A Newton-Multigrid Method for PDE Constrained Optimization

Martin Engel

Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn

Workshop on PDE Constrained Optimization
University of Hamburg
March 27-29, 2008

Outline

Outline

Problem Formulation

Optimal Control Problem

$$\begin{aligned} \text{Minimize} \quad J(y,u) &= \frac{1}{2}||y-\bar{y}||_{L^2}^2 + \frac{\sigma}{2}||u||_{L^2}^2 \\ \text{subject to} \quad C(y,u) &= Ly + N(y) - u - f = 0 \end{aligned}$$

- Y $(= H_0^1(\Omega))$ and U $(= L^2(\Omega))$ Hilbert spaces
- (Distributed) control $u \in U$
  (later: $u \in U_{ad} \subset U$ a closed and convex set)
- Target state $\bar{y} \in U$
- L second-order elliptic operator, $C : Y \times U \to W$

Existence and Uniqueness

Lagrangian

$$L(y,u,p) = J(y,u) - \langle C(y,u), p \rangle_{W,W'}$$

First-Order Optimality Conditions (Equality-Constrained Case)

$$\nabla L(y^*,u^*,p^*) = \begin{bmatrix} \nabla_y J(y^*,u^*) + (C_y(y^*,u^*))^* p^* \\ \nabla_u J(y^*,u^*) + (C_u(y^*,u^*))^* p^* \\ C(y^*,u^*) \end{bmatrix} = 0$$

Outline

Newton-Lagrange, SQP

Applying Newton's method to the first-order conditions yields a sequence of quadratic problems (QPs):

Newton Step

$$\begin{bmatrix} L_{yy}(y,u,p) & L_{yu}(y,u,p) & (C_y(y,u))^* \\ L_{uy}(y,u,p) & L_{uu}(y,u,p) & (C_u(y,u))^* \\ C_y(y,u) & C_u(y,u) & 0 \end{bmatrix} \begin{bmatrix} \delta y \\ \delta u \\ \delta p \end{bmatrix}$$

$$= - \begin{bmatrix} \nabla_y J(y,u) + (C_y(y,u))^* p \\ \nabla_u J(y,u) + (C_u(y,u))^* p \\ C(y,u) \end{bmatrix}$$

The Discrete QPs

The (discretized) QP in each Newton step is a linear

Saddle Point System

$$\begin{bmatrix} W_{11} & W_{12} & C_1^T \\ W_{21} & W_{22} & C_2^T \\ C_1 & C_2 & 0 \end{bmatrix} \begin{bmatrix} y \\ u \\ p \end{bmatrix} = \begin{bmatrix} f \\ g \\ h \end{bmatrix} \tag{1}$$

Solution Methods

- Direct factorization methods (memory, complexity)
- Iterative methods (efficient preconditioning)
- Multigrid (suitable relaxation method)

Full Space vs. Reduced Space Methods

Reduced Space

- Elimination yields reduced system in (lower dimensional) space of controls
- Feasibility in each step
- Can increase nonlinearity ([Ghattas and Bark])

Full Space

- Simultaneous solution for all unknowns
- Feasibility at minimizer
- Ill-conditioned system requires effective preconditioning

Remark: preconditioning, relaxation and different degrees of inexactness tend to blur the line between reduced and full space methods.

Multigrid in Optimization

Multigrid in Inner Iterations

- MG for Fredholm operator of the 2nd kind [Hackbusch '80]
- MG solver for state and adjoint in RSQP [Schulz '96]

MG for Outer Iteration

- Collective Relaxation [Brandt '84, Vanka '86, Borzi '02, Ascher and Haber '03]
- Relaxation as inexact block-factorization
  - Gradient Descent [Arian and Ta'asan '94]
  - Range Space Iterations [Braess and Sarazin '97, Wittum '89]
  - Null Space Iterations [Maar and Schulz '00]
  - Constraint Preconditioner (based on [Keller, Gould and Wathen '00])
- MG/Opt [Brandt '84, Lewis and Nash '00, Borzi '05]

Constraint Preconditioner

Consider a *constraint preconditioner*

$$P = \begin{pmatrix} G_{11} & G_{12} & C_1^T \\ G_{21} & G_{22} & C_2^T \\ C_1 & C_2 & 0 \end{pmatrix} \quad \text{for} \quad K = \begin{pmatrix} W_{11} & W_{12} & C_1^T \\ W_{21} & W_{22} & C_2^T \\ C_1 & C_2 & 0 \end{pmatrix}$$

