Model Reduction in PDE-Constrained Optimization

Matthias Heinkenschloss

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

June 14, 2016

Funded in part by AFOSR and NSF

2016 EU Regional School
Aachen Institute for
Advanced Study in Computational Engineering Science (AICES)
RWTH Aachen University
Outline

Overview

Example Optimization Problems
Optimization Problem
Projection Based Model Reduction
Back to Optimization
Error Estimates
Linear-Quadratic Problems
Shape Optimization with Local Parameter Dependence
Semilinear Parabolic Problems
Trust-Region Framework
Overview

- PDE constrained optimization arises in many science and engineering applications
- Numerical solution is iterative and requires the solution of many PDEs.
- Can we systematically replace the underlying PDE by a (projection based) reduced order model to reduce the computational cost?

Flow control

Control of semilinear reaction advection diffusion

Reservoir optimization

Shape optimization of biochips
Reduced order modeling has a long history in optimization

- Newton’s method uses a ’reduced order’ (reduced nonlinearity) model.
- Surrogate optimization.
- Multilevel optimization.
- ....

This course

- Focusses on projection based reduced order models.
- There is a close connection with surrogate optimization and especially with multilevel optimization.
Original problem

\[ \begin{align*}
\min J(y, u) \\
\text{s.t. } c(y, u) &= 0, \quad \text{(PDE, size } n) \\
\quad &\quad u \in U_{ad}, \quad \text{(control constraints)}
\end{align*} \]

where \( y \in \mathbb{R}^n, n \text{ large, are the states and } u \in \mathbb{R}^m \text{ are the controls.} \)

Reduced order problem

Construct \( W, V \in \mathbb{R}^{n \times r}, r \ll n. \ \text{rank}(V) = \text{rank}(W) = r. \) Reduced order problem:

\[ \begin{align*}
\min J(V\hat{y}, u) \\
\text{s.t. } W^T c(V\hat{y}, u) &= 0, \quad \text{(ROM PDE, size } r) \\
\quad &\quad u \in U_{ad}. 
\end{align*} \]

Optimization variables: states \( \hat{y} \in \mathbb{R}^r, r \ll n, \) and controls \( u \in \mathbb{R}^m. \) Reduced state equation \( W^T c(V\hat{y}, u) = 0 \in \mathbb{R}^r. \)
Rich literature on projection based ROMs in optimization, incl.:

- (Strongly convex) Linear-Quadratic Optimal Control Problems
  [Antil et al., 2010], [Chen and Quarteroni, 2014], [Gubisch and Volkwein, 2013],
  [Kammann et al., 2013], [Kärcher and Grepl, 2014b],
  [Kärcher and Grepl, 2014a], [Kärcher et al., 2014], [Negri et al., 2013]
  [Tröltzsch and Volkwein, 2009], [Volkwein, 2011], ...

- Shape/Design Optimization [Amsallem et al., 2015], [Antil et al., 2011],
  [Choi et al., 2015], [Rozza and Manzoni, 2010], [Zahr and Farhat, 2015], ...

- Flow control [Borggaard and Gugercin, 2015], [Kunisch and Volkwein, 1999],
  [Ravindran, 2000], [Rowley et al., 2004], ...

- Newton-Kantorovich type estimates [Dihlmann and Haasdonk, 2015],
  [Gohlke, 2013].

- Optimality system based [Kunisch and Volkwein, 2008], [Grimm et al., 2015],
  [Kunisch and Müller, 2015], ...

- Balanced Truncation: [Antil et al., 2010], [Antil et al., 2011],
  [Antil et al., 2012], [Sun et al., 2008].

- Trust-region based approaches [Afanasiev and Hinze, 2001], [Arian et al., 2000],
  [Fahl and Sachs, 2003], [Gohlke, 2013], [Yue and Meerbergen, 2013], ...

- Model Predictive control [Ghiglieri and Ulbrich, 2014],
  [Alla and Volkwein, 2014], ...

- Feedback control [Kunisch et al., 2004], [Kunisch and Xie, 2005],
  [Alla and Falcone, 2013], ...
Outline

Overview

Example Optimization Problems

Optimization Problem

Projection Based Model Reduction

Back to Optimization

Error Estimates

Linear-Quadratic Problems

Shape Optimization with Local Parameter Dependence

Semilinear Parabolic Problems

Trust-Region Framework
Overview

Some examples of optimization problems governed by partial differential equations (PDEs)

- Linear Quadratic Elliptic Problem
- Linear Quadratic Parabolic Problem
- Shape Optimization with Local Parameter Dependence
- Oil Reservoir Waterflooding Optimization
Linear Quadratic Elliptic Problem

\[
\text{minimize } \frac{1}{2} \int_{\Omega_s} (y(x) - \hat{y}(x))^2 dx + \frac{\alpha}{2} \int_{\partial\Omega_c} u^2(x) dx
\]

subject to

\[-\kappa_f \Delta y(x) + a(x) \cdot \nabla y(x) = 0, \quad x \in \Omega_f,\]
\[-\kappa_s \Delta y(x) = f(x), \quad x \in \Omega_s,\]
\[\kappa \frac{\partial}{\partial n} y(x) = 0, \quad x \in \partial\Omega \setminus \partial\Omega_c,\]
\[y(x) = d(x) + u(x), \quad x \in \partial\Omega_c.\]

Velocity field \(a(x)\) and control boundary \(\partial\Omega_c\) (bold vertical line)
Finite element approximation

\[
\min \frac{1}{2} y^T Q y + y^T c + \frac{1}{2} u^T R u,
\]
\[
s.t. \ A y(t) + B u(t) = f.
\]

Strongly convex problem.
Linear Quadratic Parabolic Problem
([Antil et al., 2010], modeled after [Ded´e and Quarteroni, 2005])

Minimize \( \frac{1}{2} \int_0^T \int_D (y(x, t) - d(x, t))^2 dx dt + \frac{10^{-4}}{2} \int_0^T \int_{U_1 \cup U_2} u^2(x, t) dx dt, \)

subject to
\[
\frac{\partial}{\partial t} y(x, t) - \nabla (\kappa \nabla y(x, t)) + a(x) \cdot \nabla y(x, t) = u(x, t) \chi_{U_1}(x) + u(x, t) \chi_{U_2}(x) \quad \text{in } \Omega \times (0, T),
\]

with boundary conditions \( y(x, t) = 0 \) on \( \Gamma_D \times (0, T) \), \( \kappa \frac{\partial}{\partial n} y(x, t) = 0 \) on \( \Gamma_N \times (0, T) \) and initial conditions \( y(x, 0) = 0 \) in \( \Omega \).

\( \Omega \) with boundary conditions for the advection diffusion equation

the velocity field \( a \) (obtained by solving steady Stokes equation)
Finite element discretization in space

\[ \min j(u) \equiv \frac{1}{2} \int_{0}^{T} \|Cy(t) - d(t)\|^2 + \frac{1}{2} u(t)^T D u(t) \, dt, \]

where \( y(t) = y(u; t) \) is the solution of

\[ My'(t) = Ay(t) + Bu(t), \quad t \in (0, T), \]
\[ y(0) = y_0. \]

Here \( y(t) \in \mathbb{R}^n, M \in \mathbb{R}^{n \times n} \text{ invert.}, A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}, n \text{ large}. \]

Strongly convex problem.
Shape Optimization with Local Parameter Dependence
([Antil et al., 2011, Antil et al., 2012])

Geometry motivated by biochip

Problems where the shape param. \( \theta \) only influences a (small) subdomain:

\[
\Omega(\theta) := \Omega_1 \cup \Omega_2(\theta), \quad \Omega_1 \cap \Omega_2(\theta) = \emptyset, \quad \Gamma = \Omega_1 \cap \Omega_2(\theta).
\]

Here \( \Omega_2(\theta) \) is the top left yellow, square domain.
\[
\min_{\theta_{\text{min}} \leq \theta \leq \theta_{\text{max}}} J(\theta) = \int_0^T \int_{\Omega_{\text{obs}}} \frac{1}{2} |\nabla \times \mathbf{v}(x, t; \theta)|^2 \, dx + \int_{\Omega_{2}(\theta)} \frac{1}{2} |\mathbf{v}(x, t; \theta) - \mathbf{v}^d(x, t)|^2 \, dx \, dt
\]

where \( \mathbf{v}(\theta) \) and \( p(\theta) \) solve the Stokes equations

\[
\begin{align*}
\mathbf{v}_t(x, t) - \mu \Delta \mathbf{v}(x, t) + \nabla p(x, t) &= \mathbf{f}(x, t), & \text{in } \Omega(\theta) \times (0, T), \\
\nabla \cdot \mathbf{v}(x, t) &= 0, & \text{in } \Omega(\theta) \times (0, T), \\
\mathbf{v}(x, t) &= \mathbf{v}_{\text{in}}(x, t) & \text{on } \Gamma_{\text{in}} \times (0, T), \\
\mathbf{v}(x, t) &= \mathbf{0} & \text{on } \Gamma_{\text{lat}} \times (0, T), \\
-(\mu \nabla \mathbf{v}(x, t) - p(x, t) I)\mathbf{n} &= 0 & \text{on } \Gamma_{\text{out}} \times (0, T), \\
\mathbf{v}(x, 0) &= \mathbf{0} & \text{in } \Omega(\theta).
\end{align*}
\]

Here \( \Omega(\theta) = \overline{\Omega_1} \cup \overline{\Omega_2(\theta)} \) and \( \overline{\Omega_2(\theta)} \) is the top left yellow, square domain. Observation region \( \Omega_{\text{obs}} \) is part of the two reservoirs.
The semi-discretized minimization problem is

$$\min_{\theta \in \Theta_{ad}} J(\theta) := \int_{0}^{T} \frac{1}{2} \int_{0}^{T} \|Cv(t, \theta) + Fp(t, \theta) + Du(t) - d\|^2 \, dt$$

where \(v(\cdot, \theta), p(\cdot, \theta)\) solves the semi-discretized Stokes equations

$$M(\theta) \frac{d}{dt} v(t) + A(\theta)v(t) + B(\theta)p(t) = K(\theta)u(t) \quad t \in [0, T],$$

$$B^T(\theta)v(t) = L(\theta)u(t) \quad t \in [0, T],$$

$$M(\theta)v(0) = M(\theta)v_0,$$

$$\theta \in \Theta_{ad}$$
Oil Reservoir Waterflooding Optimization

From: http://plant-engineering.tistory.com/267
Reservoir Model: Two-phase immiscible incompressible flow with capillary pressure (see, e.g., [Peaceman, 1977], [Chen et al., 2006]).
States: saturations $s_w$, pressure $p$, velocity $v$.

\[
\phi(x) \frac{\partial}{\partial t} s_w(x, t) \\
+ \nabla \left( f_w(s_w(x, t)) \left[ v(x, t) + d(s_w(x, t)) \right] \right) = q_w(x, t), \quad x \in \Omega, \ t > 0,
\]
\[
v(x, t) + K(x) \lambda(s_w(x, t)) \nabla p(x, t) = 0, \quad x \in \Omega, \ t > 0,
\]
\[
\nabla \cdot v(x, t) = q(x, t), \quad x \in \Omega, \ t > 0,
\]
\[
v(x, t) \cdot n = 0, \quad x \in \partial \Omega, \ t > 0,
\]
\[
v_w(x, t) \cdot n = 0, \quad x \in \partial \Omega, \ t > 0,
\]
\[
s_w(x, 0) = s_{w_{init}}(x), \quad x \in \Omega.
\]

▶ Porosity $\phi(x)$, diagonal permeability $K(x)$ from SPE 10 dataset.
▶ Phase mobility $\lambda_\alpha = k_{r\alpha}/\mu_\alpha$; total mobility $\lambda = \lambda_o + \lambda_w$; water fractional flow function $f_w = \lambda_w/\lambda$.
▶ Capillary pressure, Brooks-Corey formula $p_c = \frac{1}{P_d} \left( \frac{s_w - s_{wc}}{1 - s_{or} - s_{wc}} \right)^{-\frac{1}{2}}$. 
Optimization Model Problem (Well Rate Optimization)

- Maximize Net Present Value (NPV)

\[
\int_0^T (1 + r_{\text{dis}})^{-t} \left[ -r_{\text{inj}} \sum_{i \in I_{\text{inj}}} \gamma q(x_i, t) - r_{\text{oper}} \sum_{i \in I_{\text{prod}}} \gamma |q(x_i, t)| f_w(s_w(x_i, t)) \right] \ dt \\
+ r_{\text{oil}} \sum_{i \in I_{\text{prod}}} \gamma |q(x_i, t)| f_o(s_w(x_i, t)) \right] \ dt
\]

- subject to
  - two-phase immiscible incompressible flow,
  - well rates sum up to zero (reservoir is closed)
    \[
    \sum_{i \in I_{\text{inj}} \cup I_{\text{prod}}} q(x_i, t) = 0, \quad t \in (0, T),
    \]
  - well rate bounds on each well
    \[
    q_{i,\text{low}} \leq q(x_i, t) \leq q_{i,\text{upp}}, \quad i \in I_{\text{inj}} \cup I_{\text{prod}}, \quad t \in (0, T).
    \]
- Data: Daily discount rate \( r_{\text{dis}} = 2 \times 10^{-4} \), oil price \( r_{\text{oil}} = 80 \), injection cost \( r_{\text{inj}} = 5 \), production cost \( r_{\text{pro}} = 5 \).
Example Result

- 500 days;
- 1000 time steps;
- $1200 \times 600 \times 10$ ft.$^3$;
- $60 \times 60 \times 5$ grid;
- $10 + 10 = 20$ wells;
- 25-days const. rate.
Overview

Example Optimization Problems

Optimization Problem

Projection Based Model Reduction

Back to Optimization

Error Estimates

Linear-Quadratic Problems

Shape Optimization with Local Parameter Dependence

Semilinear Parabolic Problems

Trust-Region Framework
Overview

- Formulate abstract optimization problem to introduce notation.
- Introduce adjoint equation method to compute gradient and Hessian information.
- Illustrate adjoint equation method on example problems.
Abstract Optimization Problem

- **Original problem**

  \[
  \min J(y, u) \\
  \text{s.t. } c(y, u) = 0, \quad \text{(governing PDE, state eqn.)} \\
  u \in U_{ad}, \quad \text{(control constraints)}
  \]

  where \( y \): states, \( u \): controls,
  - \((\mathcal{Y}, \| \cdot \|_Y), (\mathcal{C}, \| \cdot \|_C)\) Banach spaces, \((U, \| \cdot \|_U)\) Hilbert space
  - \( U_{ad} \subset U \) nonempty, closed convex set,
  - \( J : \mathcal{Y} \times U \to \mathbb{R}, c : \mathcal{Y} \times U \to \mathcal{C} \) are smooth mappings.
  - Can think \((\mathcal{Y}, \| \cdot \|_Y) = (\mathbb{R}^n, \| \cdot \|_2), (\mathcal{C}, \| \cdot \|_C) = (\mathbb{R}^n, \| \cdot \|_2), (U, \| \cdot \|_U) = (\mathbb{R}^m, \| \cdot \|_2)\).

- **Reduced order problem**

  \[
  \min \hat{J}(\hat{y}, u) \\
  \text{s.t. } \hat{c}(\hat{y}, u) = 0, \\
  u \in U_{ad},
  \]

  where \( \hat{y} \): ROM states, \( u \): controls,
  - \((\hat{\mathcal{Y}}, \| \cdot \|_{\hat{\mathcal{Y}}}), (\hat{\mathcal{C}}, \| \cdot \|_{\hat{\mathcal{C}}})\) Banach spaces,
  - \( \hat{J} : \hat{\mathcal{Y}} \times U \to \mathbb{R}, c : \hat{\mathcal{Y}} \times U \to \hat{\mathcal{C}} \) are smooth mappings.
Problem Formulation

Original problem

\[
\begin{align*}
\min & \quad J(y, u) \\
\text{s.t.} & \quad c(y, u) = 0, \\
& \quad u \in \mathcal{U}_{ad}.
\end{align*}
\]

\[\Downarrow\]

\( y(u) \) unique sol. of \( c(y, u) = 0 \)

\[\Downarrow\]

\[
\begin{align*}
\min & \quad j(u) \\
\text{s.t.} & \quad u \in \mathcal{U}_{ad}
\end{align*}
\]

where \( j(u) \overset{\text{def}}{=} J(y(u), u) \).

Reduced order problem

\[
\begin{align*}
\min & \quad \hat{J}(\hat{y}, u) \\
\text{s.t.} & \quad \hat{c}(\hat{y}, u) = 0, \\
& \quad u \in \mathcal{U}_{ad}.
\end{align*}
\]

\[\Downarrow\]

\( \hat{y}(u) \) unique sol. of \( \hat{c}(\hat{y}, u) = 0 \)

\[\Downarrow\]

\[
\begin{align*}
\min & \quad \hat{j}(u) \\
\text{s.t.} & \quad u \in \mathcal{U}_{ad}
\end{align*}
\]

where \( \hat{j}(u) \overset{\text{def}}{=} \hat{J}(\hat{y}(u), u) \).
Compute gradient and Hessian information for

\[ j(u) = J(y(u), u), \quad \text{where } y(u) \text{ solves } c(y, u) = 0. \]

Assumption

- \( J \) and \( c \) are twice continuously differentiable,
- \( c_y(y, u) \) is continuously invertible.

