Efficient Multigrid Methods for Large-scale Optimization Problems Constrained by PDEs

UMBC
Department of MATHEMATICS & STATISTICS

ANDREI DRĂGĂNESCU (PI, UMBC), MONA HAJGHASSEM (UMBC)
former groups members:
JYOTI SARASWAT (Thomas More College), ANA MARIA SOANE (USNA)

Acknowledgment of support

This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under Award Number DE-SC0005455, and by the National Science Foundation under awards DMS-1016177 and DMS-0821311. Computations have been performed in part on UMBC's High Performance Computing Facility (HPCF).

Overview

Optimization problems constrained by partial differential equations (PDEs) is a research area to which the scientific and engineering communities have devoted an increased level of effort over the last decade. The computational revolution of the last twenty years has fostered not only high-resolution numerical computations based on PDE models, but also a shift from model based simulation to model based design. The latter translates into the question of solving optimization problems with PDEs acting as equality constraints in order to identify initial and/or boundary values, material properties, sources, and other parameters for which the PDE models behave in a desired way. However, just growth in computing power is insufficient for tackling PDE-constrained optimization problems at the same extreme scales at which the PDEs themselves can be solved: although current computing capabilities allow, in principle, for the numerical solution of PDEs with 10–100 billion unknowns, solving PDE-constrained optimization problems of comparable size still requires significant algorithmic development.

The objectives of this project are to develop, analyze, and implement efficient multigrid methods for solving large-scale optimization problems constrained by PDEs, with particular focus on the linear algebraic aspects of the solvers. Applications include **optimal design of manufacturing processes, history matching for petroleum reservoir simulations, data assimilation for weather prediction**, to name only a few.

Why multigrid?
Originating in the 1960s with work on numerical PDEs, the multigrid paradigm is that the solution process of a PDE-related numerical computation can be significantly accelerated by using multiple resolutions/discretizations of the PDE. The embodiment of the paradigm is strongly problem-dependent. The study of multigrid methods for optimization problems has increased since the early 2000s and this research area developed in several, non-equivalent directions.

Abstract formulation:

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

- 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 $e(y,u) = 0$ is a well-posed PDE, *i.e.*, for all $u \in U$ there is a unique $y \in Y$ (depending continuously on u), so that

$$e(y,u) = 0, \quad K(u) \stackrel{\text{def}}{=} y \,.$$

Basic multigrid mechanism: linear PDE, no control or state constraints

- Assume K is a linear smoothing operator (e.g., solution operator of elliptic PDE) and formulate the optimization problem in reduced form

$$\begin{cases} \text{minimize } \hat{J}(u) = \frac{1}{2}\|K(u) - y_d\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|Lu\|^2, \\ \text{subj. to} \quad u \in U_{ad} \subset U, \quad L = I \text{ or } \nabla \,. \end{cases} \tag{2}$$

- Discretization of problem (2) is equivalent to the regularized normal equations

$$G_h u \stackrel{\text{def}}{=} (\beta I + K_h^* K_h) u = K_h^* \pi_h y_d \,.$$

- Two-grid preconditioner:

$$T_h = G_{2h}\pi_{2h} + \beta(I - \pi_{2h}), \quad T_h^{-1} = G_{2h}^{-1}\pi_{2h} + \beta^{-1}(I - \pi_{2h}) \,. \tag{3}$$

Theorem 1 (2004) *For h sufficiently small and $u \in V_h$*

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

where p is the order of the discrete method.

Model problems of interest:

- **PDE-constraints of interest:**
  - Stationary: linear and semilinear elliptic equations, stationary fluid flows (see the case study)
  - Time-dependent: parabolic equations, non-stationary fluid flows (Navier-Stokes), hyperbolic equations (shallow-water equations), chaotic dynamical systems
  - Stochastic PDEs
- **Controls of interest:**
  - Distributed controls (forcing, material properties)
  - Boundary controls
  - Initial values (for time-dependent problems)
- **Regularization of interest:** classical L^2, square-variation regularization (to enforce smooth controls), L^1-regularization (to enforce sparsity)

Special Challenges:

- Inequality constraints and L^1-regularization require non-smooth optimization methods, interior point methods.
- Nonlinear and time-dependent PDE-constraints are notoriously challenging in a large-scale context.
- Hyperbolic PDEs give rise to non-smooth solution operators K.

