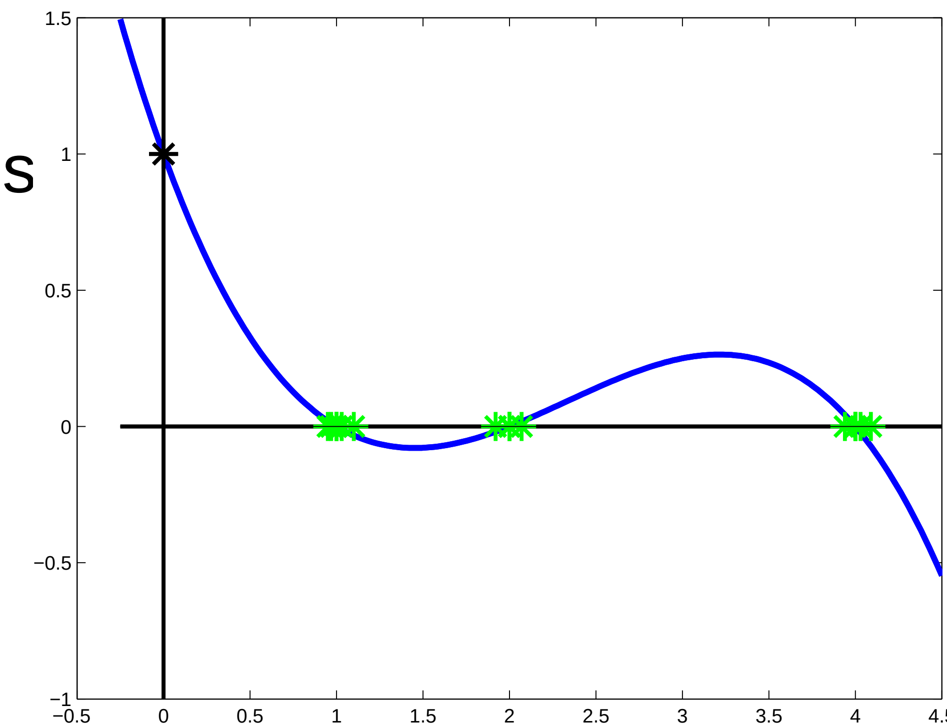
MINRES *(Paige & Saunders (1975))*:

$$\frac{\|r_k\|}{\|r_0\|} \leq \min_{p\in\Pi_k, p(0)=1} \max_{\lambda\in\sigma(\mathcal{A})} |p(\lambda)|$$

$$r_k = f - \mathcal{A}y_k$$

For fast convergence:

- few distinct eigenvalues
- cluster eigenvalues



How to achieve such properties?

How to achieve such properties?

Preconditioning: $\mathcal{A}u = f \quad \longrightarrow \quad \mathcal{P}^{-1}\mathcal{A}u = \mathcal{P}^{-1}f$

How to achieve such properties?

Preconditioning: $\boldsymbol{\mathcal{A}} \boldsymbol{u} = \boldsymbol{f} \quad \longrightarrow \quad \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{\mathcal{A}} \boldsymbol{u} = \boldsymbol{\mathcal{P}}^{-1} \boldsymbol{f}$

Block diagonal/triangular preconditioners

Block diagonal/triangular preconditioners:

based on observation (*Murphy, Golub & W (2000), Korzak(1999)*)

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}$$

preconditioned by

- $\begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix}$ has 3 distinct eigenvalues $(1, \frac{1}{2} \pm \frac{\sqrt{5}}{2})$
- $\begin{bmatrix} A & B^T \\ 0 & S \end{bmatrix}$ has 2 distinct eigenvalues

where $S = BA^{-1}B^T$ (Schur Complement)

$\Rightarrow$ MINRES /GMRES terminates in 3 / 2 iterations

$\Rightarrow$ want approximations $\widehat{A}$, $\widehat{S} \Rightarrow$ 3 / 2 clusters
$\Rightarrow$ fast convergence

For Stokes problem:

$\boldsymbol{A} = \underline{\boldsymbol{K}}$ is discrete Laplacians: use $\widehat{\boldsymbol{A}}$ =multigrid cycles

- geometric mutigrid: relaxed Jacobi smoothing, standard grid transfers
- algebraic multigrid: `HSL` routine `HSL_MI20`
  (Boyle, Mihajlovic & Scott (2007))

$\widehat{S}$ for Stokes problem:

Babuska-Brezzi stability $\Rightarrow$ Schur Complement spectrally equivalent to the finite element identity matrix, the mass matrix, M ie. the Gram matrix of the finite element basis functions $\{\phi_j, j = 1, \ldots, n\}$ in $L_2(\Omega), \Omega \subset \mathbb{R}^d$:

$$BA^{-1}B^T \approx \nabla \cdot (\nabla^2)^{-1} \nabla \approx Id \to M$$

Specifically $\exists \gamma > 0, \Gamma \leq \sqrt{d}$ such that

$$\gamma \leq \frac{x^T B A^{-1} B^T x}{x^T M x} \leq \Gamma$$

where $$M = \{m_{i,j}\}, \quad m_{i,j} = \int_\Omega \phi_i \phi_j$$

(*Bank, Welfert & Yserentant (1990), Silvester & W (1994)*)

Mass matrix is effectively preconditioned by its diagonal D=diag(M) *(W (1987))*

eg. for Q1 (bilinear) elements in 2-dimensions (rectangles)

eigenvalues of $D^{-1}M$ all in $[1/4, 9/4]$

eg. for Q1 (trilinear) finite elements in 3-dimensions (bricks)