Consider problems with large number of controls \( u \).
Use adjoint equation approach for derivative computation.
See, e.g., chapter 1 in [Hinze et al., 2009] or [Heinkenschloss, 2008].
Gradient Computation

- Derivative

\[ \langle Dj(u), v \rangle_{\mathcal{U}^*, \mathcal{U}} = \langle DyJ(y(u), u), Dy(u)v \rangle_{\mathcal{Y}^*, \mathcal{Y}} + \langle DuJ(y(u), u), v \rangle_{\mathcal{U}^*, \mathcal{U}}. \]
Gradient Computation

- **Derivative**

\[
\langle Dj(u), v \rangle_{\mathcal{U}^*, \mathcal{U}} = \langle Dy J(y(u), u), Dy(u)v \rangle_{\mathcal{Y}^*, \mathcal{Y}} + \langle Du J(y(u), u), v \rangle_{\mathcal{U}^*, \mathcal{U}}
\]

- **Implicit function theorem applied to** \(c(y(u), u) = 0\) **gives**

\[
c_y(y, u)(Dy(u)v) + c_u(y, u)v = 0 \implies Dy(u)v = -c_y(y, u)^{-1} c_u(y, u).
\]
Gradient Computation

- **Derivative**

\[
\langle Dj(u), v \rangle_{U^*,U} = \langle DyJ(y(u), u), Dy(u)v \rangle_{Y^*,Y} + \langle DuJ(y(u), u), v \rangle_{U^*,U}
\]

- **Implicit function theorem applied to** \(c(y(u), u) = 0\) **gives**

\[
c_y(y, u)(Dy(u)v) + c_u(y, u)v = 0 \implies Dy(u)v = -c_y(y, u)^{-1}c_u(y, u).
\]

- **Derivative** \((y = y(u))\)

\[
\langle Dj(u), v \rangle_{U^*,U}
= \langle DyJ(y, u), -c_y(y, u)^{-1}c_u(y, u)v \rangle_{Y^*,Y} + \langle DuJ(y, u), v \rangle_{U^*,U}
= \langle -c_y(y, u)^{-*}DyJ(y, u), c_u(y, u)v \rangle_{C^*,C} + \langle DuJ(y, u), v \rangle_{U^*,U}
= \langle c_u(y, u)^*p + DuJ(y, u), v \rangle_{U^*,U}.
\]
Connection with Lagrangian $L(y, u, p) = J(y, u) + \langle p, c(y, u) \rangle c^*, c$:

- The adjoint variable $p$ solves $c_y(y, u)^* p = -D_y J(y, u)$, which is equivalent to $D_y L(y, u, p) = 0$.

- Derivative

$$\langle D j(u), v \rangle_{U^*, U} = \langle c_u(y, u)^* p + D_u J(y, u), v \rangle_{U^*, U} = \langle D_u L(y, u, p), v \rangle_{U^*, U}.$$
Connection with Lagrangian $L(y, u, p) = J(y, u) + \langle p, c(y, u) \rangle c^*, c$:

- The adjoint variable $p$ solves $c_y(y, u)^*p = -D_yJ(y, u)$, which is equivalent to $D_yL(y, u, p) = 0$.
- Derivative

\[
\langle Dj(u), v \rangle_{U^*, U} = \langle c_u(y, u)^*p + D_uJ(y, u), v \rangle_{U^*, U}
\]

\[
= \langle D_uL(y, u, p), v \rangle_{U^*, U}, \forall v \in U.
\]

The gradient $\nabla j(u)$ is the vector in $U$ such that

\[
\langle \nabla j(u), v \rangle_U = \langle Dj(u), v \rangle_{U^*, U}
\]

\[
= \langle c_u(y, u)^*p + D_uJ(y, u), v \rangle_{U^*, U}, \forall v \in U
\]

(Riesz representation)
Connection with Lagrangian \( L(y, u, p) = J(y, u) + \langle p, c(y, u) \rangle c^*, c \):

- The adjoint variable \( p \) solves \( c_y(y, u)^* p = -D_y J(y, u), \)
  which is equivalent to \( D_y L(y, u, p) = 0. \)
- Derivative

\[
\langle D j(u), v \rangle_{U^*, U} = \langle c_u(y, u)^* p + D_u J(y, u), v \rangle_{U^*, U}
= \langle D_u L(y, u, p), v \rangle_{U^*, U},
\]

The gradient \( \nabla j(u) \) is the vector in \( U \) such that

\[
\langle \nabla j(u), v \rangle_U = \langle D j(u), v \rangle_{U^*, U}
= \langle c_u(y, u)^* p + D_u J(y, u), v \rangle_{U^*, U} \quad \forall v \in U
\]

(Riesz representation)

Gradient Computation Using Adjoints

1. Given \( u \), solve \( c(y, u) = 0 \) for \( y \) (if not done already).
2. Solve the adjoint equation \( c_y(y(u), u)^* p = -D_y J(y(u), u) \) for \( p \).
   Denote the solution by \( p(u) \).
3. Compute \( D j(u) = D_u J(y(u), u) + c_u(y(u), u)^* p(u). \)

Two PDE solves (possibly nonlinear PDE in step 1, linear PDE in step 2)
Hessian Computation

- Apply implicit differentiation to
  \[ D_j(u) = D_u J(y(u), u) + c_u(y(u), u) \cdot p(u) \]
  to compute Hessian information.
Hessian Computation

- Apply implicit differentiation to
  \[ D_j(u) = D_u J(y(u), u) + c_u(y(u), u)^* p(u) \]
  to compute Hessian information.

- Hessian–Times–Vector Computation

  1. Given \( u \), solve \( c(y, u) = 0 \) for \( y \) (if not done already).
  2. Solve adjoint eqn. \( c_y(y, u)^* p = -D_y J(y, u) \) for \( p \) (if not done already).
  3. Solve \( c_y(y, u) w = -c_u(y, u) v \).
  4. Solve \( c_y(y, u)^* q = -D_{yy} L(y, u, p) w - D_{yu} L(y, u, p) v \).
  5. Compute
     \[ D^2 j(u) v = c_u(y, u)^* q + D_{uy} L(y, u, p) w + D_{uu} L(y, u, p) v. \]
Hessian Computation

- Apply implicit differentiation to
  \[ D_j(u) = D_u J(y(u), u) + c_u(y(u), u)^* p(u) \]
  to compute Hessian information.

- Hessian–Times–Vector Computation
  1. Given \( u \), solve \( c(y, u) = 0 \) for \( y \) (if not done already).
  2. Solve adjoint eqn. \( c_y(y, u)^* p = -D_y J(y, u) \) for \( p \) (if not done already).
  3. Solve \( c_y(y, u) w = -c_u(y, u) v \).
  4. Solve \( c_y(y, u)^* q = -D_{yy} L(y, u, p) w - D_{yu} L(y, u, p) v \).
  5. Compute
     \[ D^2 j(u)v = c_u(y, u)^* q + D_{uy} L(y, u, p) w + D_{uu} L(y, u, p) v. \]

Two linear PDE solves in steps 3+4 per direction \( v \).
Hessian Computation

Apply implicit differentiation to

\[
D_j(u) = D_u J(y(u), u) + c_u(y(u), u) * p(u)
\]

to compute Hessian information.

Hessian–Times–Vector Computation

1. Given \( u \), solve \( c(y, u) = 0 \) for \( y \) (if not done already).
2. Solve adjoint eqn. \( c_y(y, u)^* p = -D_y J(y, u) \) for \( p \) (if not done already).
3. Solve \( c_y(y, u) w = -c_u(y, u) v \).
4. Solve \( c_y(y, u)^* q = -D_{yy} L(y, u, p) w - D_{yu} L(y, u, p) v \).
5. Compute

\[
D^2 j(u)v = c_u(y, u)^* q + D_{uy} L(y, u, p) w + D_{uu} L(y, u, p) v.
\]

Two linear PDE solves in steps 3+4 per direction \( v \).

Vector \( s_u \) solves Newton equation \( \nabla^2 j(u) s_u = -\nabla j(u) \)

if and only if \( (s_y, s_u) \) solves the quadratic program

\[
\min \left\langle \begin{bmatrix} D_y J(.), & D_u J(.) \end{bmatrix}, \begin{bmatrix} s_y, & s_u \end{bmatrix} \right\rangle + \frac{1}{2} \left\langle \begin{bmatrix} s_y, & s_u \end{bmatrix}, \begin{bmatrix} D_{yy} L(.,) & D_{yu} L(.,) \\ D_{uy} L(.,) & D_{uu} L(.,) \end{bmatrix} \begin{bmatrix} s_y, & s_u \end{bmatrix} \right\rangle,
\]

s.t. \( c_y(.) s_y + c_u(.) s_u = 0 \),

where \( .) = (y(u), u) \) and \( (..) = (y(u), u, p(u)) \).
Example: Elliptic Optimal Control Problem

- **Problem:**

  $$\text{Minimize } j(u) = \frac{1}{2} \int_D (y(x) - d(x))^2 dx + \frac{\alpha}{2} \int_{\Gamma_c} |\nabla u(x)|^2 d\sigma$$

  where $y$ solves

  $$-\nabla (\kappa \nabla y(x)) + a \cdot \nabla y(x) = f(x) \quad \text{in } \Omega,$$
  $$y(x) = u(x) \quad \text{on } \Gamma_c,$$
  $$y(x) = 0 \quad \text{on } \Gamma_D,$$
  $$\kappa \nabla y(x)n = 0, \quad \text{on } \Gamma_N.$$  

- **Control space** $H^1_0(\Gamma_c)$. **State space**

  $$\mathcal{V} = \{ v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_c \cup \Gamma_D \}.$$
Example: Elliptic Optimal Control Problem

Problem:

Minimize \( j(u) = \frac{1}{2} \int_D (y(x) - d(x))^2 \, dx + \frac{\alpha}{2} \int_{\Gamma_c} |\nabla u(x)|^2 \, d\sigma \)

where \( y \) solves

\[
-\nabla(\kappa \nabla y(x)) + a \cdot \nabla y(x) = f(x) \quad \text{in } \Omega,
\]

\[
y(x) = u(x) \quad \text{on } \Gamma_c,
\]

\[
y(x) = 0 \quad \text{on } \Gamma_D,
\]

\[
\kappa \nabla y(x)n = 0, \quad \text{on } \Gamma_N.
\]

Control space \( H^1_0(\Gamma_c) \). State space

\[ \mathcal{V} = \{ v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_c \cup \Gamma_D \}. \]

Handle Dirichlet boundary condition using the inverse trace theorem.

(In finite element approximation: via interpolation):

For every \( u \in H^1_0(\Gamma_c) \) there exists \( y(u; \cdot) \in H^1(\Omega) \) such that

\( y(u; x) = u(x) \) on \( \Gamma_c \).

Moreover \( H^1_0(\Gamma_c) \ni u \mapsto y(u; \cdot) \in H^1(\Omega) \) is bounded and linear.
Example: Elliptic Optimal Control Problem

- **Problem:**

  \[
  \text{Minimize } j(u) = \frac{1}{2} \int_D (y(x) - d(x))^2 dx + \frac{\alpha}{2} \int_{\Gamma_c} |\nabla u(x)|^2 d\sigma
  \]

  where \( y \) solves

  \[
  -\nabla (\kappa \nabla y(x)) + a \cdot \nabla y(x) = f(x) \quad \text{in } \Omega, \\
  y(x) = u(x) \quad \text{on } \Gamma_c, \\
  y(x) = 0 \quad \text{on } \Gamma_D, \\
  \kappa \nabla y(x)n = 0, \quad \text{on } \Gamma_N.
  \]

- **Control space** \( H^1_0(\Gamma_c) \). **State space**

  \( \mathcal{V} = \{ v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_c \cup \Gamma_D \} \).

- **Handle Dirichlet boundary condition using the inverse trace theorem.**
  (In finite element approximation: via interpolation):

  For every \( u \in H^1_0(\Gamma_c) \) there exists \( y(u; \cdot) \in H^1(\Omega) \) such that
  \[
  y(u; x) = u(x) \text{ on } \Gamma_c.
  \]
  Moreover \( H^1_0(\Gamma_c) \ni u \mapsto y(u; \cdot) \in H^1(\Omega) \) is bounded and linear.

- **Write** \( y = y_0 + y(u; \cdot) \), where \( y_0 \in \mathcal{V} \).
Lagrangian:

\[ L(y, u, p) = \frac{1}{2} \int_{D} (y_0(x) + y(u; x) - y^d(x))^2 dx + \frac{\alpha}{2} \int_{\Gamma_c} |\nabla u(x)|^2 d\sigma \]

\[ + \int_{\Omega} \kappa \nabla y_0 \nabla p + a \cdot \nabla y_0 p \, dx \]

\[ + \int_{\Omega} \kappa \nabla y(u; \cdot) \nabla p + a \cdot y(u; \cdot) \nabla p - fp \, dx \]
Lagrangian:

\[
L(y, u, p) = \frac{1}{2} \int_D \left( y_0(x) + y(u; x) - y^d(x) \right)^2 dx + \frac{\alpha}{2} \int_{\Gamma_c} |\nabla u(x)|^2 d\sigma
\]

\[
+ \int_{\Omega} \kappa \nabla y_0 \nabla p + a \cdot \nabla y_0 p \, dx
\]

\[
+ \int_{\Omega} \kappa \nabla y(u; \cdot) \nabla p + a \cdot y(u; \cdot) \nabla p - f p \, dx
\]

Adjoint equation:

\[
-\nabla (\kappa \nabla p(x)) - a \cdot \nabla p(x) = -(y_0(x) + y(u; x) - y^d(x))|_D, \quad \text{in } \Omega,
\]

\[
p(x) = 0, \quad \text{on } \Gamma_c \cup \Gamma_D,
\]

\[
\kappa \nabla p(x) \cdot n + a \cdot n p(x) = 0, \quad \text{on } \Gamma_N.
\]
Lagrangian:

\[
L(y, u, p) = \frac{1}{2} \int_D \left( y_0(x) + y(u; x) - y^d(x) \right)^2 dx + \frac{\alpha}{2} \int_{\Gamma_c} |\nabla u(x)|^2 d\sigma \\
+ \int_\Omega \kappa \nabla y_0 \nabla p + a \cdot \nabla y_0 p \, dx \\
+ \int_\Omega \kappa \nabla y(u; \cdot) \nabla p + a \cdot y(u; \cdot) \nabla p - f p \, dx
\]

Adjoint equation:

\[
-\nabla (\kappa \nabla p(x)) - a \cdot \nabla p(x) = -(y_0(x) + y(u; x) - y^d(x))|_D, \quad \text{in } \Omega,
\]

\[
p(x) = 0, \quad \text{on } \Gamma_c \cup \Gamma_D,
\]

\[
\kappa \nabla p(x) \cdot n + a \cdot n p(x) = 0, \quad \text{on } \Gamma_N.
\]

Derivative

\[
\langle D j(u), v \rangle_{H^1_0(\Gamma_c)^*, H^1(\Gamma_c)} \\
= \int_{\partial \Omega} \alpha \nabla u(x) \nabla v(x) \, d\sigma + \int_{\Omega} \kappa \nabla y(v; \cdot) \nabla p + a \cdot y(v; \cdot) \nabla p \, dx \\
+ \int_D (y_0(x) + y(u; x) - y^d(x)) y(v; x) \, dx
\]

What's the gradient?
The gradient $\nabla j(u) = g \in H^1_0(\Gamma_c)$ is a function that satisfies

$$\langle Dj(u), v \rangle_{H^1_0(\Gamma_c)^*, H^1(\Gamma_c)} = \int_{\partial \Omega} \alpha \nabla u(x) \nabla v(x) \, d\sigma + \int_{\Omega} \kappa \nabla y(v; \cdot) \nabla p + \mathbf{a} \cdot y(v; \cdot) \nabla p \, dx$$

$$+ \int_D (y_0(x) + y(u; x) - y^d(x)) y(v; x) \, dx$$

$$= \int_{\partial \Omega} g(x) v(x) + \nabla g(x) \nabla v(x) \, ds = \langle g, v \rangle_{H^1_0(\Gamma_c)} \quad \forall v \in H^1_0(\Gamma_c).$$
The gradient $\nabla j(u) = g \in H_0^1(\Gamma_c)$ is a function that satisfies

$$\langle Dj(u), v \rangle_{H_0^1(\Gamma_c)^*, H^1(\Gamma_c)} = \int_{\partial\Omega} \alpha \nabla u(x) \nabla v(x) \, d\sigma + \int_{\Omega} \kappa \nabla y(v; \cdot) \nabla p + a \cdot y(v; \cdot) \nabla p \, dx$$

$$+ \int_D (y_0(x) + y(u; x) - y^d(x)) y(v; x) \, dx$$

$$= \int_{\partial\Omega} g(x) v(x) + \nabla g(x) \nabla v(x) \, ds = \langle g, v \rangle_{H_0^1(\Gamma_c)} \quad \forall v \in H_0^1(\Gamma_c).$$

Solve Laplace equation on the boundary,

$$\int_{\Gamma_c} \nabla \tilde{g}(x) \nabla v(x) \, d\sigma = \int_{\Omega} \kappa \nabla y(v; \cdot) \nabla p + a \cdot y(v; \cdot) \nabla p \, dx$$

$$+ \int_D (y_0(x) + y(u; x) - y^d(x)) y(v; x) \, dx \quad \forall v \in H_0^1(\Gamma_c)$$

and then

$$\nabla j(u) = \alpha u + \tilde{g}.$$
The gradient $\nabla j(u) = g \in H^1_0(\Gamma_c)$ is a function that satisfies
\[
\langle Dj(u), v \rangle_{H^1_0(\Gamma_c)^*, H^1(\Gamma_c)}
= \int_{\partial \Omega} \alpha \nabla u(x) \nabla v(x) \, d\sigma + \int_{\Omega} \kappa \nabla y(v; \cdot) \nabla p + a \cdot y(v; \cdot) \nabla p \, dx
\]
\[+ \int_D (y_0(x) + y(u; x) - y^d(x))y(v; x) \, dx \]
\[= \int_{\partial \Omega} g(x)v(x) + \nabla g(x) \nabla v(x) \, ds = \langle g, v \rangle_{H^1_0(\Gamma_c)} \quad \forall v \in H^1_0(\Gamma_c).
\]

Solve Laplace equation on the boundary,
\[
\int_{\Gamma_c} \nabla \tilde{g}(x) \nabla v(x) \, d\sigma = \int_{\Omega} \kappa \nabla y(v; \cdot) \nabla p + a \cdot y(v; \cdot) \nabla p \, dx
\]
\[+ \int_D (y_0(x) + y(u; x) - y^d(x))y(v; x) \, dx \quad \forall v \in H^1_0(\Gamma_c)
\]
and then
\[
\nabla j(u) = \alpha u + \tilde{g}.
\]

Note: Other ways to incorporate Dirichlet boundary controls (Lagrange multipliers, very weak form of Laplace equation) may lead to different weak forms and to different control and state spaces.
Example: Parabolic Optimal Control Problem

Problem:

Minimize \( \frac{1}{2} \int_0^T \int_D (y(x, t) - d(x, t))^2 \, dx \, dt + \frac{\alpha}{2} \int_0^T \int_U u^2(x, t) \, dx \, dt \),

where \( y \) solves

\[
\frac{\partial}{\partial t} y(x, t) - \nabla(\kappa \nabla y(x, t)) + \mathbf{a} \cdot \nabla y(x, t) = u(x, t) \chi_U(x) \quad \text{in } \Omega \times (0, T),
\]

\( y(x, t) = 0, \) on \( \Gamma_D \times (0, T), \quad \kappa \nabla y(x, t) n = 0, \) on \( \Gamma_N \times (0, T), \quad y(x, 0) = 0, \) in \( \Omega. \)

Lagrangian (formally)

\[
L(y, u, p) = \frac{1}{2} \int_0^T \int_D (y(x, t) - d(x, t))^2 \, dx \, dt + \frac{\alpha}{2} \int_0^T \int_U u^2(x, t) \, dx \, dt \\
+ \int_0^T \int_\Omega \frac{\partial}{\partial t} y(x, t) p(x, t) + \kappa \nabla y(x, t) \nabla p(x, t) + \mathbf{a} \cdot \nabla y(x, t) p(x, t) \, dx \, dt \\
- \int_0^T \int_U u(x, t) p(x, t) \, dx \, dt
\]

(For details chapter 1 in [Hinze et al., 2009] or [Tröltzsch, 2010a].)
Adjoint equation

\[- \frac{\partial}{\partial t} p(x, t) - \nabla (\kappa \nabla p(x, t))\]

\[- \mathbf{a} \cdot \nabla p(x, t) = -(y(x, t) - d(x, t)) \chi_D(x) \quad \text{in } \Omega \times (0, T),\]

\[p(x, t) = 0, \quad \text{on } \Gamma_D \times (0, T),\]

\[(\kappa \nabla p(x, t) + \mathbf{a} p(x, t)) n = 0, \quad \text{on } \Gamma_N \times (0, T),\]

\[p(x, T) = 0, \quad \text{in } \Omega.\]

Gradient

\[\nabla j(u) = \alpha u(x, t) - p(x, t) \quad x \in U, t \in (0, T).\]
Outline

Overview

Example Optimization Problems

Optimization Problem

Projection Based Model Reduction

Back to Optimization

Error Estimates

Linear-Quadratic Problems

Shape Optimization with Local Parameter Dependence

Semilinear Parabolic Problems

Trust-Region Framework
Reduced-Order Dynamical Systems

\[
\dot{y}(t) = Ay(t) + Bu(t) + f(t)
\]

\[
s(t) = Cy(t)
\]

\[
\dot{\hat{y}}(t) = W^T AV\hat{y} + W^T Bu + W^T f(t)
\]

\[
\hat{s} = CV\hat{y}
\]

\[
\dot{y}(t) = f(y(t), u(t), t)
\]

\[
s(t) = g(y(t), t)
\]

Replace \( y(t) \in \mathbb{R}^n \) by \( V\hat{y}(t) = \sum_{i=1}^{r} v_i\hat{y}_i(t) \), \( \hat{y} \in \mathbb{R}^r \) where \( r \ll n \) and multiply the state equation by \( W^T \). (Often \( W = V \).)

\[
\hat{s} = g(V\hat{y})
\]

Two main questions:

▶ Accuracy of the reduced order model? Approximation of the input-to-output map \( u \mapsto s \).

▶ Efficiency of the reduced order model?
Projection Based Reduced Order Models (ROMs) - Overview

- Reduced Basis Method.
  See books [Hesthaven et al., 2015], [Patera and Rozza, 2007], [Quarteroni et al., 2016].

- Proper Orthogonal Decomposition (POD).
  See survey article [Hinze and Volkwein, 2005] and sections in books [Hesthaven et al., 2015], [Quarteroni et al., 2016].

- Balanced Truncation Model Reduction (BTMR) for linear time invariant problems.
  See book [Antoulas, 2005] and for connections with POD [Rowley, 2005].

- Interpolation Based Model Reduction.
  See survey articles [Antoulas et al., 2010], [Benner et al., 2015].
Reduced Order Model (ROM) of Parametric Elliptic PDE

- Given Hilbert space $\mathcal{V}$, bounded coercive bilinear form $a(\cdot, \cdot; \mu)$ on $\mathcal{V} \times \mathcal{V}$, and bounded linear functional $f(\cdot; \mu)$ on $\mathcal{V}$.
- Find $y \in \mathcal{V}$ that satisfies the variational formulation:
  $$a(y, v; \mu) = f(v; \mu) \quad \forall v \in \mathcal{V}.$$ 
- Given bounded linear functional $\ell(\cdot)$ on $\mathcal{V}$, we are often interested in
  $$s(\mu) = \ell(y(\mu)) \quad \text{(quantity of interest)}$$
- Example
  $$-\nabla^2 y(\mu) + \begin{bmatrix} \mu \\ 0 \end{bmatrix} \cdot \nabla y(\mu) = 100e^{-5\sqrt{\|x\|^2}}, \quad \text{in } \Omega = [-1, 1]^2, \quad \mu \in [-10, 10],$$
  $$y(\mu) = 0, \quad \text{on } \partial\Omega,$$
  $$s(\mu) = \int_{\Omega} y(x; \mu) dx.$$ 

Here $\mathcal{V} = H^1_0(\Omega)$ and $f(v; \mu) = \int_{\Omega} 100e^{-5\sqrt{\|x\|^2}} v(x) dx$,
$$a(y, v; \mu) = \int_{\Omega} \nabla y(x) \cdot \nabla v(x) + \begin{bmatrix} \mu \\ 0 \end{bmatrix} \cdot \nabla y(x)v(x) dx.$$
Finite Element Approximation

- $\mathcal{V}_n = \text{span}\{\phi_1, \ldots, \phi_n\} \subset \mathcal{V}$.

  Find $y = y(\mu) \in \mathcal{V}_n$ such that

  $$a(y, v; \mu) = f(v; \mu) \quad \forall v \in \mathcal{V}_n.$$  

- Linear system for $y = \sum_{i=1}^{n} y_i \phi_i$:

  $$A(\mu)y = f(\mu), \quad (n \times n)$$

  where

  $$A(\mu)_{ij} = a(\phi_j, \phi_i; \mu),$$

  $$f(\mu)_i = f(\phi_i; \mu).$$

- Well posedness: If there exists $\alpha > 0$ such that

  $$a(v, v; \mu) \geq \alpha \|v\|_\mathcal{V}^2$$

  for all $v \in \mathcal{V}$ and $\mu \in \Gamma$,

  then $y^T A(\mu)y \geq \alpha \|y\|_\mathcal{V}^2$ (norm $\|y\|_\mathcal{V}^2 = \sum_{i,j=1}^{n} y_i y_j \langle \phi_i, \phi_j \rangle_\mathcal{V}$) and

  $$\begin{align*}
  A(\mu)y &= f(\mu) \\
  \Rightarrow \alpha \|y\|_\mathcal{V}^2 &\leq y^T A(\mu)y = y^T f(\mu) \leq \|y\|_\mathcal{V} \|f(\mu)\|_{\mathcal{V}^{-1}} \\
  \Rightarrow \|y\|_\mathcal{V} &\leq \alpha^{-1} \|f(\mu)\|_{\mathcal{V}^{-1}} \Rightarrow \|A(\mu)^{-1}\|_\mathcal{V} \leq \alpha^{-1}. 
  \end{align*}$$
Reduced Order Model (ROM)

- Subspace \( \mathcal{V}_r = \text{span}\{\zeta_1, \ldots, \zeta_r\} \subset \mathcal{V}_n, r \ll n \).
  Find \( \hat{y} = \hat{y}(\mu) \in \mathcal{V}_r \) such that
  \[
  a(\hat{y}, v; \mu) = f(v; \mu) \quad \forall v \in \mathcal{V}_r.
  \]

- Linear system for \( \hat{y} = \sum_{i=1}^{r} \hat{y}_i \zeta_i \):
  - Represent \( \mathcal{V}_r \) basis: \( \zeta_i = \sum_{k=1}^{n} v_{ki} \phi_k, \mathbf{V} \equiv (v_{ij}) \in \mathbb{R}^{n \times r} \).
  - Insert into bilinear/linear forms
    \[
    a(\zeta_j, \zeta_i; \mu) = \sum_{k=1}^{n} \sum_{\hat{k}=1}^{n} v_{ki} v_{\hat{k}j} a(\phi_{\hat{k}}, \phi_k; \mu) = (\mathbf{V}^T \mathbf{A}(\mu) \mathbf{V})_{ij},
    \]
    \[
    f(\zeta_i; \mu) = \sum_{k=1}^{n} v_{ki} f(\phi_k; \mu) = (\mathbf{V}^T \mathbf{f}(\mu))_i.
    \]

- ROM
  \[
  a(\hat{y}, \zeta_i; \mu) = \sum_{j=1}^{r} \hat{y}_j a(\zeta_j, \zeta_i; \mu) = f(\zeta_i; \mu), \quad i = 1, \ldots, r,
  \]
  is equivalent to
  \[
  \mathbf{V}^T \mathbf{A}(\mu) \mathbf{V} \hat{y} = \mathbf{V}^T \mathbf{f}(\mu). \quad (r \times r)
  \]

- Well posedness inherited: \( \hat{y}^T \mathbf{V}^T \mathbf{A}(\mu) \mathbf{V} \hat{y} \geq \alpha \| \mathbf{V} \hat{y} \|_2^2 \equiv \alpha \| \hat{y} \|_\mathbf{V}^2. \)
Basic ROM Algorithm

1. **Compute Snapshots:** Given \( \{\mu_1, \ldots, \mu_r\} \) compute full solutions:
\[
A(\mu_i)y(\mu_i) = f(\mu_i)
\]

2. **Orthogonalize:** Find \( V \in \mathbb{R}^{n \times r} \) where \( V^T V = I \) and
\[
\text{Ran}(V) = \text{span}\{y(\mu_1), \ldots, y(\mu_r)\}
\]

3. **Construct Reduced Order System:**
\[
\hat{A}(\mu) = V^T A(\mu) V \in \mathbb{R}^{r \times r}, \quad \hat{f}(\mu) = V^T f(\mu) \in \mathbb{R}^r
\]

4. **ROM Solution:** Cheaply solve reduced order system for out-of-sample parameter choices \( \mu \):
\[
\hat{A}(\mu)\hat{y} = \hat{f}(\mu).
\]

Approximation \( y(\mu) \approx V\hat{y}(\mu) \).

Above is basic algorithm

- At what parameters \( \mu_i \) do we sample?
- ROM \( \hat{A}(\mu), \hat{f}(\mu) \) is smaller, but evaluation of \( \hat{A}(\mu), \hat{f}(\mu) \) not cheap
Error in the solution

\[ A(\mu)(y - V\hat{y}) = f(\mu) - A(\mu)V\hat{y} \]

implies that

\[ \|y - V\hat{y}\| \leq \|A(\mu)^{-1}\| \|f(\mu) - A(\mu)V\hat{y}\| . \]

Error in output of interest \( s(\mu) = l^T y(\mu) \):

\[ |s(\mu) - \hat{s}(\mu)| = |l^T y - \hat{l}^T \hat{y}| \]

\[ \leq \|l\| \|A(\mu)^{-1}\| \|f(\mu) - A(\mu)V\hat{y}\| . \]

But can do much better [Machiels et al., 2001].
ROM Error Estimate - Quantity of Interest

Primal:
\[ A(\mu)y(\mu) = f(\mu) \]
\[ s(\mu) = l^T y(\mu) \]

Primal RB:
\[ \hat{A}(\mu)\hat{y}(\mu) = \hat{f}(\mu) \]
\[ \hat{s}(\mu) = \hat{l}^T \hat{y}(\mu) \]

Primal Residual:
\[ \| f(\mu) - A(\mu)V\hat{y}(\mu) \| \]

Primal Bound:
\[ \| y - V\hat{y} \| = \| A^{-1}(f - AV\hat{y}) \| \leq \| A^{-1} \| \| f - AV\hat{y} \| \]

Dual:
\[ p(\mu)^T A(\mu) = -l(\mu)^T \]

Dual RB:
\[ \hat{p}(\mu)^T \hat{A}(\mu) = -\hat{l}(\mu) \]

Dual Residual:
\[ \| -l(\mu) - A(\mu)^T V\hat{p}(\mu) \| \]

Dual Bound:
\[ \| p - V\hat{p} \| = \| A^{-T}(-l - A^T V\hat{p}) \| \leq \| A^{-T} \| \| -l - A^T V\hat{p} \| \]
ROM Error Estimate - Quantity of Interest

Galerkin orthogonality

\[ a(y - \hat{y}, \hat{p}) = f(\hat{p}) - f(\hat{p}) = 0 \quad \Rightarrow \quad (V\hat{p})^T A(y - V\hat{y}) = 0 \]

\[ a(\hat{y}, p - \hat{p}) = -\ell(\hat{y}) + \ell(\hat{y}) = 0 \quad \Rightarrow \quad (p - V\hat{p})^T AV^T \hat{y} = 0 \]

Error bound

\[ |s(\mu) - \hat{s}(\mu)| = |l^T y - \hat{l}^T \hat{y}| \]

\[ = |p^T Ay - (V\hat{p})^T AV\hat{y}| \]

\[ = |p^T A(y - V\hat{y}) + (p - V\hat{p})^T AV^T \hat{y}| \]

\[ = |p^T A(y - V\hat{y}) - (V\hat{p})^T A(y - V\hat{y})| \]

\[ = |(p - V\hat{p})^T A(y - V\hat{y})| \]

\[ \leq \|A^{-T}\|\|A^T(p - V\hat{p})\|\|A(y - V\hat{y})\| \]

\[ |s(\mu) - \hat{s}(\mu)| \leq \Delta(\mu) \equiv \|A^{-1}\|\|f - AV\hat{y}\|\| - l - A^T V\hat{p}\| \]
Petrov-Galerkin Reduced Order Model (ROM)

- Could also generate two subspaces

\[ \mathcal{V}_r = \text{span}\{\zeta_1, \ldots, \zeta_r\} \subset \mathcal{V}_n, \]
\[ \mathcal{W}_r = \text{span}\{\xi_1, \ldots, \xi_r\} \subset \mathcal{V}_n, \quad r \ll n. \]

For example

\[ \mathcal{V}_r = \text{span}\{y(\mu_1), \ldots, y(\mu_r)\} \subset \mathcal{V}_n, \]
\[ \mathcal{W}_r = \text{span}\{p(\mu_1), \ldots, p(\mu_r)\} \subset \mathcal{V}_n. \]

Find \( \hat{y} = \hat{y}(\mu) \in \mathcal{V}_n \) such that

\[ a(\hat{y}, w; \mu) = f(v_n; \mu) \quad \forall w \in \mathcal{W}_r. \]

- Linear system for \( y = \sum_{j=1}^{n} \hat{y}_j \phi_j \):

\[ \mathbf{W}^T \mathbf{A}(\mu) \mathbf{V} \hat{y} = \mathbf{W}^T f(\mu). \quad (r \times r) \]

- But well posedness no longer inherited.
Not automatically guaranteed that \( \mathbf{W}^T \mathbf{A}(\mu) \mathbf{V} \) is invertible or inverse is uniformly bounded in \( r \).
Reduced Basis Method - Greedy Selection

Recall \(|s(\mu) - \hat{s}(\mu)| \leq \Delta(\mu) \equiv \|A^{-1}\| \|f - AV\hat{y}\| - 1 - A^T V\hat{p}\|.

Choose \(\Gamma_{\text{train}} \subset \Gamma\), tolerance \(\epsilon > 0\) and maximum ROM size \(r_{\text{max}}\).

Given \(r > 0\), \(V \in \mathbb{R}^{n \times r}\).

While \(r < r_{\text{max}}\)

1. **Next sample point** via greedy and error estimate

   \[
   \mu_{r+1} = \arg\max_{\mu \in \Gamma_{\text{train}}} \Delta(\mu)
   \]

2. If \(\Delta(\mu_{r+1}) < \epsilon\) stop. **Computed ROM of desired accuracy.**

3. **Compute new basis vector** \(y(\mu_{r+1})\) by solving

   \[
   A(\mu_{r+1})y = f(\mu_{r+1})
   \]

4. **Update old basis:** compute \(V \in \mathbb{R}^{n \times (r+1)}\) where \(V^T V = I\) and

   \[
   \text{Ran}(V) = \text{span}\{y(\mu_1), \ldots, y(\mu_{r+1})\}
   \]

5. **Update Reduced Order System:**

   \[
   \hat{A}(\mu) = V^T A(\mu) V \in \mathbb{R}^{(r+1) \times (r+1)}, \quad \hat{f}(\mu) = V^T f(\mu) \in \mathbb{R}^{r+1}
   \]

6. Set \(r \leftarrow r + 1\).

   Constructed \(y(\mu_1), \ldots, y(\mu_{r+1})\) are linearly independent.

Convergence of the greedy selection [Binev et al., 2011].
Example

\[-\nabla^2 y(\mu) + \begin{bmatrix} \mu \\ 0 \end{bmatrix} \cdot \nabla y(\mu) = 100e^{-5\sqrt{\|x\|^2}}, \quad \text{in } \Omega = (-1, 1)^2, \quad \mu \in [-10, 10],\]

\[y(\mu) = 0, \quad \text{on } \partial \Omega\]

\[s(\mu) = \int_{\Omega} y(\mu)\]

(a) \(\mu = -10\)  
(b) \(\mu = 0\)  
(c) \(\mu = 10\)
Figure: Convection-Diffusion Equation: Output, $s(\mu)$ vs. Parameter, $\mu$
Figure: Convection-Diffusion Equation: Output, $s(\mu)$ vs. Parameter, $\mu$
Figure: Convection-Diffusion Equation: Output, $s(\mu)$ vs. Parameter, $\mu$
Figure: Convection-Diffusion Equation: Output, $s(\mu)$ vs. Parameter, $\mu$
Figure: Convection-Diffusion Equation: Output, $s(\mu)$ vs. Parameter, $\mu$
Figure: Convection-Diffusion Equation: Output, $s(\mu)$ vs. Parameter, $\mu$
Figure: Convection-Diffusion Equation: Output, $s(\mu)$ vs. Parameter, $\mu$
Proper Orthogonal Decomposition

- Given snapshots \( y(\mu_1), \ldots, y(\mu_m) \in \mathcal{V}_n, m > r \).
- Compute orthonormal basis \( v_1, \ldots, v_r \in \mathcal{V}_n \) as solution of

\[
\min \sum_{k=1}^{m} \left\| y_k - \sum_{i=1}^{r} \langle y_k, v_i \rangle \nu \, v_i \right\|_\mathcal{V}^2
\]

s.t. \( \langle v_i, v_j \rangle_\nu = \delta_{ij} \).

