Multigrid Methods in PDE-Constrained Optimization

Andrei Draganescu
University of Maryland, Baltimore County

Workshop Exploratoriu: Teme Actuale de Cercetare in Matematici Aplicate
September 22-23, 2010

Acknowledgments

- *collaborators:*
  Cosmin Petra (Argonne National Lab),
  Todd Dupont (U Chicago), Volkan Akçelik (Exxon-Mobil),
  George Biros (Georgia Tech), Omar Ghattas (U Texas),
  Judith Hill (Oak Ridge National Lab), Jungho Lee (Oak Ridge National Lab), Bart van Bloemen Waanders (Sandia),
  Jyoti Saraswat (UMBC), *Na Rae Lee* (UMBC),

- *sponsors:*
  NSF, DOE

Outline

Problem formulation

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

Optimal control problem

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

- Motivating examples:
  1. K time-T solution operator of a parabolic equation
  2. K image-blurring (integral) operator
  3. $K = -\Delta^{-1}$ - elliptic optimal control problem

Motivating applications

1. Backward (inverse) advection-diffusion problems:

- $T > 0$ fixed "end-time", f end-time state, u_0 initial state
- $u(\cdot, t)$ transported quantity subjected to:

$$\begin{cases} \partial_t u - \nabla \cdot (a\nabla u + bu) + cu = 0 & \text{on } \Omega \\ u(x,t) = 0 & \text{for } x \in \partial\Omega,\ t \in [0,T] \\ u(x,0) = u_0(x) & \text{for } x \in \Omega \end{cases}$$

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

$$S\, u_0 = S(T)u_0 \stackrel{\text{def}}{=} u(\cdot, T)$$

Motivating applications

2. Image deblurring:

- K Fredholm first kind integral operator (blurring)

$$(Ku)(x) = \int k(x-y)u(y)$$

- f blurred image, $u : \Omega \to [0,1]$ correct image

3. Elliptic optimal control problem:

- PDE-constrained optimal control problem

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}\|y-f\|^2 + \tfrac{\beta}{2}\|u\|^2 \\ &\text{subj to:} && -\Delta y = u \\ & && a \le u \le b \end{aligned}$$

Motivating applications

2. Image deblurring:

- K Fredholm first kind integral operator (blurring)

$$(Ku)(x) = \int k(x-y)u(y)$$

- f blurred image, $u : \Omega \to [0,1]$ correct image

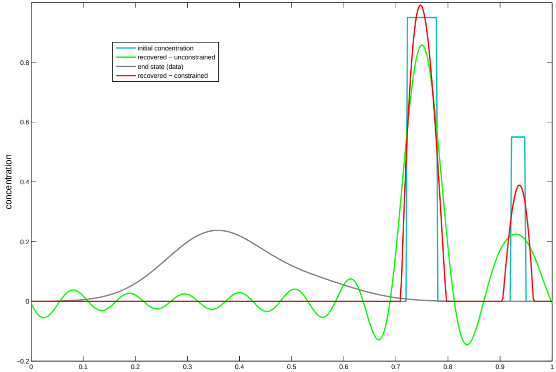
3. Elliptic optimal control problem:

- PDE-constrained optimal control problem

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

Why bound-constraints ?

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



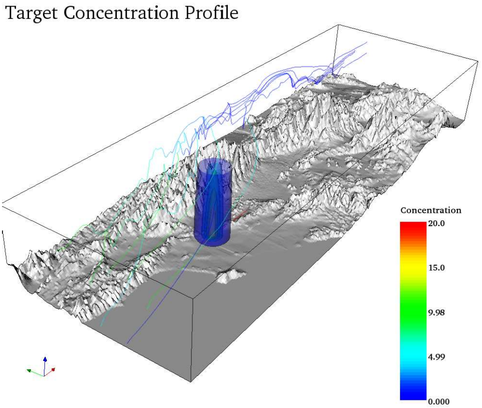
Context

- Large-scale problems:
  3D: $1000 \times 1000 \times 100 = 10^8$ spatial unknowns
- K treated as black-box
- Multiple resolutions of K are available
- Unconstrained problem can be solved efficiently using multigrid
- Goal: find a highly efficient solution method for inequality-constrained problems