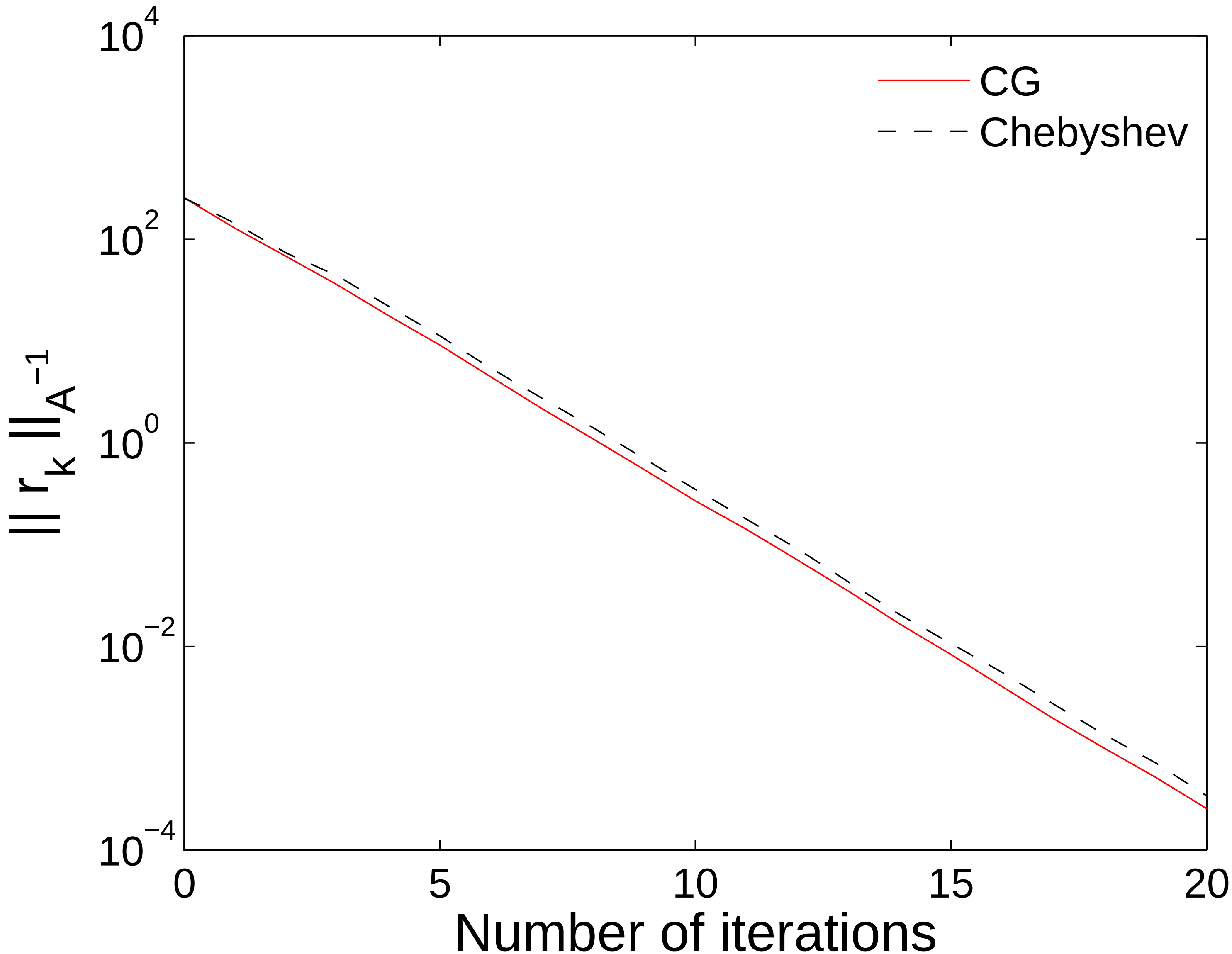
eigenvalues of $D^{-1}M$ all in $[1/8, 27/8]$

independently of mesh size h (ie. independently of discrete problem dimension)

Could use $\widehat{S} = D$ but better a few iterations of a diagonally preconditioned iteration for M:

- diagonally scaled Conjugate Gradients: leads to a nonlinear preconditioner
- diagonally scaled Chebyshev (semi-)iteration: is linear and we have precise eigenvalue inclusion intervals

(*Golub & Varga (1961), W & Rees (2008))*



CG
Chebyshev
10^4
10^2
10^0
10^{-2}
10^{-4}
$\| r_k \|_{A^{-1}}$
0
5
10
15
20
Number of iterations

For Stokes problem:

$$\boldsymbol{P} = \begin{bmatrix} \widehat{\boldsymbol{A}} & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{\boldsymbol{S}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{\boldsymbol{AMG}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_{\boldsymbol{20}}(\boldsymbol{D}^{-1}\boldsymbol{M}) \end{bmatrix}$$

Here: Q2-Q1 mixed finite element (Taylor-Hood),
$\widehat{\boldsymbol{A}} = \boldsymbol{A}_{\boldsymbol{AMG}}$: 1 AMG V-cycle

$\boldsymbol{h}$	Iterations	CPU time (s)
$\boldsymbol{2^{-2}}$ (187)	19	0.015
$\boldsymbol{2^{-3}}$ (659)	24	0.073
$\boldsymbol{2^{-4}}$ (2,467)	26	0.082
$\boldsymbol{2^{-5}}$ (9,539)	28	0.21
$\boldsymbol{2^{-6}}$ (37,507)	29	3.80
$\boldsymbol{2^{-7}}$ (148,739)	29	15.5

Results from Rene Schneider (Leeds/Chemnitz): P_2–P_1 on 1 processor of a cluster of Sun Fire 6800 with UltraSPARC II Cu 900MHz processors

		CPU time	
degrees of freedom	iterations	for solution	for setup
659	23	2.3e-2	1.5e-1
2467	25	6.4e-1	5.0e-2
9539	25	2.0	1.5e-1
37507	25	1.2e+1	5.9e-1
148739	22	6.9e+1	2.5
592387	24	3.5e+2	1.0e+1
2364419	23	1.5e+3	4.2e+1
9447427	24	6.7e+3	1.7e+2
37769219	24	2.7e+4	6.8e+2

PDE-constrained Optimization

General problem:

Given $\Omega \subset \mathbb{R}^2$ or $\mathbb{R}^3$, $\widehat{y} \in L_2(\Omega)$ as some desired state then for some (regularisation) parameter β

$$\min_{y,u} \ \frac{1}{2}\|y - \widehat{y}\|^2_{L_2(\Omega)} + \frac{\beta}{2}\|u\|^2_{L_2(\Omega)}$$

subject to

$$\mathcal{L}y = u \quad \text{in} \ \Omega, \qquad y = \widehat{y} \quad \text{on} \ \partial\Omega$$

where $\mathcal{L}$ represents a partial differential operator can also include:

bounds on the control (via a primal-dual active set strategy):

$$\underline{\mathbf{u}} \leq \mathbf{u} \leq \overline{\mathbf{u}}$$

(*Hintermueller, Ito & Kunisch (2002)*)

bounds on the state (via Moreau-Yoshida regularization):

$$\underline{\mathbf{y}} \leq \mathbf{y} \leq \overline{\mathbf{y}}$$

(*Herzog & Sachs (2009)*)

also boundary control...