- Solution
  - Compute eigenvectors \( v_1, v_2, \ldots \in \mathcal{V}_n \) and eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0 \) of the linear operator

\[
\psi \mapsto \mathcal{K}\psi = \sum_{k=1}^{m} y_k \langle y_k, \psi \rangle_\nu.
\]

- Solution \( v_1, v_2, \ldots, v_r \),

\[
\sum_{k=1}^{m} \left\| y_k - \sum_{i=1}^{r} \langle y_k, v_i \rangle \nu \, v_i \right\|_\mathcal{V}^2 = \sum_{i=r+1}^{m} \lambda_i.
\]
Finite dimensional representation of snapshots
\( y(\mu_1), \ldots, y(\mu_m) \in \mathbb{R}^n, \, m > r. \)

Inner product \( \langle v, w \rangle_V = v^T M w, \, M \) s.p.d. (not nec. mass matrix)

Solution

Define \( Y = [y(\mu_1), \ldots, y(\mu_m)] \in \mathbb{R}^{n \times m}. \)

Compute \( M \)-orthonormal eigenvectors \( v_1, v_2, \ldots \in \mathbb{R}^n \) and eigenvalues \( \lambda_1 \geq \ldots \geq \lambda_m \geq 0 \) of generalized \( n \times n \) eigenvalue prob.

\[
M Y Y^T M v_i = \lambda_i M v_i.
\]

Alternatively, if \( n > m \) compute eigenvectors \( w_1, w_2, \ldots \in \mathbb{R}^m \) and eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\min\{m,n\}} \geq 0 \) of

\[
Y^T M Y w_i = \lambda_i w_i.
\]

\( v_i = \lambda_i^{-1/2} Y w_i, \, i = 1, \ldots, m. \)

Usually, fix tolerance \( \epsilon > 0 \). Compute eigenvectors \( v_1, v_2, \ldots \in \mathbb{R}^n \) and eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\min\{m,n\}} \geq 0. \)

Find smallest \( r \) such that \( \sum_{i=r+1}^m \lambda_i < \epsilon. \)

If only some of the largest eigenvals. and vecs. are computed:

Find smallest \( r \) such that \( \lambda_{r+1}/\lambda_1 < \epsilon. \)

Reduced order model \( V = [v_1, \ldots, v_r] \in \mathbb{R}^{n \times r}. \)
POD often used to ‘compress’ solution of (linear or nonlinear) dynamical system

\[ M \mathbf{y}'(t) = A \mathbf{y}(t) + \mathbf{f}(t), \quad t \in (0, T), \]
\[ \mathbf{y}(0) = \mathbf{y}_0. \]

Solutions \( \mathbf{y}(t_0), \ldots, \mathbf{y}(t_m) \in \mathbb{R}^n \) at time steps \( 0 = t_0 < \ldots < t_m = T \) used as snapshots.

Can combine Reduced Basis Method and POD for parameterized dynamical systems

\[ M(\mu) \mathbf{y}'(t; \mu) = A(\mu) \mathbf{y}(t; \mu) + \mathbf{f}(t; \mu), \quad t \in (0, T), \]
\[ \mathbf{y}(0; \mu) = \mathbf{y}_0, \]
\[ s(\mu) = \int_0^T \mathbf{c}(t; \mu)^T \mathbf{y}(0; \mu) dt. \quad \text{(output of interest)} \]

Corresponding dual

\[ -M(\mu)^T \mathbf{p}'(t; \mu) = A(\mu)^T \mathbf{p}(t; \mu) - \mathbf{c}(t; \mu), \quad t \in (0, T), \]
\[ \mathbf{p}(T; \mu) = 0, \]
Balanced Truncation Model Reduction (BTMR)

Consider

\[
\frac{d}{dt} y(t) = Ay(t) + Bu(t), \quad t \in (0, T)
\]

\[
z(t) = Cy(t) + Du(t), \quad t \in (0, T)
\]

\[y(0) = 0.\]

Projection methods for model reduction produce \( n \times r \) matrices \( V, W \) with \( r \ll n \) and with \( W^T V = I_r \).

One obtains a reduced form by setting \( y = V\hat{y} \) and projecting so that

\[
W^T [V \frac{d}{dt}\hat{y}(t) - AV\hat{y}(t) - Bu(t)] = 0, \quad t \in (0, T).
\]

This leads to a reduced order system of order \( n \) given by

\[
\frac{d}{dt} \hat{y}(t) = \hat{A}\hat{y}(t) + \hat{B}u(t), \quad t \in (0, T)
\]

\[
\hat{z}(t) = \hat{C}\hat{y}(t) + Du(t), \quad t \in (0, T)
\]

\[\hat{y}(0) = 0.\]

with \( \hat{A} = W^T AV, \hat{B} = W^T B, \) and \( \hat{C} = CV. \)
Controllability and Observability Gramians

- Recall

\[ y'(t) = Ay(t) + Bu(t), \quad t \in (0, T) \]

\[ z(t) = Cy(t) + Du(t), \quad t \in (0, T). \]

Assume the system is stable \((\text{Re}(\lambda(A)) < 0)\), controllable and observable.
Controllability and Observability Gramians

Recall

\[ y'(t) = Ay(t) + Bu(t), \quad t \in (0, T) \]
\[ z(t) = Cy(t) + Du(t), \quad t \in (0, T). \]

Assume the system is stable (Re($\lambda(A)$) < 0), controllable and observable.

Controllability Gramian.

\[ P = \int_{0}^{\infty} e^{At} B B^T e^{A^T t} dt. \]

Eigenspaces corresponding to large eigenvalues are ‘easy’ to control (control has smaller energy).

Controllability Gramian solves the Lyapunov equation

\[ A P + P A^T + B B^T = 0. \]
Controllability and Observability Gramians

- Recall

\[ y'(t) = Ay(t) + Bu(t), \quad t \in (0, T) \]
\[ z(t) = C y(t) + D u(t), \quad t \in (0, T). \]

Assume the system is stable (Re(\lambda(A)) < 0), controllable and observable.

- Controllability Gramian.

\[ \mathcal{P} = \int_0^\infty e^{At} B B^T e^{A^T t} dt. \]

- Eigenspaces corresponding to large eigenvalues are ‘easy’ to control (control has smaller energy).
- Controllability Gramian solves the Lyapunov equation

\[ A \mathcal{P} + \mathcal{P} A^T + BB^T = 0. \]

- Observability Gramian.

\[ \mathcal{Q} = \int_0^\infty e^{A^T t} C^T C e^{A t} dt. \]

- Eigenspaces corresponding to large eigenvalues are ‘easy’ to observe.
- Observability Gramian solves the Lyapunov equation

\[ A^T \mathcal{Q} + \mathcal{Q} A + C^T C = 0. \]
Compute controllability and observability gramians \( P, Q \) \( P = UU^T \) and \( Q = LL^T \) in factored form, i.e., solve

\[
AP + PA^T + BB^T = 0, \\
A^TQ + QA + CT C = 0.
\]

Compute the SVD \( U^T L = ZSY^T \), where \( S_r = diag(\sigma_1, \sigma_2, \ldots, \sigma_r) \) with \( S = S_n \), and \( \sigma_1 \geq \sigma_2 \geq \ldots \).

Set \( V = UZ_r S_n^{-1/2} \), \( W = LY_r S_n^{-1/2} \), where \( n \) is selected to be the smallest positive integer such that \( \sigma_{r+1} < \tau \sigma_1 \). Here \( \tau > 0 \) is a prespecified constant. The matrices \( Z_r, Y_r \) consist of the corresponding leading \( r \) columns of \( Z, Y \).

It is easily verified that \( PW = VS_r \) and \( QV = WS_r \).

Hence

\[
0 = W^T (AP + PA^T + BB^T) W = \hat{A}S_r + S_r \hat{A}^T + \hat{B}\hat{B}^T,
\]

\[
0 = V^T (A^TQ + QA + CT C) V = \hat{A}^T S_r + S_r \hat{A} + \hat{C}^T \hat{C}.
\]
Two important properties of balanced truncation model reduction:

- $\hat{A}$ is stable
- For any given input $u$ we have

$$\|z - \hat{z}\|_{L_2} \leq 2\|u\|_{L_2}(\sigma_{n+1} + \ldots + \sigma_N)$$

where $\hat{z}$ is the output (response) of the reduced model [Glover, 1984].
Empirical Interpolation Method

- \( V^T A(\mu) V \in \mathbb{R}^{r \times r} \), but evaluation \( \mu \mapsto V^T A(\mu) V \) requires evaluation \( \mu \mapsto A(\mu) \mapsto V^T A(\mu) V \) at cost dependent on \( n \).

- If \( A(\mu) = A_0 + \sum_{j=1}^{k} \mu_j A_j \), then

\[
V^T A(\mu) V = V^T A_0 V + \sum_{j=1}^{k} \mu_j V^T A_j V.
\]

Precompute \( V^T A_k V \in \mathbb{R}^{r \times r} \), \( j = 0, \ldots, k \), afterwards evaluate \( \mu \mapsto V^T A(\mu) V \) at cost of \( O(r^2) \).

- For example, finite element discretization of

\[
-\nabla^2 y(\mu) + \begin{bmatrix} \mu \\ 0 \end{bmatrix} \cdot \nabla y(\mu) = 100e^{-5\sqrt{\|x\|^2}}, \quad \text{in } \Omega = (-1, 1)^2, \\
y(\mu) = 0, \quad \text{on } \partial \Omega
\]

leads to \( A(\mu) = A_{\text{diff}} + \mu A_{\text{adv}} \).
Empirical Interpolation Method

- Empirical Interpolation Method (EIM): [Barrault et al., 2004] [Eftang et al., 2010].
- Application of DEIM for finite element approximations [Antil et al., 2014], [Tiso and Rixen, 2013].
- Element based (compared to nodal/point based) version: [Farhat et al., 2015].
Outline

Overview

Example Optimization Problems

Optimization Problem

Projection Based Model Reduction

Back to Optimization

Error Estimates

Linear-Quadratic Problems

Shape Optimization with Local Parameter Dependence

Semilinear Parabolic Problems

Trust-Region Framework
Recall Optimization Problem

Original problem

\[
\begin{align*}
\text{min} & \quad J(y, u) \\
\text{s.t.} & \quad c(y, u) = 0, \\
& \quad u \in U_{ad}.
\end{align*}
\]

\[y(u) \text{ unique sol. of } c(y, u) = 0\]

\[
\begin{align*}
\text{min} & \quad j(u) \\
\text{s.t.} & \quad u \in U_{ad}
\end{align*}
\]

where \( j(u) \overset{\text{def}}{=} J(y(u), u) \).

Reduced order problem

\[
\begin{align*}
\text{min} & \quad \hat{J}(\hat{y}, u) \\
\text{s.t.} & \quad \hat{c}(\hat{y}, u) = 0, \\
& \quad u \in U_{ad}.
\end{align*}
\]

\[\hat{y}(u) \text{ unique sol. of } \hat{c}(\hat{y}, u) = 0\]

\[
\begin{align*}
\text{min} & \quad \hat{j}(u) \\
\text{s.t.} & \quad u \in U_{ad}
\end{align*}
\]

where \( \hat{j}(u) \overset{\text{def}}{=} \hat{J}(\hat{y}(u), u) \).
Gradient Computation (Using Adjoints)

1. Given $u$, solve state PDE $c(y, u) = 0$ for $y = y(u)$.
2. Solve the adjoint PDE $c_y(y(u), u)^* p = -D_y J(y(u), u)$ for $p = p(u)$.
3. Compute $Dj(u) = Du J(y(u), u) + c_u(y(u), u)^* p(u)$.

$$\langle \nabla j(u), v \rangle_U = \langle Dj(u), v \rangle_{U^*}, U$$

$$= \langle c_u(y, u)^* p + Du J(y, u), v \rangle_{U^*}, U \quad \forall v \in U$$
Outline

Overview

Example Optimization Problems

Optimization Problem

Projection Based Model Reduction

Back to Optimization

Error Estimates

Linear-Quadratic Problems

Shape Optimization with Local Parameter Dependence

Semilinear Parabolic Problems

Trust-Region Framework
▶ Optimization problem: More parameters - controls $\mathbf{u} \in \mathcal{U}_{ad}$; objective function $j(\mathbf{u})$ is quantify of interest.

▶ Approximating just the objective function $j(\mathbf{u})$ is not enough. Also need to approximate gradient information $\nabla j(\mathbf{u})$. 
Error Estimate (Strongly Convex Function) I

- $u_*$ minimizer of original objective $j$, $\hat{u}_*$ min. of reduced objective $\hat{j}$.

Want to estimate error $\|\hat{u}_* - u_*\|_U$.

- Optimality conditions

\[
\langle \nabla j(u_*), u - u_* \rangle_U \geq 0 \quad \forall u \in \mathcal{U}_{ad},
\]
\[
\langle \nabla \hat{j}(\hat{u}_*), u - \hat{u}_* \rangle_U \geq 0 \quad \forall u \in \mathcal{U}_{ad}.
\]

- Assume $j$ is strongly convex function on convex set $C \subset \mathcal{U}$: There exists $\kappa > 0$ such that

\[
\langle u - w, \nabla j(u) - \nabla j(w) \rangle_U \geq \kappa \|u - w\|^2_U \text{ for all } u, w \in C \subset \mathcal{U}.
\]

- Let $\xi \in \mathcal{U}$ be such that

\[
\langle \nabla j(\hat{u}_*) + \xi, u - \hat{u}_* \rangle_U \geq 0 \quad \forall u \in \mathcal{U}_{ad}.
\]

$\xi = \nabla \hat{j}(\hat{u}_*) - \nabla j(\hat{u}_*)$ always works; sometimes can find better $\xi$.

$(\hat{u}_* \text{ solves perturbed optimization problem } \min_{u \in \mathcal{U}_{ad}} j(u) + \langle \xi, u \rangle_U)$
Error Estimate (Strongly Convex Function) II

- Combine optimality conds. & convexity: If $u_*, \hat{u}_* \in C$,

\[
\kappa \|u_* - \hat{u}_*\|^2 \leq \langle u_* - \hat{u}_*, \nabla j(u_*) - \nabla j(\hat{u}_*) \rangle \\
\leq \langle u_* - \hat{u}_*, \nabla j(u_*) - \nabla j(\hat{u}_*) \rangle + \langle u_* - \hat{u}_*, \nabla j(\hat{u}_*) + \xi \rangle \\
= \langle u_* - \hat{u}_*, \nabla j(u_*) \rangle + \langle u_* - \hat{u}_*, \xi \rangle \leq \langle u_* - \hat{u}_*, \xi \rangle.
\]

- Hence \(\|u_* - \hat{u}_*\|_U \leq \kappa^{-1} \|\xi\|_U\) \((\leq \kappa^{-1} \|\nabla j(\hat{u}_*) - \nabla j(\hat{u}_*)\|_U)\)

- Estimate error in gradients to get estimate for error in solution.

- Applies when
  - \(j\) is strongly convex function on \(U_{ad}\) admissible set, e.g., convex-linear quadratic problems.
  - \(j\) satisfies strong second order optimality conditions at \(u_*\) and \(\hat{u}_*\) is in neighborhood of \(u_*\).
Error Estimate for Unconstrained Problems

- $\hat{u}_* = \arg\min_u \hat{j}(u)$ minimizer of *unconstrained* reduced problem.
- **Newton-Kantorovich Theorem:** Let $r > 0$ and $\nabla^2 j \in \text{Lip}_L(B_r(\hat{u}_*))$. $\nabla^2 j(\hat{u}_*)$ be nonsingular and constants $\zeta, \eta \geq 0$ such that

$$\|\nabla^2 j(\hat{u}_*)^{-1}\| = \zeta, \quad \|\nabla^2 j(\hat{u}_*)^{-1} \nabla j(\hat{u}_*)\| \leq \eta.$$ 

If $L \zeta \eta \leq \frac{1}{2}$, there is unique local minimum $u_*$ of $j$ in ball around $\hat{u}_*$ with radius $\min \left\{ r, \left(1 - \sqrt{1 - 2L\zeta \eta}\right)/(L\zeta) \right\} \leq \min \left\{ r, 2\eta \right\}$.

- **Estimate $\eta$:**

$$\|\nabla^2 j(\hat{u}_*)^{-1}(\nabla j(\hat{u}_*) - \nabla \hat{j}(\hat{u}_*))\| \leq \zeta \|\nabla j(\hat{u}_*) - \nabla \hat{j}(\hat{u}_*)\| = \eta.$$ 

- Hence $\|u_* - \hat{u}_*\|_U \leq 2\zeta \|\nabla \hat{j}(\hat{u}_*) - \nabla j(\hat{u}_*)\|_U$.
- **Estimate error in gradients to get estimate for error in solution.**
- Need $L \zeta \eta \leq \frac{1}{2}$, i.e., $\nabla j(\hat{u}_*)$ small enough.
- Can estimate error using convergence properties of Newton’s method started with $\hat{u}_*$ applied to original problem.
Outline

Overview
Example Optimization Problems
Optimization Problem
Projection Based Model Reduction
Back to Optimization
Error Estimates

Linear-Quadratic Problems
Shape Optimization with Local Parameter Dependence
Semilinear Parabolic Problems
Trust-Region Framework
Elliptic Linear-Quadratic Model Problem

▶ Original problem

\[
\min \frac{1}{2} y^T Q y + c^T y + \frac{\alpha}{2} u^T R u \\
\text{s.t. } Ay + Bu = b, \\
u \in U_{ad}
\]

where \( A \in \mathbb{R}^{n \times n} \) invertible, \( B \in \mathbb{R}^{n \times m} \), \( b, c \in \mathbb{R}^n \), \( Q = Q^T \in \mathbb{R}^{n \times n} \) positive semidef., \( R = R^T \in \mathbb{R}^{m \times m} \) positive def. \( n \) large.

