PDE Constrained Optimization

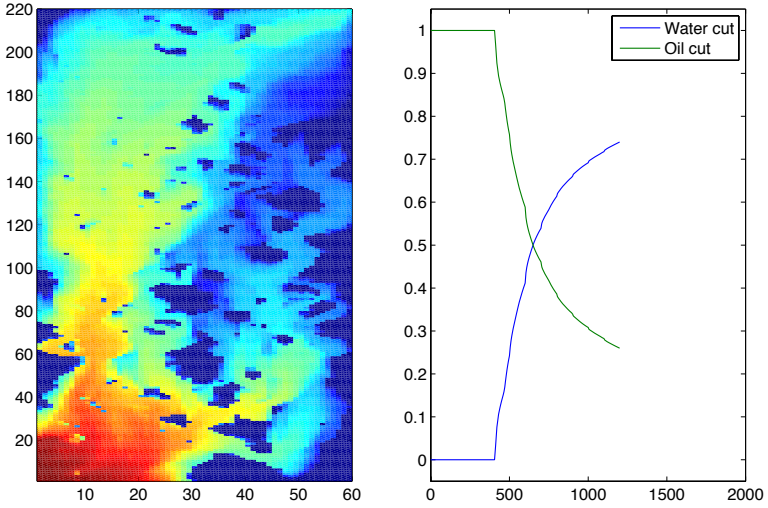
Matthias Heinkenschloss

Department of Computational and Applied Mathematics
Rice University, Houston, Texas
heinken@rice.edu

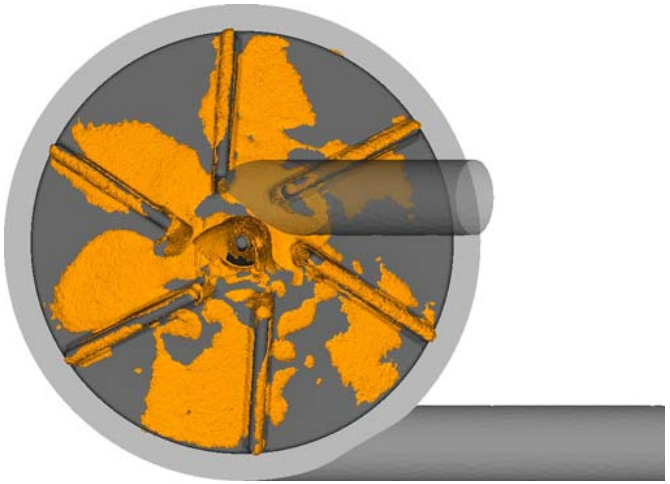
May 11, 2008

SIAM Conference on Optimization
May 10-13, 2008, Boston, Massachusetts

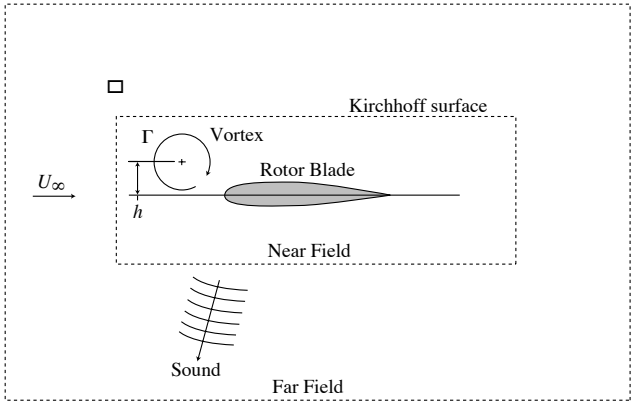
Three Applications of PDE Constrained Optimization

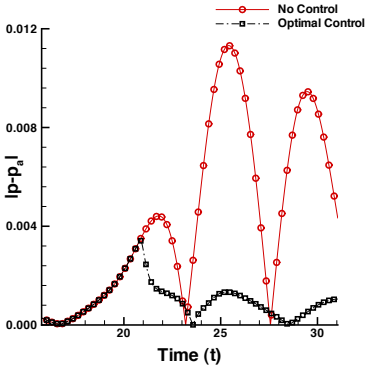


(oil) reservoir management
(with A. El Bakry, K. D. Wiegand (ExxonMobil))



shape optim. of medical devices
(with M. Behr (RWTH Aachen))



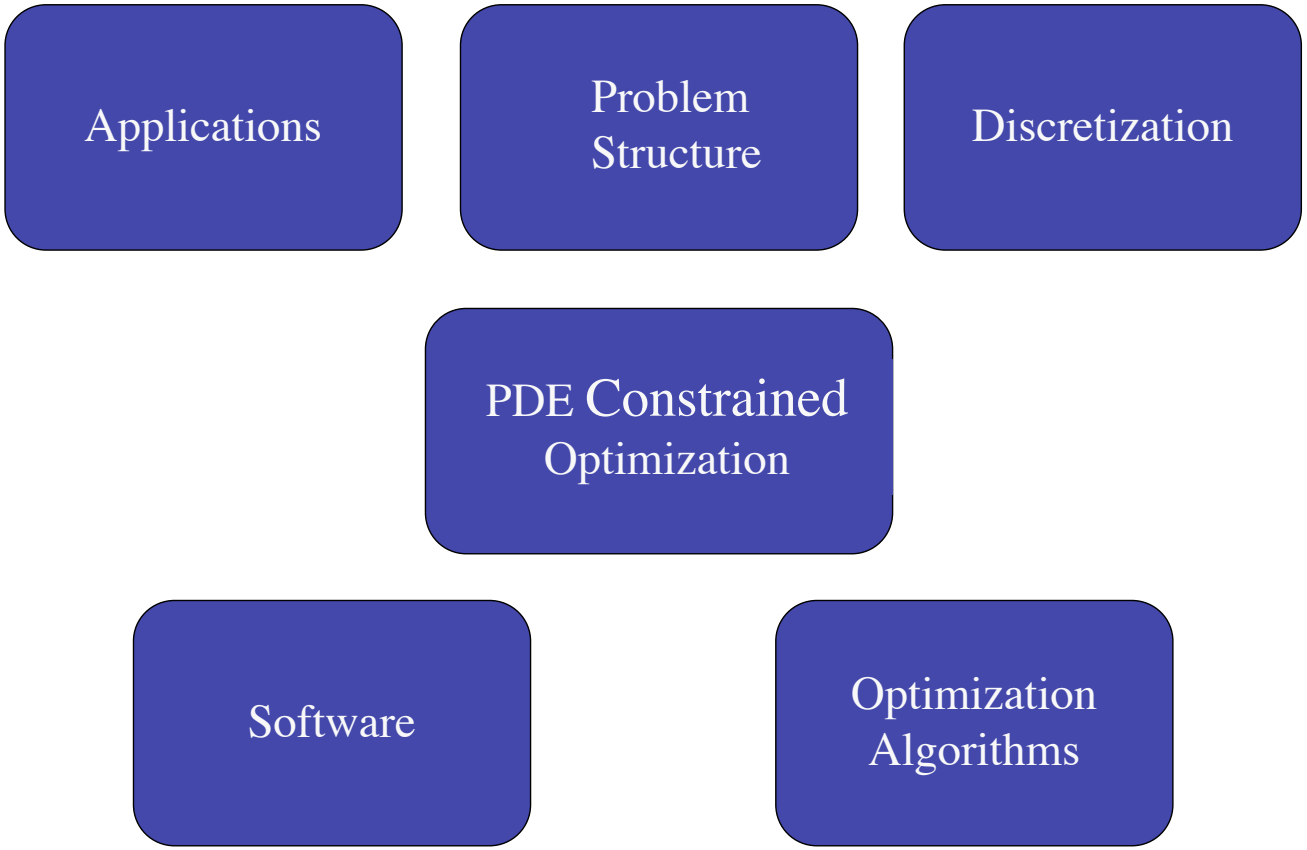


on-blade control of blade-vortex-interaction generated noise
(with S. S. Collis (Sandia Nat. Labs.), K. Ghayour (ExxonMobil))

Characteristics of PDE Constrained Optimization Problems

- All problems are PDE constrained optimization problems - there are many, many more.
- Evaluation of objective function and constraint functions involves expensive simulations (in the previous examples solution of partial differential equations (PDEs)).
- THE optimization problem does not exist. Instead each problem leads to a family of optimization problems which are closely linked. (Hierarchy of optimization problems obtained by refinement of discretization.)
- The robust and efficient solution of such optimization problems requires the integration of application specific structure, numerical simulation and optimization algorithms.

