Multigrid preconditioners for linear systems arising in PDE constrained optimization

Andrei Draganescu

Department of Mathematics and Statistics
University of Maryland, Baltimore County

March 5, 2013

Acknowledgments

- *Former collaborators:*
  Todd Dupont (U Chicago), Volkan Akçelik (Exxon),
  George Biros (U Texas), Omar Ghattas (U Texas),
  Judith Hill (ORNL), Cosmin Petra (ANL), Bart van Bloemen Waanders (Sandia).
- *Current collaborators:* (UMBC)
  Mona Hajghassem, *Jyoti Saraswat* , Ana Maria Soane
- *Grants:*
  - NSF awards DMS-1016177 and DMS-0821311.
  - DOE contract no: DE-SC0005455.

Outline

Abstract problem formulation

$$\begin{cases} \text{minimize} & J(y,u) = \frac{1}{2}||y - y_d||^2_{L^2(\Omega)} + R(u,y), \\ \text{subj. to} & u \in U_{ad} \subset U, \ \ y \in Y_{ad} \subset Y, \\ & e(y,u) = 0\ . \end{cases} \tag{1}$$

- U_{ad} and Y_{ad} – sets of admissible controls resp. states (convex, closed, non-empty).
- Ex.: $U_{ad} = \{u \in U \ : \ \underline{u} \le u \le \overline{u}\}$, $Y_{ad} = \{y \in Y \ : \ \underline{y} \le y \le \overline{y}\}$.
- Equality constraint is a well-posed PDE:
  for all $u \in U$ there is a unique $y \in Y$ (depending continuously on u), so that
  $$e(y,u) = 0, \ \ y \stackrel{\text{def}}{=} K(u)\ .$$

Reduced problem formulation

If $U_{ad} = U$ and $Y_{ad} = Y$, problem can be reformulated as unconstrained:

$$\min_{u \in U} \; J(u) = \frac{1}{2}\|K(u) - y_d\|^2 + \frac{\beta}{2}\|Lu\|^2, \quad u \in U_{\mathrm{ad}}\ . \tag{2}$$

- If $\beta \ll 1$, essentially we want solve

$$K(u) = y_d\ .$$

- However, problems of interest are ill-posed, need regularization:
  - $L = I \;\; \Rightarrow$ find u of smallest norm ;
  - $L = \nabla \Rightarrow$ find u of smallest variation.

Motivating applications

1. Reverse advection-diffusion problems (source inversion):

- $T > 0$ fixed “end-time”, y_d end-time state, u initial state
- $z(\cdot, t)$ transported quantity subjected to:

$$\begin{cases} \partial_t z - \nabla \cdot (a \nabla z + bz) + cz = 0 & \text{on } \Omega \\ z(x,t) = 0 & \text{for } x \in \partial\Omega,\ t \in [0, T] \\ z(x,0) = u(x) & \text{for } x \in \Omega \end{cases}$$

- $K = S(T)$: initial - to - final

$$K\,u = S(T)u \stackrel{\text{def}}{=} z(\cdot, T)$$

Further motivating applications

2. Elliptic optimal control problem:

- PDE-constrained optimal control problem

$$\begin{cases} \text{minimize} & \frac{1}{2}\|y - y_d\|^2 + \frac{\beta}{2}\|u\|^2 \,, \\ \text{subj to:} & -\Delta y = u \,, \quad u|_{\partial\Omega} = 0 \,, \\ & \underline{u} \leq u \leq \overline{u} \,. \end{cases}$$

- If unconstrained, then $K = (-\Delta)^{-1}$.

The case of linear constraints

Assume K linear, $U_{ad} = U$:

$$\min_u J(u) = \frac{1}{2}\|Ku - y_d\|^2 + \frac{\beta}{2}\|u\|^2$$

- Newton's method gives the solution explicitly in one step:

$$u^{\min} = u_0^{\text{guess}} - G^{-1}\,\nabla J(u_0^{\text{guess}})\ ,$$

  where

$$G = G(\beta) = I + \beta^{-1}K^* \cdot K\ ,$$
$$\nabla J(u) = u + \beta^{-1}K^*(Ku - y_d)\ .$$

- Formulation is equivalent to the regularized normal equations

$$(\beta I + K^* \cdot K)u = K^* y_d\ .$$

Strategy: discretize-then-optimize

- Natural FE discretization for the operator K:

$$\min_u J(u) = \frac{1}{2}\|K_h u - y_d\|^2 + \frac{\beta}{2}\|u\|^2 .$$

- Solution of discrete problem:

$$u_h^{\min} = u_{0,h}^{\text{guess}} - G_h^{-1} \nabla J_h(u_{0,h}^{\text{guess}}) ,$$

where

$$G_h = G_h(\beta) = I + \beta^{-1} K_h^* \cdot K_h ,$$
$$\nabla J_h(u) = u + \beta^{-1} K_h^*(K_h u - \pi_h y_d) ,$$

π_h is the orthogonal projection onto the finite element space V_h

- **Main problem: need to invert the operator G_h efficiently.**

Main issues

- The matrix representing the linear operator G_h is dense, potentially large, and not available.
- Matrix-vector product cost is comparable to two forward computations (expensive, but feasible):

$$G_h \cdot u = u + \beta^{-1} K_h^* \cdot K_h u \ .$$

- Gradient computation also costs as much as two forward computations (only done once):

$$\nabla J_h(u) = u + \beta^{-1} K_h^* (K_h u - \pi_h y_d) \ .$$

- Need iterative methods.

Solution using conjugate gradient

- Eigenvalues of G_h cluster around 1 $\Rightarrow$
  CG is a good choice for solving inverting G_h:
  the number of iterations
  - is independent of the resolution;
  - grows only logarithmically with $\beta \to 0$.
- A measure of success: speedup over CG.

