Preconditioning PDE-constrained Optimization Problems

Andy Wathen
Oxford University, UK

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

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

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

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

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

Discretisation: finite elements

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

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

$$\text{subject to} \qquad Ky = Mu + d$$

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

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

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

stationarity $\Rightarrow$ Saddle point system

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

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

in usual saddle point form

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

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

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

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

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

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

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

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 $\mathcal{P}^{-1}\mathcal{A}$

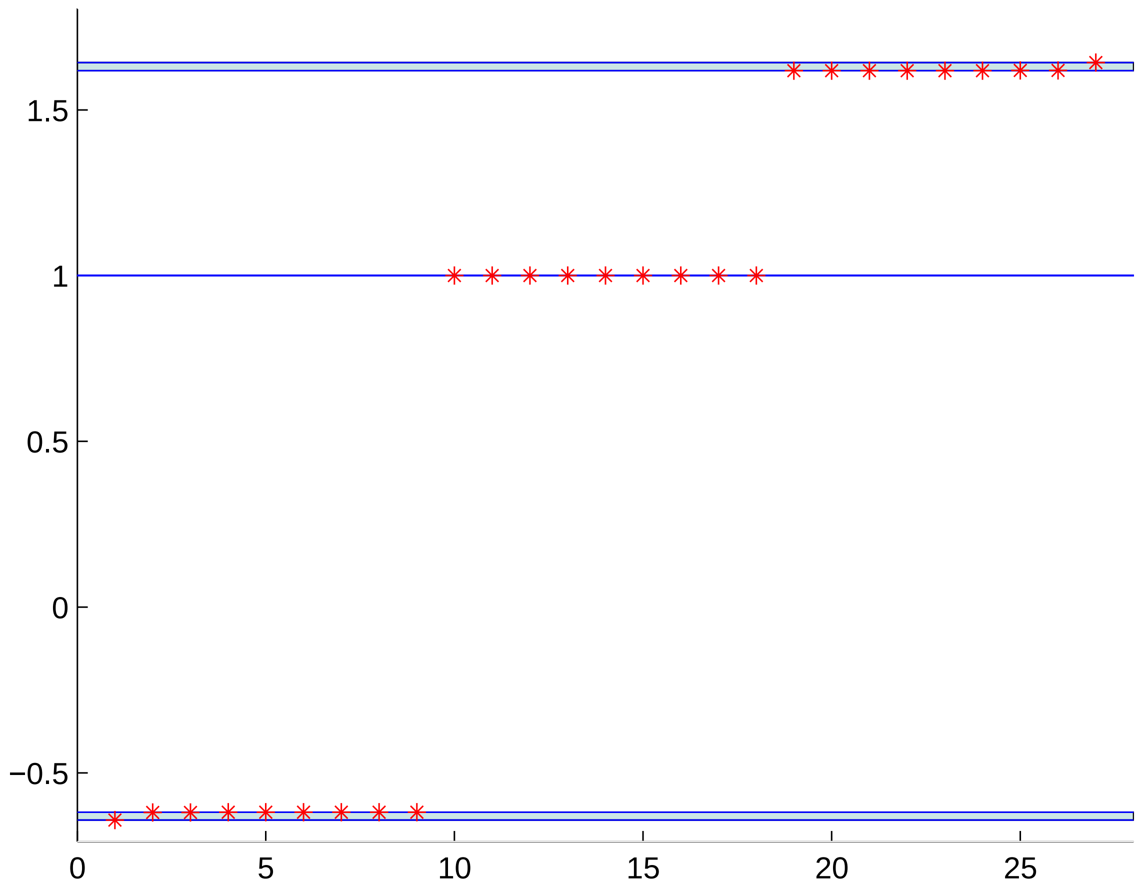
$$\nu = 1,$$

$$\frac{1}{2}\left(1 + \sqrt{5 + \frac{2\alpha_1 h^4}{\beta}}\right) \leq \nu \leq \frac{1}{2}\left(1 + \sqrt{5 + \frac{2\alpha_2}{\beta}}\right)$$

$$\text{or } \frac{1}{2}\left(1 - \sqrt{5 + \frac{2\alpha_2}{\beta}}\right) \leq \nu \leq \frac{1}{2}\left(1 - \sqrt{5 + \frac{2\alpha_1 h^4}{\beta}}\right),$$