Then the preconditioned matrix $P^{-1}K$ has ([Keller, Gould and Wathen '00]):

- an eigenvalue 1 with multiplicity 2N
- M eigenvalues which are defined by the generalized eigenvalue problem

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

  where Z is a basis for the nullspace of the linearized constraints
- the dimension of $\mathcal{K}(P)$ is at most M+2

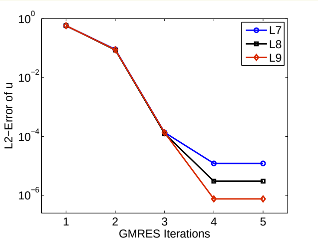
Constraint Preconditioner, cont'd

$$P_{CP} = \begin{pmatrix} 0 & 0 & C_1^T \\ 0 & B_Z & C_2^T \\ C_1 & C_2 & 0 \end{pmatrix}$$

P_{CP}-GMRES converges in three iterations, provided that the true reduced Hessian is used, i.e. $B_Z = H_Z = Z^T W Z$, where

$$H_Z = C_2^T C_1^{-T} W_{11} C_1^{-1} C_2 - C_2^T C_1^{-T} W_{12} - W_{21} C_1^{-1} C_2 + W_{22}$$

for the fundamental basis $Z = \begin{bmatrix} -C_2^T C_1^{-T} & I \end{bmatrix}$



- 2 PDE solves (C_1^{-1}, C_1^{-T})
- 2 PDE solves per MatVec with H_Z

Inexact Constraint Preconditioner as Smoother

One application of P_{CP}^{-1} results in the following algorithm:

1. solve for p: $C_1^T p = f_y$
2. solve for u: $H_z g = f_u - C_2^T p$
3. solve for y: $C_1 y = f_p - C_2 u$

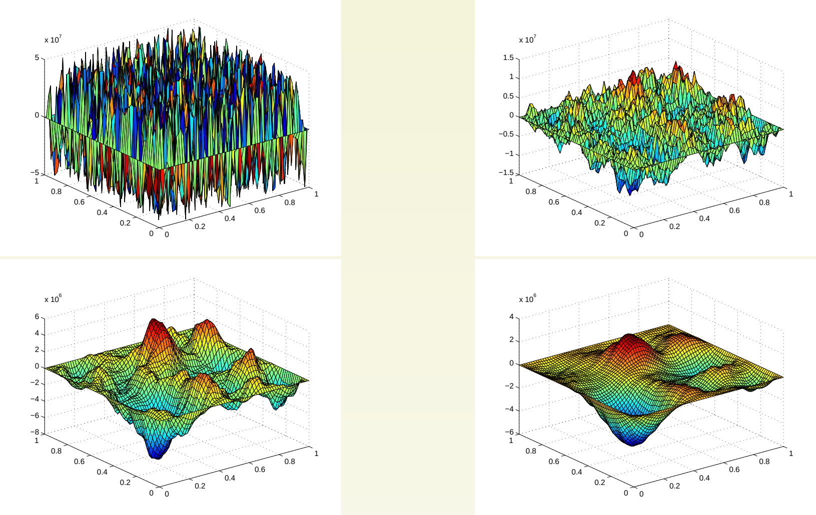
Inexact version of $\tilde{P}_{CP}$:

- replace the solve in 1. with relaxation $\tilde{C}_1^T$ for C_1^T
- replace the solve in 3. with relaxation $\tilde{C}_1$ for C_1
- solve 2. with a fixed (small) number of CG iterations.
- use relaxations $\tilde{C}_1^T$ and $\tilde{C}_1$ to approximate matrix-vector product with H_z in 2

Use $\tilde{P}_{CP}$ as preconditioner in Richardson iteration

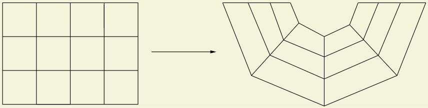
Illustration of Smoothing Behaviour

Error in controls, initial value and iterations 2,3,4.



