PDE-constrained optimisation: from linear to nonlinear constraints

Sue Thorne

STFC Rutherford Appleton Laboratory

In collaboration with

- Nick Gould (STFC Rutherford Appleton Laboratory)
- Dominique Orban (Ecole Polytechnique de Montreal)
- Tyrone Rees (University of British Columbia)
- Valeria Simoncini (Università di Bologna)
- Andy Wathen (University of Oxford)

PDE-constrained optimization

- Given u and boundary conditions g, calculate y, where

$$\mathcal{L}y = u, \quad \alpha_1 y + \alpha_2 \frac{\partial y}{\partial n} = g \text{ on } \partial\Omega$$

on some domain Ω

PDE-constrained optimization

- Given u and boundary conditions g, calculate y, where

$$\mathcal{L}y = u, \quad \alpha_1 y + \alpha_2 \frac{\partial y}{\partial n} = g \text{ on } \partial\Omega$$

  on some domain Ω

- Suppose given g and a target $\hat{y}$ on some domain $\hat{\Omega} \subset \Omega$. Want to calculate u such that $y \approx \hat{y}$: distributed control

PDE-constrained optimization

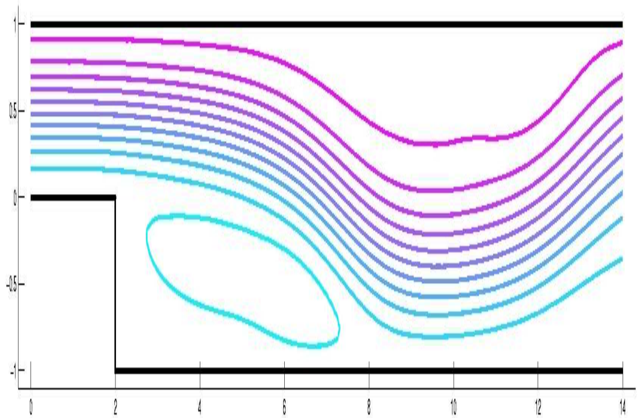
- Given u and boundary conditions g, calculate y, where

$$\mathcal{L}y = u, \quad \alpha_1 y + \alpha_2 \frac{\partial y}{\partial n} = g \text{ on } \partial\Omega$$

  on some domain Ω

- Suppose given g and a target $\hat{y}$ on some domain $\hat{\Omega} \subset \Omega$. Want to calculate u such that $y \approx \hat{y}$: distributed control
- Suppose given u and a target $\hat{y}$ on some domain $\hat{\Omega} \subset \Omega$. Want to calculate g such that $y \approx \hat{y}$: boundary control

Different target temperatures



Reduce recirculation

Distributed control

$$\min_{y,u} \frac{1}{2} \left\| \omega(x) \left(y - \widehat{y} \right) \right\|_2^2 + \beta \left\| u \right\|_2^2$$

subject to

$$\begin{aligned} \mathcal{L}y &= u \text{ in } \Omega \\ y &= \widehat{y} \text{ on } \delta\Omega \end{aligned}$$

Here

$$\omega(x) = \begin{cases} 1 & x \in \hat{\Omega} \\ 0 & \text{otherwise} \end{cases}$$

Distributed control

Discretize:

$$\mathtt{y_h} = \sum y_j \phi_j, \quad \mathtt{u_h} = \sum u_j \phi_j$$

$$\min_{\mathtt{y_h},\mathtt{u_h}} \frac{1}{2} \|\omega(x)\,(\mathtt{y_h} - \widehat{y})\|_2^2 + \beta\,\|\mathtt{u_h}\|_2^2$$

subject to

$$\begin{aligned} \mathcal{L}\mathtt{y_h} &= \mathtt{u_h} \text{ in } \Omega \\ \mathtt{y_h} &= g \text{ on } \delta\Omega \end{aligned}$$

Let $\mathcal{L} = -\nabla^2$

Distributed control

$$
\begin{aligned}
\|\mathtt{u_h}\|_2^2 &= \int_\Omega \mathtt{u_h}^2 \\
&= \sum_i \sum_j u_i u_j \int_\Omega \phi_i \phi_j \\
&= u^T M u, \\
\frac{1}{2}\|\omega(x)\,(\mathtt{y_h} - \widehat{y})\|_2^2 &= \frac{1}{2}\int_\Omega \omega(x)\,(\mathtt{y_h} - \widehat{y})^2 \\
&= \frac{1}{2}\sum_i \sum_j y_i y_j \int_\Omega \omega_i \omega_j \phi_i \phi_j - 2\sum_j y_j \int_\Omega \omega_j \phi_j \widehat{y} + \frac{1}{2}\int_{\hat{\Omega}} \widehat{y}^2 \\
&= \frac{1}{2} y^T \bar{M} y - y^T b + c, \\
Ky &= Mu + d,
\end{aligned}
$$

where M is the mass matrix, K is the stiffness matrix, $\bar{M} = WMW$ and $W = \mathsf{diag}(\omega_i)$

Distributed control

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

subject to

$$Ky - Mu \quad = \quad d$$

Distributed control

$$\min_{y,u} \frac{1}{2} y^T \bar{M} y - y^T b + c + \beta u^T M u + l^T \left(K y - M u - d \right)$$

Optimality conditions:

$$\begin{bmatrix} 2\beta M & 0 & -M \\ 0 & \bar{M} & K^T \\ -M & K & 0 \end{bmatrix} \begin{bmatrix} u \\ y \\ l \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ d \end{bmatrix}$$

Direct vs Iterative Methods

	Direct Methods		Iterative Methods
$\surd$	Black box	$\surd$	Large problems
$\surd$	Robust (large $\kappa(A)$?)	$\surd$	Preconditioning – convergence
$\times$	Memory with large problems?	$\times$	Iterative method?
		$\times$	Preconditioner?