Context

- Large-scale problems:
  3D: $1000 \times 1000 \times 100 = 10^8$ spatial unknowns
- K treated as black-box
- Multiple resolutions of K are available
- Unconstrained problem can be solved efficiently using multigrid
- Goal: find a highly efficient solution method for inequality-constrained problems

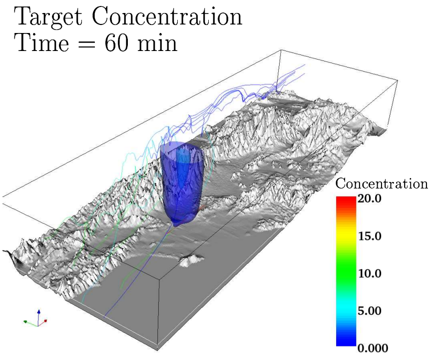
Scenario: air contamination event in the LA Basin
initial event



courtesy of Omar Ghattas

Scenario: air contamination event in the LA Basin

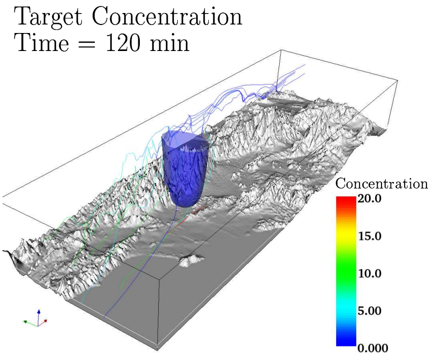
evolution: 60 minutes



courtesy of Omar Ghattas

Scenario: air contamination event in the LA Basin

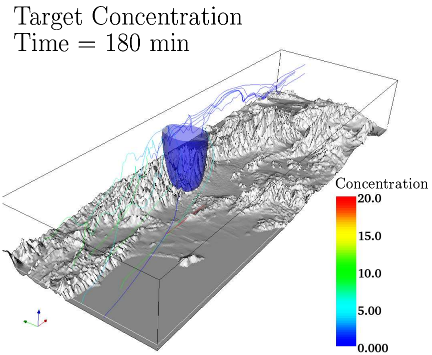
evolution: 120 minutes



courtesy of Omar Ghattas

Scenario: air contamination event in the LA Basin

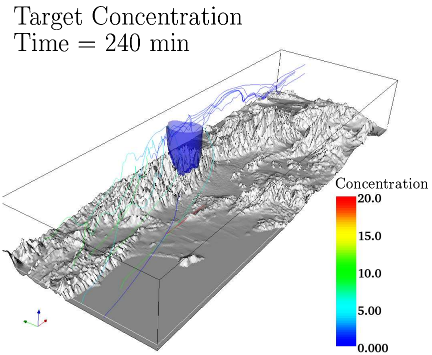
evolution: 180 minutes



courtesy of Omar Ghattas

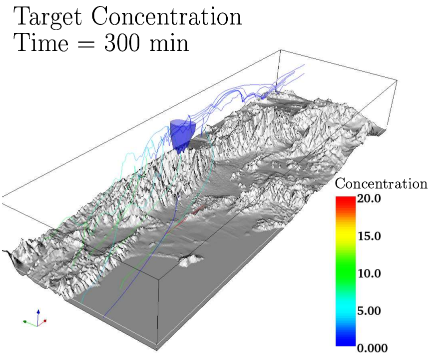
Scenario: air contamination event in the LA Basin

evolution: 240 minutes



courtesy of Omar Ghattas

Scenario: air contamination event in the LA Basin
evolution: 300 minutes



courtesy of Omar Ghattas

Related work

- Unconstrained
  Hackbusch, King (92), Rieder (97), Hanke and Vogel (99), Kaltenbacher (03), Draganescu and Dupont (08),
  Akcelik, Biros, Draganescu, Ghattas, Hill, and van Bloemen Waanders (05), Biros and Dogan (08)
- Control-constrained
  Maar and Schulz (00), Borzi and Kunisch (05), Vallejos and Borzi (08), Benzi, Haber, and Taralli (09)
- Review
  Borzi and Schulz (SIAM Review, June 09)

Discretization

Discretize-then-Optimize strategy

- Spaces: conforming FE, linear splines (V_h)
- Operators: $K_h : V_h \to V_h$ satisfy
  1. smoothing:

$$\|Ku\|_{H^m(\Omega)} \leq C \|u\|, \quad \forall u \in L^2(\Omega),\ m = 0, 1, 2$$