Multigrid Components I

$$
\begin{aligned}
-\nabla \cdot K \nabla y &= f \quad \text{in } \Omega \\
y &= 0 \quad \text{on } \partial\Omega
\end{aligned}
$$



$F : \hat{\Omega} \rightarrow \Omega$ smooth map, RT_0-spaces are reduced to cell-centered FD by applying suitable quadrature rules [Arbogast, Wheeler, Yotov]

- Standard coarsening: $h_k = h_{\text{coarse}} 2^{-k}$
- Direct coarse grid discretization. For $h_k, k = 0, \ldots, L-1$
  - Grid sequence $\hat{\Omega}_{h_k}$
  - Discretized constraints $L_{h_k} y_{h_k} = M_{h_k} u_{h_k}$, i.e. on level k we have $C_1 = L_{h_k}$ and $C_2 = M_{h_k}$

Multigrid Components II

- Coarse grid solver is 3-step constraint-preconditioned Krylov method
- V-, W-, and F-Cycles with ν_1 pre- and ν_2 postsmoothing steps

$$x_k^{m+1} \leftarrow MG^\gamma(k, b_k, x_k^m) \; (\gamma = 1 : V, \gamma = 2 : W)$$

if $k = 0$ **then**
  Solve $K_0 x_0 = b_0$
else
  Presmooth $\tilde{x}_k = S^{\nu_1} x_k^m$
  Residual restriction $b_{k-1} = \mathcal{I}_k^{k-1}(b_k - K_k \tilde{x}_k)$
  Grid recursion $v_{k-1} = MG^\gamma(k-1, b_{k-1}, 0)$
  Correction $\tilde{\tilde{x}}_k = \tilde{x}_k + \mathcal{I}_{k-1}^k v_{k-1}$
  Postsmooth $x_k^{m+1} = S^{\nu_2} \tilde{\tilde{x}}_k$
end if

- Variable V-Cycle: $\nu_k = 2^{L-1-k} \nu_f$
- F-Cycle recursively defined as F-Cycle followed by V-Cycle

Interpolation Operators for CCFD

- Blockwise definition

$$\mathcal{I}_h^H = \begin{pmatrix} I_h^H & & \\ & I_h^H & \\ & & I_h^H \end{pmatrix}$$

- I_h^H four point average (FPA)

$$I_{h,FPA}^H = \frac{1}{4}\begin{bmatrix} 1 & & 1 \\ & \cdot & \\ 1 & & 1 \end{bmatrix}$$

- I_H^h bilinear (BL) or Wesseling/Khalil (WK) interpolation

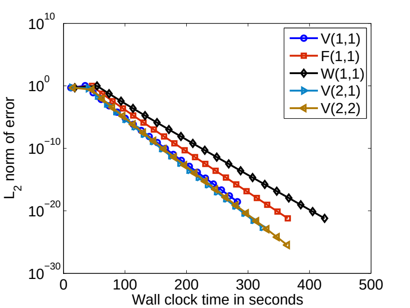
$$I_{H,BL}^h = \frac{1}{16}\left]\begin{matrix} 1 & 3 & 3 & 1 \\ 3 & 9 & 9 & 3 \\ 3 & 9 & 9 & 3 \\ 1 & 3 & 3 & 1 \end{matrix}\right[\qquad I_{H,WK}^h = \frac{1}{16}\left]\begin{matrix} 1 & 1 & & \\ 1 & 3 & 2 & \\ & 2 & 3 & 1 \\ & & 1 & 1 \end{matrix}\right[$$

- $m_p + m_r = 4 > 2$

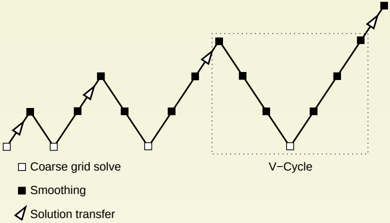
Numerical Results, MG, Linear Model Problem

Avg. rates of convergence for different cycle types on the Linear Model Problem

Cycle \ L	5	6	7	8	9	10	11	12
V(1,1)	1.02_{-1}	1.08_{-1}	1.13_{-1}	1.14_{-1}	1.14_{-1}	1.14_{-1}	1.15_{-1}	1.15_{-1}
V(2,1)	6.37_{-2}	6.70_{-2}	7.07_{-2}	7.21_{-2}	7.30_{-2}	7.31_{-2}	7.31_{-2}	7.31_{-2}
V(2,2)	4.78_{-2}	4.85_{-2}	5.15_{-2}	5.26_{-2}	5.34_{-2}	5.35_{-2}	5.35_{-2}	5.35_{-2}
F(1,1)	8.18_{-2}	8.16_{-2}	8.18_{-2}	8.20_{-2}	8.20_{-2}	8.20_{-2}	8.21_{-2}	8.21_{-2}
W(1,1)	8.18_{-2}	8.16_{-2}	8.18_{-2}	8.20_{-2}	8.20_{-2}	8.20_{-2}	8.21_{-2}	8.21_{-2}