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

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

or indeed that

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

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 $\widehat{M}$?: $T_{20}(D^{-1}M)$: Chebyshev (semi-) iterations

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

- geometric mutigrid: relaxed Jacobi smoothing, standard grid transfers
  $\rightarrow \mathcal{P}_{MG}$
- algebraic multigrid: `HSL` routine `HSL_MI20`
  (Boyle, Mihajlovic & Scott (2007))
  $\rightarrow \mathcal{P}_{AMG}$

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

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

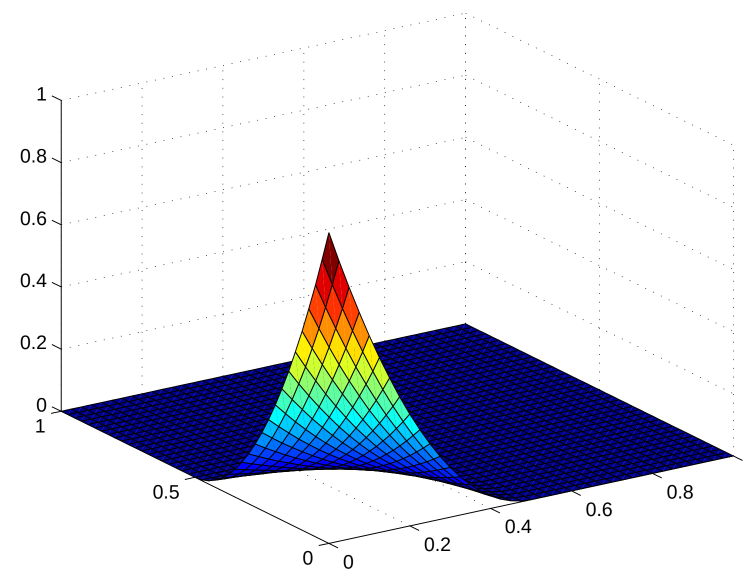
$$\min_{\mathbf{y},\mathbf{u}} \frac{1}{2}\|\mathbf{y} - \widehat{\mathbf{y}}\|^2_{L_2(\Omega)} + \frac{\beta}{2}\|\mathbf{u}\|^2_{L_2(\Omega)}$$

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, $\boldsymbol{\beta = 10^{-2}}$

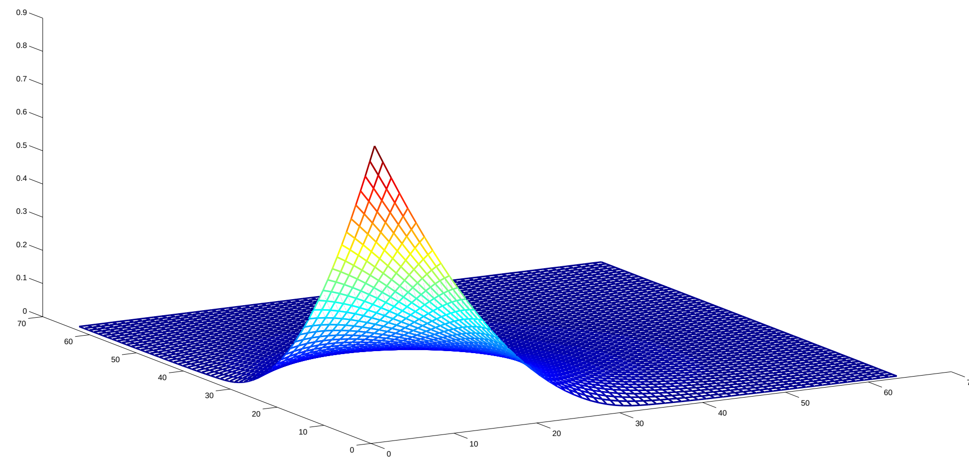
$\widehat{\mathbf{y}}$:



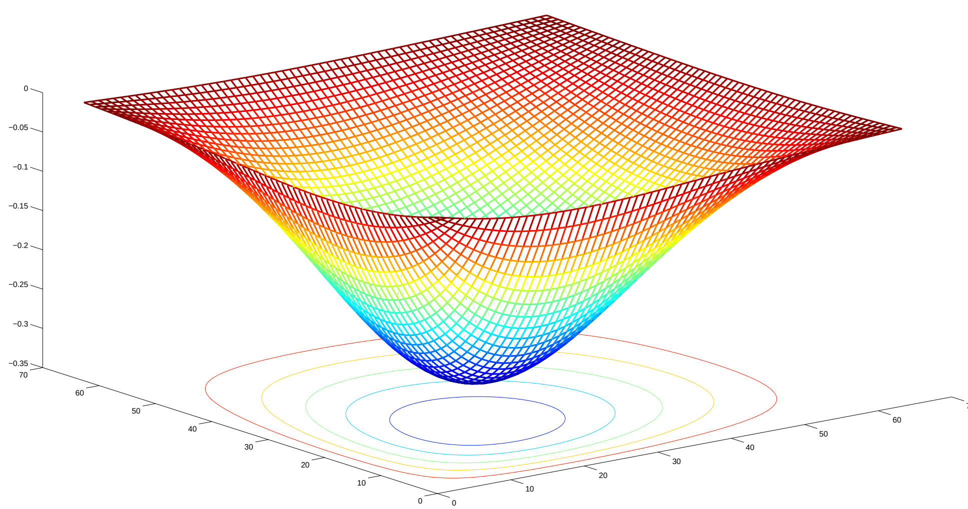
CPU times (MINRES iterations) in 2D, tol 10^{-6}

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	—	297 (9)	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	—	297 (9)	230 (10)

h	Approximation $S = S_1$ β				Approximation $S = S_2$ β			
	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

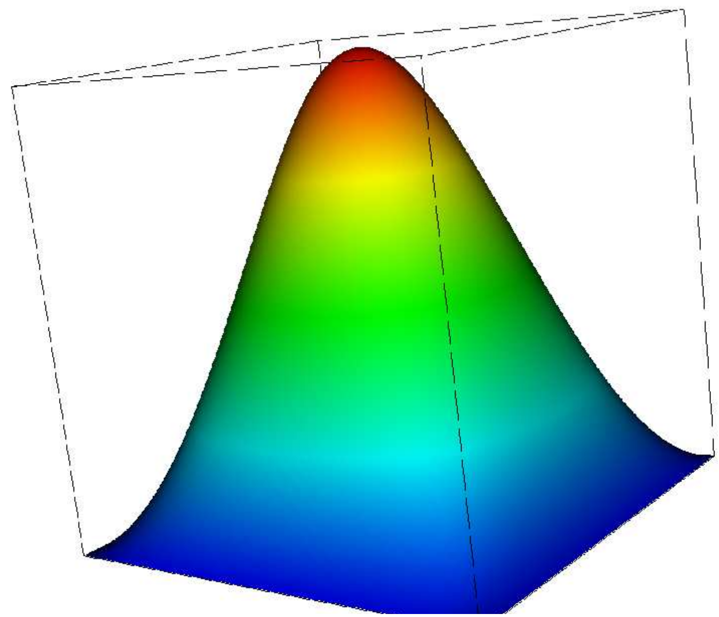
With bound constraints: an active set strategy (or projected gradient) $\Rightarrow$ an outer nonlinear loop

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

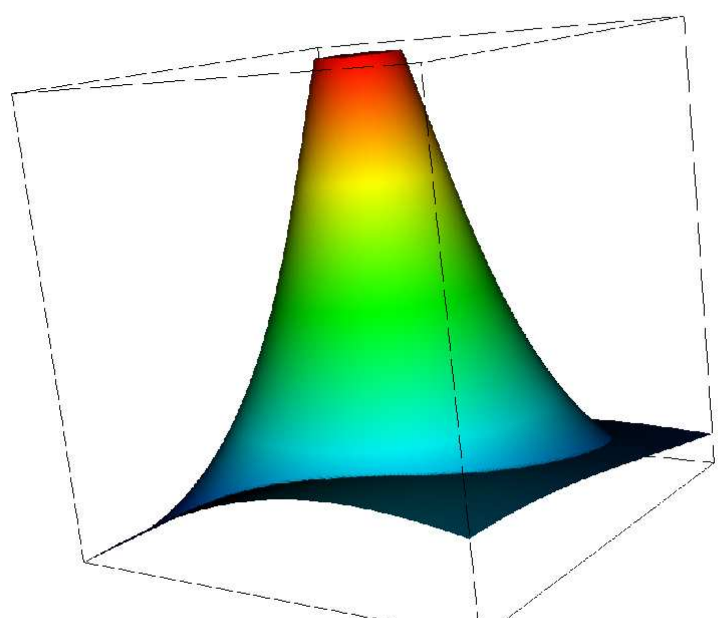
h	3n	PDE solve	Unconstrained Control prob	Bound-constrained Control prob
2^{-2}	27	4	40	68(2)
2^{-3}	147	4	40	104(3)
2^{-4}	675	4	40	160(4)
2^{-5}	2883	6	40	160(4)
2^{-6}	11907	6	44	180(4)
2^{-7}	48387	6	48	240(5)
2^{-8}	195075	6	48	280(5)
2^{-9}	783363	6	56	320(5)