Simple sample problem:

desirable $\widehat{\mathbf{y}}$, controlable body force $\mathbf{u}$

$$\min_{\mathbf{y},\mathbf{u}} \frac{1}{2}\|\mathbf{y}-\widehat{\mathbf{y}}\|^2_{L_2(\Omega)} + \frac{\beta}{2}\|\mathbf{u}\|^2_{L_2(\Omega)}$$

subject to

$$-\nabla^2\mathbf{y} = \mathbf{u} \quad \text{in } \Omega, \qquad \mathbf{y} = \widehat{\mathbf{y}} \quad \text{on } \partial\Omega$$

$$\min_{\mathbf{y},\mathbf{u}} \frac{1}{2}\|\mathrm{y}-\widehat{\mathrm{y}}\|^2 + \frac{\beta}{2}\|\mathbf{u}\|^2$$

$$\text{subject to} \qquad -\nabla^2 \mathrm{y} = \mathbf{u} \quad \text{in } \Omega, \qquad \mathrm{y} = \widehat{\mathrm{y}} \quad \text{on } \partial\Omega$$

Discretisation: finite elements

$$\mathrm{y}_h = \sum y_j \phi_j, \;\; y = (y_1, y_2, \ldots, y_n)^T$$
$$\mathbf{u}_h = \sum u_j \phi_j, \;\; u = (u_1, u_2, \ldots, u_n)^T$$

$$\min_{y,u} \;\; \frac{1}{2} y^T M y + y^T b + \frac{\beta}{2} u^T M u$$

$$\text{subject to} \qquad K y = M u + d$$

$M = \{m_{i,j}\}, m_{i,j} = \int_\Omega \phi_i \phi_j$ — mass matrix

$K = \{k_{i,j}\}, k_{i,j} = \int_\Omega \nabla\phi_i \cdot \nabla\phi_j$ — stiffness matrix as before so, Lagrangian:

$$\frac{1}{2}\boldsymbol{y}^T M \boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{b} + \frac{\beta}{2}\boldsymbol{u}^T M \boldsymbol{u} + \boldsymbol{\lambda}^T (K\boldsymbol{y} - M\boldsymbol{u} - \boldsymbol{d})$$

stationarity $\Rightarrow$ Saddle point system

$$\begin{bmatrix} M & 0 & K^T \\ 0 & \beta M & -M \\ K & -M & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{u} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{b} \\ 0 \\ \boldsymbol{d} \end{bmatrix}$$

$$\text{Note} \quad B = \begin{bmatrix} K & -M \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} M & 0 \\ 0 & \beta M \end{bmatrix}$$

in usual saddle point form

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}$$

With bound constraints Lagrangian:

$$\min_{y,u,\lambda} \frac{1}{2} y^T M y \quad + \quad y^T b + \frac{\beta}{2} u^T M u + \lambda^T (K y - M u - d)$$
$$+ \quad \underline{\mu}^T (\underline{u} - u) + \overline{\mu}^T (u - \overline{u})$$

with $\underline{\mu}, \overline{\mu} \geq 0$ and the complimentarity conditions

$$\underline{\mu}^T (\underline{u} - u) = 0 = \overline{\mu}^T (u - \overline{u})$$

Recall $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{K} & -\boldsymbol{M} \end{bmatrix}$ and $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{M} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta M} \end{bmatrix}$

so $\widehat{\boldsymbol{S}}$? $\boldsymbol{S} = \boldsymbol{B A^{-1} B^T}$ (Schur Complement)

$$= \begin{bmatrix} \boldsymbol{K} & -\boldsymbol{M} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\beta}\boldsymbol{M}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{K}^T \\ -\boldsymbol{M} \end{bmatrix}$$

$$= \frac{1}{\beta}\boldsymbol{M} + \boldsymbol{K M^{-1} K^T} = \frac{1}{\beta}\boldsymbol{M} + \boldsymbol{S}_1$$

$$\text{Recall} \quad B = \begin{bmatrix} K & -M \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} M & 0 \\ 0 & \beta M \end{bmatrix}$$

so $\widehat{S}$? $\quad S = BA^{-1}B^T$ (Schur Complement)

$$= \begin{bmatrix} K & -M \end{bmatrix} \begin{bmatrix} M^{-1} & 0 \\ 0 & \frac{1}{\beta} M^{-1} \end{bmatrix} \begin{bmatrix} K^T \\ -M \end{bmatrix}$$

$$= \frac{1}{\beta} M + K M^{-1} K^T = \frac{1}{\beta} M + S_1$$

$$= -\frac{2}{\sqrt{\beta}} K + \left(K + \frac{1}{\sqrt{\beta}} M\right) M^{-1} \left(K + \frac{1}{\sqrt{\beta}} M\right)^T = -\frac{2}{\sqrt{\beta}} K + S_2$$

Eigenvalues:

$$\lambda(S_1^{-1} S) \in \left[1 + ch^4/\beta, 1 + C/\beta\right], \quad \lambda(S_2^{-1} S) \in [1/2, 1]$$

So choose $\widehat{S} = S_1$ for all except small β or $\widehat{S} = S_2$ for all β

(*Pearson & W (2010), Schoberl & Zulehner (2007)*)