The Full Multigrid Algorithm



Full Multigrid

$k = 0$:

$\quad x_0^{FMG} \leftarrow Kx_0 = b_0$

for $k = 1, 2, \ldots, L$ **do**

$\quad x_k^0 = \Pi_{k-1}^k x_{k-1}^{FMG}$

$\quad x_k^{FMG} = MG^r(x_k^0, k+1, \gamma)$

end for

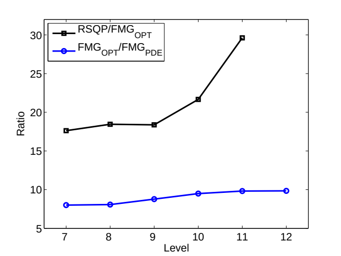
FMG Properties

- $\mathcal{O}(N)$ operations to compute x_h^{FMG} $\left(W_l^{FMG} = \frac{4}{3} r W_l \text{ in 2D}\right)$
- $||x_l - x_l^{FMG}|| \sim ||x - x_l||$ $\left(||x_l - x_l^{FMG}|| \leq C_{\rho,r} h^\alpha, \quad ||M_l|| \leq \rho < 1\right)$

Numerical Results, FMG, Linear Model Problem

L_2-norm of discretization error and wall-clock time in seconds for one FMG cycle. On each level, a V(1,1) cycle is used.

Level	h	N	Time[s]	ratio	L_2-Error	ratio
7	1/128	49152	0.1680	—	6.06389E-05	—
8	1/256	196608	0.7578	4.511	1.59350E-05	0.263
9	1/512	786432	3.6797	4.856	4.06443E-06	0.255
10	1/1024	3145728	16.4102	4.459	1.02422E-06	0.252
11	1/2048	12582912	68.3516	4.165	2.56857E-07	0.251
12	1/4096	50331648	276.5470	4.041	6.42940E-08	0.250



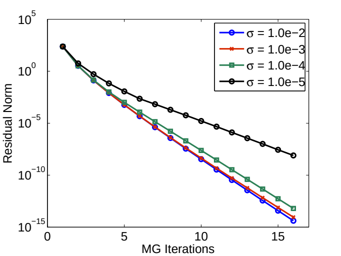
Costs of Relaxation

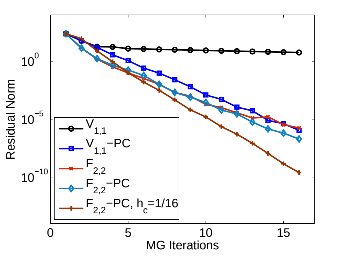
- MatVec K: $2C_N$
- Constraint blocks: $2C_N$
- MatVec B_Z (CG iter): $2C_N$
- MatVec B_Z (CG init): $2C_N$

Total: $8C_N$, plus: inner products,...

Convergence and L^2-Regularization

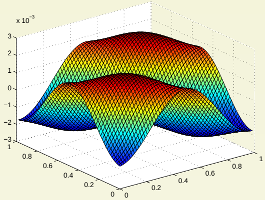
For small σ, the coarse grid correction with discrete H_Z is not effective

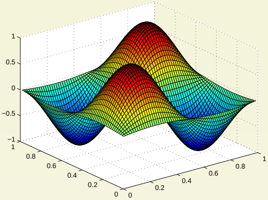


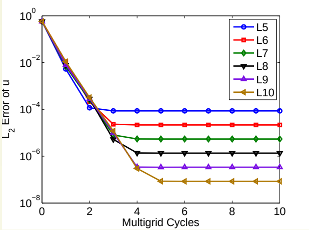


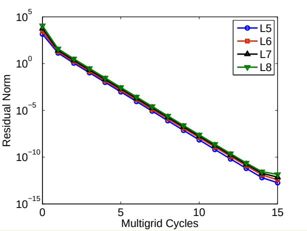
- Use more robust cycle (F, W, variable-V)
- Use MG as preconditioner for GMRES
- Preconditioning H_Z?
- Learn from Helmholtz/Convection-Diffusion?
- Choosing appropriate coarse grid size h_c only "sure thing"