Case study: Optimal control of fluids

Optimal control problem constrained by the Navier-Stokes equations:

$$\begin{cases} \text{minimize } \frac{\gamma_u}{2}\|u - u_d\|^2 + \frac{\gamma_p}{2}\|p - p_d\|^2 + \frac{\beta}{2}\|f\|^2 \\ \text{subj. to} \quad -\nu\Delta u + (u\cdot\nabla)u + \nabla p = f \,, \\ \qquad \text{div } u = 0 \,, \ u|_{\Omega} = 0 \end{cases} \tag{4}$$

- Newton-Krylov method: $f_{k+1} = f_k - G_h(f_k)^{-1}\nabla\hat{J}_h(f_k)$.
- At each Newton iteration use multigrid preconditioned CG to solve: $G_h(f_k)\delta f_k = \nabla\hat{J}_h(f_k)$.
- For a fixed force f, let $u = \mathcal{U}(f)$, $p = \mathcal{P}(f)$; define $\mathcal{L} = \mathcal{L}(f)$ by

$$\mathcal{L}\begin{bmatrix} v \\ q \end{bmatrix} \stackrel{\text{def}}{=} \begin{bmatrix} \nu\Delta v + (v\cdot\nabla)u + (u\cdot\nabla)v + \nabla q \\ \nabla\cdot v \end{bmatrix} .$$

- Define $\mathcal{E}(v,q) = v$ and $\mathcal{I}(f) = (f, 0)$.
- Gradient formula:

$$\nabla J(f) = \underbrace{\mathcal{E}(\mathcal{L}^*)^{-1}\begin{bmatrix} g_u(u - u_d) \\ g_p(p - p_d) \end{bmatrix}}_{z_f} + \beta f \,.$$

- Given z_f define the linear operator

$$\mathcal{D}v = (v\cdot\nabla)z_f + (\nabla\cdot v)z_f - \nabla^T v\cdot z_f \,.$$

- Hessian operator:

$$\mathcal{G}(f) = \mathcal{E}(\mathcal{L}^*)^{-1}\begin{bmatrix} \mathcal{D} + g_u I \\ g_p I \end{bmatrix}\mathcal{L}^{-1}\mathcal{I} + \beta I \,.$$

- Two-grid preconditioner for the Hessian at each Newton iteration:

$$T_h(f_k) = G_{2h}(\pi_{2h}f_k)\pi_{2h} + \beta(I - \pi_{2h}) \tag{5}$$

Theorem 2 (2013) *Let $f \in U_{ad} \cap X_h$, $\gamma_u = 1$, $\gamma_p = 0$. If standard finite element approximations*

$$\|(\mathcal{U} - \mathcal{U}_h)f\| \le Ch^2\|f\|, \ \|(\mathcal{P} - \mathcal{P}_h)f\| \le Ch\|f\|$$