  2. smoothed approximation: for all h

$$\|Ku - K_h u\|_{H^m(\Omega)} \leq C h^{2-m} \|u\| \quad \forall u \in V_h,\ m = 0, 1$$

Discretization

Discretize-then-Optimize strategy

- Spaces: conforming FE, linear splines (V_h)
- Operators: $K_h : V_h \to V_h$ satisfy
  1. smoothing:

$$\|Ku\|_{H^m(\Omega)} \leq C \|u\|, \quad \forall u \in L^2(\Omega), \ m = 0, 1, 2$$

  2. smoothed approximation: for all h

$$\|Ku - K_h u\|_{H^m(\Omega)} \leq Ch^{2-m} \|u\| \quad \forall u \in V_h, \ m = 0, 1$$

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

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

Discrete problem formulation

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

Discrete optimal control problem

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

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

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

Optimization methods

- Optimization algorithms (outer iteration):
  - semi-smooth Newton methods (active-set type strategies)
  - here: 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)
  - here: 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)
  - here: 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**

Primal-dual interior point methods

For fixed resolution V_h:

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

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

- Mehrotra's predictor-corrector IPM

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

- reduced system

$$(\beta W + U^{-1}V + K^T W K)\Delta u = r_c - U^{-1} r_a$$

with U, V diagonal, positive

Primal-dual interior point methods

For fixed resolution V_h:

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

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

- Mehrotra's predictor-corrector IPM

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

- reduced system

$$(\beta W + U^{-1} V + K^T W K)\Delta u = r_c - U^{-1} r_a$$

with U, V diagonal, positive

Primal-dual interior point methods

For fixed resolution V_h:

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

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

- Mehrotra's predictor-corrector IPM

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

- reduced system

$$(\beta W + U^{-1} V + K^T W K)\Delta u = r_c - U^{-1} r_a$$

with U, V diagonal, positive

The systems

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

$$(D_{\beta+\lambda} + \underbrace{W^{-1}K^T WK}_{K^*K})$$

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

- ... and further

$$D_{\sqrt{\beta+\lambda}}(I + \underbrace{W^{-1}AK^T WKA}_{(KA)^*(KA)})D_{\sqrt{\beta+\lambda}}$$

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

The systems

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

$$(D_{\beta+\lambda} + \underbrace{W^{-1}K^T W K}_{K^*K})$$

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

- ... and further

$$D_{\sqrt{\beta+\lambda}}(I + \underbrace{W^{-1}AK^T WKA}_{(KA)^*(KA)})D_{\sqrt{\beta+\lambda}}$$

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

The systems

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

$$(D_{\beta+\lambda} + \underbrace{W^{-1}K^T WK}_{K^*K})$$

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

- ... and further

$$D_{\sqrt{\beta+\lambda}}(I + \underbrace{W^{-1}AK^T WKA}_{(KA)^*(KA)})D_{\sqrt{\beta+\lambda}}$$

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

- Scale separation:

$$V_h = \overbrace{V_{2h}}^{\text{“smooth“ functions}} \oplus \overbrace{W}^{\text{“rough“ functions}}$$

- $\pi = \pi_{2h} = L^2-$projection onto V_{2h}

$$G_h = (\pi + \rho)(I + L_h{}^* L_h)(\pi + \rho) \approx \rho + \underbrace{\pi(I + L_h{}^* L_h)\pi}_{\approx G_{2h}\pi}$$

Preconditioner

$$G_h \approx M_h \stackrel{\text{def}}{=} \rho \oplus G_{2h}\pi$$
$$M_h{}^{-1} = \rho + G_{2h}{}^{-1}\pi$$

Two-grid preconditioner

- Scale separation:

$$V_h = \overbrace{V_{2h}}^{\text{“smooth“ functions}} \oplus \overbrace{W}^{\text{“rough“ functions}}$$

- $\pi = \pi_{2h} = L^2-$projection onto V_{2h}

$$G_h = (\pi + \rho)(I + L_h{}^* L_h)(\pi + \rho) \approx \rho + \underbrace{\pi(I + L_h{}^* L_h)\pi}_{\approx G_{2h}\pi}$$

Preconditioner

$$G_h \approx M_h \stackrel{\text{def}}{=} \rho \oplus G_{2h}\pi$$
$$M_h{}^{-1} = \rho + G_{2h}{}^{-1}\pi$$