Need to look at the big picture, not only at one component



Optimization Approach

- Selection of suitable optimization algorithm depends on the properties of the optimization problem, properties of the PDE simulator, ...
- I focus on PDE constrained optimization problems with many control variables/parameters u.
- I focus on derivative based, Newton-type algorithms for PDE constrained optimization problems.
  - $+$ fast convergence,
  - $+$ often mesh independent convergence behavior,
  - $+$ efficiency from integration of optimization and simulation,
  - $-$ require insight into simulator.
- PDE constrained optimization is a very active area, as indicated by the large number of talks/minisymposia in the area of PDE constrained optimization at this meeting. There are many interesting developments that I do not have time to cover.

Outline

Overview

Problem Formulation

Optimality Conditions

Discretization and Optimization

Optimization Algorithms

Conclusions

Abstract Optimization Problem

$$
\begin{aligned}
&\min J(y,u) \\
&\text{s.t. } c(y,u) = 0, && \text{(the governing PDE)} \\
&\quad\;\; g(y,u) = 0, && \text{(additional equality constr.)} \\
&\quad\;\; h(y,u) \in -K && \text{(additional inequality constr.)} \\
&\quad\;\; y \in \mathcal{Y}_{ad}, u \in \mathcal{U}_{ad}.
\end{aligned}
$$

where

- $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$, $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$, $(\mathcal{C}, \|\cdot\|_{\mathcal{C}})$ are real Banach spaces,
- $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ is a real normed space,
- $\mathcal{Y}_{ad} \subset \mathcal{Y}$, $\mathcal{U}_{ad} \subset \mathcal{U}$ are nonempty, closed convex sets,
- $K \subset \mathcal{H}$ is a nonempty, closed convex cone,
- $J : \mathcal{Y} \times \mathcal{U} \to \mathbb{R}$, $c : \mathcal{Y} \times \mathcal{U} \to \mathcal{C}$, $h : \mathcal{Y} \times \mathcal{U} \to \mathcal{H}$ are smooth mappings.

Notation:
y: states, $\mathcal{Y}$: state space, u: controls, $\mathcal{U}$: control space,
$c(y,u) = 0$ state equation.

Problem Formulation

$$\begin{array}{ll} \min & J(y,u) \\ \text{s.t.} & c(y,u) = 0, \\ & g(y,u) = 0, \\ & h(y,u) \in -K \end{array}$$

$$\Downarrow$$

$y(u)$ is the unique solution of $c(y,u) = 0$

$$\Downarrow$$

$$\left.\begin{array}{ll} \min & \widehat{J}(u) \\ \text{s.t.} & \widehat{g}(u) = 0, \\ & \widehat{h}(u) \in -K, \end{array}\right\} \text{reduced problem}$$

where $\widehat{J}(u) \stackrel{\text{def}}{=} J(y(u),u)$, $\widehat{g}(u) \stackrel{\text{def}}{=} g(y(u),u)$, $\widehat{h}(u) \stackrel{\text{def}}{=} h(y(u),u)$.

The reduced problem formulation is often used, but it is not always clear that it can be used.

- The problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\int_\Omega (y(x)-\hat{y}(x))^2 dx + \frac{\alpha}{2}\int_{\partial\Omega} u^2(x) ds, \\ \text{subject to} \quad & -\Delta y(x) = l(x) \text{ in } \Omega, \\ & \frac{\partial}{\partial n} y(x) = u(x) \text{ on } \partial\Omega \end{aligned}$$

is well-posed and has a unique solution, but for given u the state equation does not have a solution or it has infinitely many solutions.

- The problem

$$\text{minimize } \frac{1}{2}\int_D \left(\frac{\partial \mathbf{v}_2}{\partial x_1} - \frac{\partial \mathbf{v}_1}{\partial x_2}\right)^2 dx + \frac{\alpha}{2}\|\mathbf{g}\|^2_{\mathbf{H}^{1/2}(\partial\Omega)}$$

subject to

$$\begin{aligned} -\nu\Delta\mathbf{v}(x) + (\mathbf{v}(x)\cdot\nabla)\mathbf{v}(x) + \nabla p(x) &= \mathbf{f}(x) && x \in \Omega, \\ \text{div } \mathbf{v}(x) &= 0 && x \in \Omega, \\ \mathbf{v}(x) &= \mathbf{g}(x) && x \in \partial\Omega, \end{aligned}$$

$$\int_\Omega \mathbf{g}(x)\cdot\mathbf{n}(x) dx = 0$$

is well-posed, but the Navier-Stokes equation is only guaranteed to have a *unique* solution if ν is large (Reynolds number is small) relative to $\mathbf{g}$ and $\mathbf{f}$.

Outline

Overview

Problem Formulation

Optimality Conditions

Discretization and Optimization

Optimization Algorithms

Conclusions

Karush-Kuhn-Tucker Theorem in Banach Spaces

Recall the optimization problem in Banach spaces

$$
\begin{aligned}
&\min J(y,u) \\
&\text{s.t. } c(y,u) = 0, && \text{(the governing PDE)} \\
&\quad h(y,u) \in -K && \text{(additional inequality constr.)} \\
&\quad y \in \mathcal{Y}_{ad}, u \in \mathcal{U}_{ad}.
\end{aligned}
$$

If (y_*, u_*) is a local minimizer and if a regularity condition (CQ) holds, then there exist continuous linear functionals (Lagrange multipliers) $\lambda_* \in \mathcal{C}^*$,

$$\mu_* \in K^* \equiv \{\ell \in \mathcal{H}^* : \ell(v) \geq 0 \text{ for all } v \in K\}$$

such that

$$
\begin{aligned}
(D_y J(y_*, u_*) + \lambda_* \circ D_y c(y_*, u_*) + \mu_* \circ D_y h(y_*, u_*))\,(y - y_*) \geq 0 \quad &\text{for all } y \in \mathcal{Y}_{a} \\
(D_u J(y_*, u_*) + \lambda_* \circ D_u c(y_*, u_*) + \mu_* \circ D_u h(y_*, u_*))\,(u - u_*) \geq 0 \quad &\text{for all } u \in \mathcal{U}_{a} \\
\mu_*(h(y_*, u_*)) = 0 &
\end{aligned}
$$

See Zowe/Kurcyusz (1979) and the books by J. Jahn (Springer Verlag, 2nd ed., 1996), J. Werner (Friedrich Vieweg & Sohn Verlag, 1984), and the books by D. G. Luenberger (John Wiley & Sons, 1969).

Karush-Kuhn-Tucker Theorem in Banach Spaces

- The previous KKT Theorem is a good guideline, but often cannot be applied directly to PDE constrained optimization.
- The choice of state and control space are important.
- Precise characterization of Lagrange multipliers is important for design and analysis of optimization algorithms.
- Precise characterization of Lagrange multipliers requires (a lot of) extra work.
- Optimality conditions for optimal control problems with control and state constraints have been derived by Casas, Bonnans, Kunisch, Bergounioux, Raymond, Tröltzsch,.....