Hence (with $\widehat{S} = S_1$) preconditioner for

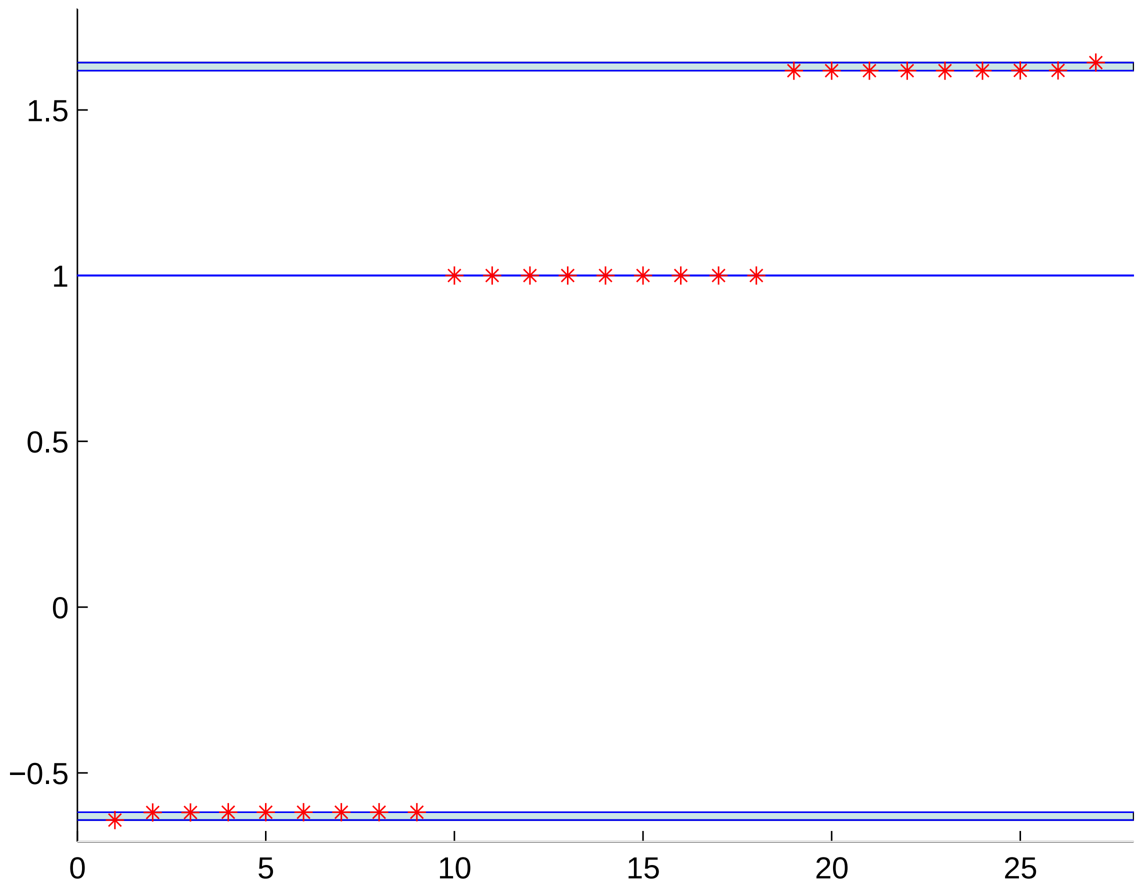
$$\mathcal{A} = \begin{bmatrix} M & 0 & K^T \\ 0 & \beta M & -M \\ K & -M & 0 \end{bmatrix} \text{ is } \mathcal{P} = \begin{bmatrix} M & 0 & 0 \\ 0 & \beta M & 0 \\ 0 & 0 & KM^{-1}K^T \end{bmatrix}$$

Eigenvalues ν of $\mathcal{P}^{-1}\mathcal{A}$

$$\begin{aligned} \nu &= 1, \\ \frac{1}{2}\left(1+\sqrt{5+\frac{2\alpha_1 h^4}{\beta}}\right) \le \ &\nu \ \le \frac{1}{2}\left(1+\sqrt{5+\frac{2\alpha_2}{\beta}}\right) \\ \text{or } \frac{1}{2}\left(1-\sqrt{5+\frac{2\alpha_2}{\beta}}\right) \le \ &\nu \ \le \frac{1}{2}\left(1-\sqrt{5+\frac{2\alpha_1 h^4}{\beta}}\right), \end{aligned}$$

where $\alpha_1, \ \alpha_2$ are positive constants independent of h.

$$\mathcal{P} = \begin{bmatrix} M & 0 & 0 \\ 0 & \beta M & 0 \\ 0 & 0 & KM^{-1}K^T \end{bmatrix}, \quad \beta = 10^{-2}$$



But

$$\mathcal{P} = \begin{bmatrix} \boldsymbol{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta\boldsymbol{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{K}\boldsymbol{M}^{-1}\boldsymbol{K}^T \end{bmatrix}$$

still expensive to use in practice so employ approximations

$$\widehat{\boldsymbol{M}} \simeq \boldsymbol{M} \quad \text{and} \quad \widehat{\boldsymbol{K}} \simeq \boldsymbol{K}$$

giving

$$\mathcal{P} = \begin{bmatrix} \widehat{\boldsymbol{M}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta\widehat{\boldsymbol{M}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \widehat{\boldsymbol{K}}\boldsymbol{M}^{-1}\widehat{\boldsymbol{K}^T} \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{A}} & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{S}} \end{bmatrix}$$