▶ Reduced order problem

\[
\min \frac{1}{2} \hat{y}^T V^T Q V \hat{y} + c^T V \hat{y} + \frac{\alpha}{2} u^T R u \\
\text{s.t. } W^T A V \hat{y} + W^T B u = W^T b, \\
u \in U_{ad}
\]

where \( V, W \in \mathbb{R}^{n \times r} \), \( r \ll n \), \( \hat{A} \overset{\text{def}}{=} W^T A V \) invertible.
Gradient original problem

\[ Ay + Bu = b, \]
\[ A^T p = -Qy - d, \]
\[ \nabla j(u) = \alpha Ru + B^T p. \]

Gradient reduced order problem

\[ W^T AV \hat{y} + W^T Bu = W^T b, \]
\[ V^T A^T W \hat{p} = -V^T QV \hat{y} - V^T d, \]
\[ \nabla \hat{j}(u) = \alpha Ru + B^T W \hat{p}. \]

Error

\[ \nabla \hat{j}(u) - \nabla j(u) = B^T (W \hat{p} - p) = B^T (W \hat{p} - \tilde{p} + \tilde{p} - p), \]
\[ \tilde{p} \text{ solves } A^T \tilde{p} = -QV \hat{y} - d \text{ (full adjoint with reduced } V \hat{y} \text{ input).} \]

Error bound

\[ \| \nabla \hat{j}(u) - \nabla j(u) \| \leq \| B \| \left( \| A^{-T} \| \| A^T W \hat{p} + QV \hat{y} + d \| \right. \]
\[ \left. + \| A^{-T} \| \| A^{-1} \| \| Q \| \| AV \hat{y} + Bu - b \| \right). \]
Basis $V$ represents state information, $W$ represents adjoint information. In principle can construct $V \neq W$, but must have $W^T AV$ invertible.

Often compute $V = W$ from samples/snapshots of states and adjoints.

Careful, states and adjoints represent different objects and have different scales (more later).

Application to parameterized optimal control problems $\mu \in \mathcal{D}$

$$\min \frac{1}{2} y^T Q(\mu) y + c(\mu)^T y + \frac{\alpha}{2} u^T R(\mu) u$$

s.t. $A(\mu) y + B(\mu) u = b(\mu)$.

Assume computable uniform bounds, e.g., $\|A(\mu)^{-1}\| \leq a$, $\mu \in \mathcal{D}$.

Have error estimates, can now use ROM machinery developed for equations.

- Use greedy procedure to sample $\mu \in \Gamma$. Use error estimate.
- Add state $y(\mu)$ and adjoint $p(\mu)$ at sample $\mu$ to basis $V = W$. 
Parabolic Linear-Quadratic Model Problem

- Consider optimal control probl. governed by advection diffusion PDE

\[
\frac{\partial}{\partial t} y(x, t) - \nabla (k(x) \nabla y(x, t)) + \mathbf{a}(x) \cdot \nabla y(x, t)) = f(x, t)
\]

in \(\Omega \times (0, T)\). Optimization variables are related to the right hand side \(f\) or to boundary data.

- After (finite element) discretization in space the optimal control problems are of the form

\[
\min j(u) \equiv \frac{1}{2} \int_0^T \|Cy(t) + Du(t) - d(t)\|^2 dt,
\]

where \(y(t) = y(u; t)\) is the solution of

\[
M y'(t) = A y(t) + B u(t), \quad t \in (0, T),
\]

\[
y(0) = y_0.
\]

Here \(y(t) \in \mathbb{R}^n, M \in \mathbb{R}^{n \times n}\) invert., \(A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}, n\) large.
\(D \in \mathbb{R}^{m \times m}\) invertible. Strongly convex problem.
Reduced optimal control problem

\[
\min \tilde{j}(u) \equiv \frac{1}{2} \int_{0}^{T} \| \hat{C} \hat{V} \hat{y}(t) + Du(t) - d(t) \|^2 dt
\]

where \( \hat{y}(t) = \hat{y}(u; t) \) solves

\[
\begin{align*}
W^T M V \hat{y}'(t) &= \hat{M} \hat{y}'(t) + \hat{A} \hat{y}(t) + \hat{B} u(t), & t \in (0, T), \\
\hat{y}(0) &= \hat{y}_0.
\end{align*}
\]

Here \( \hat{y}(t) \in \mathbb{R}^r, \hat{M}, \hat{A} \in \mathbb{R}^{r \times r}, \hat{B} \in \mathbb{R}^{r \times m}, \) with \( r \ll n \) small.
Gradient computation original problem

\[
\begin{aligned}
M y'(t) &= Ay(t) + Bu(t), & t \in (0, T), & y(0) = y_0, \\
z(t) &= Cy(t) + Du(t) - d(t), & t \in (0, T),
\end{aligned}
\]

\[-M^T p'(t) = A^T p(t) + C^T z(t), & t \in (0, T), & p(T) = 0,
\]

\[
\nabla j(u) = q(t) = B^T p(t) + D^T z(t), & t \in (0, T)
\]

Gradient computation reduced problem

\[
\begin{aligned}
\hat{M} \hat{y}'(t) &= \hat{A} \hat{y}(t) + \hat{B} u(t), & t \in (0, T) & \hat{y}(0) = \hat{y}_0, \\
\hat{z}(t) &= \hat{C} \hat{y}(t) + Du(t) - d(t), & t \in (0, T),
\end{aligned}
\]

\[-\hat{M}^T \hat{p}'(t) = \hat{A}^T \hat{p}(t) + \hat{C}^T \hat{z}(t), & t \in (0, T) & \hat{p}(T) = 0,
\]

\[
\nabla \hat{j}(u) = \hat{q}(t) = \hat{B}^T \hat{p}(t) + D^T \hat{z}(t), & t \in (0, T)
\]

‘Duality’ in input-output maps \( u \mapsto z, w \mapsto q \) of state, adjoint sys.

Need to approximate input-to-output maps \( u \mapsto z, w \mapsto q \).
**Balanced Truncation Model Reduction (BTMR) error bound:**
If system is stable ($\text{Re}(\lambda(A)) < 0$), controllable and observable (true for model problem), can use BTMR to compute $W, V \in \mathbb{R}^{N \times n}$:

$$
\|z - \hat{z}\|_{L^2} \leq 2(\sigma_{r+1} + \ldots + \sigma_n) \|u\|_{L^2} \quad \forall u,
$$

$$
\|q - \hat{q}\|_{L^2} \leq 2(\sigma_{r+1} + \ldots + \sigma_n) \|w\|_{L^2} \quad \forall w,
$$

where $\sigma_1 \geq \ldots \geq \sigma_r \geq \sigma_{r+1} \geq \ldots \sigma_n \geq 0$ are Hankel singular vals.

**Introduce auxiliary adjoint for error estimate**

- **Original problem**

  $$
  -Mp'(t) = A^T p(t) + C^T z(t), \quad t \in (0, T), \quad p(T) = 0,
  $$

  $$
  \nabla j(u) = q(t) = B^T p(t) + D^T z(t), \quad t \in (0, T)
  $$

- **Reduced order problem**

  $$
  -\hat{p}'(t) = \hat{A}^T \hat{p}(t) + \hat{C}^T \hat{z}(t), \quad t \in (0, T) \quad \hat{p}(T) = 0,
  $$

  $$
  \nabla \hat{j}(u) = \hat{q}(t) = \hat{B}^T \hat{p}(t) + \hat{D}^T \hat{z}(t), \quad t \in (0, T)
  $$

- **BTMR bound requires same input in full and reduced adjoint system.**

- **Easy to fix:** Introduce auxiliary adjoint $\tilde{p}$ as solution of the original adjoint, but with input $\hat{z}$ instead of $z$. 

---

Matthias Heinkenschloss  
June 14, 2016

78
Assume that there exists $\gamma > 0$ such that

$$v^T Av \leq -\gamma v^T M v, \quad \forall v \in \mathbb{R}^n.$$  

(Satisfied for model problem).

Gradient error:

$$\|\nabla j(u) - \nabla \hat{j}(u)\|_{L^2} \leq 2 \left( c\|u\|_{L^2} + \|\hat{z}(u)\|_{L^2} \right) (\sigma_{r+1} + \ldots + \sigma_n)$$

for all $u \in L^2$! (\(\hat{z}(u)\) output of reduced order state with input $u$.)

Solution error:

$$\|u_* - \hat{u}_*\|_{L^2} \leq \frac{2}{K} \left( c\|\hat{u}_*\|_{L^2} + \|\hat{z}_*\|_{L^2} \right) (\sigma_{r+1} + \ldots + \sigma_n).$$
Example Problem (modeled after [Dedé and Quarteroni, 2005])

Minimize \( \frac{1}{2} \int_0^T \int_D (y(x, t) - d(x, t))^2 \, dx \, dt + \frac{10^{-4}}{2} \int_0^T \int_{U_1 \cup U_2} u^2(x, t) \, dx \, dt \),

subject to

\[
\frac{\partial}{\partial t} y(x, t) - \nabla (\kappa \nabla y(x, t)) + a(x) \cdot \nabla y(x, t) = u(x, t) \chi_{U_1}(x) + u(x, t) \chi_{U_2}(x) \quad \text{in } \Omega \times (0, T),
\]

with boundary conditions \( y(x, t) = 0 \) on \( \Gamma_D \times (0, T) \), \( \frac{\partial}{\partial n} y(x, t) = 0 \) on \( \Gamma_N \times (0, T) \) and initial conditions \( y(x, 0) = 0 \) in \( \Omega \).

\( \Omega \) with boundary conditions for the advection diffusion equation

the velocity field \( a \)
<table>
<thead>
<tr>
<th>grid</th>
<th>$k$</th>
<th>$m$</th>
<th>$n$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>168</td>
<td>9</td>
<td>1545</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>283</td>
<td>16</td>
<td>2673</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>618</td>
<td>29</td>
<td>6036</td>
<td>9</td>
</tr>
</tbody>
</table>

Number $k$ of observations, number $m$ of controls, size $n$ of full order system, and size $r$ of reduced order system for three discretizations.

Largest Hankel singular values and threshold $10^{-4}\sigma_1$ (grid # 3)
Integrals $\int_{U_1} u^*_2(x, t) dx$ (solid blue line) and $\int_{U_1} \hat{u}^*_2(x, t) dx$ (dashed red line) of the optimal controls computed using the full and and the reduced order model.

Full and reduced order model sols. in excellent agreement: $\|u_* - \hat{u}_*\|_{L^2}^2 = 6 \cdot 10^{-3}$. 
Convergence histories of the Conjugate Gradient algorithm applied to full (+) and reduced (o) order optimal control problems.

Recall error bound for the gradients:

$$\| \nabla j(u) - \nabla \hat{j}(u) \|_{L^2} \leq 2 \left( c \| u \|_{L^2} + \| \hat{z}(u) \|_{L^2} \right) (\sigma_{r+1} + \ldots + \sigma_n) \quad \forall u \in L^2!$$
Consider minimization problem

\[
\min_{\theta \in \Theta_{ad}} j(\theta) := \int_0^T \int_{\Omega(\theta)} \ell(y(x,t;\theta), t, \theta) dx \; dt
\]

where \( y(x,t;\theta) \) solves

\[
\frac{\partial}{\partial t} y(x,t) - \nabla (\kappa(x) \nabla y(x,t)) + V(x) \cdot \nabla y(x,t)) = f(x,t) \quad (x,t) \in \Omega(\theta) \times (0,T),
\]

\[
\kappa(x) \nabla y(x,t) \cdot n = g(x,t) \quad (x,t) \in \Gamma_N(\theta) \times (0,T),
\]

\[
y(x,t) = d(x,t) \quad (x,t) \in \Gamma_D(\theta) \times (0,T),
\]

\[
y(x,0) = y_0(x) \quad x \in \Omega_D(\theta)
\]

Semidiscretization in space leads to

\[
\min_{\theta \in \Theta_{ad}} j(\theta) := \int_0^T \ell(y(t;\theta), t, \theta) \; dt
\]

where \( y(t;\theta) \) solves

\[
M(\theta) \frac{d}{dt} y(t) + A(\theta) y(t) = B(\theta) u(t), \quad t \in [0,T],
\]

\[
M(\theta)y(0) = M(\theta)y_0.
\]
We would like to replace the large scale problem

\[
\min_{\theta \in \Theta_{ad}} j(\theta) := \int_0^T \ell(y(t; \theta), t, \theta) \, dt
\]

where \( y(t; \theta) \) solves

\[
M(\theta) \frac{d}{dt} y(t) + A(\theta)y(t) = B(\theta)u(t), \quad t \in [0, T],
\]

\[
M(\theta)y(0) = M(\theta)y_0
\]

by a reduced order problem

\[
\min_{\theta \in \Theta_{ad}} \hat{J}(\theta) := \int_0^T \ell(\hat{y}(t; \theta), t, \theta) \, dt
\]

where \( \hat{y}(t; \theta) \) solves

\[
\hat{M}(\theta) \frac{d}{dt} \hat{y}(t) + \hat{A}(\theta)y(t) = \hat{B}(\theta)u(t), \quad t \in [0, T],
\]

\[
\hat{M}(\theta)\hat{y}(0) = \hat{M}(\theta)\hat{y}_0.
\]

Problem is that we need a reduced order model that approximates the full order model for all \( \theta \in \Theta_{ad} \)! Cannot be done using BTMR. I am not aware of any MR method that can do this with guaranteed error bounds.
Localized parameters (nonlinearity)

- Consider classes of problems where the shape parameter $\theta$ only influences a (small) subdomain:

$$\bar{\Omega}(\theta) := \bar{\Omega}_1 \cup \bar{\Omega}_2(\theta), \quad \Omega_1 \cap \Omega_2(\theta) = \emptyset \Gamma = \bar{\Omega}_1 \cap \bar{\Omega}_2(\theta).$$

- The FE stiffness matrix times vector can be decomposed into

$$Ay = \begin{pmatrix} A^{II}_1 & A^{I\Gamma} & 0 \\ A^{\Gamma I}_1 & A^{\Gamma\Gamma}(\theta) & A^{I\Gamma}_2(\theta) \\ 0 & A^{\Gamma\Gamma}_2(\theta) & A^{II}_2(\theta) \end{pmatrix} \begin{pmatrix} y^I_1 \\ y^\Gamma \\ y^I_2 \end{pmatrix}$$

where $A^{\Gamma\Gamma}(\theta) = A^{\Gamma\Gamma}_1 + A^{\Gamma\Gamma}_2(\theta)$.

The matrices $M, B$ admit similar representations.

- Consider objective functions of the type

$$\int_0^T \ell(y(t), t, \theta) dt = \frac{1}{2} \int_0^T \| C^I_1 y^I_1 - d^I_1(t) \|_2^2 + \bar{\ell}(y^\Gamma(t), y^I_2(t), t, \theta) dt.$$
Our Optimization problem

$$\min_{\theta \in \Theta_{ad}} j(\theta) := \int_{0}^{T} \ell(y(t; \theta), t, \theta) \, dt$$

where $y(t; \theta)$ solves

$$M(\theta) \frac{d}{dt} y(t) + A(\theta)y(t) = B(\theta)u(t), \quad t \in [0, T],$$

$$M(\theta)y(0) = M(\theta)y_0$$

can now be written as

$$\min_{\theta \in \Theta_{ad}} j(\theta) := \frac{1}{2} \int_{0}^{T} \|C_1^I y_1^I - d_1^I(t)\|_2^2 + \ell(y^\Gamma(t), y_2^I(t), t, \theta) \, dt.$$
First Order Optimality Conditions

▶ Lagrangian

\[ L(y, p, \theta) = \int_0^T \ell(y(t), t, \theta) \, dt + \int_0^T p(t)^T \left( M(\theta) \frac{d}{dt} y(t) + A(\theta) y(t) - B(\theta) u(t) \right) \, dt \]
First Order Optimality Conditions

- **Lagrangian**

\[ L(y, p, \theta) = \int_0^T \ell(y(t), t, \theta) \, dt + \int_0^T p(t)^T \left( M(\theta) \frac{d}{dt} y(t) + A(\theta) y(t) - B(\theta) u(t) \right) \, dt \]

- **The first order necessary optimality conditions are**

\[ M(\theta) \frac{d}{dt} y(t) + A(\theta) y(t) = B(\theta) u(t) \quad t \in [0, T], \]

\[ M(\theta) y(0) = y_0, \]

\[-M(\theta) \frac{d}{dt} p(t) + A^T(\theta) p(t) = -\nabla_y \ell(y(t), t, \theta) \quad t \in [0, T],\]

\[ M(\theta) p(T) = 0. \]

\[ \nabla_\theta L(y, p, \theta)(\tilde{\theta} - \theta) \geq 0, \quad \tilde{\theta} \in \Theta_{ad} \]
First Order Optimality Conditions

- **Lagrangian**

\[
L(y, p, \theta) = \int_0^T \ell(y(t), t, \theta) \, dt + \int_0^T p(t)^T \left( M(\theta) \frac{d}{dt} y(t) + A(\theta) y(t) - B(\theta) u(t) \right) dt
\]

- The first order necessary optimality conditions are

\[
M(\theta) \frac{d}{dt} y(t) + A(\theta) y(t) = B(\theta) u(t) \quad t \in [0, T],
\]

\[
M(\theta) y(0) = y_0,
\]

\[-M(\theta) \frac{d}{dt} p(t) + A^T(\theta) p(t) = -\nabla_y \ell(y(t), t, \theta) \quad t \in [0, T],
\]

\[
M(\theta) p(T) = 0.
\]

\[
\nabla_\theta L(y, p, \theta)(\tilde{\theta} - \theta) \geq 0, \quad \tilde{\theta} \in \Theta_{ad}
\]