Definition: let $\kappa(\mathcal{A}) = ||\mathcal{A}||_2||\mathcal{A}^{-1}||_2$ be the condition number of $\mathcal{A}$

Spectral properties of linear system

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

If A is symmetric and positive definite, then $\lambda(A) \in I^- \cup I^+$, where

$$\begin{aligned} I^- &= \left[\frac{1}{2}\left(\lambda_{\min}(A) - \sqrt{\lambda_{\min}^2(A) + 4\,\|B\|^2}\right), \frac{1}{2}\left(\|A\| - \sqrt{\|A\|^2 + 4\sigma_{\min}^2(B)}\right)\right], \\ I^+ &= \left[\lambda_{\min}(A), \frac{1}{2}\left(\|A\| + \sqrt{\|A\|^2 + 4\,\|B\|^2}\right)\right], \end{aligned}$$

[Rusten and Winther 1992]

Spectral properties of linear system

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

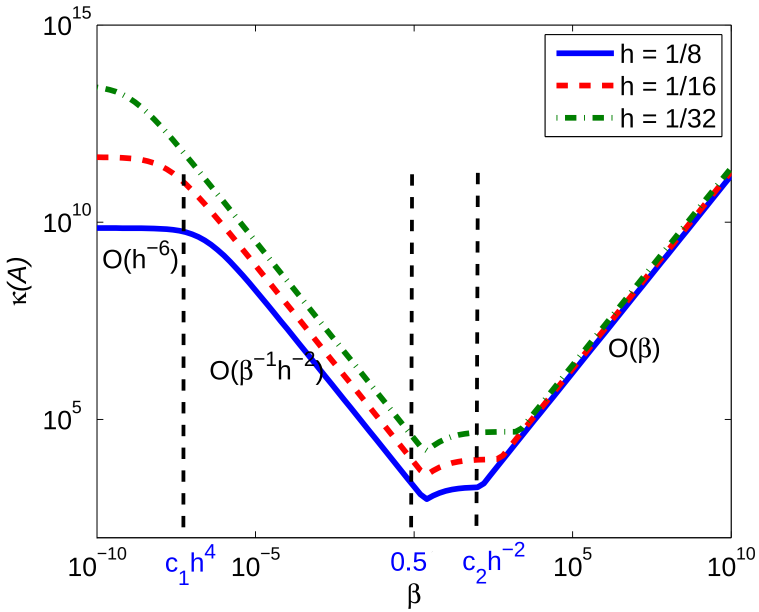
If A is symmetric and positive semi-definite, then $\lambda(H) \in I^- \cup I^+$, where

$$\begin{aligned} I^- &= \left[\frac{1}{2}\left(\lambda_{\min}(A) - \sqrt{\lambda_{\min}^2(A) + 4\|B\|^2}\right), \frac{1}{2}\left(\|A\| - \sqrt{\|A\|^2 + 4\sigma_{\min}^2(B)}\right)\right], \\ I^+ &= \left[l(A,B), \frac{1}{2}\left(\|A\| + \sqrt{\|A\|^2 + 4\|B\|^2}\right)\right], \end{aligned}$$

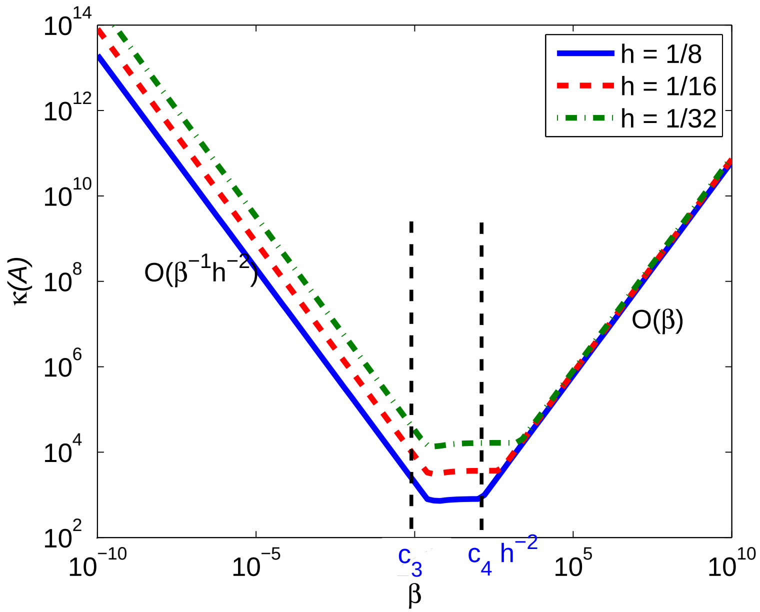
$l(A,B)$ defined in Dollar 2009 (revised)