Example 1 (Only PDE Constraint)

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\int_\Omega (y(x)-\hat{y}(x))^2 dx + \frac{\alpha}{2}\int_\Omega u^2(x)dx, \\ \text{subject to} \quad & -\Delta y(x) = u(x) + l(x),\ x \in \Omega, \quad y(x) = 0\ x \in \partial\Omega. \end{aligned}$$

Problem fits into previous framework if we define $\mathcal{Y} = H_0^1(\Omega)$, $\mathcal{U} = L^2(\Omega)$,

- $J(y,u) = \frac{1}{2}\int_\Omega (y(x)-\hat{y}(x))^2 dx + \frac{\alpha}{2}\int_\Omega u^2(x)dx$,
- $c : H_0^1(\Omega) \times L^2(\Omega) \to H^{-1}(\Omega)$, where

$$\langle c(y,u), \phi\rangle_{\mathcal{Y}',\mathcal{Y}} = \int_\Omega \nabla y \nabla \phi dx - \int_\Omega u\phi dx - \int_\Omega l\phi dx.$$

If $(y_*, u_*) \in H_0^1(\Omega) \times L^2(\Omega)$ is a local minimizer, then there exists $\lambda_* \in H_0^1(\Omega)$ such that

$$\begin{aligned} -\Delta\lambda_*(x) &= -(y_*(x) - \widehat{y}(x)), && x \in \Omega, \\ \lambda_*(x) &= 0 && x \in \partial\Omega, \\ \lambda_*(x) + \alpha u_*(x) &= 0 && \text{a.e. in } \Omega. \end{aligned}$$

Example 1 (Only PDE Constraint)

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\int_\Omega (y(x)-\hat{y}(x))^2 dx + \frac{\alpha}{2}\int_\Omega u^2(x)dx, \\ \text{subject to} \quad & -\Delta y(x) = u(x) + l(x),\ x \in \Omega, \quad y(x) = 0\ x \in \partial\Omega. \end{aligned}$$

Problem fits into previous framework if we define $\mathcal{Y} = H_0^1(\Omega)$, $\mathcal{U} = L^2(\Omega)$,

- $J(y,u) = \frac{1}{2}\int_\Omega (y(x)-\hat{y}(x))^2 dx + \frac{\alpha}{2}\int_\Omega u^2(x)dx$,
- $c : H_0^1(\Omega) \times L^2(\Omega) \to H^{-1}(\Omega)$, where

$$\langle c(y,u), \phi\rangle_{\mathcal{Y}',\mathcal{Y}} = \int_\Omega \nabla y \nabla \phi dx - \int_\Omega u\phi dx - \int_\Omega l\phi dx.$$

If $(y_*, u_*) \in H_0^1(\Omega) \times L^2(\Omega)$ is a local minimizer, then there exists $\lambda_* \in H_0^1(\Omega)$ such that

$$\begin{aligned} -\Delta\lambda_*(x) &= -(y_*(x) - \widehat{y}(x)), && x \in \Omega, \\ \lambda_*(x) &= 0 && x \in \partial\Omega, \\ \alpha u_*(x) - \lambda_*(x) &= 0 && \text{a.e. in } \Omega. \end{aligned}$$

Optimality conditions involve another linear PDE, the adjoint PDE.

Example 2 (Pointwise Control Constraints)

$$\begin{aligned}
\text{Minimize} \quad & \frac{1}{2}\int_\Omega (y(x)-\hat{y}(x))^2 dx + \frac{\alpha}{2}\int_\Omega u^2(x)dx, \\
\text{subject to} \quad & -\Delta y(x) = u(x) + l(x),\ x \in \Omega, \quad y(x) = 0\ x \in \partial\Omega, \\
& u_{\text{low}}(x) \le u(x) \le u_{\text{upp}}(x) \text{ a.e. in } \Omega.
\end{aligned}$$

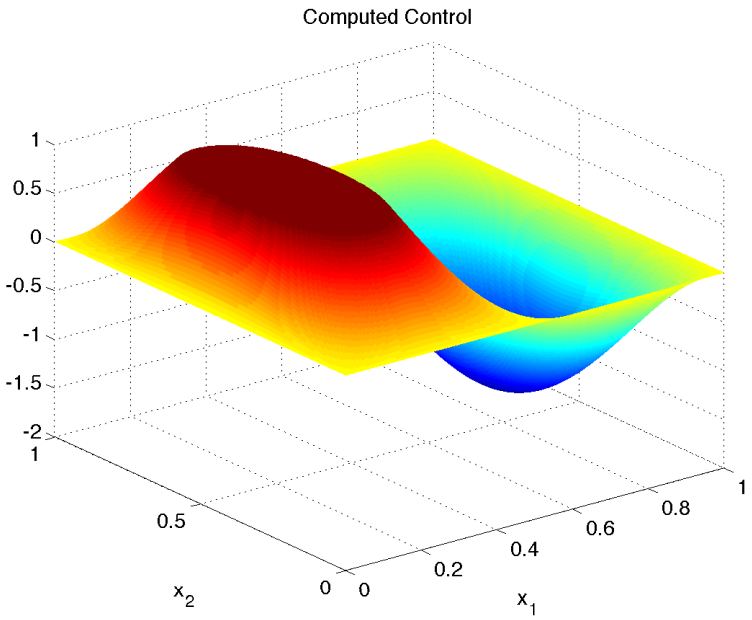
If $(y_*, u_*) \in H_0^1(\Omega) \times L^2(\Omega)$ is a local minimizer, then there exist $\lambda_* \in H_0^1(\Omega)$ and $\mu_{low,*}, \mu_{upp,*} \in L^2(\Omega)$, with $\mu_{low,*}, \mu_{upp,*} \ge 0$ a.e. in Ω such that

$$\begin{aligned}
-\Delta\lambda_*(x) &= -(y_*(x) - \widehat{y}(x)), && x \\
\lambda_*(x) &= 0 && x \\
\alpha u_*(x) - \lambda_*(x) - \mu_{low,*}(x) + \mu_{upp,*}(x) &= 0, && \text{a.e.} \\
\int_\Omega (u_{low,*} - u_*)\mu_{low,*} dx &= \int_\Omega (u_* - u_{upp,*})\mu_{upp,*} dx = 0.
\end{aligned}$$

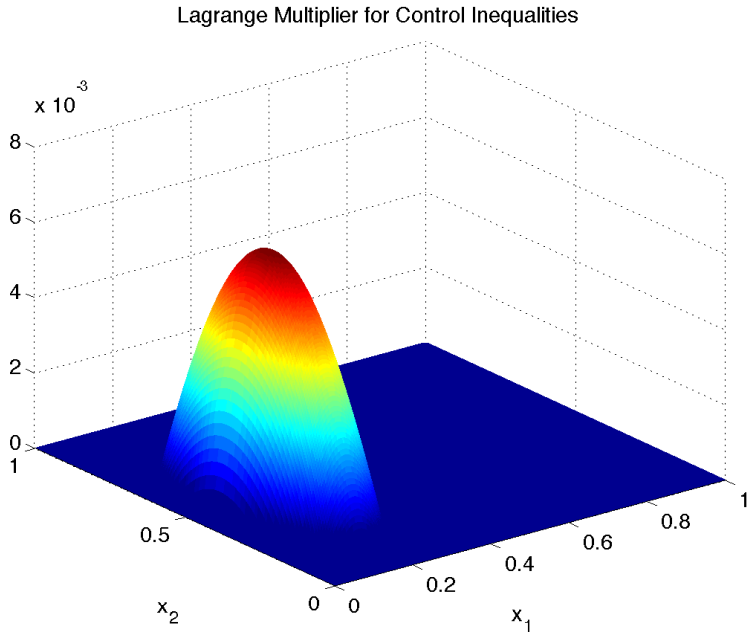
Lagrange multipliers corresponding to pointwise control constraints are L^2 functions.