Important subtlety:

$$\widehat{\boldsymbol{K}} \simeq \boldsymbol{K}$$

does *not* imply that

$$\widehat{\boldsymbol{K}} M^{-1} \widehat{\boldsymbol{K}^{\boldsymbol{T}}} \simeq \boldsymbol{K} M^{-1} \boldsymbol{K}^{\boldsymbol{T}}$$

or indeed that

$$\widehat{\boldsymbol{K}} \widehat{\boldsymbol{K}^{\boldsymbol{T}}} \simeq \boldsymbol{K} \boldsymbol{K}^{\boldsymbol{T}}$$

without further conditions which are satisfied in this case
(Braess & Peisker (1986))

$$\mathcal{P} = \begin{bmatrix} \widehat{M} & 0 & 0 \\ 0 & \beta\widehat{M} & 0 \\ 0 & 0 & \widehat{K}M^{-1}\widehat{K^T} \end{bmatrix} = \begin{bmatrix} \widehat{A} & 0 \\ 0 & \widehat{S} \end{bmatrix}$$

so $\widehat{M}$?: $T_{20}(D^{-1}M)$ as before

For $\mathcal{L} = -\nabla^2$, K is a discrete Laplacian: use multigrid cycles as before

In our examples:

$\widehat{K}$ is the action of 2 V-cycles

$\widehat{M}$ is the action of 20 Chebyshev semi-iterative steps

Example problem: $\Omega = [0,1]^d, d = 2, 3$

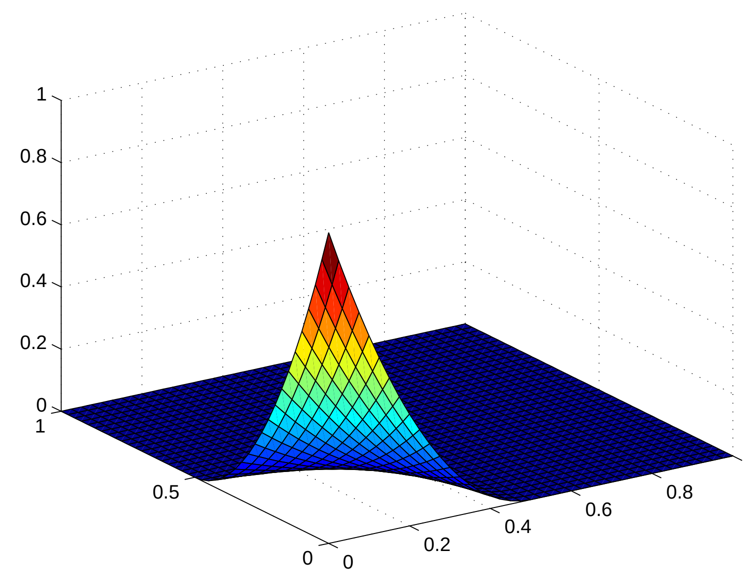
$$\min_{\mathbf{y},\mathbf{u}} \frac{1}{2}\|\mathbf{y} - \widehat{\mathbf{y}}\|^2_{L_2(\Omega)} + \frac{\beta}{2}\|\mathbf{u}\|^2_{L_2(\Omega)}$$

subject to

$$-\nabla^2 \mathbf{y} = \mathbf{u} \quad \text{in } \Omega, \qquad \mathbf{y} = \widehat{\mathbf{y}} \quad \text{on } \partial\Omega$$

Q1 (bilinear) finite elements, $\beta = 10^{-2}$

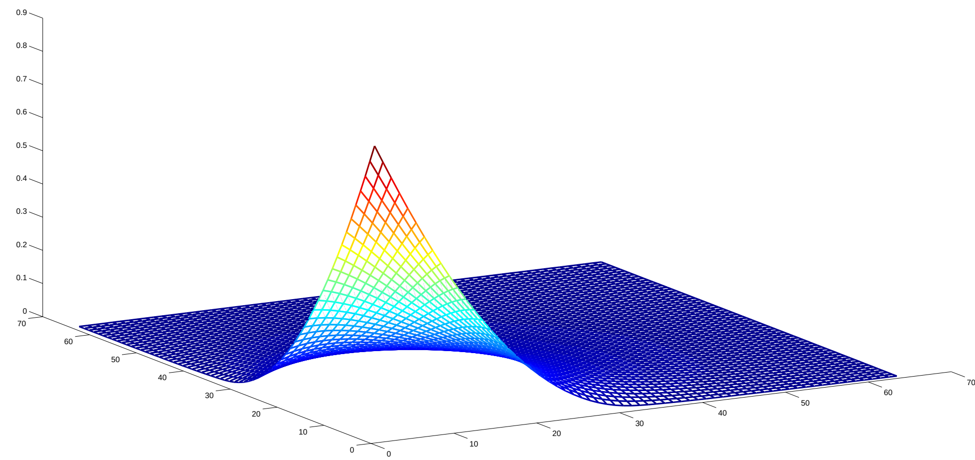
$\widehat{\mathbf{y}}$:



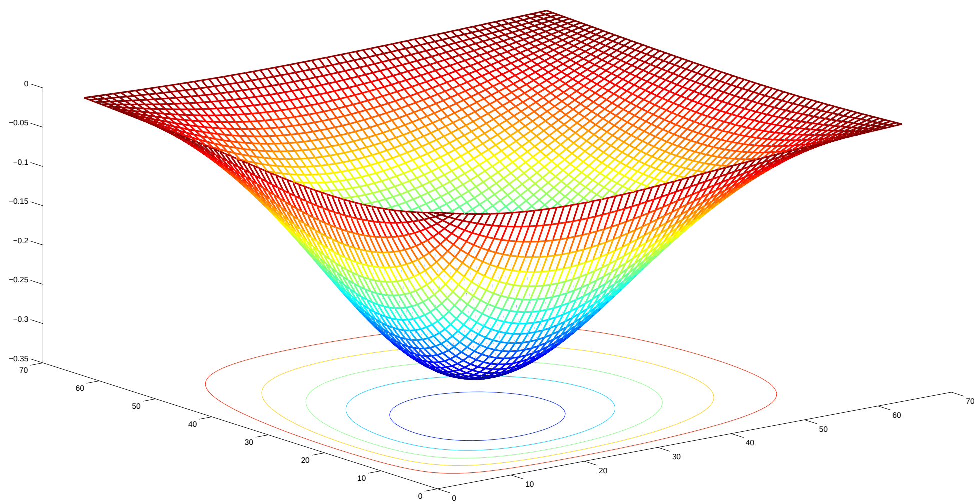
CPU times (MINRES iterations) in 2D, tol 10^{-6}

h	3n	backslash	MINRES ($\mathcal{P}_{AMG}$)	MINRES ($\mathcal{P}_{MG}$)
2^{-2}	27	0.0003	0.02 (7)	0.13 (7)
2^{-3}	147	0.002	0.03 (9)	0.16 (9)
2^{-4}	675	0.01	0.05 (9)	0.21 (9)
2^{-5}	2883	0.08	0.14 (9)	0.41 (9)
2^{-6}	11907	0.46	0.61 (9)	1.29 (9)
2^{-7}	48387	3.10	2.61 (9)	5.09 (9)
2^{-8}	195075	15.5	15.0 (11)	23.6 (9)
2^{-9}	783363	—	75.6 (11)	136 (9)