Spectral properties of linear system

$\hat{\Omega} = \Omega$



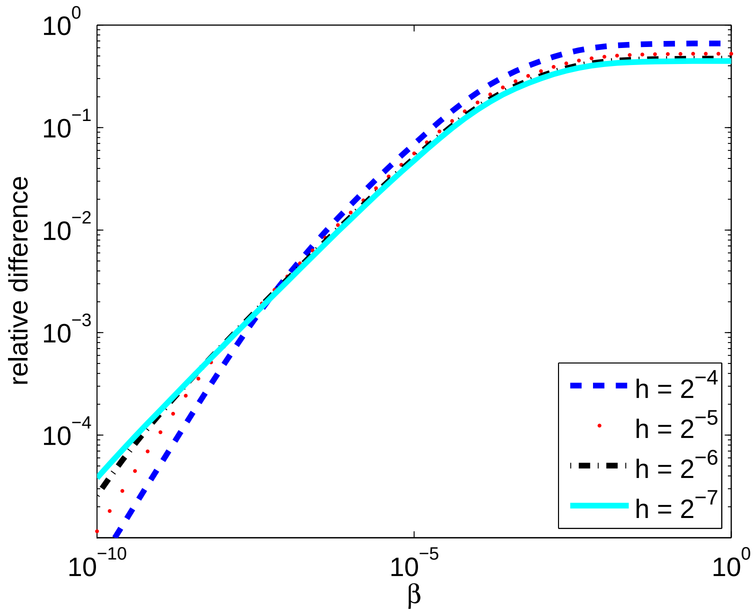
$\hat{\Omega} \neq \Omega$



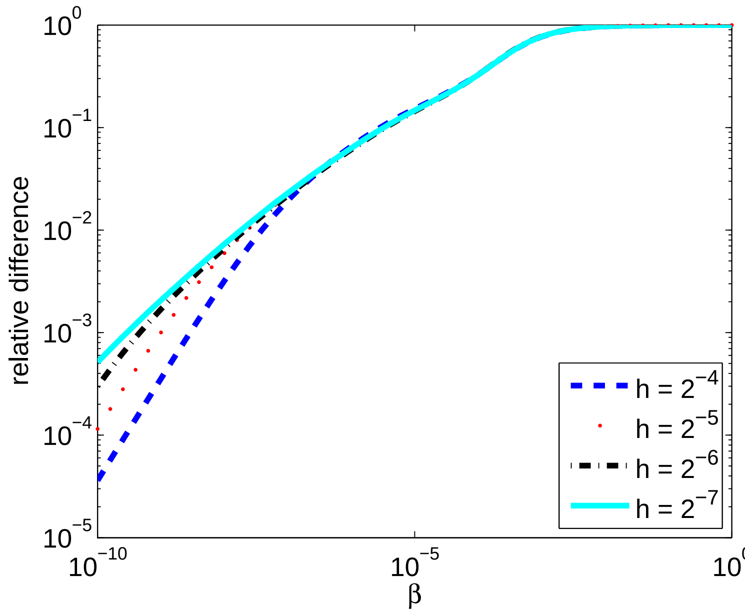
$\hat{\Omega}$	$\hat{\Omega}_1$	$\hat{\Omega}_2$	$\hat{y}(x_1,x_2)\|_{\hat{\Omega}_1}$	$\hat{y}(x_1,x_2)\|_{\hat{\Omega}_2}$
$\hat{\Omega}_1 \cup \hat{\Omega}_2$	$\left[0, \frac{1}{2}\right]^2$	$\Omega/\hat{\Omega}_1$	$(2x-1)^2\,(2y-1)^2$	0
$\hat{\Omega}_1 \cup \hat{\Omega}_2$	$\left\{(x,y) : (x-\frac{5}{8})^2 + (y-\frac{3}{4})^2 \leq \frac{1}{25}\right\}$	$\partial\Omega$	2	0

Spectral properties of linear system

$\hat{\Omega} = \Omega$



$\hat{\Omega} \neq \Omega$

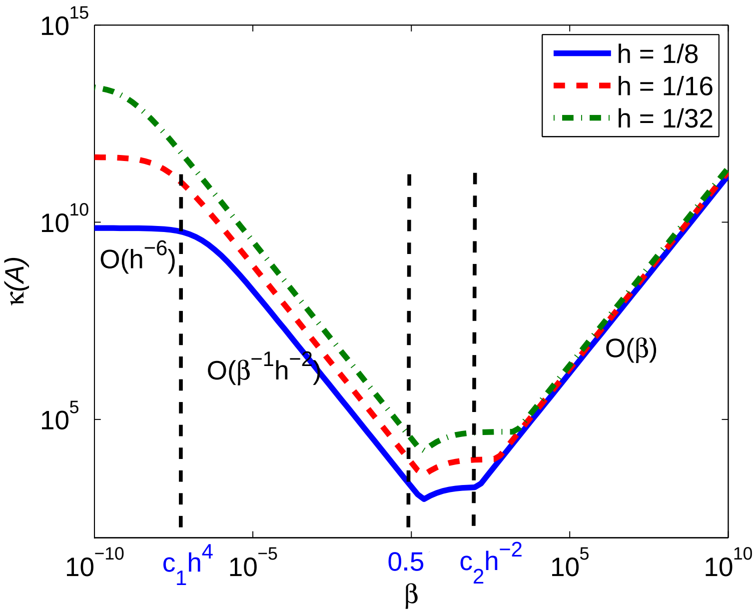


$\hat{\Omega}$	$\hat{\Omega}_1$	$\hat{\Omega}_2$	$\hat{y}(x_1,x_2)\|_{\hat{\Omega}_1}$	$\hat{y}(x_1,x_2)\|_{\hat{\Omega}_2}$
$\hat{\Omega}_1 \cup \hat{\Omega}_2$	$\left[0, \frac{1}{2}\right]^2$	$\Omega / \hat{\Omega}_1$	$(2x-1)^2\,(2y-1)^2$	0
$\hat{\Omega}_1 \cup \hat{\Omega}_2$	$\left\{(x,y) : (x-\frac{5}{8})^2 + (y-\frac{3}{4})^2 \leq \frac{1}{25}\right\}$	$\partial\Omega$	2	0