- Gradient of \( j \) is given by \( \nabla j(\theta) = \nabla_\theta \ell(y(t), p(t), \theta) \).
Using DD structure, state and adjoint equations can be written as

\[
\begin{align*}
M_{II}^1 \frac{d}{dt} y_I^1(t) + M_{I\Gamma}^1 \frac{d}{dt} y_{\Gamma}^1(t) + A_{II}^1 y_I^1(t) + A_{I\Gamma}^1 y_{\Gamma}^1(t) &= B_I^1 u_1^I(t) \\
M_{II}^2(\theta) \frac{d}{dt} y_I^2(t) + M_{I\Gamma}^2(\theta) \frac{d}{dt} y_{\Gamma}^2(t) + A_{II}^2(\theta)y_I^2(t) + A_{I\Gamma}^2(\theta)y_{\Gamma}^2(t) &= B_I^2(\theta)u_2^I(t) \\
M_{I\Gamma}^1 \frac{d}{dt} y_I^1(t) + M_{I\Gamma}^{\Gamma}(\theta) \frac{d}{dt} y_{\Gamma}^1(t) + M_{II}^2(\theta) \frac{d}{dt} y_I^2(t)
&+ A_{I\Gamma}^1 y_I^1(t) + A_{I\Gamma}^{\Gamma}(\theta) \frac{d}{dt} y_{\Gamma}^1(t) + A_{II}^2(\theta)y_I^2(t) = B_{\Gamma}^1(\theta)u_{\Gamma}^I(t),
\end{align*}
\]

\[
M_{I\Gamma}^1 \frac{d}{dt} p_I^1(t) + M_{I\Gamma}^2(\theta) \frac{d}{dt} p_{\Gamma}^1(t) + A_{I\Gamma}^1 p_I^1(t)
&+ A_{I\Gamma}^{\Gamma}(\theta) \frac{d}{dt} p_{\Gamma}^1(t) + A_{II}^2(\theta)p_I^2(t)
= -\nabla y_{\Gamma}^2 \tilde{\ell}(.),
\]

To apply model reduction to the system corresponding to fixed subdomain \(\Omega\), we have to identify how \(y_I^1\) and \(p_I^1\) interact with other components.
Using DD structure, state and adjoint equations can be written as

\[
\begin{align*}
M^{II}_1 \frac{d}{dt} y^I_1(t) + M^{I\Gamma}_1 \frac{d}{dt} y^\Gamma(t) + A^{II}_1 y^I_1(t) + A^{I\Gamma}_1 y^\Gamma(t) &= B^I_1 u^I_1(t) \\
M^{II}_2(\theta) \frac{d}{dt} y^I_2(t) + M^{I\Gamma}_2(\theta) \frac{d}{dt} y^\Gamma(t) + A^{II}_2(\theta) y^I_2(t) + A^{I\Gamma}_2(\theta) y^\Gamma(t) &= B^I_2(\theta) u^I_2(t) \\
M^{I\Gamma}_1 \frac{d}{dt} y^I_1(t) + M^{\Gamma\Gamma}(\theta) \frac{d}{dt} y^\Gamma(t) + M^{I\Gamma}_2(\theta) \frac{d}{dt} y^I_2(t) \\
+ A^{I\Gamma}_1 y^I_1(t) + A^{\Gamma\Gamma}(\theta) \frac{d}{dt} y^\Gamma(t) + A^{I\Gamma}_2(\theta) y^I_2(t) &= B^\Gamma(\theta) u^\Gamma(t),
\end{align*}
\]

\[
\begin{align*}
-M^{II}_1 \frac{d}{dt} p^I_1(t) - M^{I\Gamma}_1 \frac{d}{dt} p^\Gamma(t) + A^{II}_1 p^I_1(t) + A^{I\Gamma}_1 p^\Gamma(t) &= -(C^I_1)^T (C^I_1 y^I_1(t) - d^I_1) \\
-M^{II}_2(\theta) \frac{d}{dt} p^I_2(t) - M^{I\Gamma}_2(\theta) \frac{d}{dt} p^\Gamma(t) + A^{II}_2(\theta) p^I_2(t) + A^{I\Gamma}_2(\theta) p^\Gamma(t) &= -\nabla y^I_2 \tilde{\ell}(.) \\
-M^{I\Gamma}_1 \frac{d}{dt} p^I_1(t) - M^{\Gamma\Gamma}(\theta) \frac{d}{dt} p^\Gamma(t) - M^{I\Gamma}_2(\theta) \frac{d}{dt} p^I_2(t) \\
+ A^{I\Gamma}_1 p^I_1(t) + A^{\Gamma\Gamma}(\theta) \frac{d}{dt} p^\Gamma(t) + A^{I\Gamma}_2(\theta) p^I_2(t) &= -\nabla y^\Gamma \tilde{\ell}(.),
\end{align*}
\]

To apply model reduction to the system corresponding to fixed subdomain \(\Omega_1\), we have to identify how \(y^I_1\) and \(p^I_1\) interact with other components.
Model Reduction of Fixed Subdomain Problem

We need to reduce

\[ M_{II} \frac{d}{dt} y_I(t) = -A_{II} y_I(t) - M_{I\Gamma} \frac{d}{dt} y_{\Gamma}(t) + B_{I} u_{I}(t) - A_{I\Gamma} y_{\Gamma}(t) \]

\[ z_I = C_{I} y_I(t) - d_{I} \]

\[ z_{\Gamma} = -M_{I\Gamma} \frac{d}{dt} y_I - A_{I\Gamma} y_{\Gamma} , \]

\[-M_{II} \frac{d}{dt} p_I(t) = -A_{II} p_I(t) + M_{I\Gamma} \frac{d}{dt} p_{\Gamma}(t) - (C_{I})^T z_I - A_{I\Gamma} p_{\Gamma}(t) \]

\[ q_I = (B_{I})^T p_I \]

\[ q_{\Gamma} = M_{I\Gamma} \frac{d}{dt} p_I - A_{I\Gamma} p_{I} \]

For simplicity we assume that

\[ M_{I\Gamma} = 0 \quad M_{II} = 0, \]
we get

\[
M_1^{II} \frac{d}{dt} y_1^I(t) = -A_1^{II} y_1^I(t) + \left( B_1^I \mid - A_1^{II} \right) \begin{pmatrix} u_1^I \\ y_1^\Gamma \end{pmatrix},
\]

\[
\begin{pmatrix} z_1^I \\ z_1^\Gamma \end{pmatrix} = \begin{pmatrix} -C_1^I \\ -A_1^{II} \end{pmatrix} y_1^I + \begin{pmatrix} I \\ 0 \end{pmatrix} d_1^I,
\]

\[
-M_1^{II} \frac{d}{dt} p_1^I(t) = -A_1^{II} p_1^I(t) + \left( -(C_1^I)^T \mid - A_1^{II} \right) \begin{pmatrix} z_1^I \\ p_1^\Gamma \end{pmatrix},
\]

\[
\begin{pmatrix} q_1^I \\ q_1^\Gamma \end{pmatrix} = \begin{pmatrix} (B_1^I)^T \\ -A_1^{II} \end{pmatrix} p_1^I.
\]

System is exactly of form needed for balanced truncation model red.
Reduced Optimization Problem

- We apply BTMR to the fixed subdomain problem with inputs and output determined by the original inputs to subdomain 1 as well as the interface conditions.
- In optimality conditions replace fixed subdomain problem by its reduced order model.
- We can interpret the resulting reduced optimality system as the optimality system of the following reduced optimization problem

\[
\min \int_0^T \frac{1}{2} \| \hat{C}_1 \hat{y}_1^I - d_1^I(t) \|^2 + \tilde{\ell}(y^\Gamma(t), y_2^I(t), t, \theta) dt
\]

subject to

\[
\begin{align*}
\hat{M}_{1\Gamma} \frac{d}{dt} \hat{y}_1^I(t) + \hat{M}_{1\Gamma} \frac{d}{dt} y^\Gamma(t) + \hat{A}_{1\Gamma} \hat{y}_1^I(t) + \hat{A}_{1\Gamma} y^\Gamma(t) & = \hat{B}_1 u_1^I(t) \\
M_{2\Gamma}(\theta) \frac{d}{dt} y_2^I(t) + M_{2\Gamma}(\theta) \frac{d}{dt} y^\Gamma(t) + A_{2\Gamma}(\theta) y_2^I(t) + A_{2\Gamma}(\theta) y^\Gamma(t) & = B_2(\theta) u_2^I(t) \\
\hat{M}_{1\Gamma} \frac{d}{dt} y_1^I(t) + M_{\Gamma\Gamma}(\theta) \frac{d}{dt} y^\Gamma(t) + M_{2\Gamma}(\theta) \frac{d}{dt} y_2^I(t) + A_{1\Gamma} \hat{y}_1^I(t) + A_{\Gamma\Gamma}(\theta) \frac{d}{dt} y^\Gamma(t) + A_{2\Gamma}(\theta) y_2^I(t) & = B(\theta) u^\Gamma(t) \\
\hat{y}_1^I(0) & = \hat{y}_{1,0}^I \quad y_2^I(0) = y_{2,0}^I, \quad y^\Gamma(0) = y_0^\Gamma, \quad \theta \in \Theta_{ad}
\end{align*}
\]
Error Estimate

If there exists $\alpha > 0$ such that
\[ v^T A v \leq -\alpha v^T M v, \quad \forall v \in \mathbb{R}^N, \]

the gradients $\nabla_{y_I^{(2)}} \tilde{\ell}(y_I^{(2)}, y_{\Gamma}, t, \theta)$, $\nabla_{y_{\Gamma}} \tilde{\ell}(y_I^{(2)}, y_{\Gamma}, t, \theta)$, $\nabla_{\theta} \tilde{\ell}(y_I^{(2)}, y_{\Gamma}, t, \theta)$, are Lipschitz continuous in $y_I^{(2)}$, $y_{\Gamma}$

for all $\|\tilde{\theta}\| \leq 1$ and all $\theta \in \Theta$ the following bound holds
\[ \max \left\{ \| D_{\theta} M^{(2)}(\theta) \tilde{\theta} \|, \| D_{\theta} A^{(2)}(\theta) \tilde{\theta} \|, \| D_{\theta} B^{(2)}(\theta) \tilde{\theta} \| \right\} \leq \gamma, \]

then there exists $c > 0$ dependent on $u$, $\hat{y}$, and $\hat{p}$ such that
\[ \| \nabla J(\theta) - \nabla \hat{J}(\theta) \|_{L^2} \leq \frac{c}{\alpha} (\sigma_{r+1} + \ldots + \sigma_n). \]

If we assume the convexity condition
\[ (\nabla J(\hat{\theta}_*) - \nabla J(\theta_*))^T (\hat{\theta}_* - \theta_*) \geq \kappa \| \hat{\theta}_* - \theta_* \|^2, \]

then we obtain the error bound
\[ \| \theta_* - \hat{\theta}_* \| \leq \frac{c}{\alpha \kappa} (\sigma_{r+1} + \ldots + \sigma_n). \]
Example

- Reference domain $\Omega_{\text{ref}}$

- Optimization problem

$$
\min \int_0^T \int_{\Gamma_L \cup \Gamma_R} |y - y^d|^2 dsdt + \int_0^T \int_{\Omega_2(\theta)} |y - y^d|^2 dxdt
$$

subject to the differential equation

$$
y_t(x, t) - \Delta y(x, t) + y(x, t) = 100 \quad \text{in } \Omega(\theta) \times (0, T),
$$

$$
n \cdot \nabla y(x, t) = 0 \quad \text{on } \partial \Omega(\theta) \times (0, T),
$$

$$
y(x, 0) = 0 \quad \text{in } \Omega(\theta)
$$

and design parameter constraints $\theta_{\text{min}} \leq \theta \leq \theta_{\text{max}}$.

- We use $k_T = 3, k_B = 3$ Bézier control points to specify the top and the bottom boundary of the variable subdomain $\Omega_2(\theta)$.

The desired temperature $y^d$ is computed by specifying the optimal parameter $\theta_*$ and solving the state equation on $\Omega(\theta_*)$. 
We use automatic differentiation to compute the derivatives with respect to the design variables $\theta$.

The semi-discretized optimization problems are solved using a projected BFGS method with Armijo line search. The optimization algorithm is terminated when the norm of projected gradient is less than $\epsilon = 10^{-4}$.

The optimal domain
\begin{tabular}{|c|c|c|}
\hline
$N_{dof}^{(1)}$ & $N_{dof}$ \\
\hline
Reduced & 147 & 581 \\
Full & 4280 & 4714 \\
\hline
\end{tabular}

Sizes of the full and the reduced order problems

The largest Hankel singular values and the threshold $10^{-4} \sigma_1$

Error in solution between full and reduced order problem:
$$\|\theta^* - \hat{\theta}^*\|_2 = 2.325 \cdot 10^{-4}$$

Optimal shape parameters $\theta^*$ and $\hat{\theta}^*$ (rounded to 5 digits) computed by minimizing the full and the reduced order model.

$\theta^*$: (1.00, 2.0000, 2.0000, -2.0000, -2.0000, -1.00)

$\hat{\theta}^*$: (1.00, 1.9999, 2.0001, -2.0001, -1.9998, -1.00)
The convergence histories of the projected BFGS algorithm applied to the full and the reduced order problems.

convergence history of the objective functionals for the full (+) and reduced (o) order model.

convergence history of the projected gradients for the full (+) and reduced (o) order model.
Example - Stokes

Geometry motivated by biochip

Problems where the shape param. $\theta$ only influences a (small) subdomain:

$$\bar{\Omega} (\theta) := \bar{\Omega}_1 \cup \bar{\Omega}_2 (\theta), \quad \Omega_1 \cap \Omega_2 (\theta) = \emptyset, \quad \Gamma = \bar{\Omega}_1 \cap \bar{\Omega}_2 (\theta).$$

Here $\bar{\Omega}_2 (\theta)$ is the top left yellow, square domain.
\[
\min_{\theta_{\text{min}} \leq \theta \leq \theta_{\text{max}}} J(\theta) = \int_0^T \int_{\Omega_{\text{obs}}} \frac{1}{2} |\nabla \times \mathbf{v}(x, t; \theta)|^2 dx + \int_{\Omega_2(\theta)} \frac{1}{2} |\mathbf{v}(x, t; \theta) - \mathbf{v}^d(x, t)|^2 dx dt
\]

where \(\mathbf{v}(\theta)\) and \(p(\theta)\) solve the Stokes equations

\[
\begin{align*}
\mathbf{v}_t(x, t) - \mu \Delta \mathbf{v}(x, t) + \nabla p(x, t) &= \mathbf{f}(x, t), & \text{in } \Omega(\theta) \times (0, T), \\
\nabla \cdot \mathbf{v}(x, t) &= 0, & \text{in } \Omega(\theta) \times (0, T), \\
\mathbf{v}(x, t) &= \mathbf{v}_{\text{in}}(x, t) & \text{on } \Gamma_{\text{in}} \times (0, T), \\
\mathbf{v}(x, t) &= \mathbf{0} & \text{on } \Gamma_{\text{lat}} \times (0, T), \\
-(\mu \nabla \mathbf{v}(x, t) - p(x, t) I)n &= 0 & \text{on } \Gamma_{\text{out}} \times (0, T), \\
\mathbf{v}(x, 0) &= \mathbf{0} & \text{in } \Omega(\theta).
\end{align*}
\]

Here \(\overline{\Omega(\theta)} = \overline{\Omega_1} \cup \overline{\Omega_2(\theta)}\) and \(\overline{\Omega_2(\theta)}\) is the top left yellow, square domain. Observation region \(\Omega_{\text{obs}}\) is part of the two reservoirs.

Stokes equation requires additional care:

- Domain decomposition ([Pavarino and Widlund, 2002]).
- Balanced truncation ([Stykel, 2006], [Heinkenschloss et al., 2008])
- See [Antil et al., 2011]
We have 12 shape parameters, $\theta \in \mathbb{R}^{12}$.
<table>
<thead>
<tr>
<th>grid</th>
<th>$m$</th>
<th>$N_{v,dof}^{(1)}$</th>
<th>$N_{\tilde{v},dof}^{(1)}$</th>
<th>$N_{v,dof}$</th>
<th>$N_{\tilde{v},dof}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>149</td>
<td>4752</td>
<td>23</td>
<td>4862</td>
<td>133</td>
</tr>
<tr>
<td>2</td>
<td>313</td>
<td>7410</td>
<td>25</td>
<td>7568</td>
<td>183</td>
</tr>
<tr>
<td>3</td>
<td>361</td>
<td>11474</td>
<td>26</td>
<td>11700</td>
<td>252</td>
</tr>
<tr>
<td>4</td>
<td>537</td>
<td>16472</td>
<td>29</td>
<td>16806</td>
<td>363</td>
</tr>
</tbody>
</table>

The number $m$ of observations in $\Omega_{obs}$, the number of velocities $N_{v,dof}^{(1)}, N_{\tilde{v},dof}^{(1)}$ in the fixed subdomain $\Omega_1$ for the full and reduced order model, the number of velocities $N_{v,dof}, N_{\tilde{v},dof}$ in the entire domain $\Omega$ for the full and reduced order model for five discretizations.

The largest Hankel singular values and the threshold $10^{-3} \sigma_1$
Error in optimal parameter computed using the full and the reduced order model (rounded to 5 digits)


The convergence histories of the projected BFGS algorithm applied to the full and the reduced order problems.

Convergence history of the objective functionals for the full (+) and reduced (o) order model.

Convergence history of the projected gradients for the full (+) and reduced (o) order model.
Recap of Part I

- Reviewed projection based model reduction for simulation.
- Reviewed adjoint eqn. approach for gradient and Hessian computation.
  - Gradient computation requires solution of adjoint PDE.
  - Hessian time vector computation requires solution of linearized state PDE and 2nd order adjoint PDE.
- Reduced order models for optimization must approximate the objective function $j(u)$ and its gradient $\nabla j(u)$.
- Considered two classes of optimization problems
  - Parameterized linear quadratic problems.
    Sample optimization problems to generate reduced order model that allows fast on-line solution of linear quadratic problem at out of sample parameter.
  - Linear quadratic problems, or problems with localized nonlinearity for which reduced order models can be computed that are good approximations for all controls $u$. 
Review Part I

- Overview
- Example Optimization Problems
- Optimization Problem
- Projection Based Model Reduction
- Back to Optimization
- Error Estimates
- Linear-Quadratic Problems
- Shape Optimization with Local Parameter Dependence
Original problem

\[
\min_j j(u) \\
\text{s.t. } u \in \mathcal{U}_{ad},
\]

where \( j(u) = J(y(u), u) \), \( y(u) \in \mathbb{R}^n \) solves \( c(y, u) = 0 \). \( n \) large.