Example 2 (Pointwise Control Constraints)

$$\begin{aligned}
\text{Minimize} \quad & \frac{1}{2}\int_\Omega (y(x)-\hat{y}(x))^2 dx + \frac{\alpha}{2}\int_\Omega u^2(x)dx, \\
\text{subject to} \quad & -\Delta y(x) = u(x),\ x \in \Omega, \quad y(x) = 0\ x \in \partial\Omega, \\
& u(x) \le 1 \text{ a.e. in } \Omega.
\end{aligned}$$



u_*



$\mu_{upp,*}$

Example 3 (Pointwise State Constraints)

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\int_\Omega (y(x)-\hat{y}(x))^2 dx + \frac{\alpha}{2}\int_\Omega u^2(x)dx, \\ \text{subject to} \quad & -\Delta y(x) = u(x) + l(x),\ x \in \Omega, \quad y(x) = 0\ x \in \partial\Omega, \\ & y_{\text{low}}(x) \le y(x) \le y_{\text{upp}}(x) \text{ a.e. in } \Omega. \end{aligned}$$

- Need more regular states y to make sense out of $y_{\text{low}}(x) \le y(x) \le y_{\text{upp}}(x)$ a.e. in Ω. Require $y \in C(\overline{\Omega})$.
- Lagrange multipliers $\nu_{low,*}, \nu_{upp,*}$ are in $y \in C(\overline{\Omega})^*$. Lagrange multipliers are measures.
- Optimality conditions

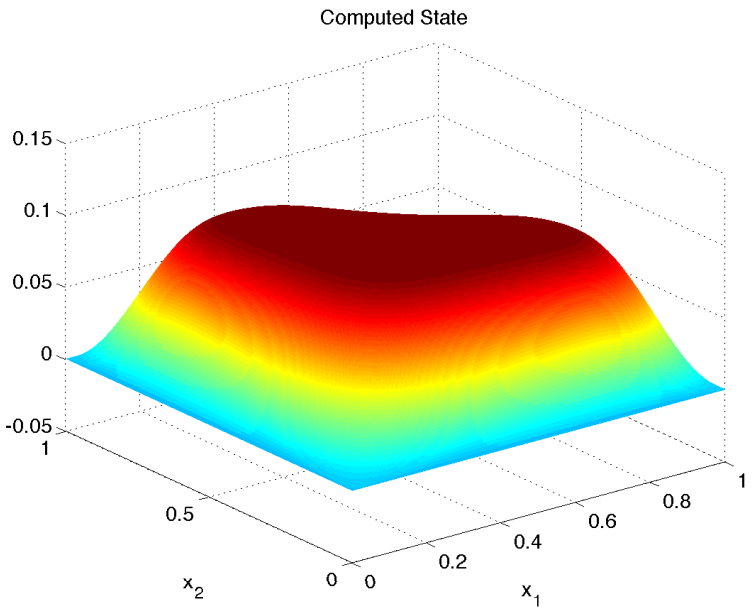
$$\begin{aligned} -\Delta\lambda_* &= -(y_* - \widehat{y}) + \nu_{upp,*} - \nu_{low,*}, && x \in \Omega, \\ \lambda_* &= 0 && x \in \partial\Omega, \\ \alpha u_* - \lambda_* &= 0, && \text{a.e. in } \Omega. \end{aligned}$$

$$\int_\Omega (y_{low,*} - y_*) d\nu_{low,*} = \int_\Omega (y_* - y_{upp,*}) d\nu_{upp,*} = 0.$$

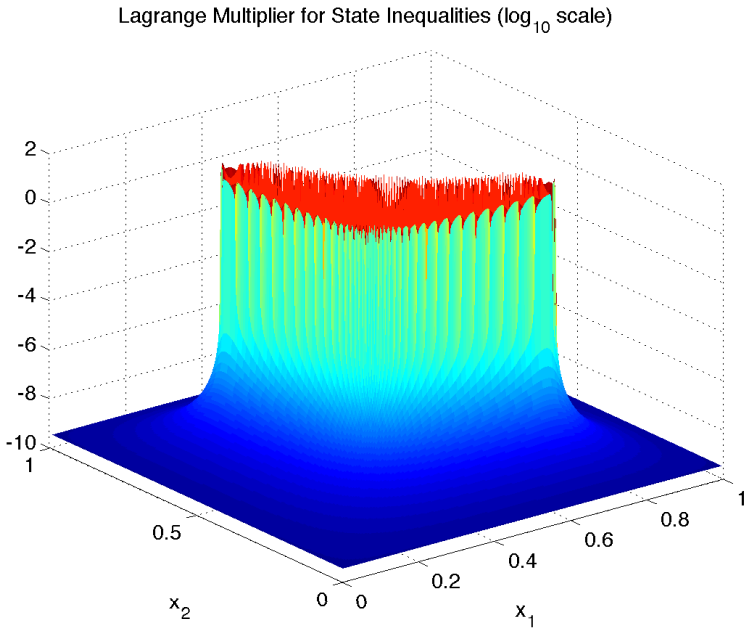
  Adjoint equation involves measures on the right hand side.
- Often, more can be said about the structure of $\nu_{upp,*}$, $\nu_{low,*}$. See Casas, Kunisch, Bergounioux, Raymond, ...

Example 3 (Pointwise State Constraints)

$$\begin{aligned}
\text{Minimize} \quad & \frac{1}{2}\int_\Omega (y(x) - \sin(2\pi x_1 x_2))^2 dx + \frac{\alpha}{2}\int_\Omega u^2(x)dx, \\
\text{subject to} \quad & -\Delta y(x) = u(x),\ x \in \Omega, \quad y(x) = 0\ x \in \partial\Omega, \\
& y(x) \le 0.1 \text{ a.e. in } \Omega.
\end{aligned}$$



y_*



$\nu_{upp,*}$

Outline

Overview

Problem Formulation

Optimality Conditions

Discretization and Optimization

Optimization Algorithms

Conclusions

- The problem we want to solve

$$
\begin{aligned}
&\min J(y,u) \\
&\text{s.t. } c(y,u) = 0, && \text{(the governing PDE)} \\
&\quad\;\; g(y,u) = 0, && \text{(additional equality constr.)} \\
&\quad\;\; h(y,u) \in -K && \text{(additional inequality constr.)}
\end{aligned}
$$

where $\mathcal{Y}, \mathcal{U}, \mathcal{C}, \mathcal{G}, \mathcal{H}$ are Banach spaces, $K \subset \mathcal{H}$ is a cone, and

$$
\begin{aligned}
J &: \mathcal{Y} \times \mathcal{U} \to \mathbb{R}, & c &: \mathcal{Y} \times \mathcal{U} \to \mathcal{C}, \\
g &: \mathcal{Y} \times \mathcal{U} \to \mathcal{G}, & h &: \mathcal{Y} \times \mathcal{U} \to \mathcal{H}.
\end{aligned}
$$

- The problem we can solve

$$
\begin{aligned}
\min \quad & J_h(y_h, u_h) \\
\text{s.t.} \quad & c_h(y_h, u_h) = 0, \\
& g_h(y_h, u_h) = 0, \\
& h_h(y_h, u_h) \in -K_h
\end{aligned}
$$

where $\mathcal{Y}_h, \mathcal{U}_h, \mathcal{C}_h, \mathcal{G}_h, \mathcal{H}_h$ are finite dimensional Banach spaces,

$$
\begin{aligned}
J_h &: \mathcal{Y}_h \times \mathcal{U}_h \to \mathbb{R}, & c_h &: \mathcal{Y}_h \times \mathcal{U}_h \to \mathcal{C}_h, \\
g_h &: \mathcal{Y}_h \times \mathcal{U}_h \to \mathcal{G}_h, & h_h &: \mathcal{Y}_h \times \mathcal{U}_h \to \mathcal{H}_h.
\end{aligned}
$$

- Does the solution (u_h, y_h) of the discretized problem converge to the solution (u, y) of the original problem? How fast?