Multigrid strategies

- major differences between G_h and an elliptic operator A_h:

G_h	A_h
smoothing	roughening
nonlocal	local
cond(G_h) bounded	cond$(A_h) \to \infty$ as $h \to 0$

- Related multigrid work:
  - Hackbusch (1981), King (1992), Rieder (1997), Hanke and Vogel (1999), Kaltenbacher (2003), Donatelli (2005), Biros and **Dogan** (2008), Draganescu and Dupont (2008), Borzi and Kunisch (2005).
  - Lewis and Nash (2000)
  - overview: Borzi and Schultz (SIAM review, 2009)
  - more recent: Wathen, Stoll, Rees, Dollar, Draganescu and Soane, etc

Two-grid approximation (heuristics)

- $V_h = \overbrace{V_{2h}}^{\text{“smooth“ functions}} \oplus \overbrace{W}^{\text{“rough“ functions}}$
- denote $\pi = \pi_{2h}$, $\rho = I - \pi_{2h}$

$$G_h = \underbrace{\pi G_h \pi}_{M_1} + \underbrace{\rho G_h \pi}_{M_2} + \underbrace{\pi G_h \rho}_{M_3} + \underbrace{\rho G_h \rho}_{M_4}$$

- since $G_h \rho = \left(I + \beta^{-1} K_h{}^* K_h\right) \rho \approx \rho$

$$M_2 \approx 0 \qquad M_3 \approx 0$$
$$M_1 \approx G_{2h}\pi \qquad M_4 \approx \rho$$

- conclusion:

$$G_h \approx M_h \stackrel{\text{def}}{=} G_{2h}\pi_{2h} \oplus (I - \pi_{2h}) .$$

Two-grid approximation (heuristics)

- $V_h = \overbrace{V_{2h}}^{\text{“smooth“ functions}} \oplus \overbrace{W}^{\text{“rough“ functions}}$
- denote $\pi = \pi_{2h}$, $\rho = I - \pi_{2h}$

$$G_h = \underbrace{\pi G_h \pi}_{M_1} + \underbrace{\rho G_h \pi}_{M_2} + \underbrace{\pi G_h \rho}_{M_3} + \underbrace{\rho G_h \rho}_{M_4}$$

- since $G_h\rho = \left(I + \beta^{-1} K_h{}^* K_h\right)\rho \approx \rho$

$$M_2 \approx 0 \qquad M_3 \approx 0$$
$$M_1 \approx G_{2h}\pi \qquad M_4 \approx \rho$$

- conclusion:

$$G_h \approx M_h \stackrel{\text{def}}{=} G_{2h}\pi_{2h} \oplus (I - \pi_{2h}) .$$

Two-grid approximation (heuristics)

- $V_h = \overbrace{V_{2h}}^{\text{“smooth“ functions}} \oplus \overbrace{W}^{\text{“rough“ functions}}$
- denote $\pi = \pi_{2h}$, $\rho = I - \pi_{2h}$

$$G_h = \underbrace{\pi G_h \pi}_{M_1} + \underbrace{\rho G_h \pi}_{M_2} + \underbrace{\pi G_h \rho}_{M_3} + \underbrace{\rho G_h \rho}_{M_4}$$

- since $G_h \rho = \left(I + \beta^{-1} K_h{}^* K_h\right) \rho \approx \rho$

$$M_2 \approx 0 \qquad M_3 \approx 0$$

$$M_1 \approx G_{2h}\pi \qquad M_4 \approx \rho$$

- conclusion:

$$G_h \approx M_h \stackrel{\text{def}}{=} G_{2h}\pi_{2h} \oplus (I - \pi_{2h}) .$$

Two-grid approximation (heuristics)

- $V_h = \overbrace{V_{2h}}^{\text{“smooth“ functions}} \oplus \overbrace{W}^{\text{“rough“ functions}}$
- denote $\pi = \pi_{2h}$, $\rho = I - \pi_{2h}$

$$G_h = \underbrace{\pi G_h \pi}_{M_1} + \underbrace{\rho G_h \pi}_{M_2} + \underbrace{\pi G_h \rho}_{M_3} + \underbrace{\rho G_h \rho}_{M_4}$$

- since $G_h \rho = \left(I + \beta^{-1} K_h{}^* K_h\right) \rho \approx \rho$

$$M_2 \approx 0 \qquad M_3 \approx 0$$
$$M_1 \approx G_{2h}\pi \qquad M_4 \approx \rho$$

- conclusion:

$$G_h \approx M_h \stackrel{\text{def}}{=} G_{2h}\pi_{2h} \oplus \left(I - \pi_{2h}\right).$$

Two-grid approximation (heuristics)

- $$V_h = \overbrace{V_{2h}}^{\text{“smooth“ functions}} \oplus \overbrace{W}^{\text{“rough“ functions}}$$
- denote $\pi = \pi_{2h}$, $\rho = I - \pi_{2h}$

$$G_h = \underbrace{\pi G_h \pi}_{M_1} + \underbrace{\rho G_h \pi}_{M_2} + \underbrace{\pi G_h \rho}_{M_3} + \underbrace{\rho G_h \rho}_{M_4}$$

- since $G_h \rho = \left(I + \beta^{-1} K_h{}^* K_h\right) \rho \approx \rho$

$$M_2 \approx 0 \qquad M_3 \approx 0$$
$$M_1 \approx G_{2h}\pi \qquad M_4 \approx \rho$$

- conclusion:

$$G_h \approx M_h \stackrel{\text{def}}{=} G_{2h}\pi_{2h} \oplus (I - \pi_{2h}) .$$

Multigrid for our problem
two-grid approximation (results)

Proposed preconditioner:

$$L_h \stackrel{\text{def}}{=} (M_h)^{-1} = {G_{2h}}^{-1}\pi_{2h} + (I - \pi_{2h}) \ .$$

Theorem (A.D., Dupont 2004):

For h sufficiently small and $u \in V_h$

$$1 - C\frac{h^p}{\beta} \leq \frac{\langle (M_h)^{-1}u, u\rangle}{\langle (G_h)^{-1}u, u\rangle} \leq 1 + C\frac{h^p}{\beta} \ ,$$

where p is the order of the discrete method.