Numerical Results, Diffusion Problem





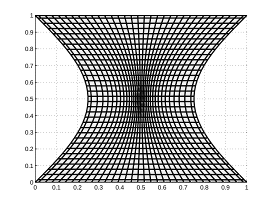


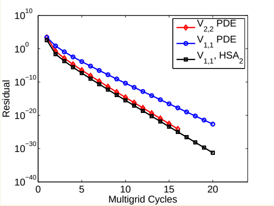


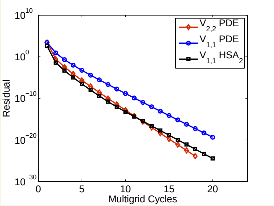
$$K = \begin{bmatrix} 11 & 9 \\ 9 & 13 \end{bmatrix}$$

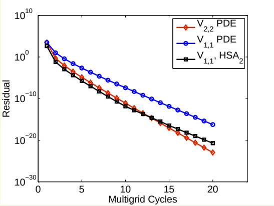
Target state $\bar{y}$ and controls u (top left to right), L_2-error of y and KKT residual. ALGS-Smoother for state and adjoint. Avg. residual reduction rate 0.0987.

Numerical Results, Non-Uniform Grid









Mesh deformation (top left), convergence history on Level 11 for $\delta = 0.1$ (top right), $\delta = 0.15$ (bottom left), $\delta = 0.2$ (bottom right). Table: convergence rates for $\delta = 0.2$.

PDE relaxation with ALGS (similar results are obtained for ILU smoothing.)

	7	8	9	10	11
$V_{2,2}$ PDE	0.0390	0.0502	0.0571	0.0622	0.0654
$V_{1,1}$ PDE	0.0997	0.1155	0.1217	0.1301	0.1313
$V_{1,1}$ KKT	0.1332	0.1352	0.1363	0.1350	0.1316

Inexact Newton Method

Fully converging iterative solvers is neither reasonable (*oversolving*) nor feasible (*too costly*), thus solve the Newton system

$$\nabla F(x_k)\Delta x_k = -F(x)$$

to some tolerance $\longrightarrow$ inexact Newton method.

Stopping Criterion

$$||r_k^{(i)}|| \leq \eta_k ||r_k||$$

- nonlinear residual $r_k = F(x_k)$
- linear (inner) residual $r_k^{(i)} = \nabla F(x_k)\Delta x_k + F(x_k)$

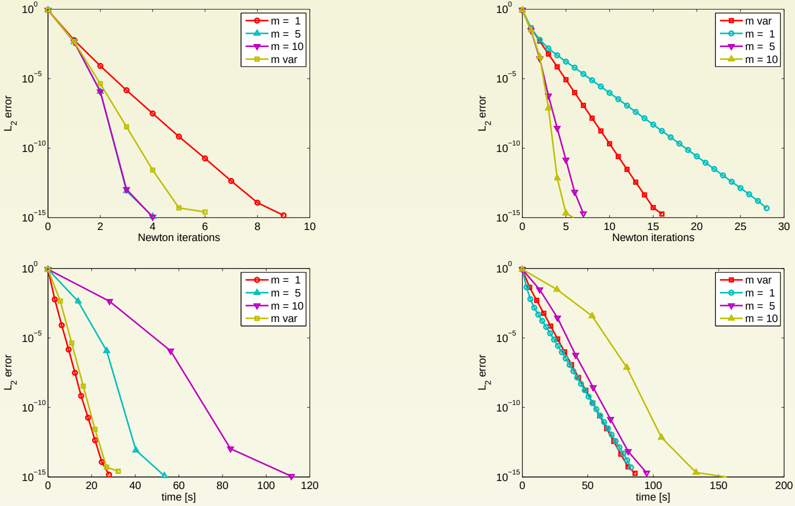
Choosing the *forcing sequence* η_k as

- $\eta_k \leq \eta < 1$ guarantees convergence
- $\eta_k \rightarrow 0$ yields superlinear convergence
- $\eta_k = \mathcal{O}(||r_k||)$ yields quadratic convergence

E.g., choose $\eta_k = \min(c||F(x_k)||^p, 0.5), \quad 0 < p \leq 1$

Semilinear Problem

$C(y,u) = -\Delta y + \gamma y e^y - u$. Convergence history for $\gamma = 1$ (left) and $\gamma = 10$.