State:



Control:



CPU times (MINRES iterations) in 3D, tol 10^{-6}

h	3n	backslash	MINRES ($\mathcal{P}_{AMG}$)	MINRES ($\mathcal{P}_{MG}$)
2^{-2}	81	0.001	0.02 (7)	0.14 (8)
2^{-3}	1029	0.013	0.13 (9)	0.26 (8)
2^{-4}	10125	25.1	1.89 (8)	1.69 (8)
2^{-5}	89373	—	22.1 (8)	15.9 (8)
2^{-6}	750141	—	—	230 (10)

CPU times (MINRES iterations) in 3D, tol 10^{-6}

h	3n	backslash	MINRES ($\mathcal{P}_{AMG}$)	MINRES ($\mathcal{P}_{MG}$)
2^{-2}	81	0.001	0.02 (7)	0.14 (8)
2^{-3}	1029	0.013	0.13 (9)	0.26 (8)
2^{-4}	10125	25.1	1.89 (8)	1.69 (8)
2^{-5}	89373	—	22.1 (8)	15.9 (8)
2^{-6}	750141	—	—	230 (10)

	Approximation $S = S_1$ β				Approximation $S = S_2$ β			
h	10^{-1}	10^{-3}	10^{-5}	10^{-7}	10^{-1}	10^{-3}	10^{-5}	10^{-7}
2^{-2}	8	12	26	28	10	14	–	–
2^{-3}	8	12	42	130	10	16	14	–
2^{-4}	8	12	48	272	12	17	15	13
2^{-5}	10	14	49	341	12	18	16	16

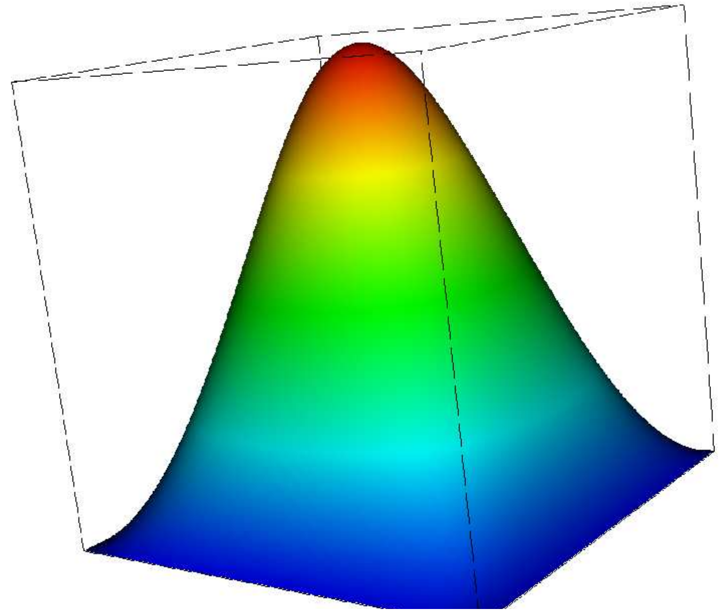
With bound constraints: an active set strategy (or projected gradient) $\Rightarrow$ an outer nonlinear loop

MINRES solution: Number of AMG V-cycles for Laplacian in 2D, tol 10^{-6}

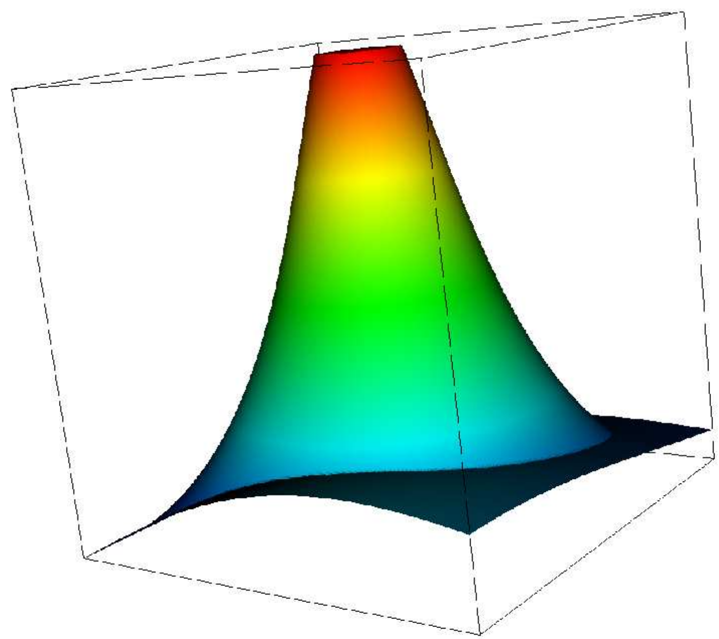
h	3n	PDE solve	Unconstrained Control prob	Bound-constrained Control prob
2^{-2}	27	4	40	68(2)
2^{-3}	147	4	40	104(3)
2^{-4}	675	4	40	160(4)
2^{-5}	2883	6	40	160(4)
2^{-6}	11907	6	44	180(4)
2^{-7}	48387	6	48	240(5)
2^{-8}	195075	6	48	280(5)
2^{-9}	783363	6	56	320(5)

energy: unconstrained $\searrow 1.483 \times 10^{-3}$
bound constrained $\searrow 1.732 \times 10^{-3}$

State:

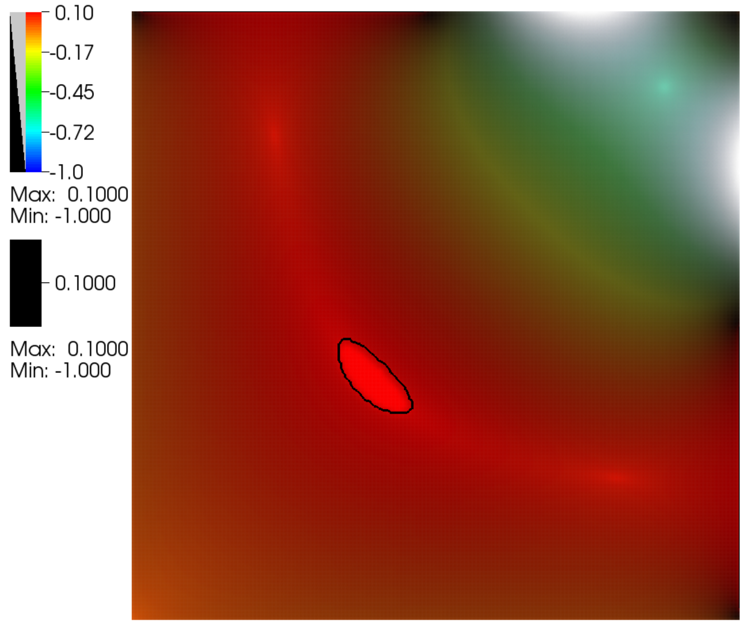


Control:

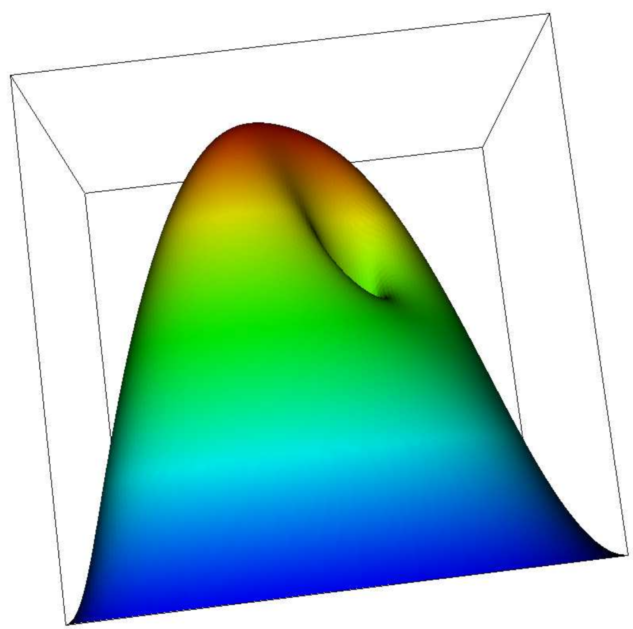


State constraints

State:



Control:



Stokes Control

$$\min_{\mathbf{y},\mathbf{p},\mathbf{u}} \frac{1}{2}\|\vec{\mathbf{y}} - \hat{\mathbf{y}}\|^2_{L^2(\Omega)} + \frac{1}{2}\|\mathbf{p} - \hat{\mathbf{p}}\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|\mathbf{u}\|^2_{L^2(\Omega)}$$

subject to
$$\begin{aligned} -\nabla^2\vec{\mathbf{y}} + \nabla\mathbf{p} &= \mathbf{u} \\ \nabla\cdot\vec{\mathbf{y}} &= 0 \end{aligned}$$

$\vec{\mathbf{y}}$: velocity, $\mathbf{p}$: pressure.

Mixed finite elements for (forward) Stokes problem:

$$\begin{bmatrix} \underline{K} & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} y \\ p \end{bmatrix} = \begin{bmatrix} u \\ g \end{bmatrix}, \; \underline{K} = \begin{bmatrix} K & 0 \\ 0 & K \end{bmatrix} \text{ in } \mathbb{R}^2$$

Cost functional

$$\frac{1}{2} y^T M_y y - y^T b + \frac{1}{2} p^T M_p p - p^T d + \frac{\beta}{2} u^T M_u u$$

combined with constraint via the Lagrangian $\Rightarrow$

$$\begin{bmatrix} M_y & 0 & 0 & \underline{K} & B^T \\ 0 & M_p & 0 & B & 0 \\ 0 & 0 & \beta M_u & -M_u & 0 \\ \underline{K} & B^T & -M_u & 0 & 0 \\ B & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ p \\ u \\ \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} b \\ d \\ 0 \\ h \\ k \end{bmatrix}.$$

Schur Complement:

$$\begin{bmatrix} \underline{\boldsymbol{K}} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{M_y^{-1}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M_p^{-1}} \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{K}} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{0} \end{bmatrix} + \frac{\boldsymbol{1}}{\boldsymbol{\beta}} \boldsymbol{M_u}$$

and again ignore 2nd term for moderate $\boldsymbol{\beta}$ (or use $\boldsymbol{S_2}$?)

So overall block diagonal preconditioner requires:

$\widehat{\boldsymbol{M_y}}, \widehat{\boldsymbol{M_p}}, \widehat{\boldsymbol{M_u}} \rightarrow$ Chebyshev

and 2 Stokes approximations

Stokes preconditioners:

$$\widehat{\begin{bmatrix} \underline{\boldsymbol{K}} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{0} \end{bmatrix}} = \begin{bmatrix} \widehat{\underline{\boldsymbol{K}}} & \boldsymbol{0} \\ \boldsymbol{B} & \widehat{\boldsymbol{M}_p} \end{bmatrix}$$

and

$$\widehat{\begin{bmatrix} \underline{\boldsymbol{K}} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{0} \end{bmatrix}} = \begin{bmatrix} \widehat{\underline{\boldsymbol{K}}} & \boldsymbol{B}^T \\ \boldsymbol{0} & \widehat{\boldsymbol{M}_p} \end{bmatrix}$$

on left and right respectively where $\widehat{\underline{\boldsymbol{K}}}$ is multigrid cycles for each discrete scalar Laplacian as before
(*Silvester & W (1993), Klawonn (1998), Elman, Silvester & W (2005)*)