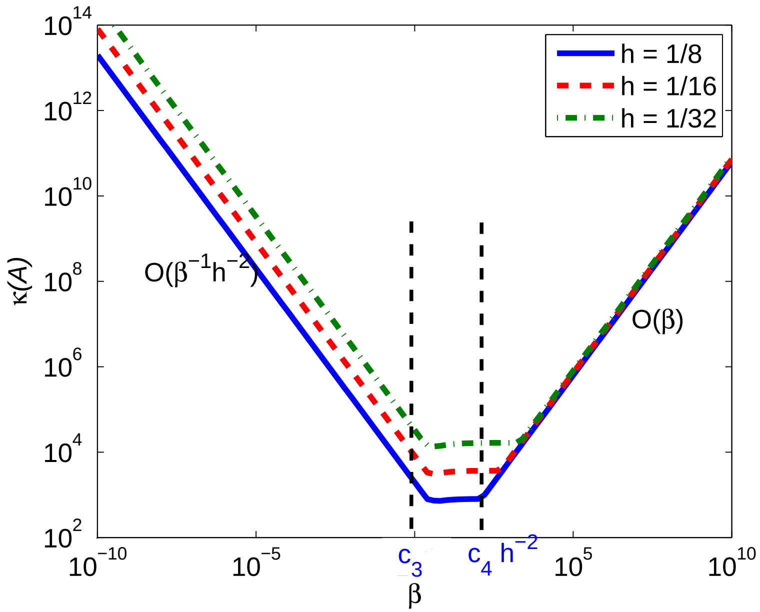
Science & Technology Facilities Council
Rutherford Appleton Laboratory

Spectral properties of linear system

$\hat{\Omega} = \Omega$



$\hat{\Omega} \neq \Omega$



$\hat{\Omega}$	$\hat{\Omega}_1$	$\hat{\Omega}_2$	$\hat{y}(x_1,x_2)\vert_{\hat{\Omega}_1}$	$\hat{y}(x_1,x_2)\vert_{\hat{\Omega}_2}$
$\hat{\Omega}_1 \cup \hat{\Omega}_2$	$\left[0, \frac{1}{2}\right]^2$	$\Omega / \hat{\Omega}_1$	$(2x-1)^2\,(2y-1)^2$	0
$\hat{\Omega}_1 \cup \hat{\Omega}_2$	$\left\{(x,y) : (x-\frac{5}{8})^2 + (y-\frac{3}{4})^2 \leq \frac{1}{25}\right\}$	$\partial\Omega$	2	0

Distributed control - iterative methods

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

subject to

$$Ky - Mu = d$$

$$\begin{bmatrix} 2\beta M & 0 & -M \\ 0 & \bar{M} & K^T \\ -M & K & 0 \end{bmatrix} \begin{bmatrix} u \\ y \\ l \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ d \end{bmatrix}$$

Projected Preconditioned CG Method

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}$$

Write

$$x = Yx_y + Zx_z,$$

where columns Z span nullspace of B and $[Y, Z]$ spans $\mathbb{R}^n$

$$\begin{aligned} BYx_y &= d, \\ Z^TAZx_z &= Z^T\left(b - AYx_y\right), \\ Y^TBw &= Y^T(b - Ax). \end{aligned}$$

If Z^TAZ is SPD, then use PCG with preconditioner Z^TGZ.

$$\|e_k\|_{Z^TAZ} \leq 2\,\|e_0\|_{Z^TAZ} \left(\frac{\sqrt{\kappa((Z^TGZ)^{-1}Z^TAZ)} - 1}{\sqrt{\kappa((Z^TGZ)^{-1}Z^TAZ)} + 1} \right)^k$$

Projected Preconditioned CG Method

Remove references to Z by making substitutions (Gould, Hribar, Nocedal, 2001):

Choose initial point x satisfying $Bx = d$
Compute $r = Ax - b$
Solve

$$\begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} g \\ v \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}$$

Set $p = -g$
repeat
Set $\alpha = r^T g / p^T A p$
Set $x = x + \alpha p$ and $r^+ = r + \alpha A p$
Solve

$$\begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} g^+ \\ v^+ \end{bmatrix} = \begin{bmatrix} r^+ \\ 0 \end{bmatrix}$$

Set $\beta = (r^+)^T g^+ / r^T g$
Set $p = -g^+ + \beta p$, $r = r^+$ and $g = g^+$
until converged

Projected Preconditioned CG Method

Remove references to Z by making substitutions:

Choose initial point x satisfying $Bx = d$

Compute $r = Ax - b$

Solve

$$\begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} g \\ v \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}$$

Set $p = -g$

repeat

Set $\alpha = r^T g / p^T A p$

Set $x = x + \alpha p$ and $r^+ = r + \alpha A p$

Solve

$$\begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} g^+ \\ v^+ \end{bmatrix} = \begin{bmatrix} r^+ \\ 0 \end{bmatrix}$$

Set $\beta = (r^+)^T g^+ / r^T g$

Set $p = -g^+ + \beta p$, $r = r^+$ and $g = g^+$

until converged

Projected Preconditioned CG Method

Solve

$$\begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ d \end{bmatrix}$$

Compute $r = Ax - b$
Solve

$$\begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} g \\ v \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}$$

Set $p = -g$
repeat
  Set $\alpha = r^T g / p^T A p$
  Set $x = x + \alpha p$ and $r^+ = r + \alpha A p$
  Solve

$$\begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} g^+ \\ v^+ \end{bmatrix} = \begin{bmatrix} r^+ \\ 0 \end{bmatrix}$$

  Set $\beta = (r^+)^T g^+ / r^T g$
  Set $p = -g^+ + \beta p$, $r = r^+$ and $g = g^+$
until converged

Projected Preconditioned CG Method

(Dollar 2005) Can be generalised to

$$\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}$$

Constraint preconditioners

$$\mathcal{A} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \qquad \mathcal{P} = \begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix}$$

Constraint preconditioners

$$\mathcal{A} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \qquad \mathcal{P} = \begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix}$$

Theorem (Keller, Gould, Wathen, 2000): If $A, G \in \mathbb{R}^{n\times n}$ are symmetric and $B \in \mathbb{R}^{m\times n}$ has full row rank, then $\mathcal{P}^{-1}\mathcal{A}$ has

- $2m$ eigenvalues at 1
- remaining $n-m$ are defined by

$$Z^T AZx = \lambda Z^T GZx,$$

where the columns of $Z \in \mathbb{R}^{n\times(n-m)}$ span nullspace of B.