From two-grid to multigrid

natural extension (V-cycle)

- Natural extension to multigrid is suboptimal:

$$L_h = G_{2h}{}^{-1}\pi_{2h} + (I - \pi_{2h}) \approx G_h^{-1}$$

$$\Downarrow \ (\text{since } L_{2h} \approx G_{2h}{}^{-1})$$

$$L_h \stackrel{\text{def}}{=} L_{2h}\pi_{2h} + (I - \pi_{2h})$$

Corollary:

For h, h_0 small enough and $u \in V_h$

$$1 - C\frac{h_0^p}{\beta} \leq \frac{\langle L_h u, u\rangle}{\langle (G_h)^{-1}u, u\rangle} \leq 1 + C\frac{h_0^p}{\beta}.$$

From two-grid to multigrid

natural extension (V-cycle)

- Natural extension to multigrid is suboptimal:

$$L_h = {G_{2h}}^{-1}\pi_{2h} + (I - \pi_{2h}) \approx G_h^{-1}$$

$$\Downarrow \ (\text{since } L_{2h} \approx {G_{2h}}^{-1})$$

$$L_h \stackrel{\text{def}}{=} L_{2h}\pi_{2h} + (I - \pi_{2h})$$

Corollary:

For h, h_0 small enough and $u \in V_h$

$$1 - C\frac{h_0^p}{\beta} \leq \frac{\langle L_h u, u\rangle}{\langle (G_h)^{-1}u, u\rangle} \leq 1 + C\frac{h_0^p}{\beta}.$$

From two-grid to multigrid

natural extension (V-cycle)

- Natural extension to multigrid is suboptimal:

$$L_h = G_{2h}{}^{-1}\pi_{2h} + (I - \pi_{2h}) \approx G_h^{-1}$$

$$\Downarrow \ (\text{since } L_{2h} \approx G_{2h}{}^{-1})$$

$$L_h \stackrel{\text{def}}{=} L_{2h}\pi_{2h} + (I - \pi_{2h})$$

Corollary:

For h, h_0 small enough and $u \in V_h$

$$1 - C\frac{h_0^p}{\beta} \leq \frac{\langle L_h u, u\rangle}{\langle (G_h)^{-1}u, u\rangle} \leq 1 + C\frac{h_0^p}{\beta}.$$

From two-grid to multigrid

Newton extension (W-cycle)

- essential ingredient: use Newton's method for the nonlinear operator equation

$$X^{-1} - G_h = 0$$

- basic idea:
  X_1 (below) is an improved approximation of $(G_h)^{-1}$ over X_0

$$X_1 = \mathcal{N}_{G_h}(X_0) \stackrel{\text{def}}{=} 2X_0 - X_0 \cdot G_h \cdot X_0$$

-

$$L_h \stackrel{\text{def}}{=} \mathcal{N}_{G_h}(L_{2h}\pi_{2h} + (I - \pi_{2h}))$$

From two-grid to multigrid

Newton extension (W-cycle)

- essential ingredient: use Newton's method for the nonlinear operator equation

$$X^{-1} - G_h = 0$$

- basic idea:
  X_1 (below) is an improved approximation of $(G_h)^{-1}$ over X_0

$$X_1 = \mathcal{N}_{G_h}(X_0) \stackrel{\text{def}}{=} 2X_0 - X_0 \cdot G_h \cdot X_0$$

- $$L_h \stackrel{\text{def}}{=} \mathcal{N}_{G_h}(L_{2h}\pi_{2h} + (I - \pi_{2h}))$$

From two-grid to multigrid

Newton extension (W-cycle)

- essential ingredient: use Newton's method for the nonlinear operator equation

$$X^{-1} - G_h = 0$$

- basic idea:
  X_1 (below) is an improved approximation of $(G_h)^{-1}$ over X_0

$$X_1 = \mathcal{N}_{G_h}(X_0) \stackrel{\text{def}}{=} 2X_0 - X_0 \cdot G_h \cdot X_0$$

-

$$L_h \stackrel{\text{def}}{=} \mathcal{N}_{G_h}(L_{2h}\pi_{2h} + (I - \pi_{2h}))$$

From two-grid to multigrid

Newton extension (result)

Theorem (A.D., Dupont 2004):

For h, h_0 sufficiently small and $u \in V_h$

$$1 - C\frac{h^p}{\beta} \le \frac{\langle L_h u, u\rangle}{\langle (G_h)^{-1} u, u\rangle} \le 1 + C\frac{h^p}{\beta} \ .$$

Numerical results

First test case: one dimensional advection-diffusion equation

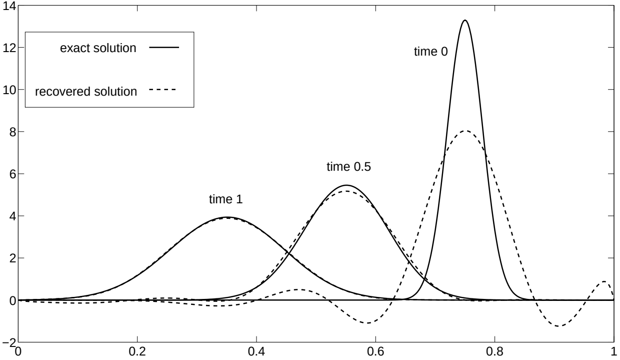
- Forward problem:

$$\partial_t z - \partial_x(a\partial_x z + bz) + cz = 0\ , \quad \text{on } (0,1), \quad z(\cdot,0) = u\ .$$

- We will test multigrid with up to 6 levels vs. conjugate gradient.
- Measures of success:
  - measure 1: cost(inverse problem) / cost(forward problem)
  - measure 2: cost(inverse problem) / cost(CG solve)

Numerical results

First test case: one dimensional advection-diffusion equation



Numerical results

First test case: one dimensional advection-diffusion equation

Table: Iteration count (I/F) for the W-cycle; $\beta = 10^{-3}$.

N	1		2		3		4		5		6	
200	15	(32.3)	11	(61.1)	9	(29.6)	7	(19.4)	6	(16.2)	5	(13.7)
400	16	(34.1)	9	(48)	7	(22.8)	6	(16.8)	5	(13.8)		
800	16	(34)	7	(38)	6	(19.8)	5	(14.4)				
1600	16	(34)	6	(32)	5	(16.9)						
3200	17	(36)	5	(26.7)								

Outline

Semilinear elliptic constraints (with Jyoti Saraswat)

- Optimal control problem:

$$\begin{aligned} &\text{minimize} && \frac{1}{2}\|y - y_d\|^2 + \frac{\beta}{2}\|u\|^2\ , \\ &\text{subj to:} && Ay + c_0 y + f(y) = u\ ,\quad u \in L^2(\Omega)\ . \end{aligned} \tag{3}$$

Assumptions and basic facts

- A is a uniformly elliptic operator on $\Omega \subset \mathbb{R}^d$ ($d = 2, 3$) with sufficiently smooth coefficients, $c_0 \geq 0$ is in L^∞.
- $f : \mathbb{R} \to \mathbb{R}$ is increasing, sufficiently smooth (C^3 will do).
- Monotone operator theory guarantees unique solution $u \to y(u) \in H_0^1$.
- Stampacchia technique produces L^∞-estimates for $y(u)$ independent of c_0, f: $\|y(u)\|_{L^\infty} \leq C_\infty \|u\|_{L^2}$.
- Full elliptic regularity is assumed: $y(u) \in H^2$.
- Mesh to allow for discrete FE maximum principle.