Standard Approach

Discretize-then-optimze

$$\begin{array}{ll} \min & \mathcal{J}(y,u) \\ \text{s.t.} & c(y,u) = 0 \\ & (y,u) \in K \end{array}$$

$\xrightarrow{\text{discretize}}$ large-scale nonlinear programming problem

$\downarrow$ optimize

apply nonlinear programming

Example (W.W. Hager, 2000)

Optimal Control Problem

$$\text{Minimize } \tfrac{1}{2}\int_0^1 u^2(t) + 2y^2(t)dt$$

where

$$\begin{aligned} \dot{y}(t) &= \tfrac{1}{2}y(t) + u(t),\ t \in [0,1], \\ y(0) &= 1. \end{aligned}$$

Solution

$$y^*(t) = \frac{2e^{3t} + e^3}{e^{3t/2}(2+e^3)},$$

$$u^*(t) = \frac{2(e^{3t} - e^3)}{e^{3t/2}(2+e^3)}.$$

DOES NOT CONVERGE! WHY?

Discretization using a 2nd order Runge Kutta method

$$\text{Minimize } \frac{h}{2}\sum_{k=0}^{K-1} u_{k+1/2}^2 + 2y_{k+1/2}^2$$

where

$$\begin{aligned} y_{k+1/2} &= y_k + \frac{h}{2}(\tfrac{1}{2}y_k + u_k), \\ y_{k+1} &= y_k + h(\tfrac{1}{2}y_{k+1/2} + u_{k+1/2}), \end{aligned}$$

$k = 0, \ldots, K$.

Solution of the discretized problem:

$$y_k = 1, \quad y_{k+1/2} = 0,$$

$$u_k = -\frac{4+h}{2h}, \quad u_{k+1/2} = 0,$$

$k = 0, \ldots, K$.

Discretization of state equation and objective function implies a discretization for the adjoint equation, which may have different convergence properties than the discretization scheme applied to state equation and objective function.

For the example problem

$$\dot{y}(t) = \tfrac{1}{2}y(t) + u(t), \qquad y(0) = 1,$$

$$y_{k+1/2} = y_k + \frac{\Delta t}{2}(\tfrac{1}{2}y_k + u_k),$$
$$y_{k+1} = y_k + \Delta t(\tfrac{1}{2}y_{k+1/2} + u_{k+1/2}),$$

$$\dot{\lambda}(t) = -\tfrac{1}{2}\lambda(t) + 2y(t), \qquad \lambda(1) = 0,$$

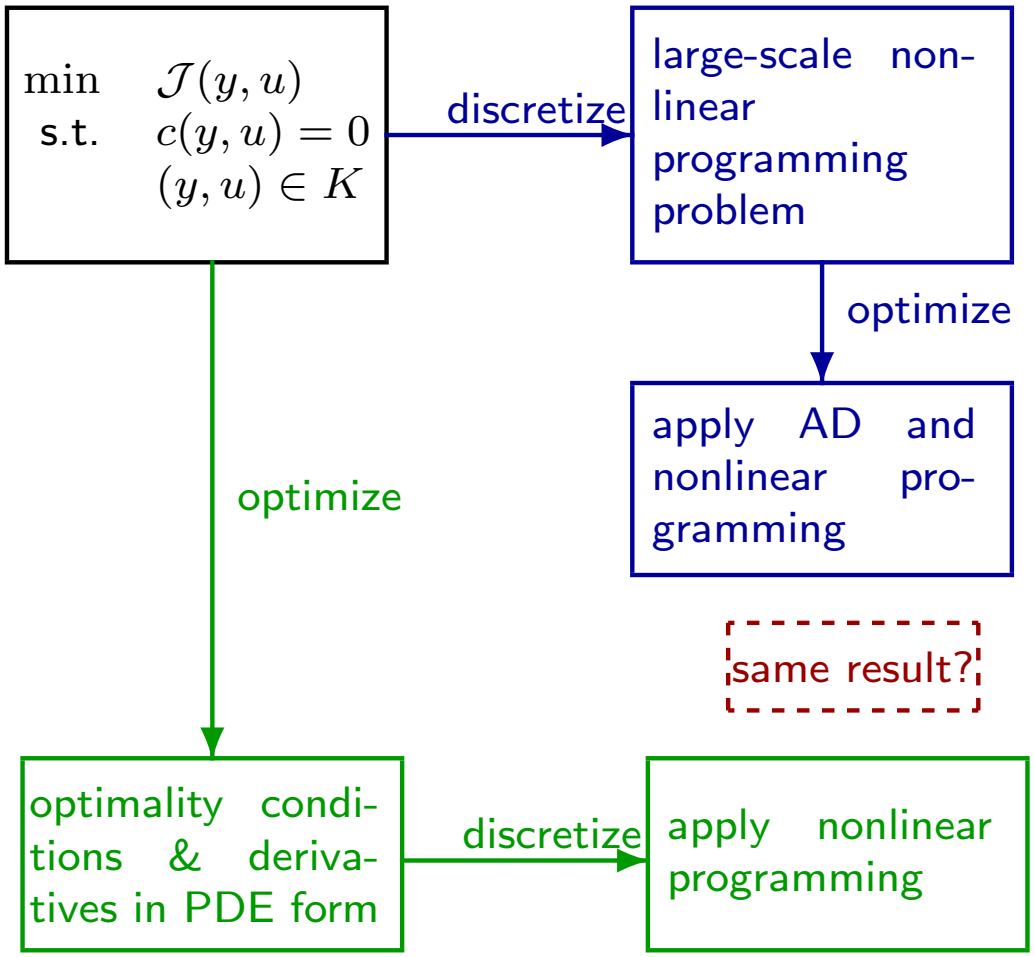
$$\lambda_{k+1/2} = \Delta t(\tfrac{1}{2}\lambda_{k+1} - 2y_{k+1/2}),$$
$$\lambda_k = \lambda_{k+1} + (1 + \Delta t/4)\lambda_{k+1/2},$$

$$u(t) - \lambda(t) = 0.$$

$$-\lambda_{k+1/2} = 0,$$
$$u_{k+1/2} - \lambda_{k+1} = 0.$$

Note, this is a discretization issue, not an issue of how the discretized optimization problem is solved!

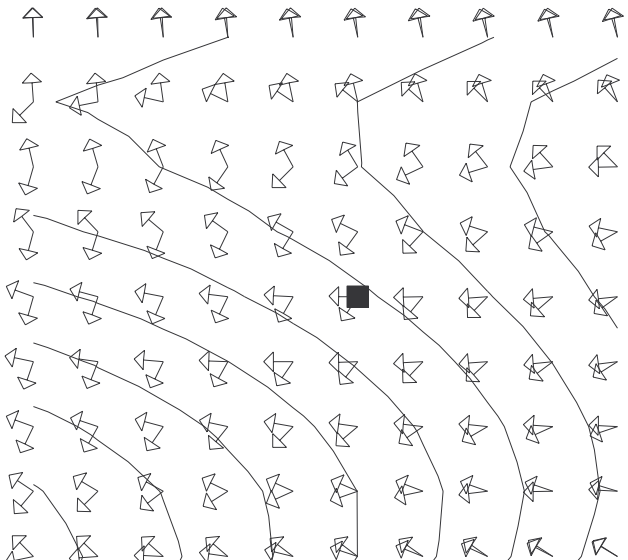
Discretize-then-optimize



Optimize-then-discretize

- To analyze the convergence of the discretization scheme for the optimization problem we need to investigate convergence of the discretization schemes for the state equation, the adjoint equation, and the gradient equation.
- (Local) convexity of the optimization problem is important (second order sufficient optimality conditions).
- Discretize-then-optimize and optimize-then-discretize are two approaches. One is not universally better than the other.
  It is important to understand the whole picture (state PDE, adjoint PDE, ...)