Preconditioner

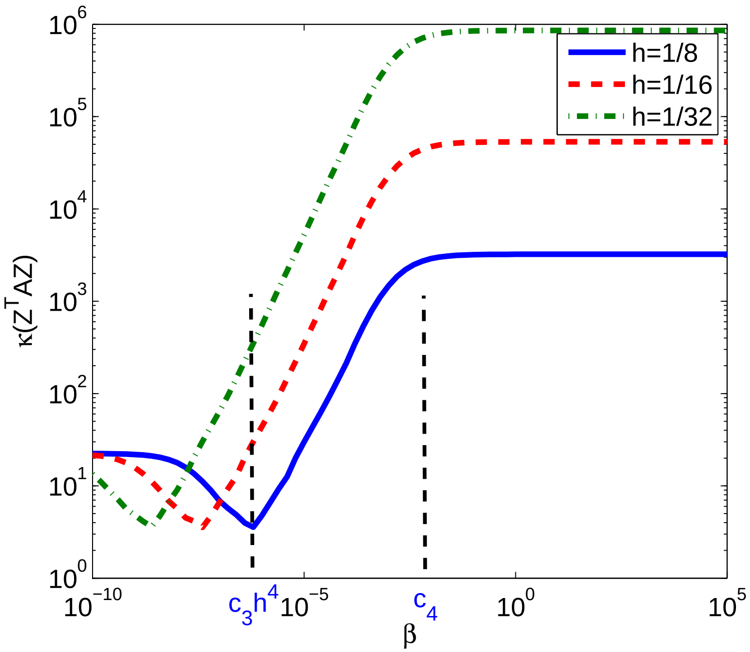
$$\mathcal{A} = \begin{bmatrix} 2\beta M & 0 & -M \\ 0 & \bar{M} & K^T \\ -M & K & 0 \end{bmatrix} \quad Z = \begin{bmatrix} M^{-1}K \\ I \end{bmatrix}$$

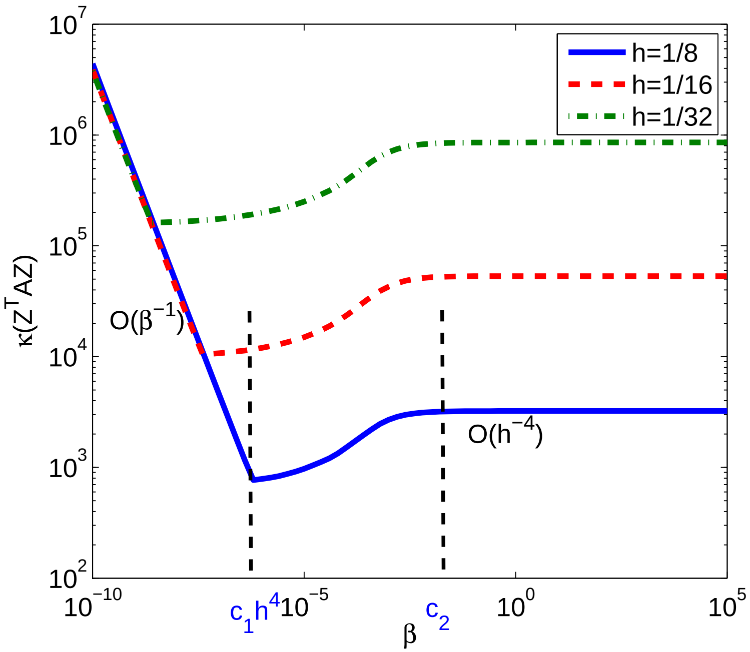
$$Z^T A Z = 2\beta K^T M^{-1} K + \bar{M}$$

Preconditioner

$$\mathcal{A} = \begin{bmatrix} 2\beta M & 0 & -M \\ 0 & \bar{M} & K^T \\ -M & K & 0 \end{bmatrix} \quad Z = \begin{bmatrix} M^{-1}K \\ I \end{bmatrix}$$

$$Z^T AZ = 2\beta K^T M^{-1} K + \bar{M}$$





Preconditioner

$$\mathcal{A} = \begin{bmatrix} 2\beta M & 0 & -M \\ 0 & \bar{M} & K^T \\ -M & K & 0 \end{bmatrix} \quad Z = \begin{bmatrix} M^{-1}K \\ I \end{bmatrix}$$

$$Z^T A Z = 2\beta K^T M^{-1} K + \bar{M}$$

$$\mathcal{P} = \begin{bmatrix} 2\beta M & 0 & -M \\ 0 & 0 & K^T \\ -M & K & 0 \end{bmatrix} ?$$

$$Z^T G Z = 2\beta K^T M^{-1} K$$

$\bar{M} = M$	$\bar{M} \neq M$
$1 + \frac{ch^4}{2\beta} \leq \lambda \leq 1 + \frac{C}{2\beta}$ $c \leq \bar{c}$ and $\bar{C} \leq C$	$1 + \frac{\bar{c}h^4}{2\beta} \leq \lambda \leq 1 + \frac{\bar{C}}{2\beta}$ $\lambda = 1$