Reduced form of control problem

- Unconstrained optimal control problem:

$$\text{minimize} \quad \tfrac{1}{2}\|y(u) - y_d\|^2 + \tfrac{\beta}{2}\|u\|^2 \tag{4}$$

- Existence of optimal control $\bar{u} \in L^2(\Omega)$ guaranteed by standard techniques: optimal state $\bar{y} = y(\bar{u}) \in H^2(\Omega) \cap H_0^1(\Omega)$.
- Uniqueness of the optimal control $\bar{u}$ is not guaranteed in general.
- The optimal control problem may not be convex.

Solving the control problem

- The state is twice differentiable with respect to the control so the cost functional is twice differentiable.
- Apply Newton's method to solve the control problem:

$$u_{n+1} = u_n - \text{Hessian}^{-1}\text{gradient} .$$

- *Grid-sequencing* used to obtain good initial guess.
- *Adjoint methods* used to obtain gradients and the Hessian-vector multiplication.

Gradient and Hessian using adjoints

- $L = L(u) = A + f'(u)$ is the linearization of the semilinear operator at y.
- Gradient: $\nabla_u J(u) = (L^*)^{-1}(y(u) - y_d) + \beta u$.
- Hessian-vector multiplication:

$$G(u)v = {L^*}^{-1}(1 - f''(u)q(u))L^{-1}v + \beta v \ ,$$

where

$$q = q(u) = (L^*)^{-1}(y(u) - y_d) \ .$$

- Cost of Hessian-vector multiplication is equivalent to two linear elliptic solves.

Mesh independence of Newton's method

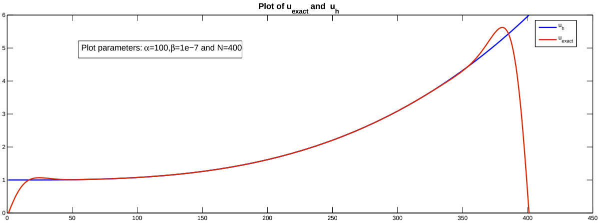


Table: Newton iterations

Resolution	50	100	200	250	300	350
Newton's iterations	4	4	4	4	4	4

Hessian and preconditioner

The Hessian:

$$G(u) = L^{*-1}(1 - f''(u)q(u))L^{-1} + \beta I$$

- As before, the Hessian is smoothing.
- Proposed two grid preconditioner:

$$M_h = \beta\rho + G_{2h}(\pi u)\pi$$

Two-grid preconditioner

Theorem (J. Saraswat, A.D., 2012)

On a quasi-uniform mesh and under usual elliptic regularity assumptions

$$\|(G_h(u) - M_h(u))v\| \leq Ch^2\|v\|, \ \ \forall v \in L^2(\Omega) \ ,$$

C independent of h.

Remark:

- Optimal order in h

One dimensional, in-vitro experiments

Table: Joint spectrum analysis in 1D: $f(y) = \alpha y^3$

N	$z_k = \max(\text{abs}(\ln d))$	ratio $= \frac{z_k}{z_{k+1}}$
10	2.426486	N/A
20	0.569206	4.262924
40	0.134355	4.236559
80	0.034536	3.890306
160	0.008709	3.965544
320	0.002182	3.990972
640	0.000545	3.997717

Here $d = eig(G_h, T_h)$.
The spectral distance between constructed preconditioner and Hessian is $O(h^2)$, which is the optimal rate.

Two-dimensional, in-vivo experiments with $f(y) = \alpha y^3$

Table: $\alpha = 1, \beta = 10^{-4}$

iterate N	16	32	64	128
1	7 (12)	6 (12)	4 (12)	4 (12)
2	7 (11)	5 (11)	4 (11)	4 (11)
3	4 (5)	3 (5)	2 (6)	1 (6)

Table: $\alpha = 1, \beta = 10^{-5}$

iterate N	16	32	64	128
1	11 (21)	8 (21)	5 (21)	4 (21)
2	10 (20)	8 (20)	5 (20)	4 (20)
3	5 (9)	4 (9)	2 (9)	2 (9)

Two-dimensional, in-vivo experiments

Table: $\alpha = 10, \beta = 10^{-5}$

iterate N	16	32	64	128
1	11 (21)	8 (21)	5 (21)	4 (21)
2	11 (20)	8 (20)	5 (20)	4 (20)
3	10 (16)	5 (16)	5 (16)	4 (16)
4	4 (8)	2 (8)	2 (8)	1 (8)

Table: $\alpha = 10, \beta = 10^{-7}$

iterate N	16	32	64	128
1	40 (76)	21 (93)	9 (99)	5 (98)
2	39 (65)	16 (72)	6 (71)	5 (71)
3	33 (50)	13 (48)	6 (49)	5 (46)
4	13 (12)	2 (12)	2 (12)	2 (12)

Outline

1. Model problems
2. Unconstrained problems with linear PDE constraints
3. Nonlinear constraints, control constraints
   - A semilinear elliptic constrained problem
   - Control-constrained problems
   - Optimal control problems constrained by the Stokes equations

Problem formulation

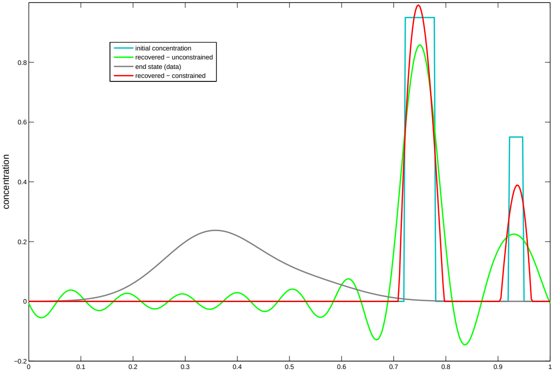
- Model problem:
  $K : L^2(\Omega) \to L^2(\Omega)$ compact, linear, $f \in L^2(\Omega)$

Optimal control problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ku - y_d\|^2 + \frac{\beta}{2}\|u\|^2 \\ \text{subj to:} & u \in L^2(\Omega), \quad a \le u \le b \end{array} \tag{5}$$

Why bound-constraints ?