Two-grid preconditioner

Theorem (A.D. and C.P., 2009)

On a uniform grid

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

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

Multigrid preconditioner

- Multigrid preconditioner definition:

$$L_h \stackrel{\text{def}}{=} \mathcal{N}_{G_h}(L_{2h}\pi + \rho)$$

where

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

In vitro system I: backwards advection-diffusion

$$K = S(T), \quad \lambda(x) = \sin(\pi x)$$

$$d_\sigma(G, M) = \max\{ |\ln \lambda| \;:\; \lambda \in \sigma(G, M) \}$$

Table: approx. $d_\sigma \approx \rho(I - M_h^{-1} G_h)$

$h \setminus \beta$	1		0.1		0.01	
	d_σ	rate	d_σ	rate	d_σ	rate
1/80	0.0206		0.1127		0.2812	
1/160	0.0066	3.1342	0.0363	3.1078	0.1270	2.2140
1/320	0.0020	3.3140	0.0102	3.5488	0.0445	2.8535
1/640	0.0006	3.5199	0.0027	3.7365	0.0123	3.6284

In vitro system II: elliptic-constrained problem

$$K = -\Delta^{-1}, \quad \lambda(x) = \sin(\pi x)$$

$$d_\sigma(G, M) = \max\{ |\ln \lambda| \ : \ \lambda \in \sigma(G, M) \ \}$$

Table: approx. $d_\sigma \approx \rho(I - M_h^{-1} G_h)$

$h \setminus \beta$	1		0.1		0.01	
	d_σ	rate	d_σ	rate	d_σ	rate
1/80	8.66e-003		5.03e-002		2.21e-001	
1/160	2.15e-003	4.0318	9.20e-003	5.4691	8.23e-002	2.6782
1/320	5.36e-004	4.0106	1.18e-003	7.7656	2.61e-002	3.1517
1/640	1.34e-004	4.0039	1.91e-004	6.2046	6.51e-003	4.0161
1/1280	3.35e-005	4.0016	4.86e-005	3.9275	1.07e-003	6.0762

Backwards advection-diffusion problem example

Optimal control problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|S(T)u - f\|^2 + \frac{\beta}{2}\|u\|^2 \\ \text{subj to:} & u \in L^2(\Omega), \quad 0 \le u \le 1 \end{array} \qquad (3)$$

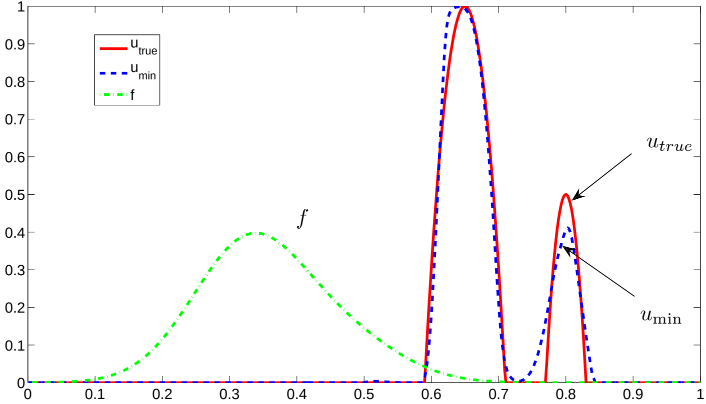
- $u(\cdot, t)$ transported quantity subjected to:

$$\begin{cases} \partial_t u - \nabla \cdot (a\nabla u + bu) + cu = 0 & \text{on } \Omega \\ u(x,t) = 0 & \text{for } x \in \partial\Omega, \ t \in [0,T] \\ u(x,0) = u_0(x) & \text{for } x \in \Omega \end{cases}$$