Biros and Ghattas (2000)

Preconditioner

$$\mathcal{A} = \begin{bmatrix} 2\beta M & 0 & -M \\ 0 & \bar{M} & K^T \\ -M & K & 0 \end{bmatrix} \quad Z = \begin{bmatrix} M^{-1}K \\ I \end{bmatrix}$$

$$Z^T A Z = 2\beta K^T M^{-1} K + \bar{M}$$

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

$$Z^T G Z = 2\beta K^T M^{-1} K$$

$\bar{M} = M$	$\bar{M} \neq M$
$1 + \frac{ch^4}{2\beta} \leq \lambda \leq 1 + \frac{C}{2\beta}$ $c \leq \bar{c}$ and $\bar{C} \leq C$	$1 + \frac{\bar{c}h^4}{2\beta} \leq \lambda \leq 1 + \frac{\bar{C}}{2\beta}$ $\lambda = 1$

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Numerical Example

Using bilinear **Q1** elements and setting $\beta = 5 \times 10^{-5}$:

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

- Solves with M : Direct method (`HSL_MA57`) or 20 Chebyshev semi-iterations
- Solves with K : Direct method (`HSL_MA57`) or two(three) V-cycles of AMG (`HSL_MI20`)
- PPCG: relative tolerance 10^{-9} for $r^T Z(Z^T G Z)^{-1} Z^T r$, `HSL_MI27`
- Fortran 95, NAG f95 compiler
- Hardware: Dell Precision T340, single Core2 Quad Q9550 processor (2.83GHz, 1333MHz FSB, 12MB L2 Cache), 4GB RAM

Numerical Example

Using bilinear **Q1** elements and setting $\beta = 5 \times 10^{-5}$:

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

- Solves with M : Direct method (`HSL_MA57`) or 20 Chebyshev semi-iterations
- Solves with K : Direct method (`HSL_MA57`) or two(three) V-cycles of AMG (`HSL_MI20`)
- PPCG: relative tolerance 10^{-9} for $r^T Z (Z^T G Z)^{-1} Z^T r$, `HSL_MI27`
- Fortran 95, NAG f95 compiler
- Hardware: Dell Precision T340, single Core2 Quad Q9550 processor (2.83GHz, 1333MHz FSB, 12MB L2 Cache), 4GB RAM

2D

N	n	Direct	PPCG(direct)	PPCG(approx)
8	147	0.002	0.001 (8)	0.003 (9)
16	675	0.01	0.006 (8)	0.011 (9)
32	2883	0.04	0.025 (8)	0.044 (9)
64	11907	0.19	0.12 (8)	0.17 (8)
128	48487	1.59	0.55 (7)	0.72 (8)
256	195075	8.82	3.27 (6)	3.18 (8)
512	783363	53.5	21.5 (6)	14.2 (8)

3D

N	n	Direct	PPCG(direct)	PPCG(approx)
4	81	0.001	0.002 (7)	0.002 (7)
8	1029	0.04	0.02 (8)	0.05 (8)
16	10125	1.25	0.33 (8)	0.64 (8)
32	89373	38.0	6.61 (7)	7.32 (7)
64	750141	1000+	217 (5)	59.0 (6)

Numerical Example

Using bilinear **Q1** elements and setting $\beta = 5 \times 10^{-5}$:

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

- Solves with M : Direct method (`HSL_MA57`) or 20 Chebyshev semi-iterations
- Solves with K : Direct method (`HSL_MA57`) or two(three) V-cycles of AMG (`HSL_MI20`)
- PPCG: relative tolerance 10^{-9} for $r^T Z(Z^T G Z)^{-1} Z^T r$, `HSL_MI27`
- Fortran 95, NAG f95 compiler
- Hardware: Dell Precision T340, single Core2 Quad Q9550 processor (2.83GHz, 1333MHz FSB, 12MB L2 Cache), 4GB RAM

2D

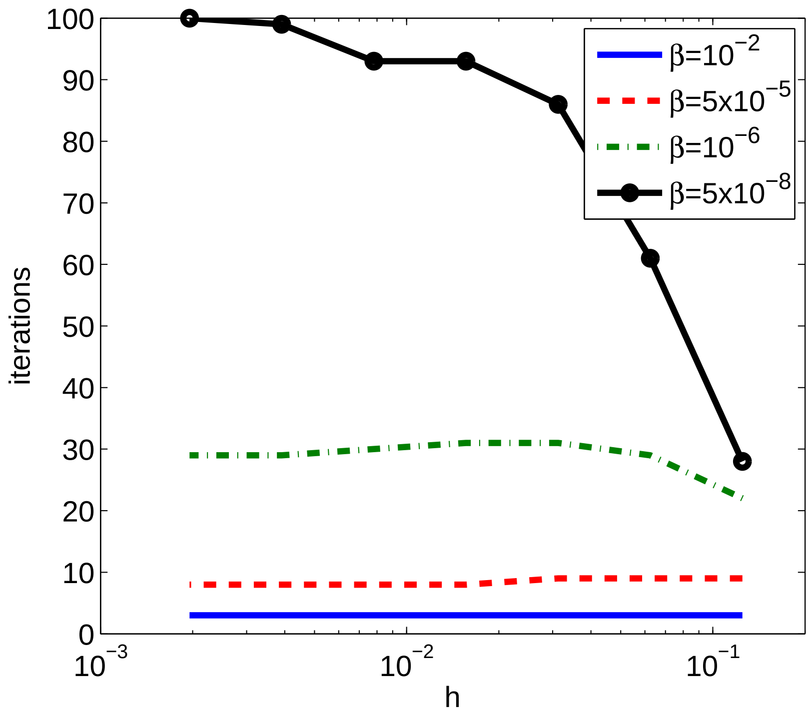
N	n	Direct	PPCG(direct)	PPCG(approx)
8	147	0.002	0.001 (4)	0.002 (4)
16	675	0.01	0.01 (4)	0.007 (4)
32	2883	0.10	0.02 (4)	0.03 (4)
64	11907	0.35	0.10 (4)	0.13 (5)
128	48487	2.78	0.50 (5)	0.53 (5)
256	195075	16.8	3.11 (5)	2.36 (5)
512	783363	147	20.5 (5)	10.3 (5)