- Physically meaningful, other qualitative considerations
- Example: solution is localized if the “true” solution is so



Discrete problem formulation

- Norms: discrete norm $\|u\|_h^2 = \sum w_i\, u^2(P_i)$
- Inequality constraints: $a \leq u \leq b$, enforced at nodes (strong enforcement)

Discrete optimal control problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|K_h u - y_{d,h}\|_h^2 + \frac{\beta}{2}\|u\|_h^2 \\ \text{subj to:} & u \in V_h, \quad a_h(P) \leq u(P) \leq b_h(P), \quad \forall \text{ node } P \end{array} \tag{6}$$

Discrete problem formulation

- Norms: discrete norm $\|u\|_h^2 = \sum w_i\, u^2(P_i)$
- Inequality constraints: $a \leq u \leq b$, enforced at nodes (strong enforcement)

Discrete optimal control problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|K_h u - y_{d,h}\|_h^2 + \frac{\beta}{2}\|u\|_h^2 \\ \text{subj to:} & u \in V_h, \quad a_h(P) \leq u(P) \leq b_h(P), \quad \forall \text{ node } P \end{array} \tag{6}$$

Discrete problem formulation

- Norms: discrete norm $\|u\|_h^2 = \sum w_i\, u^2(P_i)$
- Inequality constraints: $a \leq u \leq b$, enforced at nodes (strong enforcement)

Discrete optimal control problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|K_h u - y_{d,h}\|_h^2 + \frac{\beta}{2}\|u\|_h^2 \\ \text{subj to:} & u \in V_h, \quad a_h(P) \leq u(P) \leq b_h(P), \quad \forall \text{ node } P \end{array} \tag{6}$$

Optimization methods

- Optimization algorithms (outer iteration):
  - Semi-smooth Newton methods (active-set type strategies)
  - Interior point methods (IPM)
- Require: solving few linear systems at each outer iteration
  - semi-smooth Newton: subsystem (principal minor)
  - IPM: modified, same-size system
- Goals:
  - small # of outer iterations (prefer mesh-independence)
  - here: fast solvers for the linear systems:
    # of linear iterations **to decrease with increasing resolution**

Optimization methods

- Optimization algorithms (outer iteration):
  - Semi-smooth Newton methods (active-set type strategies)
  - Interior point methods (IPM)
- Require: solving few linear systems at each outer iteration
  - semi-smooth Newton: subsystem (principal minor)
  - IPM: modified, same-size system
- Goals:
  - small # of outer iterations (prefer mesh-independence)
  - here: fast solvers for the linear systems: # of linear iterations **to decrease with increasing resolution**

Optimization methods

- Optimization algorithms (outer iteration):
  - Semi-smooth Newton methods (active-set type strategies)
  - Interior point methods (IPM)
- Require: solving few linear systems at each outer iteration
  - semi-smooth Newton: subsystem (principal minor)
  - IPM: modified, same-size system
- Goals:
  - small # of outer iterations (prefer mesh-independence)
  - here: fast solvers for the linear systems:
    # of linear iterations **to decrease with increasing resolution**

A. Primal-dual interior point methods (with Cosmin Petra)

For fixed resolution V_h and uniform grids:

- solve perturbed KKT system for $\mu \downarrow 0$:

$$\begin{aligned} (\beta\mathbf{I} + \mathbf{K}^T\mathbf{K})\mathbf{u} - \mathbf{v} &= -\mathbf{K}^T\mathbf{y}_d \\ \mathbf{u}\cdot\mathbf{v} &= \mu\mathbf{e} \\ \mathbf{u}, \mathbf{v} &> \mathbf{0} \end{aligned}$$

- Mehrotra's predictor-corrector IPM

$$\begin{aligned} (\beta\mathbf{I} + \mathbf{K}^T\mathbf{K})\Delta\mathbf{u} - \Delta\mathbf{v} &= r_c \\ \mathbf{V}\Delta\mathbf{u} + \mathbf{U}\Delta\mathbf{v} &= r_a \end{aligned}$$

- reduced system

$$(\beta\mathbf{I} + \mathbf{U}^{-1}\mathbf{V} + \mathbf{K}^T\mathbf{K})\Delta\mathbf{u} = \mathbf{r}_c - \mathbf{U}^{-1}\mathbf{r}_a$$

with $\mathbf{U}, \mathbf{V}$ diagonal, positive

A. Primal-dual interior point methods (with Cosmin Petra)

For fixed resolution V_h and uniform grids:

- solve perturbed KKT system for $\mu \downarrow 0$:

$$\begin{aligned} (\beta \mathbf{I} + \mathbf{K}^T\mathbf{K})\mathbf{u} - \mathbf{v} &= -\mathbf{K}^T\mathbf{y}_d \\ \mathbf{u}\cdot\mathbf{v} &= \mu\mathbf{e} \\ \mathbf{u}, \mathbf{v} &> \mathbf{0} \end{aligned}$$

- Mehrotra's predictor-corrector IPM

$$\begin{aligned} (\beta \mathbf{I} + \mathbf{K}^T\mathbf{K})\Delta\mathbf{u} - \Delta\mathbf{v} &= r_c \\ \mathbf{V}\Delta\mathbf{u} + \mathbf{U}\Delta\mathbf{v} &= r_a \end{aligned}$$

- reduced system

$$(\beta \mathbf{I} + \mathbf{U}^{-1}\mathbf{V} + \mathbf{K}^T\mathbf{K})\Delta\mathbf{u} = \mathbf{r}_c - \mathbf{U}^{-1}\mathbf{r}_a$$

with $\mathbf{U}, \mathbf{V}$ diagonal, positive

A. Primal-dual interior point methods (with Cosmin Petra)

For fixed resolution V_h and uniform grids:

- solve perturbed KKT system for $\mu \downarrow 0$:

$$\begin{aligned} (\beta \mathbf{I} + \mathbf{K}^T\mathbf{K})\mathbf{u} - \mathbf{v} &= -\mathbf{K}^T\mathbf{y}_d \\ \mathbf{u}\cdot\mathbf{v} &= \mu\mathbf{e} \\ \mathbf{u},\mathbf{v} &> \mathbf{0} \end{aligned}$$

- Mehrotra's predictor-corrector IPM

$$\begin{aligned} (\beta \mathbf{I} + \mathbf{K}^T\mathbf{K})\Delta\mathbf{u} - \Delta\mathbf{v} &= r_c \\ \mathbf{V}\Delta\mathbf{u} + \mathbf{U}\Delta\mathbf{v} &= r_a \end{aligned}$$

- reduced system

$$(\beta \mathbf{I} + \mathbf{U}^{-1}\mathbf{V} + \mathbf{K}^T\mathbf{K})\Delta\mathbf{u} = \mathbf{r}_c - \mathbf{U}^{-1}\mathbf{r}_a$$

  with $\mathbf{U}, \mathbf{V}$ diagonal, positive

The systems

- the matrix: $(\beta\mathbf{I} + \mathbf{U}^{-1}\mathbf{V} + \mathbf{K}^T\mathbf{K})$
- $\mathbf{U}^{-1}\mathbf{V}$ represents a relatively smooth function
- need to invert

$$(D_{\beta+\lambda} + \underbrace{\mathbf{K}^T\mathbf{K}}_{K*K})$$

with $D_{\beta+\lambda} = \beta I + \mathbf{U}^{-1}V$

- ... and further

$$D_{\sqrt{\beta+\lambda}}(I + \underbrace{A\mathbf{K}^T\mathbf{K}A}_{(KA)^*(KA)})D_{\sqrt{\beta+\lambda}}$$

with $A = D_{\sqrt{1/(\beta+\lambda)}}$

The systems

- the matrix: $(\beta\mathbf{I} + \mathbf{U}^{-1}\mathbf{V} + \mathbf{K}^T\mathbf{K})$
- $\mathbf{U}^{-1}\mathbf{V}$ represents a relatively smooth function
- need to invert

$$(D_{\beta+\lambda} + \underbrace{\mathbf{K}^T\mathbf{K}}_{K^*K})$$

with $D_{\beta+\lambda} = \beta I + \mathbf{U}^{-1}V$

- ... and further

$$D_{\sqrt{\beta+\lambda}}(I + \underbrace{A\mathbf{K}^T\mathbf{K}A}_{(KA)^*(KA)})D_{\sqrt{\beta+\lambda}}$$

with $A = D_{\sqrt{1/(\beta+\lambda)}}$

The systems

- the matrix: $(\beta \mathbf{I} + \mathbf{U}^{-1}\mathbf{V} + \mathbf{K}^T\mathbf{K})$
- $\mathbf{U}^{-1}\mathbf{V}$ represents a relatively smooth function
- need to invert

$$(D_{\beta+\lambda} + \underbrace{\mathbf{K}^T\mathbf{K}}_{K*K})$$

with $D_{\beta+\lambda} = \beta I + \mathbf{U}^{-1}V$

- ... and further

$$D_{\sqrt{\beta+\lambda}}(I + \underbrace{A\mathbf{K}^T\mathbf{K}A}_{(KA)^*(KA)})D_{\sqrt{\beta+\lambda}}$$

with $A = D_{\sqrt{1/(\beta+\lambda)}}$

The systems

- Need good preconditioner for

$$G_h = I + (K_h A_h)^*(K_h A_h) = I + (L_h)^*(L_h)$$

with $A_h = D_{\sqrt{1/(\beta+\lambda_h)}}$

- Assume $\lambda_h = \text{interpolate}(\lambda)$

$$\begin{aligned} L_h &\stackrel{\text{def}}{=} K_h A_h \\ L &\stackrel{\text{def}}{=} K D_{\sqrt{1/(\beta+\lambda)}} \end{aligned}$$

Key facts

- $G_h = I + L_h{}^* L_h$ is **dense**, available only matrix-free
- $\mathrm{cond}(I + L_h{}^* L_h) = O(\beta^{-1})$, mesh-independent, large
- $A_h = D_{\sqrt{1/(\beta+\lambda_h)}}$ neutral with respect to smoothing
- $L_{(h)} = K_{(h)} A_{(h)}$ same smoothing properties as $K_{(h)}$

Two-grid preconditioner

Theorem (A.D. and Petra, 2009)

On a uniform grid

$$\rho(I - M_h{}^{-1} G_h) \leq C h^2 \|(\beta + \lambda)^{-\frac{1}{2}}\|_{W^2_\infty}$$

Remarks:

- optimal order in h
- quality expected to decay as $\mu \downarrow 0$ since λ only L^2 in general
- for fixed β # linear iterations/outer iteration expected to decrease with $h \downarrow 0$
- M_h is slightly non-symmetric

Backwards advection-diffusion problem example

Optimal control problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|S(T)u - y_d\|^2 + \frac{\beta}{2}\|u\|^2 \\ \text{subj to:} & u \in L^2(\Omega), \;\; 0 \leq u \leq 1 \end{array} \tag{7}$$

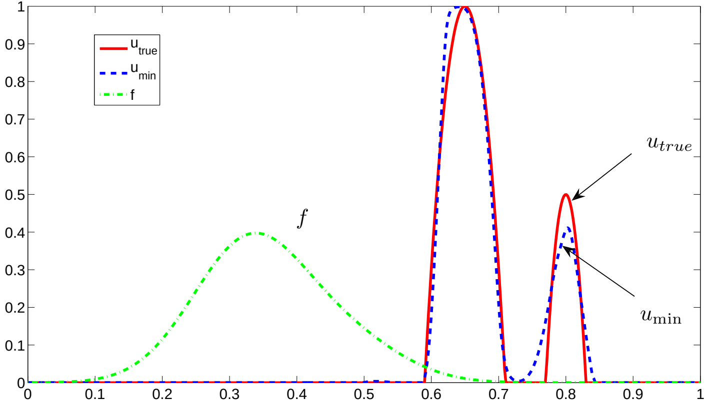
- $z(\cdot, t)$ transported quantity subjected to:

$$\begin{cases} \partial_t z - \nabla \cdot (a\nabla z + bz) + cz = 0 & \text{on } \Omega \\ z(x,t) = 0 & \text{for } x \in \partial\Omega,\ t \in [0,T] \\ z(x,0) = u(x) & \text{for } x \in \Omega \end{cases}$$