Reduced order problem

Construct \( V \in \mathbb{R}^{n \times r} \), \( r \ll n \), \( \text{rank}(V) = r \).

Reduced order problem:

\[
\min \hat{j}(u) \\
\text{s.t. } u \in \mathcal{U}_{ad},
\]

where \( \hat{j}(u) = J(V\hat{y}(u), u) \), \( \hat{y}(u) \in \mathbb{R}^r \) solves reduced state equation \( V^T c(V\hat{y}, u) = 0 \in \mathbb{R}^r \).
Review Proper Orthogonal Decomposition I

▶ Finite dimensional representation of snapshots
\( \mathbf{y}(t_1), \ldots, \mathbf{y}(t_m) \in \mathbb{R}^n, m > r. \)

▶ Inner product \( \mathbf{v}^T \mathbf{M} \mathbf{w} \) and norm \( \| \cdot \|_\mathbf{M} \). \( \mathbf{M} \) s.p.d. but not nec. mass matrix.

▶ Compute orthonormal basis \( \mathbf{v}_1, \ldots, \mathbf{v}_r \) as solution of

\[
\min \sum_{k=1}^{m} \| \mathbf{y}(t_k) - \sum_{i=1}^{r} \mathbf{y}(t_k)^T \mathbf{M} \mathbf{v}_i \mathbf{v}_i \|_\mathbf{M}^2 \\
\text{s.t. } \mathbf{v}_i^T \mathbf{M} \mathbf{v}_j = \delta_{ij}.
\]

▶ Solution

▶ Define \( \mathbf{Y} = [\mathbf{y}(t_1), \ldots, \mathbf{y}(t_m)] \in \mathbb{R}^{n \times m}. \)

▶ Compute \( \mathbf{M} \)-orthonormal eigenvecs. \( \mathbf{v}_1, \mathbf{v}_2, \ldots \in \mathbb{R}^n \) and eigenvals.

\( \lambda_1 \geq \ldots \geq \lambda_m \geq 0 \) of generalized \( n \times n \) eigenvalue prob.

\[ \mathbf{M} \mathbf{Y} \mathbf{Y}^T \mathbf{M} \mathbf{v}_i = \lambda_i \mathbf{M} \mathbf{v}_i. \]
Alternatively, if \( n > m \) compute eigenvectors \( \mathbf{w}_1, \mathbf{w}_2, \ldots \in \mathbb{R}^m \) and eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\min\{m,n\}} \geq 0 \) of

\[
\mathbf{Y}^T \mathbf{M} \mathbf{Y} \mathbf{w}_i = \lambda_i \mathbf{w}_i.
\]

\[
\mathbf{v}_i = \lambda_i^{-1/2} \mathbf{Y} \mathbf{w}_i, \ i = 1, \ldots, m.
\]

Usually, fix tolerance \( \epsilon > 0 \). Compute eigenvectors \( \mathbf{v}_1, \mathbf{v}_2, \ldots \in \mathbb{R}^n \) and eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\min\{m,n\}} \geq 0 \).

Find smallest \( r \) such that \( \sum_{i=r+1}^{m} \lambda_i < \epsilon \).

If only some of the largest eigenvals. and vecs. are computed:

Find smallest \( r \) such that \( \lambda_{r+1}/\lambda_1 < \epsilon \).

Reduced order model \( \mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_r] \in \mathbb{R}^{n \times r} \).

Error

\[
\sum_{k=1}^{m} \| \mathbf{y}(t_k) - \sum_{i=1}^{r} \mathbf{y}(t_k)^T \mathbf{M} \mathbf{v}_i \mathbf{v}_i \|^2_{\mathbf{M}} = \sum_{i=r+1}^{\min\{m,n\}} \lambda_i
\]
Review Error Estimate for Unconstrained Problems

- \( \widehat{u}_* = \arg\min_u \widehat{j}(u) \) minimizer of *unconstrained* reduced problem.

- Newton-Kantorovich Theorem: Let \( r > 0 \) and \( \nabla^2 j \in \text{Lip}_L(B_r(\widehat{u}_*)) \). \( \nabla^2 j(\widehat{u}_*) \) be nonsingular and constants \( \zeta, \eta \geq 0 \) such that

\[
\left\| \nabla^2 j(\widehat{u}_*)^{-1} \right\| = \zeta, \quad \left\| \nabla^2 j(\widehat{u}_*)^{-1} \nabla j(\widehat{u}_*) \right\| \leq \eta.
\]

If \( L\zeta \eta \leq \frac{1}{2} \), there is unique local minimum \( u_* \) of \( j \) in ball around \( \widehat{u}_* \) with radius \( \min \left\{ r, \left( 1 - \sqrt{1 - 2L\zeta \eta} \right) / (L\zeta) \right\} \leq \min \left\{ r, 2\eta \right\} \).

- Estimate \( \eta \):

\[
\left\| \nabla^2 j(\widehat{u}_*)^{-1} \left( \nabla j(\widehat{u}_*) - \nabla \widehat{j}(\widehat{u}_*) \right) \right\| \leq \zeta \left\| \nabla j(\widehat{u}_*) - \nabla \widehat{j}(\widehat{u}_*) \right\| = \eta.
\]

- Hence

\[
\left\| u_* - \widehat{u}_* \right\|_U \leq 2\zeta \left\| \nabla \widehat{j}(\widehat{u}_*) - \nabla j(\widehat{u}_*) \right\|_U
\]

- Estimate error in gradients to get estimate for error in solution.

- Need \( L\zeta \eta \leq \frac{1}{2} \), i.e., \( \nabla j(\widehat{u}_*) \) small enough.

- Can estimate error using convergence properties of Newton’s method started with \( \widehat{u}_* \) applied to original problem.
Outline

Overview

Example Optimization Problems

Optimization Problem

Projection Based Model Reduction

Back to Optimization

Error Estimates

Linear-Quadratic Problems

Shape Optimization with Local Parameter Dependence

Semilinear Parabolic Problems

Trust-Region Framework
Semilinear Parabolic Model Problems

- **Distributed control**

\[
\min \frac{1}{2} \int_0^T \int_\Omega |y(x, t) - z(x, t)|^2 \, dx \, dt + \frac{\alpha}{2} \int_0^T \int_\Omega |u(x, t)|^2 \, dx \, dt,
\]

where \( y = y(u) \) solves

\[
y_t(x, t) - \nu \Delta y(x, t) + y(x, t)^3 + u(x, t) = f(x, t), \quad (x, t) \in \Omega \times (0, T),
\]

\[
y(x, t) = 0, \quad (x, t) \in \Gamma \times (0, T),
\]

\[
y(x, 0) = y_0(x), \quad x \in \Omega.
\]

- **Robin boundary control of Burgers equation**

\[
\min \frac{1}{2} \int_0^1 \int_0^T |y(x, t) - z(x, t)|^2 \, dx \, dt + \frac{\alpha}{2} \int_0^T (|u_0(t)|^2 + |u_1(t)|^2) \, dt,
\]

where \( y = y(u) \) solves

\[
y_t(x, t) - \nu y_{xx}(x, t) + y(x, t)y_x(x, t) = f(x, t), \quad (x, t) \in (0, 1) \times (0, T),
\]

\[
\nu y_x(0, t) + \sigma_0 y(0, t) = u_0(t), \quad t \in (0, T),
\]

\[
\nu y_x(1, t) + \sigma_1 y(1, t) = u_1(t), \quad t \in (0, T),
\]

\[
y(x, 0) = y_0(x), \quad x \in (0, 1).
\]
Semidiscrete optimization problem

\[
\min_{u} j(u) = \int_{0}^{T} \left( \frac{1}{2} y(t)^T Q y(t) + c(t)^T y(t) + \frac{\alpha}{2} u(t)^T R u(t) \right) dt
\]

where \( y \) is the solution of

\[
M \frac{d}{dt} y(t) + A y(t) + N(y(t)) + B u(t) = f(t), \quad t \in (0, T),
\]

\[
y(0) = y_0.
\]

State \( y(t) \in \mathbb{R}^n, n \) large, control \( u(t) \in \mathbb{R}^m \).
Assume (satisfied for our two model problems)

- \( M \in \mathbb{R}^{n \times n}, R \in \mathbb{R}^{m \times m} \) are spd,
- \( Q \in \mathbb{R}^{n \times n} \) is spsd,
- there exist \( \gamma > 0 \) such that \( v^T A v \geq \gamma v^T M v \) \( \forall v \in \mathbb{R}^n \),
- \( N \) smooth; \( N, N' \) locally Lipschitz cont.; \( v^T N(v) \geq 0 \) \( \forall v \in \mathbb{R}^n \)
Well-posedness

- If $v^T N(v) \geq 0$ for all $v \in \mathbb{R}^n$, state equation has unique $y = y(u)$:

$$\|y(t)\|_M^2 \leq e^{-\frac{\gamma}{2} t} \|y_0\|_M^2 + \frac{1}{2\gamma} e^{-\frac{\gamma}{2} t} \left( \int_0^T e^{\frac{\gamma}{2} s} \|f(s) - Bu(s)\|_{M-1}^2 ds \right) \quad \forall t$$

and

$$\int_0^T \|y(t)\|_M^2 \leq \frac{2}{\gamma} (1 - e^{-\frac{\gamma}{2} T}) \|y_0\|_M^2 + \frac{1}{\gamma^2} (e^{\frac{\gamma}{2} T} - 1) \int_0^T \|f(t) - Bu(t)\|_{M-1}^2 dt.$$ 

- If there is constant $c_0$ such that for all pairs $y, u$ satisfying state eqn.,

$$\int_0^T \frac{1}{2} y(t)^T Q y(t) + c(t)^T y(t) \, dt \geq c_0,$$

(in model problems satisfied with $c_0 = -\int_0^T \int_{\Omega} z^2(x, t) \, dx \, dt$), then optimal control problem has a solution.
Reduced Order Problem

- Construct a reduced basis matrix, \( V \in \mathbb{R}^{n \times r} \)

\[
\min \hat{j}(u) = \int_0^T \frac{1}{2} (V\hat{y}(t))^T Q(V\hat{y}(t)) + (c(t))^T V\hat{y}(t) + \frac{\alpha}{2} u(t)^T R u(t) \, dt
\]

where \( \hat{y} \) solves

\[
V^T M V \frac{d}{dt} \hat{y}(t) + V^T A V \hat{y}(t) + V^T N(V\hat{y}(t)) + V^T B u(t) = V^T f(t), \quad t \in (0, T),
\]

\[
y(0) = V^T M y_0.
\]

- Evaluation of nonlinear term \( (V^T N(V\hat{y}(t))). \)
  - In model problems, \( N \) is (low order) polynomial nonlinearity.
    - Precompute tensors - evaluation of \( \hat{y}(t) \mapsto V^T N(V\hat{y}(t)) \) costs \( O(r) \).
  - Otherwise use (D)EIM to build approx. whose eval. costs \( O(r) \).
Gradient evaluation original problem

- Solve state PDE

\[ M \frac{d}{dt} y(t) + A y(t) + N(y(t)) + Bu(t) = f(t), \quad t \in (0, T), \quad y(0) = y_0. \]

- Solve adjoint PDE

\[ -M \frac{d}{dt} p(t) + A^T p(t) + N'(y(t))^T p(t) = -(Qy(t) + c(t)), \quad t \in (0, T), \quad p(T) = 0. \]

- Gradient: \( \nabla j(u) = \alpha Ru + B^T p. \)

Gradient evaluation reduced order problem

- Solve reduced order state PDE

\[ \frac{d}{dt} \hat{y}(t) + V^T A V \hat{y}(t) + V^T N(V \hat{y}(t)) + V^T Bu(t) = V^T f(t), \quad t \in (0, T), \quad \hat{y}(0) = V^T M y_0. \]

- Solve reduced order adjoint PDE

\[ -\frac{d}{dt} \hat{p}(t) + V^T A^T V \hat{p}(t) + V^T N'(V \hat{y}(t))^T V \hat{p}(t) = -(V^T Q V \hat{y}(t) + V^T c(t)), \quad t \in (0, T), \quad \hat{p}(T) = 0. \]

- Gradient: \( \nabla j(u) = \alpha Ru + B^T V \hat{p}. \)
Error between gradients:

\[
\|\nabla j(u) - \nabla \hat{j}(u)\|_{L^2} \\
\leq \|B\|_2 \|M^{-\frac{1}{2}}\|_2 \left( C_1 \|y - VV^TMy\|_{L^2} + C_2 \|\tilde{p} - VV^T\tilde{M}\tilde{p}\|_{L^2} \right),
\]

where \(C_1, C_2\) are constants and \(\tilde{p}\) solves aux. adjoint equation

\[
-M \frac{d}{dt} \tilde{p}(t) + A^T\tilde{p}(t) + N'(V\hat{y}(t))^T\tilde{p}(t) = -(QV\hat{y}(t) + c(t)) \\
\tilde{p}(T) = 0
\]

Error between solutions: If \(L\zeta\|\nabla j(u) - \nabla \hat{j}(u)\|_{L^2} < \frac{1}{2}\),

\[
\|\hat{u}^* - u^*\|_{L^2} \\
\leq 2\zeta \|B\|_2 \|M^{-\frac{1}{2}}\|_2 \left( C_1 \|y - VV^TMy\|_{L^2} + C_2 \|\tilde{p} - VV^T\tilde{M}\tilde{p}\|_{L^2} \right).
\]
Numerical Experiments

- POD basis computations
  - States $y(t)$ and adjoints $p(t)$ represent different quantities.
  - $y(t)$ and $p(t)$ can have very different scales - for least squares problems often $\|p(t)\| \ll \|y(t)\|$.
  - POD is scale dependent:
    E.g., for snapshots $S = \text{diag}(e_1, \ldots, e_k, \delta e_{k+1}, \ldots, \delta e_{2k})$,
    POD basis is $V = (e_1, \ldots, e_k)$.
  - Apply POD to state and adjoint snapshots separately:
    - Compute POD basis $V_S$ from state snapshots.
    - Compute POD basis $V_A$ from adjoint snapshots.
    - Compute $(M\text{-})$ orthogonal matrix $V \in \mathbb{R}^{n \times r}$ with $R(V) = R(V_S V_A)$.

- Algorithm
  - Given control $\hat{u}_c$.
  - Solve full state and equations to construct state and adjoint snapshots.
  - Compute $V$ as above.
  - Solve reduced optimization problem to get new control $\hat{u}_+$.
  - If $\|\hat{u}_c - \hat{u}_+\|_{L^2(0,T)}$ and relative state, adjoint, gradient residuals are small, stop. Otherwise set $\hat{u}_c = \hat{u}_+$ and repeat.

Often works very well, but no convergence guarantee
Numerical Experiments

- Full and reduced order optimization problems solved using Newton-CG

- Semilinear heat equation ($\alpha = 10^{-2}$)

<table>
<thead>
<tr>
<th>iter</th>
<th>$r$</th>
<th>$\frac{|\hat{u}^<em>-u^</em>|}{|u^*|}$</th>
<th>$\hat{j}(\hat{u}^*)$</th>
<th>$\frac{|y-VV^TMy|}{|y|}$</th>
<th>$\frac{|\tilde{p}-VV^T\tilde{M}\tilde{p}|}{|\tilde{p}|}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>7.95e-03</td>
<td>-1.6825</td>
<td>5.47e-03</td>
<td>1.49e-03</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>6.54e-04</td>
<td>-1.6844</td>
<td>4.36e-03</td>
<td>7.13e-04</td>
</tr>
</tbody>
</table>

> 60 times faster than full order problem

- Burgers equation ($\alpha = 10^{-3}$)

<table>
<thead>
<tr>
<th>iter</th>
<th>$r$</th>
<th>$\frac{|\hat{u}^<em>-u^</em>|}{|u^*|}$</th>
<th>$\hat{j}(\hat{u}^*)$</th>
<th>$\frac{|y-VV^TMy|}{|y|}$</th>
<th>$\frac{|\tilde{p}-VV^T\tilde{M}\tilde{p}|}{|\tilde{p}|}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>9.05e-01</td>
<td>-0.1451</td>
<td>1.14e-02</td>
<td>1.11e-03</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>4.15e-02</td>
<td>-0.1302</td>
<td>3.04e-03</td>
<td>9.74e-05</td>
</tr>
</tbody>
</table>

$\approx$ 4 times faster than full order problem
Semilinear heat equation

\[ u^*(1) \]

\[ \hat{u}^*(1) \]

Burgers equation

\[ u_1^*(t) \]

\[ \hat{u}_1^*(t) \]
In both examples
\[
\| \nabla j(\hat{u}^*) - \nabla \hat{j}(\hat{u}^*) \|
\]
\[
\approx \| B \|_2 \| M^{-\frac{1}{2}} \|_2 \left( \| y - VV^T M y \|_{L^2} + \| \tilde{p} - VV^T M \tilde{p} \|_{L^2} \right).
\]

In both examples
\[
\| \hat{u}^* - u^* \|_{L^2}
\]
\[
\leq \alpha^{-1} \| B \|_2 \| M^{-\frac{1}{2}} \|_2 \left( \| y - VV^T M y \|_{L^2} + \| \tilde{p} - VV^T M \tilde{p} \|_{L^2} \right).
\]

For zero-residual nonlinear least squares problems
\[
\alpha^{-1} = \zeta = \| \nabla^2 j(\hat{u}^*)^{-1} \|.
\]

Since \( \alpha > 0 \) is control penalty, nonlinear least squares problem usually has non-zero residual.

For semilinear heat eqn., our numerical estimate of \( \zeta = \| \nabla^2 j(u^*)^{-1} \| \gg \alpha^{-1} \).