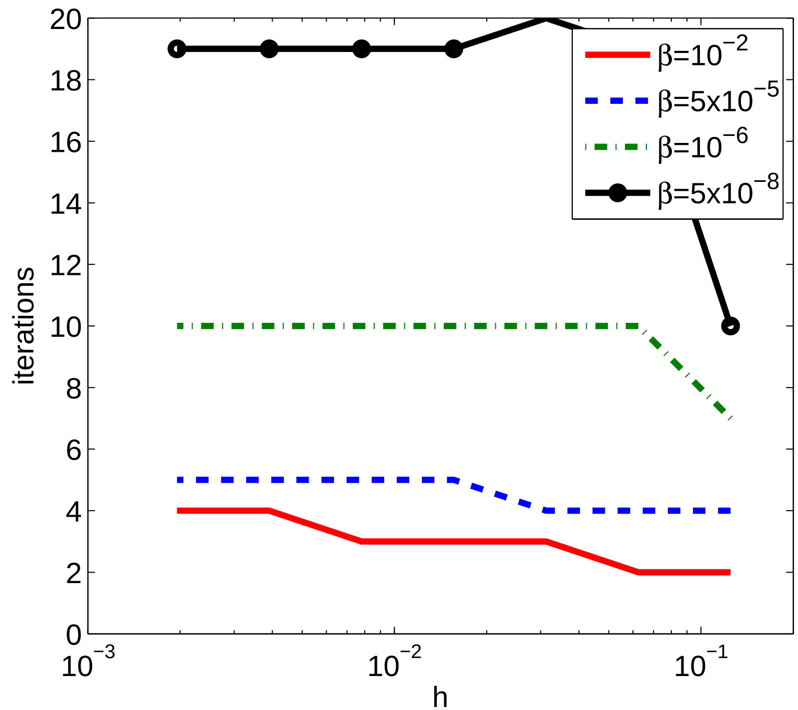
3D

N	n	Direct	PPCG(direct)	PPCG(approx)
4	81	0.001	0.001 (3)	0.001 (3)
8	1029	0.05	0.02 (4)	0.03 (4)
16	10125	1.19	0.31 (5)	0.49 (5)
32	89373	59.2	6.32 (5)	6.00 (5)
64	750141	1000+	219 (5)	58.9 (5)

Behaviour of preconditioner with β

$\bar{M} = M$	$\bar{M} \neq M$
$1 + \frac{ch^4}{2\beta} \leq \lambda \leq 1 + \frac{C}{2\beta}$ $c \leq \bar{c}$ and $\bar{C} \leq C$	$1 + \frac{\bar{c}h^4}{2\beta} \leq \lambda \leq 1 + \frac{\bar{C}}{2\beta}$ $\lambda = 1$





Distributed control with nonlinear PDEs

$$\min_{y,u} \frac{1}{2} \|y - \widehat{y}\|_2^2 + \beta \|u\|_2^2$$

subject to

$$\begin{aligned} \mathcal{L}(y) &= u \text{ in } \Omega \\ y &= \widehat{y} \text{ on } \delta\Omega \end{aligned}$$

Optimality conditions:

$$\begin{aligned} My + J(y)^T l &= b, \\ 2\beta Mu - Ml &= 0, \\ F(y) - Mu &= d. \end{aligned}$$

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Trust-funnel method (Gould and Toint)

$$\min_x f(x) \quad \text{subject to} \quad c(x) = 0$$

Attempts to consider the objective function and constraints as independently as possible

Trust-funnel method (Gould and Toint)

$$\min_x f(x) \quad \text{subject to} \quad c(x) = 0$$

Attempts to consider the objective function and constraints as independently as possible

$$\text{Find } n \text{ to reduce} \quad \|c_k + J_k n\| \quad \text{subject to} \quad \|n\| \le \Delta_1$$

$$\text{Find } l \text{ to reduce} \quad \left\|g_k + J_k^T l\right\|$$

$$\text{Find } t \text{ to reduce} \quad g_k^T t + \frac{1}{2} t^T H_k t \quad \text{subject to} \quad J_k t = 0 \quad \text{and} \quad \|t\| \le \Delta_2$$

$$x_{k+1} = x_k + n + t$$

Trust-funnel method (Gould and Toint)

$$\min_x f(x) \quad \text{subject to} \quad c(x) = 0$$

Attempts to consider the objective function and constraints as independently as possible

$$\text{Find } n \text{ to reduce} \quad \|c_k + J_k n\| \quad \text{subject to} \quad \|n\| \leq \Delta_1$$

$$\text{Find } l \text{ to reduce} \quad \left\|g_k + J_k^T l\right\|$$

$$\text{Find } t \text{ to reduce} \quad g_k^T t + \frac{1}{2} t^T H_k t \quad \text{subject to} \quad J_k t = 0 \quad \text{and} \quad \|t\| \leq \Delta_2$$

$$x_{k+1} = x_k + n + t$$

- Adjust Δ_1 and Δ_2 for convergence
- Only require matrix-vector multiplications (preconditioning?)
- Alternative matrix-free method by Curtis, Nocedal and Wächter