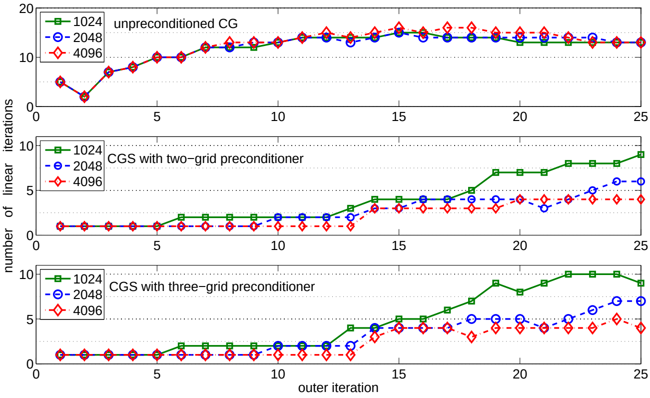
- $K = S(T)$: initial - to - final

$$S\, u_0 = S(T)u_0 \stackrel{\text{def}}{=} u(\cdot, T)$$

Solution

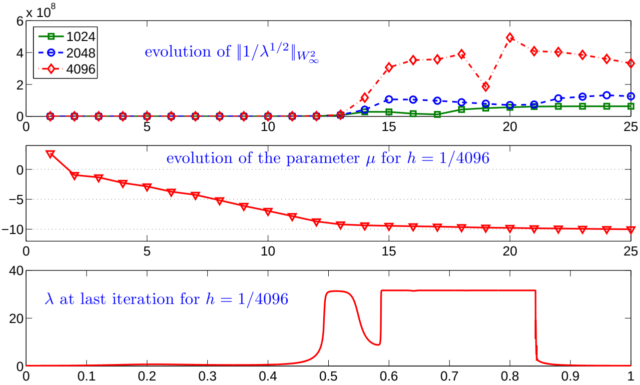

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
u_{true}
u_{min}
f
f
u_{true}
$u_{\min}$

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

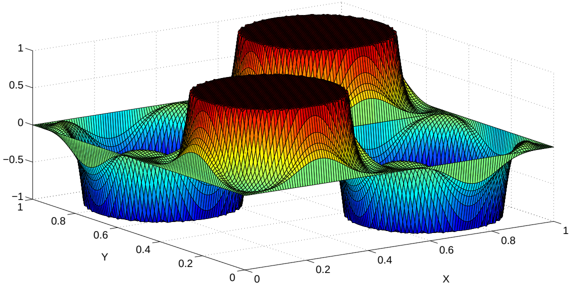
Total number of finest-level mat-vecs (application of K)

$h \setminus$ 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

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

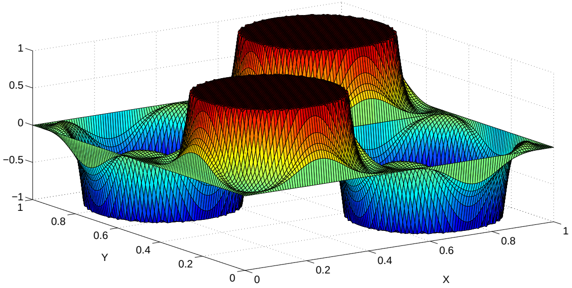
$$\Delta f = \frac{3}{2}\sin(2\pi x)\,\sin(2\pi y),\ \beta = 10^{-6}$$



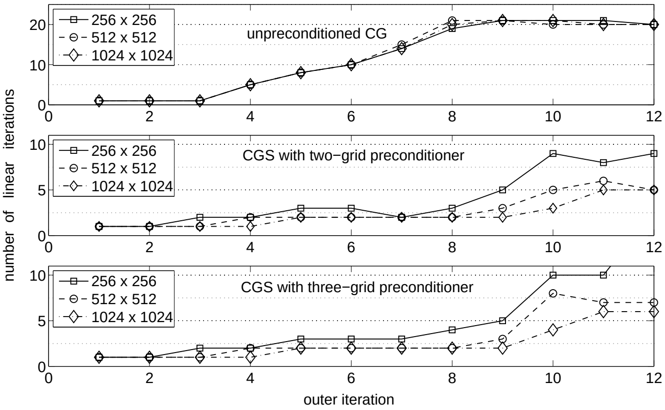
Elliptic-constrained problem

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

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



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

Current and future directions

- Inverse systems of semilinear advection-reaction-diffusion equations (Saraswat, N. R. Lee)
- Inverse hyperbolic problems (Hill, J. Lee)
- State constrained problems
- Long-term goal: efficient solution of large-scale data assimilation problems (4D Var method for weather and climate modeling)

Multigrid preconditioner

- Multigrid preconditioner definition:

$$L_h \stackrel{\text{def}}{=} \mathcal{N}_{G_h}(L_{2h}\pi + \rho)$$

where

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