Outline

Control Constraints

- Unilateral constraints on the control

$$u \in U_{ad} = \{v \in U | v \leq u_{\text{bnd}} \text{ a.e. in } \Omega\}$$

- Variational inequality

$$\hat{J}(u^*)(u - u^*) \geq 0 \quad \text{for all } \ u \in U_{ad}$$

- Using $\hat{J}(u) = H_Z u - L^{-1}\bar{y}$ and substituting p yields

$$(\sigma u^* - p^*, u - u^*) \geq 0 \quad \text{for all } \ u \in U_{ad}$$

- Complimentarity condition

$$\sigma u^* - p^* + \lambda = 0, \quad \lambda(u^* - u_{\text{bnd}}) = 0, \quad \lambda \geq 0 \ \text{ a.e.}$$

Primal Dual Active Set Strategy

The equivalent form $\lambda = c\max(0, u^* + \frac{\lambda}{c} - u_{\text{bnd}}),\quad c > 0$ for the complementarity condition gives rise to

Primal-Dual Active Set Method

$k = 0$, choose y^0, u^0, p^0, λ^0
while Not converged **do**
  Predict $\mathcal{A}_k, \mathcal{I}_k$
  if $n \geq 2$ and $\mathcal{A}_k = \mathcal{A}_{k+1}$ **then**
    Converged
  else
    Solve EQP
  end if
  $k = k + 1$
end while

Active Set Prediction

$$\begin{aligned}\mathcal{A}_k &= \{x | u_{k-1} + \frac{\lambda_{k-1}}{\sigma} > u_{\text{bnd}}\} \\ \mathcal{I}_k &= \Omega \backslash \mathcal{A}_k\end{aligned}$$

[Hintermüller, Ito, Kunisch]

The PDAS EQP

The EQP (semismooth Newton step) is given by

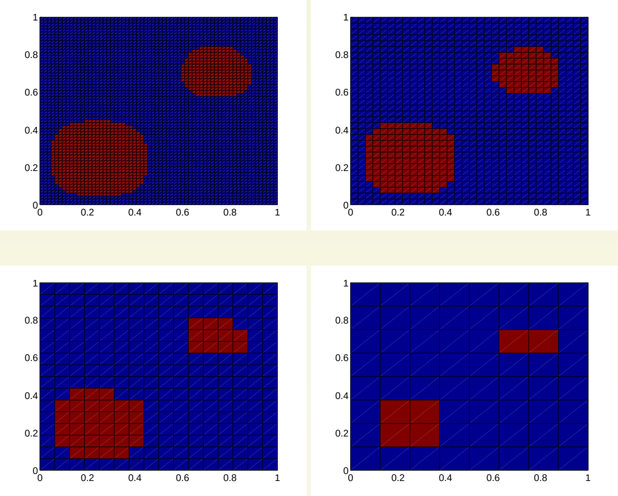
PDAS System

$$
\begin{aligned}
My^{k+1} + L^T p^{k+1} &= \bar{y} \\
\sigma M u^{k+1} + M^T p^{k+1} + \lambda^{k+1} &= 0 \\
Ly^{k+1} + Mu^{k+1} &= 0 \\
\lambda^{k+1} &= 0 \ \text{ on } \ \mathcal{I}_k \\
u^{k+1} &= u_{\text{bnd}} \ \text{ on } \ \mathcal{A}_k
\end{aligned}
$$

- Use splitting $u^k = [u_{\mathcal{I}_k} \ u_{\mathcal{A}_k}]$ to reduce EQP to a system for $y^k, u_{\mathcal{I}_k}, p^k$
- Solve rEQP with our Multigrid method
- $B_{Z,\mathcal{I}} = \sigma M_{\mathcal{I},\mathcal{I}} + M^{\mathcal{I}} L^{-T} M L^{-1} M_{\mathcal{I}}$
- Set $\lambda_{\mathcal{A}}^{k+1} = R_{\mathcal{A}}(\sigma M_{\mathcal{I}} u_{\mathcal{I}}^{k+1} + Mp^{k+1})$