hold, and under standard regularity assumptions,

$$\|(G_h(u) - T_h(u))v\| \le Ch^2\|v\| \ \ \forall v \in X_h \,,$$

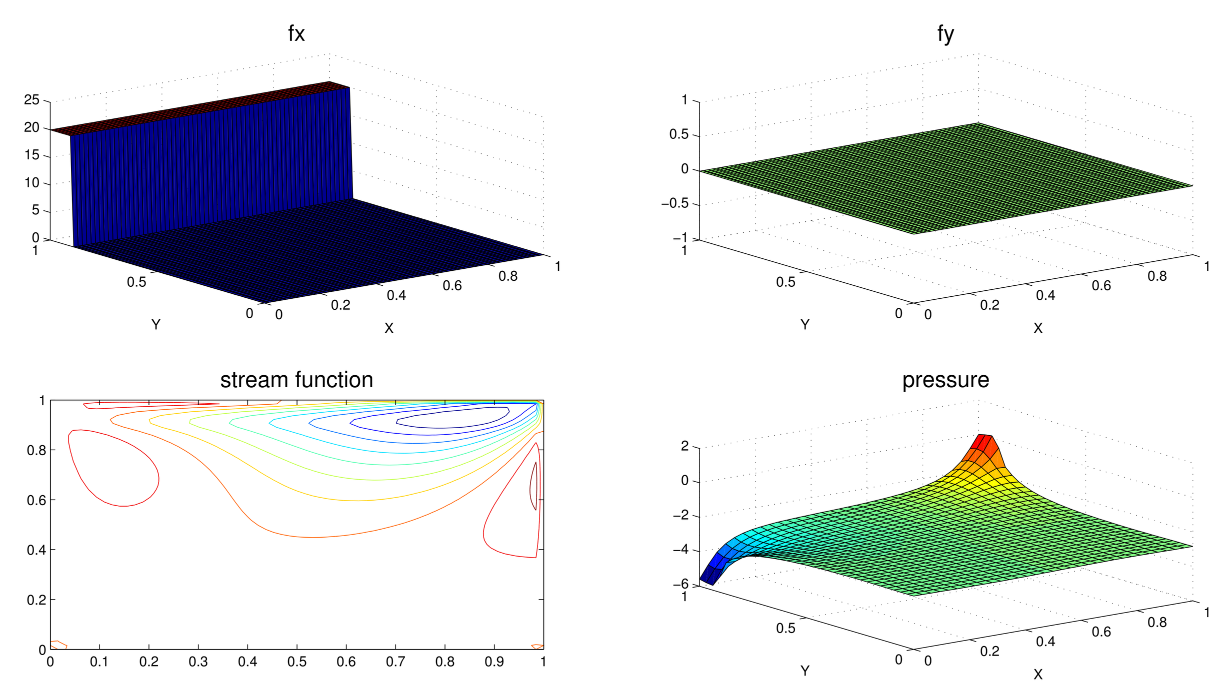
with C independent of h.

Numerical results:

- Use $f_1(x,y) = g$ if $y > 0.9$, 0 otherwise $f_2(x,y) = 0$ to generate velocity-pressure data (u_d, p_d) to resemble lid-driven cavity.
- Samples of comparison of iteration counts and runtimes for multigrid vs. unpreconditioned CG $g = 1$, $Re \approx 0.55$.

h	2^{-5}	2^{-6}	2^{-7}	2^{-8}
$g_p = 0, \beta = 10^{-4}$ (velocity control only)				
# cg its	40, 46	40, 39	41, 44	41, 36
t_{cg} (s)	35.58	179.66	1196.67	7386.29
# mg its ($n_0 = 16$)	5, 5	5, 4	3, 3	2, 2
t_{mg} (s)	98.47	135.93	324.46	1283.44
$g_p = 0, \beta = 10^{-5}$ (velocity control only)				
# cg its	96, 113	96, 97	96, 104	96, 89
t_{cg} (s)	105.48	518.83	2789.97	17294.9
# mg its ($n_0 = 16$)	10, 10	10, 9	5, 5	3, 3
t_{mg} (s)	95.65	164.41	408.62	1577.79
$g_p = 10^{-2}, \beta = 10^{-4}$ (pressure/velocity control)				
# cg its	47, 51	49, 53	47, 57, 21	48, 55
t_{cg} (s)	40.76	220.40	1527.53	9641.7
# mg its ($n_0 = 16$)	9, 9	9, 9	9, 9, 6	12, 13
t_{mg} (s)	104.01	162.58	724.42	3684.8

"Target" data ($Re \approx 5.5$)



Recovered data ($\gamma_u = 1$, $\gamma_p = 10^{-3}$, $\beta = 10^{-5}$)