Gives symmetric Schur complement approximation

Here: Q2-Q1 mixed finite elements for cavity flow

4 AMG V-cycles to approx each $\boldsymbol{K}$

20 Chebyshev semi-iterations for each $\boldsymbol{M}$

$\boldsymbol{h}$	Iterations	CPU time (s)
$\mathbf{2^{-2}}$ (344)	26	0.48
$\mathbf{2^{-3}}$ (1,512)	31	1.05
$\mathbf{2^{-4}}$ (6,344)	33	3.69
$\mathbf{2^{-5}}$ (25,992)	33	18.0
$\mathbf{2^{-6}}$ (105,224)	34	84.2
$\mathbf{2^{-7}}$ (423,432)	34	342

recall: 29 MINRES iterations for forward Stokes solve (but only 1 AMG V-cycle) $\Rightarrow$ approx 10 times more work to solve the control problem than a single PDE solve.

Time-dependent problems

$$\frac{1}{2}\int_0^T\int_{\Omega_1}(\mathrm{y}(\mathrm{x},\mathrm{t})-\bar{\mathrm{y}}(\mathrm{x},\mathrm{t}))^2\,dxdt+\frac{\beta}{2}\int_0^T\int_{\Omega_2}(\mathrm{u}(\mathrm{x},\mathrm{t}))^2\,dxdt$$

subject to

$$y_t-\nabla^2 y=u \qquad \text{in } \Omega\times[0,T]$$

with boundary conditions $y=0$ on $\partial\Omega$ and initial condition $y(x,0)=y_0(x)$.

Adjoint PDE:

$$-p_t-\nabla^2 p=y-\bar{y}$$

with $p=0$ on $\partial\Omega, p(x,T)=y(x,T)-\bar{y}(x,T)$.

Backwards Euler in time, Galerkin finite elements in space

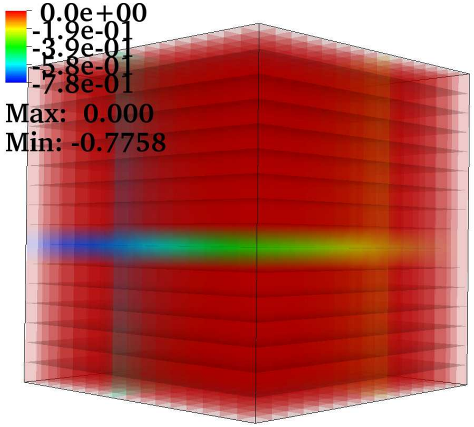
Main issue now is that differential operator is $\underline{K}$ where *all* time step values $\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_N$ are involved:

$$\underbrace{\begin{bmatrix} M+\tau K & & & & \\ -M & M+\tau K & & & \\ & -M & M+\tau K & & \\ & & \ddots & \ddots & \\ & & & -M & M+\tau K \end{bmatrix}}_{\underline{K}} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}$$

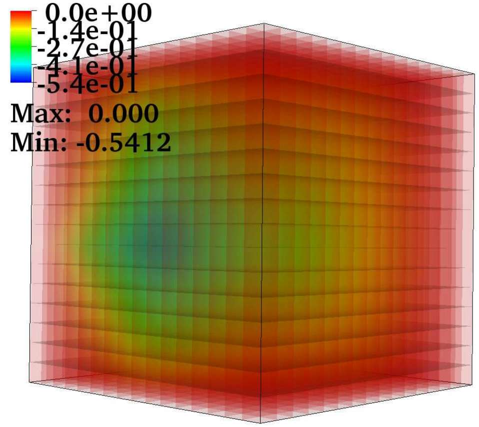
so

$$\begin{bmatrix} \underline{M} & 0 & -\underline{K}^T \\ 0 & \beta\underline{M} & \underline{M} \\ -\underline{K} & \underline{M} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \underline{M}\mathbf{y} \\ 0 \\ d \end{bmatrix}.$$

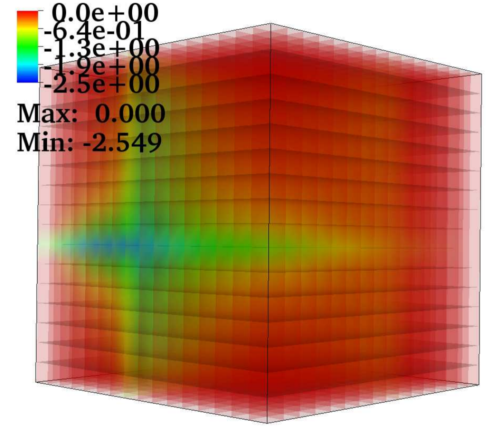
is an even larger system, but similar ideas go through...



Desired state



State at $t = 1$



Control at $t = 1$

References

Rees, T., Dollar, H.S. & Wathen, A.J., 'Optimal solvers for PDE-constrained Optimization', SISC (2010)

Rees, T. & Wathen, A.J., 'Preconditioned iterative solvers for the optimal control of the Stokes equations', Oxford NA Report 10/04.

Stoll, M. & Wathen, A.J., 'All-at once solution of time-dependent PDE-constrained Optimization problems', Oxford NA Report 10/13

Murphy, M.F., Golub, G.H. & Wathen, A.J., 'A note on preconditioning for indefinite linear systems', SISC (2000)

Pearson, J.W. & Wathen, A.J., 'A new approximation of the Schur complement in preconditioners for PDE-constrained Optimization', Oxford NA report 10/14

Wathen, A.J. & Rees, T., 2009, 'Chebyshev semi-iteration in preconditioning', Victor Pereyra special issue of ETNA.

Benzi, M, Golub, G.H. & Liesen, J., 2005, 'Numerical Solution of Saddle Point Problems', Acta Numerica 2005.

Acknowledgements

This work is partially supported by Award No. KUK-C1-013-04 made by King Abdullah University of Science and Technology (KAUST)