- For convex problems one can solve the system of optimality conditions. In this case the optimize-then-discretize gives approximation properties that are at least as good as those given by discretize-then-optimize approach.
  But, the optimality system is now potentially non-symmetric.
- For nonlinear problems, the optimize-then-discretize may lead to inexact gradients:
  $$(\nabla \widehat{J}(u_h))_h \neq \nabla \widehat{J}_h(u_h).$$
  But, usually one can show $\|(\nabla \widehat{J}(u_h))_h - \nabla \widehat{J}_h(u_h)\| \to 0$.
  Need to use optimization carefully! At a fixed discretization the (gradient based) optimization algorithm will get suck if the stopping tolerance is too fine relative to the accuracy in the computed gradient $(\nabla \widehat{J}(u_h))_h$.



$(\nabla \widehat{J}(u_h))_h$ and $\nabla \widehat{J}_h(u_h)$ for a shape design problem from Burkardt, Gunzburger, Peterson (2002).

- Approaches to coordinate choice of discretization level and optimization.
  - Consistent approximations (Polak (1997)):
    How accurately does one solve the discretized optimization prolem before increasing the discretization level? Requires only asymptotic error estimates.
  - Trust-region based model management approaches (Carter (1989/91), Alexandrov/Dennis/Lewis/Torczon (1998), ...):
    At a given iterate y_k, u_k select an approximate problem based on function and derivative information for the original problem. Can go back to approximate model. Requires error estimates.
  - Adaptive mesh refinement for elliptic/parabolic optimal control problems
    Becker/Rannacher (2001,...), Liu et. al. (2003,...), Günther/Hinze (2008), Hintermüller/Hoppe (2005,..), S. Ulbrich (2008), Vexler (2005,...). Applies mostly to linear-quadratic or convex optimal control problems.
- Efficient solution of optimization subproblems at fixed level.
  - Interior point
  - multigrid
  - domain decomposition
  - ....

Outline

Overview

Problem Formulation

Optimality Conditions

Discretization and Optimization

Optimization Algorithms

Conclusions

Infinite Dim. Problem and Optimization

- Consider the unconstrained optimization problem

$$\min_{\mathcal{U}} \widehat{J}(u), \qquad (P)$$

  where $\mathcal{U}$ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{U}}$.
- After a discretization, this leads to

$$\min_{\mathcal{U}^h} \widehat{J}(u_h) \qquad (P_h)$$

  for some finite dimensional subspace $\mathcal{U}^h \subset \mathcal{U}$.
- If we select a basis $\phi_1, \ldots, \phi_n$ of $\mathcal{U}^h$, then we can write every $u \in \mathcal{U}$ as $u = \sum_{i=1}^n u_i \phi_i$.
  (P_h) becomes an optimization problem in $\mathbf{u} = (u_1, \ldots, u_n)^T \in \mathbb{R}^n$, but the inner product leads to a weighted Euclidean product

$$\langle u_1, u_2 \rangle_{\mathcal{U}} = \mathbf{u}_1^\top \mathbf{T} \mathbf{u}_2$$

  with positive definite $\mathbf{T} \in \mathbb{R}^{n \times n}$ given by $\mathbf{T}_{ij} = \langle \phi_i, \phi_j \rangle_{\mathcal{U}}$.

The discretized problem can be viewed as problem in $\mathbb{R}^n$

$$\min_{\mathbb{R}^n} \widehat{J}(\mathbf{u}),$$

but $\mathbb{R}^n$ is equipped with the weighted Euclidean product

$$\mathbf{u}_1^\top \mathbf{T} \mathbf{u}_2$$

not with $\mathbf{u}_1^\top \mathbf{u}_2$.

This introduces a scaling that depends on the basis chosen and on the mesh. Optimization algorithms that are not scaling invariant, will be affected by this.

Gradient Computation

- Let $\widehat{J} : \mathbb{R}^n \to \mathbb{R}$. Denote the derivative of $\widehat{J}$ by $D\widehat{J}$.
- The gradient $\nabla \widehat{J}(\mathbf{u})$ is defined to be the vector that satisfies

$$\langle \nabla \widehat{J}(\mathbf{u}), \mathbf{u}' \rangle = D\widehat{J}(\mathbf{u})\mathbf{u}' \quad \forall \mathbf{u}'$$

  (Riesz representation). $\nabla \widehat{J}(\mathbf{u})$ depends on the inner product.
- If we use

$$\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \mathbf{u}_1^\top \mathbf{u}_2,$$

  then

$$\nabla \widehat{J}(\mathbf{u}) = \nabla_1 \widehat{J}(\mathbf{u}) := \left(\frac{\partial}{\partial \mathbf{u}_j} \widehat{J}(\mathbf{u}) \right)_{j=1,\ldots,n},$$

  i.e., $\nabla \widehat{J}(\mathbf{u})$ is the vector of partial derivatives.
- If we use

$$\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \mathbf{u}_1^\top \mathbf{T} \mathbf{u}_2,$$

  then

$$D\widehat{J}(\mathbf{u})\mathbf{u}' = \nabla_1 \widehat{J}(\mathbf{u})^\top \mathbf{u}' = \left(\mathbf{T}^{-1} \nabla_1 \widehat{J}(\mathbf{u}) \right)^\top \mathbf{T}\mathbf{u}'$$

  i.e., $\nabla \widehat{J}(\mathbf{u}) = \mathbf{T}^{-1} \nabla_1 \widehat{J}(\mathbf{u})$.
- Same result as scaling of the $\mathbf{u}$-variable by $\mathbf{T}^{1/2}$.

- If we discretize the optimal control and solve the discretized problem as a nonlinear problem in $\mathbb{R}^n$ with standard Euclidean inner product, the convergence of
  - gradient
  - quasi-Newton
  - conjugate gradient (CG)
  - Newton CG
  - ...

  methods depend on the mesh size.
- Often, the finer the mesh size, the more poorly scaled the discretized nonlinear programming problems become.
- It is important to analyze the optimization algorithm for the infinite dimensional problem and to apply it properly to the discretized problems.
- If one can prove convergence of the optimization algorithm for the infinite dimensional problem, if one choses an appropriate discretization, and if one applies the optimization algorithm properly to the discretized problem, then the convergence of (Newton-type) optimization algorithms applied to the discretized problems if often (nearly) independent of the size of the discretization.

Example: Mesh Independence of Newton's Method

Newton's method applied to two elliptic optimal control problems.
α penalty parameter for the control.
h mesh size

Number of Iterations												
	Example 1											
	TOL $= 10^{-8}$						TOL $= 10^{-6}$					
$\alpha \quad h^{-1}$	12	24	48	96	192	384	12	24	48	96	192	384
10^{-6}	7	7	7	7	7	7	7	7	7	7	7	7
10^{-4}	8	8	8	8	8	8	7	7	7	7	7	7
10^{-2}	10	10	10	10	10	10	8	8	8	8	8	8
	Example 2											
10^{-6}	7	7	7	7	7	7	6	6	6	6	6	6
10^{-4}	9	9	9	9	9	9	7	7	7	7	7	7
10^{-2}	9	9	9	9	9	9	7	7	7	7	7	7

Number of Newton iterations is independent of mesh size h!