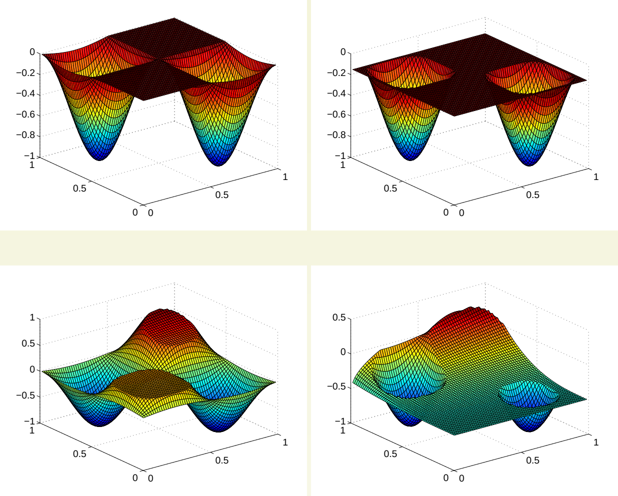
Restriction of Active Sets

Active Sets on coarse grids are constructed by piecewise constant restriction of $\chi_{\mathcal{A}}$ with the condition $\mathcal{A}_{2h} \subset \mathcal{A}_h$ and $\mathcal{I}_{2h} \supset \mathcal{I}_h$ for each h.



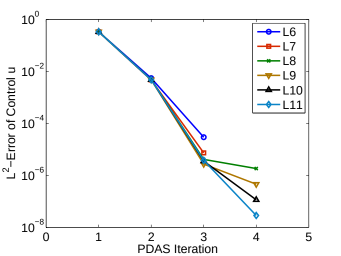
Active (red) and inactive (blue) sets for $h = 2^{-k}$, $k = 6, 5, 4, 3$.

Control-Constrained Model Problems

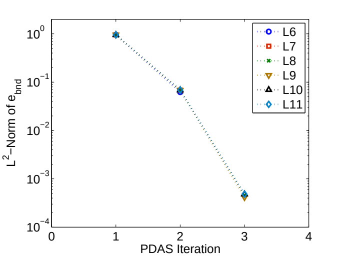


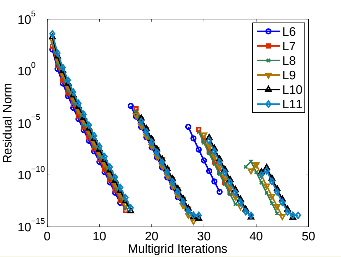
Computed optimal controls u on $h = 2^{-6}$ mesh for different upper bound functions u_b.

Numerical Results for PDAS



Level	e_{L^2}	ratio
6	2.9225_{-5}	—
7	7.3051_{-6}	4.0005
8	1.8262_{-6}	4.0001
9	4.5655_{-7}	4.0000
10	1.1414_{-7}	4.0000
11	2.8534_{-8}	4.0000



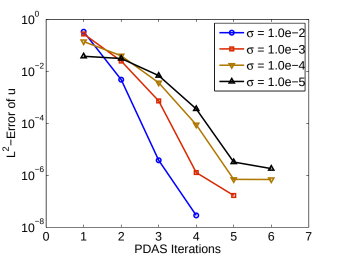


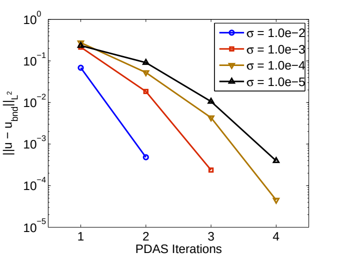
Numerical Results II

σ	L	1	2	3	4	5
	9	124952	+6096	+24	—	—
		6.843_{-2}	4.149_{-4}	0.0	—	—
1.0_{-2}	10	499670	+24474	+144	—	—
		6.848_{-2}	4.800_{-4}	0.0	—	—
	11	1998192	+98254	+706	—	—
		6.856_{-2}	4.820_{-4}	0.0	—	—
		101450	+20528	+7744	+1330	+20
	9	2.338_{-1}	9.117_{-2}	8.090_{-3}	3.379_{-4}	0.0
		—	1.667	9.732_{-1}	5.163	—
		405590	+82330	+30916	+5342	+110
1.0_{-5}	10	2.326_{-1}	8.836_{-2}	9.050_{-3}	3.848_{-4}	0.0
		—	1.632	1.158	4.699	—
		1622360	+329360	+123534	+21352	+546
	11	2.331_{-1}	9.109_{-2}	1.055_{-2}	3.960_{-4}	0.0
		—	1.675	1.271	3.555	—

Size of active sets, $||u_h - u_{\mathrm{bnd},h}||$ and ratio e_k/e_{k-1}^2 (for $\sigma = 1.0_{-5}$ only) for levels 9,10,11 and PDAS iteration k.