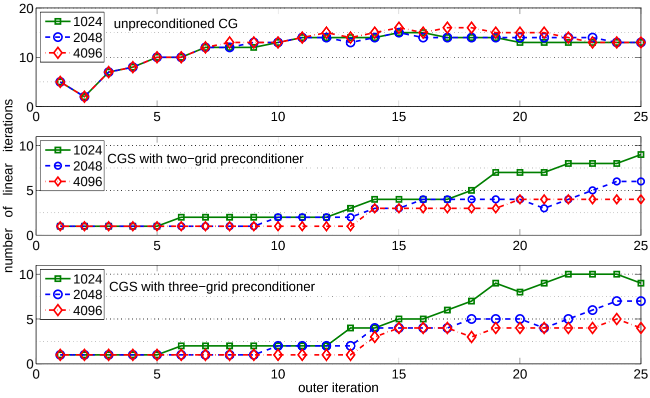
- $K = S(T)$: initial - to - final

$$K\,u = S(T)u \stackrel{\text{def}}{=} z(\cdot, T)$$

Solution

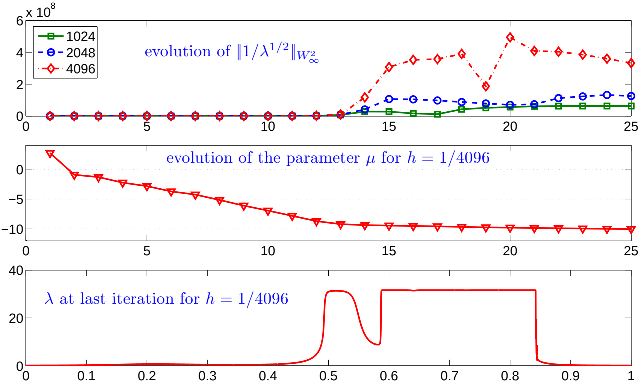


Iteration count / predictor-step linear systems



Evolution of quantities of interest

- Evolution of $\|\lambda^{-\frac{1}{2}}\|_{W^2_\infty}$, μ, and last λ_h:



Another measure of success

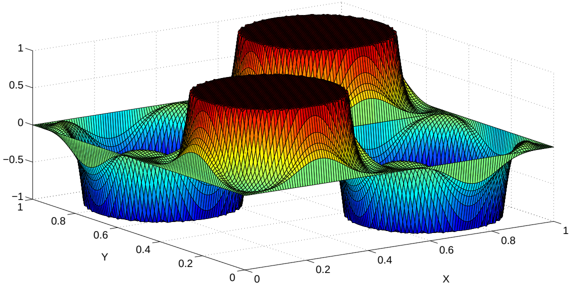
Total number of finest-level mat-vecs (application of K)

h \ levels	1	2	3
1/1024	728	581	661
1/2048	740	463	489
1/4096	764	403	425
1/8192	768	377	403

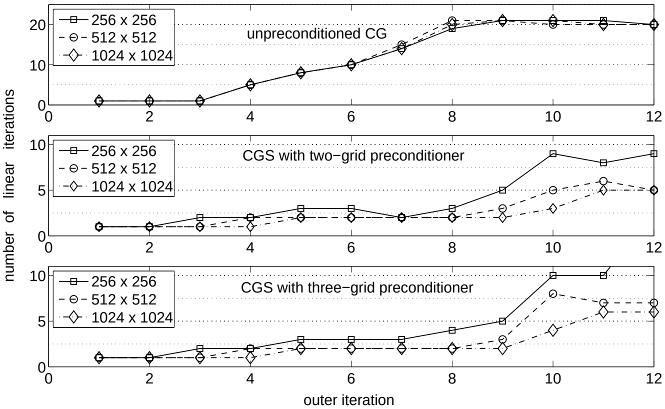
Elliptic-constrained problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|y-f\|^2 + \frac{\beta}{2}\|u\|^2 \\ \text{subj to:} & -\Delta y = u, \quad -1 \le u \le 1 \end{array}$$

$$\Delta f = \frac{3}{2}\sin(2\pi x)\,\sin(2\pi y),\ \beta = 10^{-6}$$



Iteration count / predictor-step linear systems



Mat-vecs count

Total number of finest-level mat-vecs (Poisson solves)

$h \setminus$ levels	1	2	3	4
1/256	354	282	572	–
1/512	355	220	250	452
1/1024	355	198	210	224
1/2048	363	172	174	174

B. Semismooth Newton methods

- KKT system (unperturbed):

$$\begin{aligned} (\beta \mathbf{I} + \mathbf{K}^T \mathbf{K})\mathbf{u} - \mathbf{v} &= -\mathbf{K}^T \mathbf{y}_d \\ \mathbf{u} \cdot \mathbf{v} &= \mathbf{0} \\ \mathbf{u}, \mathbf{v} &\geq \mathbf{0} \end{aligned}$$

- Reformulate as a semismooth nonlinear system:

$$\begin{aligned} (\beta \mathbf{I} + \mathbf{K}^T \mathbf{K})\mathbf{u} - \mathbf{v} &= -\mathbf{K}^T \mathbf{y}_d \\ \mathbf{v} - \max(\mathbf{0}, \mathbf{v} - \beta \mathbf{u}) &= \mathbf{0}\ . \end{aligned}$$

Active set strategy

- Define the *active index-set* by

$$\mathcal{A} = \{i \in \{1, \ldots, N\} \ : \ (\mathbf{v} - \beta \mathbf{u})_i > 0\}$$

and the *inactive index-set* by

$$\mathcal{I} = \{i \in \{1, \ldots, N\} \ : \ (\mathbf{v} - \beta \mathbf{u})_i \leq 0\} \ .$$

- The semismooth Newton method produces a sequence of active/inactive sets $(\mathcal{A}_k, \mathcal{I}_k)_{k=1,2,\ldots}$ that approximate $(\mathcal{A}, \mathcal{I})$.