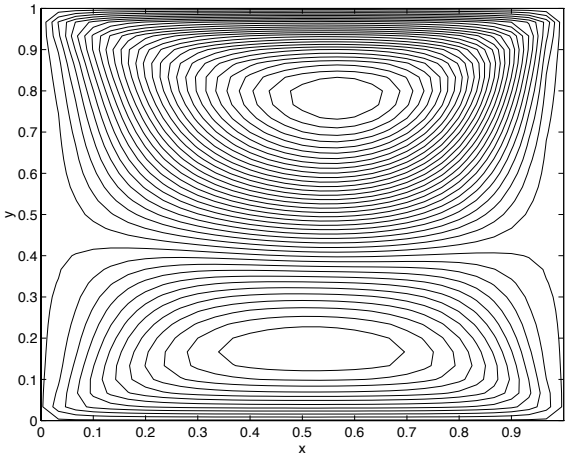
Example: Flow Separation in Driven Cavity

$$\min J(\mathbf{u}, \mathbf{g}) = \frac{1}{2} \int_{\{x_2=0.4\}} |u_2(x)|^2 \, dx + \frac{\gamma}{2} \, \|\mathbf{g}\|^2_{H^1(\Gamma_c)}$$

subject to

$$\begin{aligned} -\tfrac{1}{Re} \, \Delta \mathbf{u} + (\mathbf{u} \cdot \nabla)\mathbf{u} + \nabla p &= \mathbf{f} && \text{in } \Omega = (0,1)^2, \\ \text{div } \mathbf{u} &= 0 && \text{in } \Omega, \\ \mathbf{u} &= \mathbf{b} && \text{on } \Gamma_u, \\ \mathbf{u} &= \mathbf{g} && \text{on } (0,1) \times \{1\}. \end{aligned}$$

$\longrightarrow$ Velocity $\mathbf{g}$ (control) $\longrightarrow$



$\longrightarrow$ Velocity $\mathbf{b}$ (given 0.5) $\longrightarrow$

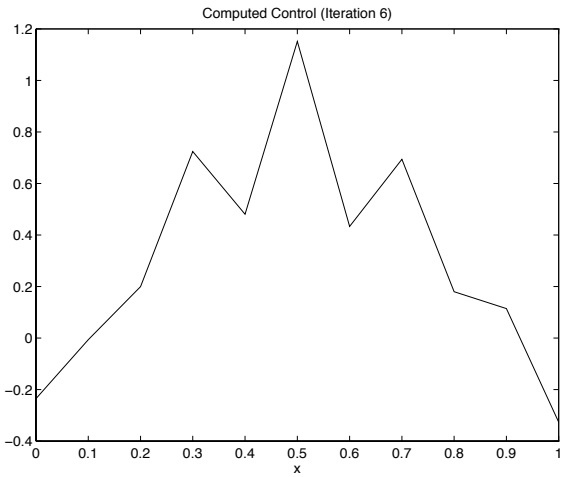
Note change of notation:
$\mathbf{u}, p$ (velocities, pressure) states,
$\mathbf{g}$ control. Control space $H^1(\Gamma_c)$.

We solve the optimal control problem using a reduced SQP method with BFGS approximation of the reduced Hessian.

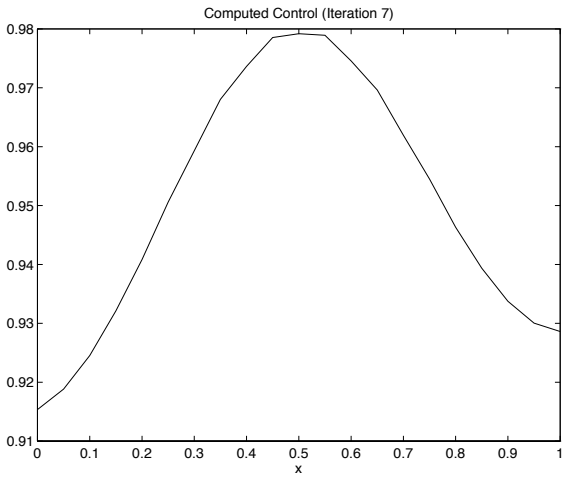
We use two variants:

1. Discretize the problem and treat the discretized problem as an NLP in $\mathbb{R}^n$.

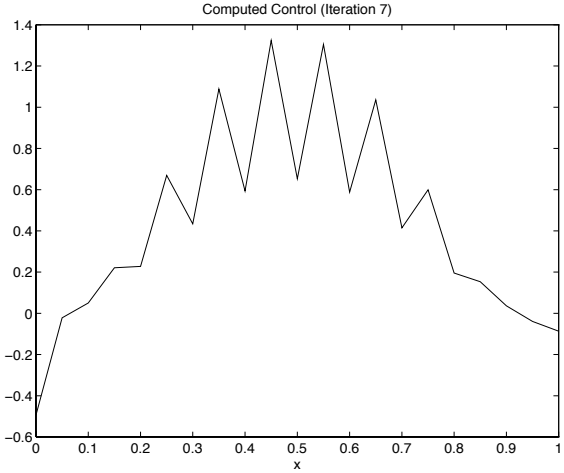
2. Use the infinite dimensional structure, i.e., control space is a subspace of $H^1(\Gamma_c)$ with $H^1(\Gamma_c)$-inner product,....



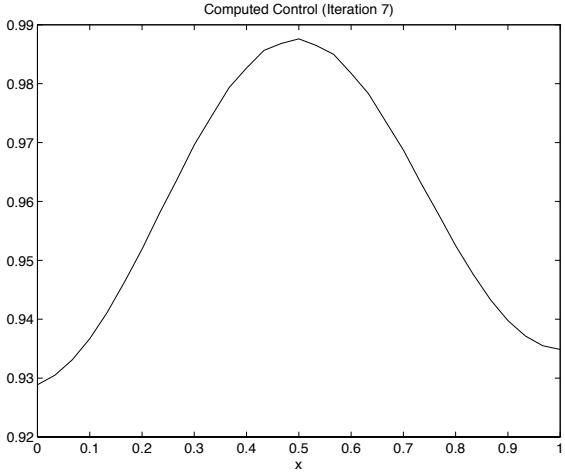
(Grid size $h = 1/5$).



(Grid size $h = 1/10$).



(Grid size $h = 1/10$).



(Grid size $h = 1/15$).

Dealing with Control and State Constraints

- Semismooth Newton Methods
  M. Ulbrich (2001), Kunisch/Hintermüller (2003),..
- Interior Point Methods
  Ito/Kelley/Sachs (1995), Prüfert/Tröltzsch/Weiser (2008), Schiela (2004,...), Schiela/Weiser (2004), Ulbrich/Ulbrich/H. (1999), Ulbrich/Ulbrich (2006),....
- Primal-Dual Active Set Methods
  Bergounioux/Ito/Kunisch (1997), Hintermüller/Kunisch (2001,...), ...
- Regularization Methods for State Constrained Problems
  Meyer/Rösch/Tröltzsch (2006), Hintermüller/Kunisch (2001,...),..

- Many (modifications of) algorithms motivated by PDE constrained optimization problems.
- Convergence analyses are available for infinite dimensional problems, but often only for small classes of problems (especially when state constraints are present).