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

State:

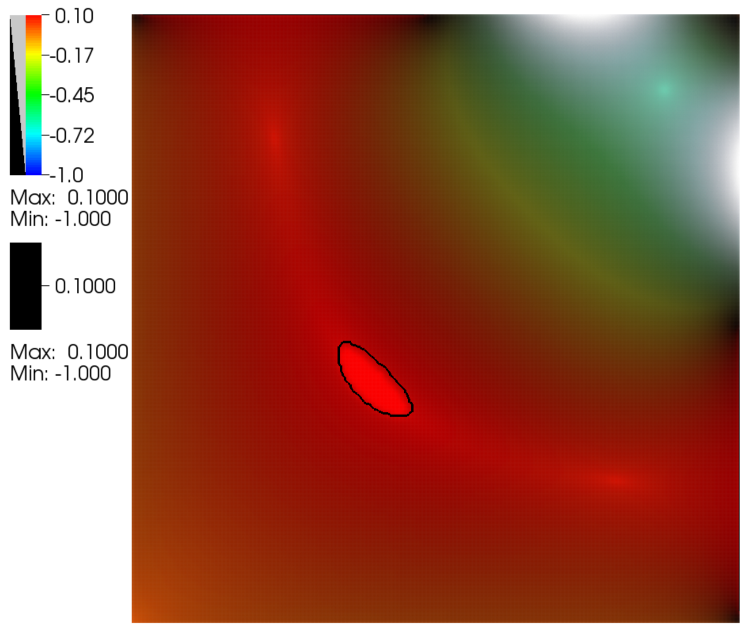


Control:

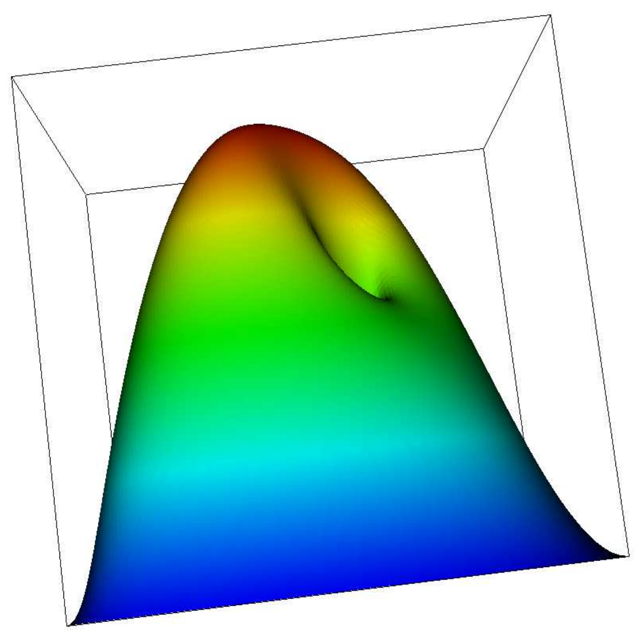


State constraints

State:



Control:



Stokes Control

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

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

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

Mixed finite elements for (forward) Stokes problem:

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

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:

$$\begin{bmatrix} \underline{\boldsymbol{K}} & \boldsymbol{B}^T \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_{\boldsymbol{y}}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{M}_{\boldsymbol{p}}^{-1} \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{K}} & \boldsymbol{B}^T \\ \boldsymbol{B} & \mathbf{0} \end{bmatrix} + \frac{\mathbf{1}}{\boldsymbol{\beta}} \boldsymbol{M}_{\boldsymbol{u}}$$

and again ignore 2nd term for moderate $\boldsymbol{\beta}$ (or use $\boldsymbol{S}_{\mathbf{2}}$)

So overall block diagonal preconditioner requires:

$\widehat{\boldsymbol{M}_{\boldsymbol{y}}}, \widehat{\boldsymbol{M}_{\boldsymbol{p}}}, \widehat{\boldsymbol{M}_{\boldsymbol{u}}} \rightarrow$ Chebyshev

and 2 Stokes approximations

Stokes preconditioners:

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

and

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

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 K

20 Chebyshev semi-iterations for each M

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

It takes 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

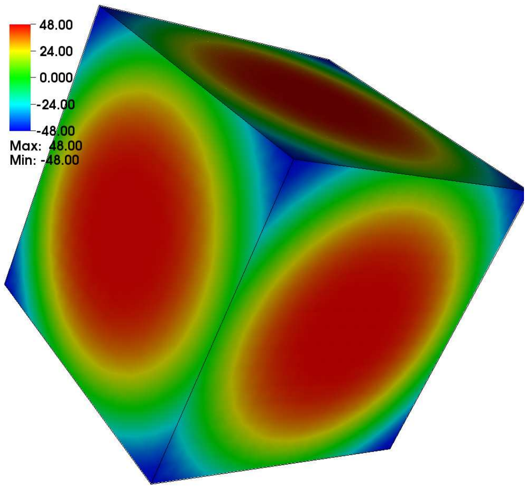
Main issue now is that differential operator is $\underline{K}$ where *all* time step values $\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_N$ are involved:

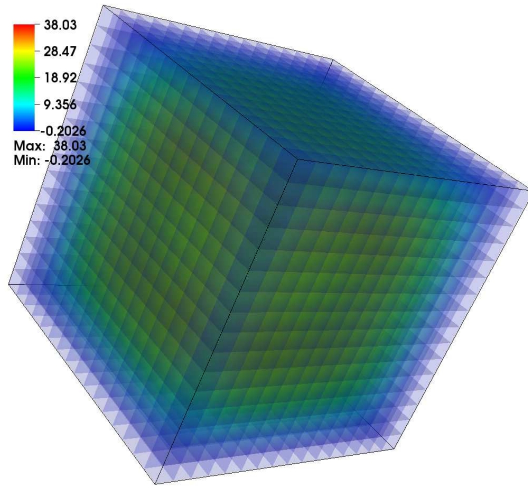
$$\underbrace{\begin{bmatrix} M+\tau K & & & & \\ -M & M+\tau K & & & \\ & -M & M+\tau K & & \\ & & \ddots & \ddots & \\ & & & -M & M+\tau K \end{bmatrix}}_{\underline{K}} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}$$

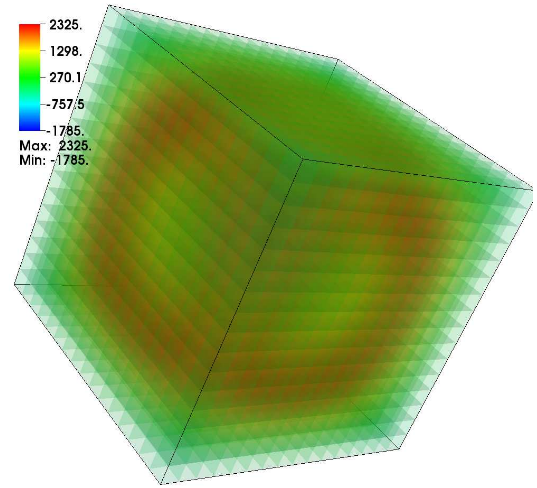
so

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

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







Desired state | State | Control

Solution at 15th time step, $\beta = 10^{-4}$

References

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

Rees, T. & Wathen, A.J., 'Preconditioned iterative methods for the optimal control of the Stokes equations', SISC (2011)

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', NLAA (2012)

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

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

Acknowledgement

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