Newton-Kantorovich was likely never applicable because
\( L \zeta \eta \not\leq \frac{1}{2} \).
Naive approach works well in this case, but no theoretical guarantees.
Outline

Overview
Example Optimization Problems
Optimization Problem
Projection Based Model Reduction
Back to Optimization
Error Estimates
Linear-Quadratic Problems
Shape Optimization with Local Parameter Dependence
Semilinear Parabolic Problems
Trust-Region Framework
Trust-Region Framework

- Want to solve $\min j(u)$.
- At given $u_k$ can construct reduced order model $\hat{j}_k(u)$.
- Approximately solve sequence of problems

$$\min \ \hat{j}_k(u_k + s),$$
$$\text{s.t. } \|s\|_{\mathcal{U}} \leq \delta_k,$$

where $\{u_k + s : \|s\|_{\mathcal{U}} \leq \delta_k\}$ is the region over which $\hat{j}_k$ is trusted to be a good model of $j$.
Trust region parameter $\delta_k > 0$ is adjusted by algorithm.

- Originally applied with

$$\hat{j}_k(u_k + s) = j(u_k) + \langle \nabla j(u_k), s \rangle + \frac{1}{2} \langle \nabla^2 j(u_k)s, s \rangle.$$

- Use trust-region framework to manage the quality of the model $\hat{j}_k$
  - Trust-region for model management: [Alexandrov et al., 1998].
  - Trust-region methods and theory book: [Conn et al., 2000].
  - Applied to reduced order modeling: [Fahl and Sachs, 2003], [Yue and Meerbergen, 2013], [Gohlke, 2013], S. Ulbrich.
At the $k^{th}$ iteration of the trust region algorithm, we approx. solve

$$\min \quad \hat{j}_k(u_k + s),$$
$$\text{s.t.} \quad \|s\|_U \leq \delta_k$$

(Theoretically, could also solve

$$\min \quad \langle \nabla \hat{j}_k(u_k), s \rangle + \frac{1}{2} \langle \nabla^2 \hat{j}_k(u_k)s, s \rangle,$$
$$\text{s.t.} \quad \|s\|_U \leq \delta_k.$$ )

Traditionally, decide on acceptance of step and update of TR radius based on the ratio between actual and predicted reduction

$$\rho_k = \frac{j(u_k) - j(u_k + s_k)}{\langle \nabla \hat{j}_k(u_k), s_k \rangle + \frac{1}{2} \langle \nabla^2 \hat{j}_k(u_k)s_k, s_k \rangle}.$$ 

Issue: This requires exact function and gradient evaluations (original convergence theory required $\nabla \hat{j}_k(u_k) = \nabla j(u_k)$).
Want to relax $\nabla \hat{j}_k(u_k) = \nabla j(u_k)$ and replace actual reduction

$$\text{ared}_k = j(u_k) - j(u_k + s_k) \quad \rightarrow \quad \text{cred}_k = \hat{j}_k(u_k) - \hat{j}_k(u_k + s_k).$$

How accurate do function and gradient approximations needs to be?

- Earlier theory required known bounds on errors; asymptotic estimates were not enough.
- Strengthened theory based on results by [Heinkenschloss and Vicente, 2001] and [Ziems and Ulbrich, 2011] to allow implementable bounds, if asymptotic error estimates are available [Kouri et al., 2014].

With our implementable relaxations on function and gradient approximations, iterates $\{u_k\}$ generated by trust-region algorithm satisfy

$$\liminf_{k \to \infty} \|\nabla \hat{j}_k(u_k)\|_U = \liminf_{k \to \infty} \|\nabla j(u_k)\|_U = 0.$$
Inexact Gradient Condition

- Model gradient needs to satisfy
  \[
  \| \nabla \hat{j}_k(u_k) - \nabla j(u_k) \|_U \leq \xi_g \min \{ \| \nabla \hat{j}_k(u_k) \|_U, \delta_k \}
  \]
  for \( \xi_g > 0 \) independent of \( k \).

- If we have an estimator \( \theta_k = \theta(u_k, s_k) \) so that for a constant \( K > 0 \),
  \[
  \left| (j(u_k) - j(u_k + s_k)) - (\hat{j}_k(u_k) - \hat{j}_k(u_k + s_k)) \right| \leq K \theta_k \quad \forall k
  \]
  then we require that
  \[
  \theta_k^\omega \leq \eta \min \{ \text{pred}_k, r_k \},
  \]
  where \( \omega \in (0, 1) \),
  \[
  \eta < \min \{ \eta_1, 1 - \eta_2 \} \quad \text{and} \quad \{ r_k \}_{k=1}^\infty \subset [0, \infty) \ \text{satisfies} \ \lim_{k \to \infty} r_k = 0.
  \]
Trust-Region Algorithm

1. **Initialization:** Given $u_k$, $\delta_k$, $0 < \gamma_1 \leq \gamma_2 < 1$, $\delta_{\text{max}} > 0$, and $0 < \eta_1 < \eta_2 < 1$.

2. **Model Selection:** Choose a model $j_k$ which satisfies
   \[
   \| \nabla \hat{j}_k(u_k) - \nabla j(u_k) \|_U \leq \xi_g \min\{ \| \nabla \hat{j}_k(u_k) \|_U, \delta_k \}. \]

3. **Step Computation:** Compute an approximate solution $s_k$ of
   \[
   \min_{\|s\|_U \leq \delta_k} \langle \nabla \hat{j}_k(u_k), s \rangle + \frac{1}{2} \langle \nabla^2 \hat{j}_k(u_k)s, s \rangle.
   \]

4. **Objective Function Update:** Determine objective function approximation $\hat{j}_k$ such that corresponding error estimate $\theta_k$ satisfies
   \[
   \theta_k^\omega \leq \eta \min \{ \text{pred}_k, \text{r}_k \}.
   \]

5. **Step Acceptance:** Compute $\varrho_k = \text{cred}_k / \text{pred}_k$.
   
   if $\varrho_k \geq \eta_1$ then $z_{k+1} = u_k + s_k$ else $z_{k+1} = u_k$ end if

6. **Trust-Region Update:**
   
   if $z_{k+1} = u_k$ then $\delta_{k+1} \in (0, \gamma_1 \|s_k\|_U]$ else Update $\delta_{k+1}$ by
   
   if $\varrho_k \leq \eta_1$ then $\delta_{k+1} \in (0, \gamma_2 \|s_k\|_U]$ end if
   
   if $\varrho_k \in (\eta_1, \eta_2)$ then $\delta_{k+1} \in [\gamma_2 \|s_k\|_U, \delta_k]$ end if
   
   if $\varrho_k \geq \eta_2$ then $\delta_{k+1} \in [\delta_k, \delta_{\text{max}}]$ end if
Application to ROM
Recall Error Bounds

▶ Error between states:

\[ \| y(u) - \hat{y}(u) \|_{L^2} \leq C_0 \| y - VV^T My \|_{L^2} \]

([Gohlke, 2013] for semilinear parabolic problem, [Chaturantabut and Sorensen, 2012] for more general case) implies

\[ | j(u) - \hat{j}(u) | \leq \tilde{C}_0 \| y - VV^T My \|_{L^2}. \]

▶ Error between gradients:

\[ \| \nabla j(u) - \nabla \hat{j}(u) \|_{L^2} \]

\[ \leq C_1 \| y - VV^T My \|_{L^2} + C_2 \| \tilde{p} - VV^T M\tilde{p} \|_{L^2} \]

where \( C_1, C_2 \) are constants and \( \tilde{p} \) solves aux. adjoint equation

\[ -M \frac{d}{dt} \tilde{p}(t) + A^T \tilde{p}(t) + N'(V\hat{y}(t))^T \tilde{p}(t) = -(QV\hat{y}(t) + c(t)) \]

\[ \tilde{p}(T) = 0 \]
Objective Function and Gradient Approximation

Given $u_k$ and tolerance $\epsilon > 0$ want to compute reduced order model, i.e., $V$ such that

$$|j(u_k) - \hat{j}(u_k)| \leq C\epsilon,$$
$$\|\nabla j(u_k) - \nabla \hat{j}(u_k)\|_{L^2} \leq C\epsilon$$

for some constant $C$ independent of $\epsilon$.

1. Solve PDE to get $y(t; u_k)$.
2. Apply POD to compute $V^y_r$ such that

$$\|y - V^y_r(V^y_r)^TMy\|_{L^2} = \sum_{i=r+1}^{n} \lambda_i \leq \epsilon. \quad (*)$$

Note for all $M$-orthogonal $V$ with $\text{Range}(V^y_r) \subset \text{Range}(V)$

$$\|y - VV^TMy\|_{L^2} \leq \epsilon.$$ 

$(*)$ implies

$$|j(u_k) - \hat{j}(u_k)| \leq \tilde{C}_0 \epsilon$$
3. Solve adjoint PDE

\[ -M \frac{d}{dt} \tilde{p}(t) + A^T \tilde{p}(t) + N'(y(t))^T \tilde{p}(t) = -(Qy(t) + c(t)) \]

\[ \tilde{p}(T) = 0 \]

(or with \( y \) replaced by \( V^y_r(V^y_r)^T M y \)) to get \( \tilde{p}(t) \).

4. Apply POD to compute \( V^p_r \) such that

\[ \| \tilde{p} - V^p_r(V^p_r)^T M \tilde{p} \|_{L^2} = \sum_{i=r+1}^{n} \lambda_i \leq \epsilon. \]

\( (r \text{ and } \lambda_i \text{'s different than } r \text{ and } \lambda_i \text{'s in 3.}). \)

Note for all \( M \)-orthogonal \( V \) with Range(\( V^y_r \)) ⊂ Range(\( V \))

\[ \| \tilde{p} - VV^T M \tilde{p} \|_{L^2} \leq \epsilon. \]

5. Compute \( M \)-orthogonal \( V \) with Range(\( V \)) = Range([\( V^y_r, V^p_r \)]).
6. Error between gradients:

\[
\| \nabla j(u_k) - \nabla \hat{j}(u_k) \|_{L^2} \\
\leq C_1 \| y - VV^T My \|_{L^2} + C_2 \| \tilde{p} - VV^T M\tilde{p} \|_{L^2} \leq C_3 \varepsilon .
\]
Numerical Experiments

- Semilinear heat equation ($\alpha = 10^{-2}$)

<table>
<thead>
<tr>
<th>iter</th>
<th>$r_k$</th>
<th>$|\hat{u}^* - u^<em>|_{L^2}/|u^</em>|_{L^2}$</th>
<th>$\hat{j}_k(\hat{u}^k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>5.66e-01</td>
<td>-1.2654</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>8.44e-02</td>
<td>-1.6746</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>3.77e-03</td>
<td>-1.6843</td>
</tr>
</tbody>
</table>

> 30 times faster than full order problem

- Burgers equation ($\alpha = 10^{-3}$)

<table>
<thead>
<tr>
<th>iter</th>
<th>$r_k$</th>
<th>$|\hat{u}^* - u^<em>|_{L^2}/|u^</em>|_{L^2}$</th>
<th>$\hat{j}_k(\hat{u}^k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>9.08e-01</td>
<td>0.0547</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>6.46e-01</td>
<td>-0.0290</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>5.55e-01</td>
<td>-0.0873</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>3.00e-01</td>
<td>-0.0152</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>1.19e-01</td>
<td>-0.1260</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>1.68e-02</td>
<td>-0.1293</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>2.32e-03</td>
<td>-0.1301</td>
</tr>
</tbody>
</table>

$\approx 5$ times faster than full order problem

- More iterations than iterative application of POD-ROM, but convergence guarantee and sometimes smaller ROMs.
- Didn’t use all of the latest features yet
Recap of Part II

- ROM for nonlinear, nonconvex problems.
- No one ROM valid over entire control space. Need to adjust ROM based on progress of optimization.
- For simple problems simple strategies can work well.
- Introduced trust-region framework for the general case.

Some Open Questions

- Handling of simple control constraints relatively easy.
- Handling of state constraints (e.g., point-wise bounds on temperature, stress, ...)?
- Not clear for what $u$ the ROM computed at $u_k$ can be used. In some examples ROM computed at $u_k$ can be used only good for $u$ very close to $u_k$. Checking validity of ROM expensive.
THE END
Adaptive control of a wake flow using proper orthogonal decomposition.
In Shape optimization and optimal design (Cambridge, 1999), volume 216 of

[Alexandrov et al., 1998] Alexandrov, N., Dennis Jr., J. E., Lewis, R. M., and
A trust region framework for managing the use of approximation models in
optimization.
Structural Optimization, 15:16–23.
Appeared also as ICASE report 97–50.

An adaptive POD approximation method for the control of advection-diffusion
equations.
In Bredies, K., Clason, C., Kunisch, K., and von Winckel, G., editors, Control and
Optimization with PDE Constraints, volume 164 of International Series of
Asymptotic stability of POD based model predictive control for a semilinear parabolic PDE.  
online first.

Design optimization using hyper-reduced-order models.  

Domain decomposition and balanced truncation model reduction for shape optimization of the Stokes system.  

Reduced order modeling based shape optimization of surface acoustic wave driven microfluidic biochips.  
Domain decomposition and model reduction for the numerical solution of PDE constrained optimization problems with localized optimization variables.

Application of the discrete empirical interpolation method to reduced order modeling of nonlinear and parametric systems.

*Approximation of large-scale dynamical systems*, volume 6 of *Advances in Design and Control*.
Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Interpolatory model reduction of large-scale dynamical systems.


Model reduction for daes with an application to flow control.
In King, R., editor, *Active Flow and Combustion Control 2014*, volume 127 of
Springer International Publishing.

Nonlinear model reduction via discrete empirical interpolation.

A state space error estimate for pod-deim nonlinear model reduction.

Weighted reduced basis method for stochastic optimal control problems with
elliptic PDE constraint.

*Computational methods for multiphase flows in porous media.*


Gradient-based constrained optimization using a database of linear reduced-order models.

*arXiv:1506.07849v1.*


*Trust–Region Methods.*

SIAM, Philadelphia.


Optimal control and numerical adaptivity for advection-diffusion equations.

Literature VII


Parallel-In-Time Gradient-Type Method For Optimal Control Problems.
PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, TX.
in progress.

Certified PDE-constrained parameter optimization using reduced basis surrogate models for evolution problems.

A posteriori error bounds for the empirical interpolation method.

Reduced order modelling approaches to PDE–constrained optimization based on proper orthogonal decomposition.
Structure-preserving, stability, and accuracy properties of the energy-conserving sampling and weighting method for the hyper reduction of nonlinear finite element dynamic models.

Optimal flow control based on POD and MPC and an application to the cancellation of Tollmien-Schlichting waves.

All optimal Hankel-norm approximations of linear multivariable systems and their $L^\infty$-error bounds.

Reduced order modeling for optimization of large scale dynamical systems.
Master’s thesis, Department of Computational and Applied Mathematics, Rice University, Houston, TX.
   Numerical analysis of optimality-system pod for constrained optimal control.
   In Mehl, M., Bischoff, M., and Schäfer, M., editors, Recent Trends in
   Computational Engineering - CE2014, volume 105 of Lecture Notes in
   Computational Science and Engineering, pages 297–317. Springer International
   Publishing.

   Proper Orthogonal Decomposition for linear-quadratic optimal control.
   Technical Report Konstanzer Schriften in Mathematik, FB Mathematik & Statistik,
   Universität Konstanz, Universitätsstrasse 10, D-78457 Konstanz, Germany.

   Numerical solution of implicitly constrained optimization problems.
   Technical Report TR08–05, Department of Computational and Applied
   Mathematics, Rice University, Houston, TX 77005–1892.

[Heinkenschloss et al., 2008] Heinkenschloss, M., Sorensen, D. C., and Sun, K.
   (2008).
   Balanced truncation model reduction for a class of descriptor systems with
   application to the Oseen equations.
Analysis of inexact trust–region SQP algorithms.

*Certified Reduced Basis Methods for Parametrized Partial Differential Equations*.

Springer Verlag, Heidelberg, New York, Berlin.

Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control.
A posteriori error estimation for semilinear parabolic optimal control problems with application to model reduction by POD.

A posteriori error estimation for reduced order solutions of parametrized parabolic optimal control problems.

A certified reduced basis method for parametrized elliptic optimal control problems.

Certified reduced basis methods for parametrized distributed optimal control problems.
Technical report, Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Germany.
Inexact objective function evaluations in a trust-region algorithm for pde-constrained optimization under uncertainty.

Uniform convergence of the pod method and applications to optimal control.
*Discrete and Continuous Dynamical Systems*, 35(9):4477–4501.

Control of Burger’s equation by a reduced order approach using proper orthogonal decomposition.

Proper orthogonal decomposition for optimality systems.

HJB-POD-based feedback design for the optimal control of evolution problems.
POD-based feedback control of the Burgers equation by solving the evolutionary HJB equation.

Output bounds for reduced-order approximations of elliptic partial differential equations.

*Projection-Based Model Reduction in the Context of Optimization with Implicit PDE Constraints.*
PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, TX.
in progress.

Reduced basis method for parametrized elliptic optimal control problems.
Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations.
MIT Pappalardo Graduate Monographs in Mechanical Engineering, Cambridge, MA.

Balancing Neumann-Neumann methods for incompressible Stokes equations.

Fundamentals of Numerical Reservoir Simulation.
Elsevier Science Inc., New York, NY, USA.

Springer, Cham.
A reduced-order approach for optimal control of fluids using proper orthogonal decomposition.

Model reduction for fluids, using balanced proper orthogonal decomposition.

Model reduction for compressible flows using POD and Galerkin projection.

Model order reduction by geometrical parametrization for shape optimization in computational fluid dynamics.

Balanced truncation model reduction for semidiscretized Stokes equation.
Literature XVI

Domain decomposition and model reduction of systems with local nonlinearities.

Discrete empirical interpolation method for finite element structural dynamics.

American Mathematical Society, Providence, RI.

Optimale Steuerung partieller Differentialgleichungen. Grundlagen, Optimalitätsbedingungen und ihre Anwendungen.
Vieweg Verlag, Braunschweig, second edition.
POD a-posteriori error estimates for linear-quadratic optimal control problems.

Optimality system POD and a-posteriori error analysis for linear-quadratic problems.
*Control Cybernet.*, 40(4):1109–1124.

A numerical study of an adjoint based method for reservoir optimization.
Master's thesis, Department of Computational and Applied Mathematics, Rice University, Houston, TX.

Accelerating optimization of parametric linear systems by model order reduction.

Progressive construction of a parametric reduced-order model for PDE-constrained optimization.
Technical report, Fachbereich Mathematik, Technische Universität Darmstadt,
Schloßgartenstr. 7, 64289 Darmstadt, Germany.