Linear systems

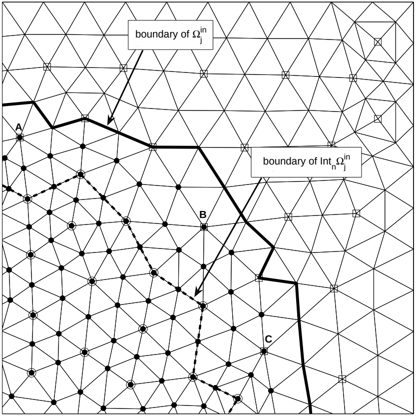
- The critical system to be solved is at each semismooth Newton iterate has the form

$$\mathbf{G}^{\mathcal{I}}\mathbf{u}_{\mathcal{I}} \stackrel{\text{def}}{=} (\beta\mathbf{I} + \mathbf{K}^{\mathsf{T}}\mathbf{K})^{\mathcal{I}\mathcal{I}}\mathbf{u}_{\mathcal{I}} = \mathbf{b}_{\mathcal{I}}\ .$$

  where $\mathcal{I}$ is the current guess at the inactive set.
- Similar preconditioning ideas can be applied: need a coarse space ${V_{2h}}^{\mathcal{I}} \subset {V_h}^{\mathcal{I}}$ then preconditioner is

$$\mathbf{M}_h = \beta(\mathbf{I} - {\pi_{2h}}^{\mathcal{I}}) + {\mathbf{G}_h}^{\mathcal{I}}{\pi_{2h}}^{\mathcal{I}}$$

Coarse space



Analysis

Theorem (A.D., 2011)

$$\rho(I - {M_h}^{-1}G_h) \leq C\beta^{-1}\left(h^2 + \sqrt{{\mu_h}^{\mathrm{in}}}\right) , \tag{8}$$

where ${\mu_h}^{\mathrm{in}}$ is the Lebesgue measure of $\partial_{\mathrm{n}}{\Omega_h}^{\mathrm{in}}$

- Preconditioner is expected to be of suboptimal quality:

$$\rho(I - {M_h}^{-1}G_h) \leq Ch^{\frac{1}{2}} .$$

Outline

Stokes control (with Ana Maria Soane)

- Model optimal control problem:

$$\begin{array}{ll} \text{minimize} & \frac{\gamma_u}{2}\|\vec{u}-\vec{u}_d\|^2 + \frac{\gamma_p}{2}\|p-p_d\|^2 + \frac{\beta}{2}\|\vec{f}-\vec{f}_0\|^2 \\ \text{subj to:} & -\nu\Delta\vec{u} + \nabla p = \vec{f}\ , \\ & \operatorname{div}\vec{u} = 0\ ,\ \vec{u}|_{\Omega} = \vec{0} \end{array}$$

- Identify force $\vec{f}$ closest to reference force $\vec{f}_0$ leading to given velocity and/or pressure “measurements” $\vec{u}_d,\ p_d$

The Hessian

- The Hessian:

$$G_h = \beta I + \gamma_u {U_h}^* U_h + \gamma_p {P_h}^* P_h$$

- The proposed two-grid preconditioner:

$$M_h = \beta\rho + G_{2h}\pi$$

$$L_h = (M_h)^{-1} = \beta^{-1}\rho + (G_{2h})^{-1}\pi$$

Two-grid preconditioner: Analysis

Theorem (A.D., A. Soane 2011)

With a Taylor-Hood $Q_2 - Q_1$ discretization and under regularity assumptions allowing for

$$\|(U - U_h)(f)\| \leq Ch^2\|f\|, \ \|(P - P_h)(f)\| \leq Ch\|f\|$$

we have

$$d_\sigma(G_h, M_h) \leq \frac{C}{\beta}\left(\gamma_u h^2 + \gamma_p h\right) \ ,$$

C independent of h, β, provided the coarsest grid is sufficiently fine.

Numerical Experiments – Pressure control

Table: Pressure measured only ($\gamma_u = 0, \ \gamma_p = 1$)

N	16			32			64			128			256	
no. levels	1	2	3	1	2	3	1	2	3	1	2	3	1	4
$\beta = 10^{-2}$	29	15	-	29	12	16	29	10	12	30	-	10	30	15
$\beta = 10^{-3}$	59	35	-	62	21	-	66	14	22	71	-	16	70	21

Time comparison at $n = 256$, number of state variables (velocity and pressure): 588290, number of control variables: 261121

no. levels	1	4
$\beta = 10^{-2}$	3460 s	2156 s
$\beta = 10^{-3}$	8457 s	2866 s

Matlab on 2$\times$ Intel (Nehalem) Xeon E5540 Quad Core (8M Cache, 2.53 GHz) CPUs with 24Gig RAM

Numerical Experiments – Velocity control

Table: Velocity measured only ($\gamma_u = 1, \ \gamma_p = 0$)

N	32			64			128			256		
no. levels	1	2	3	1	2	3	1	2	3	1	2	4
$\beta = 10^{-4}$	11	3	3	11	3	3	-	-	-	-	-	-
$\beta = 10^{-5}$	20	4	4	20	3	3	21	-	3	22	-	2
$\beta = 10^{-6}$	42	6	8	44	4	4	45	-	3	45	-	3

Time comparison at $n = 256$, number of state variables (velocity and pressure): 588290, number of control variables: 261121

no. levels	1	4
$\beta = 10^{-5}$	2622 s	393 s
$\beta = 10^{-6}$	5303 s	599 s

Conclusions

- Multigrid techniques open the possibility of solving an increasing class of large-scale PDE constrained optimal control problems at a reasonable cost.
- Main ingredients: a fast and reliable outer iteration (Newton, IPM, semismooth Newton), fast methods for the linear systems involved.
- Current techniques do not work as well for control-constrained problems (require special formulation, linear elements).

Future work and open problems

- Good preliminary results for steady-state Navier-Stokes controlled problems.
- Space-time PDEs and controls.
- Hyperbolic PDE constrained problems.
- Control-constrained problems: reconcile multigrid preconditioners for IPM and SSNM; handle higher order elements.
- State-constrained problems: will any of this work?