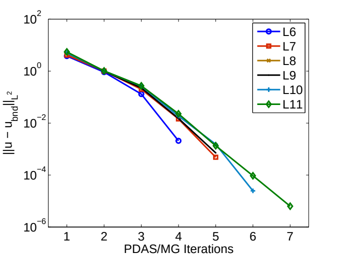
Numerical Results III

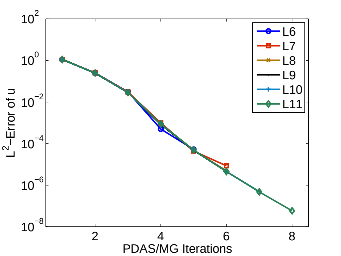




		1	2	3
bnd 1	$\mathcal{A}$	1998192	+98254	+706
	e_{bnd}	6.856_{-2}	4.820_{-4}	0.0
bnd 2	$\mathcal{A}$	2536106	+ 106944	+742
	e_{bnd}	9.4103_{-2}	7.106_{-4}	0.0
bnd 3	$\mathcal{A}$	2684033	+128150	+808
	e_{bnd}	1.081_{-1}	7.547_{-4}	0.0
bnd 4	$\mathcal{A}$	2856365	+123806	+797
	e_{bnd}	1.305_{-1}	8.618_{-4}	0.0

Numerical Results IV

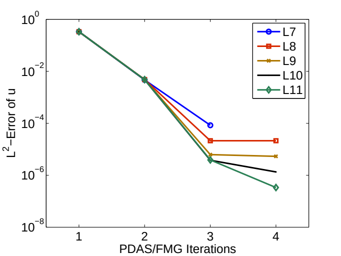




	1	2	3	4	5	6	7	total
1	2044753	+703069	+104690	+12471	+877	+80	+2	
	1	1	1	1	1	1	1	7
2	2007484	+185614	+3921	+27				
	2	2	2	2				8
$\varepsilon = 10_{-10}$	1998192	+98254	+706					
	13	10	9					32

Numerical Results V

One FMG cycle per PDAS step recovers mesh-independent behaviour.



Level	e_{L^2}	ratio
7	8.36991_{-5}	—
8	2.12820_{-5}	3.9329
9	5.35098_{-6}	3.9772
10	1.34025_{-7}	3.9925
11	3.35271_{-7}	3.9975

L	1	2	3
7	7794	+398	0
8	31158	+1604	+6
9	124952	+6092	+28
10	499670	+24474	+144
11	1998192	+98242	+718

Outline

Conclusions

Summary:

- Multigrid for optimal control problems
- Control-constraints via PDAS
- Full multigrid solves optimal control problems in $\mathcal{O}(n)$
- State/adjoint-specific smoothers (ALGS, ILU)

Future work:

- Extension to 3D
- Parallelization

References I

E. Arian and S. Ta'asan.
Multigrid one-shot methods for optimal control problems: infinite dimensional control.
Technical Report 94-52, ICASE, NASA Langley Research Center, 1994.

U. M. Ascher and E. Haber.
A multigrid method for distributed parameter estimation problems.
Electronic transactions on Numerical Analysis, 15:1–17, 2002.

A. Borzi and K. Kunisch.
A multigrid scheme for elliptic constrained optimal control problems.
Computational Optimization and Applications, 31:309–333, 2005.

D. Braess and R. Sarazin.
An efficient smoother for the Stokes problem.
Applied Numerical Mathematics, 23:3–19, 1997.

References II

W. Hackbusch.
Fast solution of elliptic control problems.
Journal of Optimization Theory and Applications, 31(4):565–581, 1980.

C. Keller, N. I. Gould, and A. J. Wathen.
Constraint preconditioning for indefinite linear systems.
SIAM Journal on Matrix Analysis and Applications, 21(4):1300–1317, 2000.

R. M. Lewis and S. G. Nash.
A multigrid approach to the optimization of systems governed by differential equations.
In *8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA, 2000. AIAA.

References III

B. Maar and V. Schulz.
Interior point multigrid methods for topology optimization.
Structural Optimization, 19(3):214–224, 2000.

S. Vanka.
Block-implicit multigrid solution of navier-stokes equations in primitive variables.
Journal of Computational Physics, 65:138–158, 1986.

G. Wittum.
Multi-Grid methods for Stokes and Navier-Stokes equations. Transforming smoothers: algorithms and numerical results.
Numerische Mathematik, 54:543–563, 1989.