Inexact Linear System Solves in Optimization Algorithms

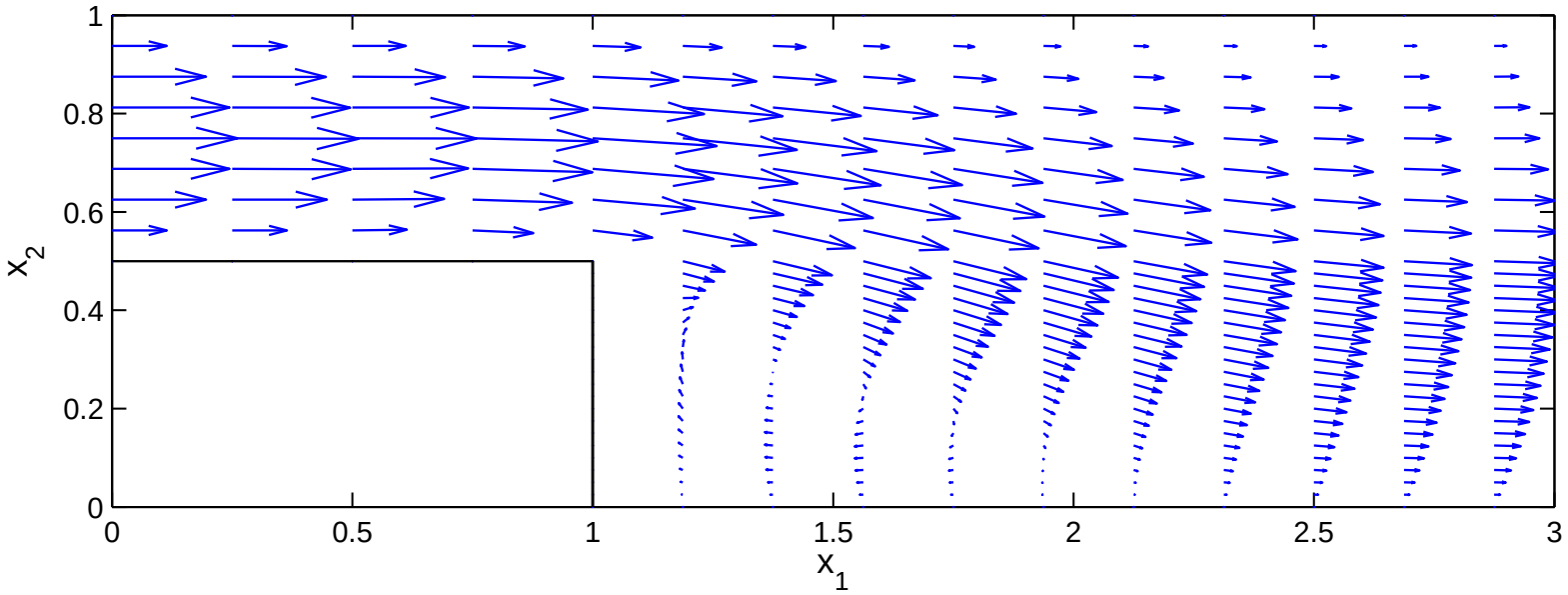
- In Newton-type methods for PDE constrained optimization most of the computing time is spent on solving linear systems.
- We need efficient iterative solvers and (matrix-free) preconditioners.
- Ho do we rigorously incorporate iterative (and therefore inexact) linear systems solves into the optimization algorithm?
  Adjust the accuracy of the iterative linear system solves based on the progress of the optimization iteration (avoid over-solving the linear systems).
  Use only quantities that can actually be computed (no Lipschitz constants).
- Old question (Ypma (1984), Dembo, Eisenstat, Steihaug (1982,...), Dennis, Walker (1984),....), but still not complely solved (H./Vicente (2001), H./Ridzal (2007), Curtis/Nocedal (2007),....).
- Some ideas can be extended to inexactness coming, e.g., for discretization.

Example: Navier Stokes Problem in 2D

$$\text{Minimize} \quad \frac{1}{2}\int_D |\boldsymbol{\nabla} \times \mathbf{u}|^2 \, d\mathbf{x} + \frac{\alpha}{2}\int_{\Gamma_c} |\nabla_s \mathbf{g}|^2 d\mathbf{x},$$

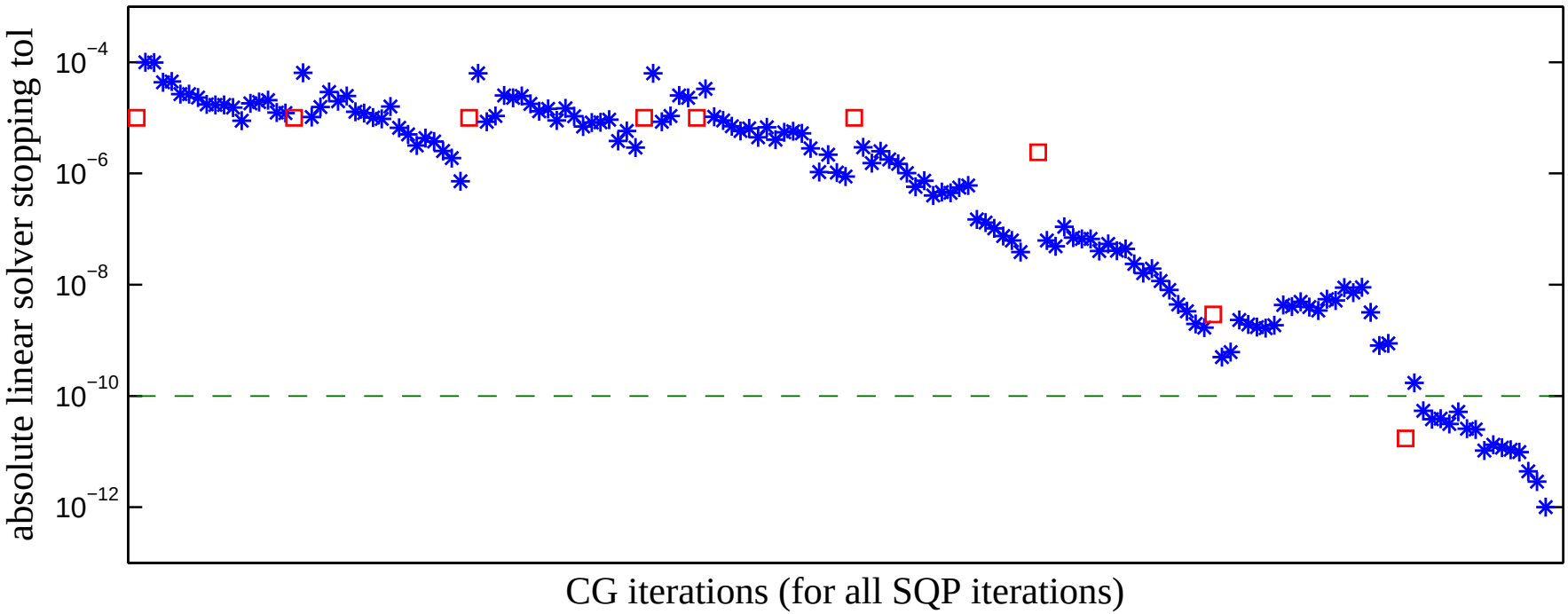
subject to

$$\begin{aligned} -\nu\Delta\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla p = \mathbf{f} \quad & \text{in} \quad \Omega, \\ \text{div } \mathbf{u} = 0 \quad & \text{in} \quad \Omega, \\ (\nu\nabla\mathbf{u} - pI)\mathbf{n} = \mathbf{0} \quad & \text{on} \quad \Gamma_{out}, \\ \mathbf{u} = \mathbf{g} \quad & \text{on} \quad \Gamma_c, \quad \mathbf{u} = \mathbf{b} \ \text{ on } \ \partial\Omega \setminus (\Gamma_c \cup \Gamma_{out}). \end{aligned}$$



Solve using composite-step trust-region method, using GMRES with ILU preconditioning (Ridzal (2006))

SQP Convergence history (□ beginning of SQP iteration, * absolute solver tolerances for linear systems solves within SQP iteration



	inx. ctrl	1e-12	1e-11	1e-10	1e-9	1e-8
converges	YES	YES	YES	YES	NO	NO
GMRES iter's	2672	4020	3728	3404	>10000	>10000
CG iter's	162	142	142	142	>500	>500
SQP iter's	9	8	8	8	>50	>50

Optimization algorithm must control stopping tolerances if linear systems have to be solved iteratively - which is the case for most PDE constrained problems.

Outline

Overview

Problem Formulation

Optimality Conditions

Discretization and Optimization

Optimization Algorithms

Conclusions

Conclusions

- PDE constrained optimization problems arise in many applications.
- The robust and efficient solution of such optimization problems requires the integration of application specific structure, numerical simulation and optimization algorithms.
- Much progress has been made in the areas of
  - existence and characterization of solutions,
  - handling of inequality constraints, especially inequality constraints on the state variables,
  - analysis of discretization errors in the optimization context and adaptation of the discretization
  - handling of inexactness,
  - efficient solution of optimization subproblems (KKT systems),
  - efficient integration of multiple meshes
  - ....
- .... but often only for model problems. All of these areas still pose many interesting and challenging research questions.

Thanks

- Thanks to
  - Collaborators, graduate students, colleagues,
  - AFOSR, DOE, NSF, Texas ATP, ExxonMobil, Bell Textron for past and/or current support.
- Thank you for your attention