Reduce $g_k^T t + \frac{1}{2} t^T H_k t$ subject to $J_k t = 0$ and $\|t\| \leq \Delta_2$,

where

$$
\begin{aligned}
g_k &= \nabla f(x_k) + H_k n_k, \\
H_k &= \nabla^2 f(x_k) + \sum_{i=1}^{m} [l_{k-1}]_i \, C_{ik}, \\
C_{ik} &= C_{ik}^T \approx \nabla_{xx} c_i(x_k)
\end{aligned}
$$

Reduce $g_k^T t + \frac{1}{2} t^T H_k t$ subject to $J_k t = 0$ and $\|t\| \leq \Delta_2$,

where

$$\begin{aligned} g_k &= \nabla f(x_k) + H_k n_k, \\ H_k &= \nabla^2 f(x_k) + \sum_{i=1}^{m} [l_{k-1}]_i \, C_{ik}, \\ C_{ik} &= C_{ik}^T \approx \nabla_{xx} c_i(x_k) \end{aligned}$$

Apply PPCG to

$$\begin{bmatrix} H_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} t \\ s \end{bmatrix} = \begin{bmatrix} g_k \\ 0 \end{bmatrix}$$

Initialise $t = 0$. Iterate until convergence or $\|t\| \geq \Delta_2$. If $\|t\| \geq \Delta_2$, back-track to boundary.

Distributed control with nonlinear PDEs

$$\min_{y,u} \frac{1}{2} \left\| \omega(x) \left(y - \widehat{y} \right) \right\|_2^2 + \beta \left\| u \right\|_2^2$$

subject to

$$\begin{aligned} -\nabla \cdot \left[\left(1 + y^2 \right) \nabla y \right] &= u \text{ in } \Omega \\ y &= \widehat{y} \text{ on } \delta\Omega \end{aligned}$$

Distributed control with nonlinear PDEs

$$\min_{y,u} \frac{1}{2} \left\|\omega(x)\,(y-\widehat{y})\right\|_2^2 + \beta \left\|u\right\|_2^2$$

subject to

$$\begin{aligned} -\nabla \cdot \left[\left(1+y^2\right) \nabla y\right] &= u \text{ in } \Omega \\ y &= \widehat{y} \text{ on } \delta\Omega \end{aligned}$$

$$\left[\begin{array}{cc|c} 2\beta M & 0 & -M \\ 0 & M+\sum_{i=1}^{m} \left[l_{k-1}\right]_i \nabla^2 F_j(y_k) & J(y_k)^T \\ \hline -M & J(y_k) & 0 \end{array}\right]$$

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Distributed control with nonlinear PDEs

$$\min_{y,u} \frac{1}{2} \|\omega(x)\,(y-\widehat{y})\|_2^2 + \beta\,\|u\|_2^2$$

subject to

$$\begin{aligned} -\nabla \cdot \left[\left(1+y^2\right)\nabla y\right] &= u \text{ in } \Omega \\ y &= \widehat{y} \text{ on } \delta\Omega \end{aligned}$$

$$\left[\begin{array}{cc|c} 2\beta M & 0 & -M \\ 0 & M + \sum_{i=1}^{m} [l_{k-1}]_i \, \nabla^2 F_j(y_k) & K^T + L(y_k)^T \\ \hline -M & K + L(y_k) & 0 \end{array}\right]$$

Distributed control with nonlinear PDEs

$$\min_{y,u} \frac{1}{2} \|\omega(x)\,(y-\widehat{y})\|_2^2 + \beta\,\|u\|_2^2$$

subject to

$$\begin{aligned} -\nabla \cdot \left[\left(1+y^2\right)\nabla y\right] &= u \text{ in } \Omega \\ y &= \widehat{y} \text{ on } \delta\Omega \end{aligned}$$

$$\left[\begin{array}{cc|c} 2\beta M & 0 & -M \\ 0 & M + \sum_{i=1}^m [l_{k-1}]_i \nabla^2 F_j(y_k) & K^T + L(y_k)^T \\ \hline -M & K + L(y_k) & 0 \end{array}\right]$$

$$P_1 = \left[\begin{array}{cc|c} I & 0 & -M \\ 0 & I & K^T + L(y_k)^T \\ \hline -M & K + L(y_k) & 0 \end{array}\right], P_2 = \left[\begin{array}{cc|c} 0 & 0 & -M \\ 0 & 2\beta K^T M^{-1} K & K^T + L(y_k)^T \\ \hline -M & K + L(y_k) & 0 \end{array}\right]$$

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Preliminary results

N		T-F iterations	PPCG calls	Total PPCG its	Max PPCG its	Average PPCG its
2^3	P_1	19	14	490	$50^*(3)$	35
	P_2	5	3	34	12	11
2^4	P_1	24	15	1388	$226^*(5)$	93
	P_2	5	3	39	20	13

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Conclusions and Future Work

- PDE-constrained problems difficult to solve
- Avoid accurate solves with discretized PDE
- Use block structure
- Mesh size independent convergence
- Nonlinear PDEs - sequence of more challenging saddle-point systems
- Nonlinear PDEs, time-dependent PDEs, different regularization terms

- `HSL_MA57` and `HSL_MI20` are part of `HSL2007`, which is free for all academics
- `HSL_MI27` will be part of `HSL2011`
- ‘Nonlinear programming without a penalty function or filter’ Gould, Toint, Math. Prog 2010
- ‘Optimal solvers for PDE-constrained optimization’ Rees, Dollar, Wathen, SISC 2010
- ‘Properties of linear systems in PDE-constrained optimization. Part I: Distributed control’, Dollar, RAL TR-2009-017
- ‘PDE-constrained optimization and constraint preconditioners’, Thorne, RAL TR-2010-016