PDE-CONSTRAINED OPTIMIZATION UNDER UNCERTAINTY USING LOW-RANK METHODS

Peter Benner

Joint work with Sergey Dolgov (U Bath), Akwum Onwunta and Martin Stoll (both MPI DCTS).

August 26, 2017

Topical Lecture

S 15: Uncertainty Quantification

GAMM Jahrestagung Ilmenau@Weimar

March 06–10, 2017
Overview

1. Introduction
2. Unsteady Heat Equation
3. Unsteady Navier-Stokes Equations
4. Numerical experiments
5. Conclusions
Physical, biological, chemical, etc. processes involve uncertainties.
PDEs with stochastic coefficients for UQ

- Physical, biological, chemical, etc. processes involve uncertainties.
- Models of these processes should account for uncertainties.
### PDEs with stochastic coefficients for UQ

- Physical, biological, chemical, etc. processes involve uncertainties.
- Models of these processes should account for uncertainties.
- PDEs governing the processes can involve uncertain coefficients, or uncertain sources, or uncertain geometry.
Introduction

PDEs with stochastic coefficients for UQ

- Physical, biological, chemical, etc. processes involve uncertainties.
- Models of these processes should account for uncertainties.
- PDEs governing the processes can involve uncertain coefficients, or uncertain sources, or uncertain geometry.
- Uncertain parameters modeled as stochastic variables \(\sim\) random PDEs with uncertain inputs.

©Peter Benner, benner@mpi-magdeburg.mpg.de
PDEs with stochastic coefficients for UQ

- Physical, biological, chemical, etc. processes involve uncertainties.
- Models of these processes should account for uncertainties.
- PDEs governing the processes can involve uncertain coefficients, or uncertain sources, or uncertain geometry.
- Uncertain parameters modeled as stochastic variables $\rightsquigarrow$ random PDEs with uncertain inputs.

Uncertainty arises because

- available data are incomplete;
Introduction

PDEs with stochastic coefficients for UQ

- Physical, biological, chemical, etc. processes involve uncertainties.
- Models of these processes should account for uncertainties.
- PDEs governing the processes can involve uncertain coefficients, or uncertain sources, or uncertain geometry.
- Uncertain parameters modeled as stochastic variables \( \sim \) random PDEs with uncertain inputs.

Uncertainty arises because

- available data are incomplete;
- data are predictable, but difficult to measure, e.g., porosity above oil reservoirs;
PDEs with stochastic coefficients for UQ

- Physical, biological, chemical, etc. processes involve uncertainties.
- Models of these processes should account for uncertainties.
- PDEs governing the processes can involve uncertain coefficients, or uncertain sources, or uncertain geometry.
- Uncertain parameters modeled as stochastic variables \( \leadsto \) random PDEs with uncertain inputs.

Uncertainty arises because

- available data are incomplete;
- data are predictable, but difficult to measure, e.g., porosity above oil reservoirs;
- data are unpredictable, e.g., wind shear.
Curse of Dimensionality

Increase matrix size of discretized differential operator for $h \rightarrow \frac{h}{2}$ by factor $2^d$.

Rapid Increase of Dimensionality, called **Curse of Dimensionality** ($d > 3$).
### Motivation I: Low-Rank Solvers

#### Curse of Dimensionality

[Bellman '57]

Increase matrix size of discretized differential operator for $h \to \frac{h}{2}$ by factor $2^d$.

\[ \rightsquigarrow \text{Rapid Increase of Dimensionality}, \text{ called } \underline{\text{Curse of Dimensionality}} \ (d > 3). \]

Consider $-\Delta u = f$ in $[0, 1] \times [0, 1] \subset \mathbb{R}^2$, uniformly discretized as

\[
(I \otimes A + A \otimes I)x =: Ax = b \quad \iff \quad AX + XA^T = B
\]

with $x = \text{vec} (X)$ and $b = \text{vec} (B)$ with low-rank right hand side $B \approx b_1 b_2^T$. 

\[ \text{Low-rankness of } \tilde{X} = VW^T \approx X \text{ follows from properties of } A \text{ and } B, \text{ e.g., } \text{[Penzl '00, Grasedyck '04].} \]

We solve this using low-rank Krylov subspace solvers. These essentially require matrix-vector multiplication and vector computations.

\[ \text{Hence, } A \text{vec} (X_k) = A \text{vec} (V_k W_k^T) = \text{vec} \left( \begin{bmatrix} AV_k & V_k \end{bmatrix} \begin{bmatrix} W_k & AW_k \end{bmatrix}^T \right) \]

The rank of $\begin{bmatrix} AV_k & V_k \end{bmatrix}$ $\in \mathbb{R}^{n, 2r}$, $\begin{bmatrix} W_k & AW_k \end{bmatrix}$ $\in \mathbb{R}^{n, 2r}$ increases but can be controlled using truncation.

\[ \text{\[ Kressner/Tobler, B/Breiten, Savostyanov/Dolgov, \ldots. } \]

\[ \odot \text{Peter Benner, benner@mpi-magdeburg.mpg.de} \]

\[ \text{PDE-constrained optimization under uncertainty} \]

4/37
Motivation I: Low-Rank Solvers

Curse of Dimensionality

Increase matrix size of discretized differential operator for $h \to \frac{h}{2}$ by factor $2^d$.

Rapid Increase of Dimensionality, called **Curse of Dimensionality** ($d > 3$).

Consider $-\Delta u = f$ in $[0, 1] \times [0, 1] \subset \mathbb{R}^2$, uniformly discretized as

$$(I \otimes A + A \otimes I)x =: Ax = b \iff AX + XA^T = B$$

with $x = \text{vec}(X)$ and $b = \text{vec}(B)$ with low-rank right hand side $B \approx b_1b_2^T$.

Low-rankness of $\tilde{X} := VW^T \approx X$ follows from properties of $A$ and $B$, e.g.,

[Penzl '00, Grasedyck '04].
Motivation I: Low-Rank Solvers

Curse of Dimensionality

Increase matrix size of discretized differential operator for $h \to \frac{h}{2}$ by factor $2^d$. $\Rightarrow$ Rapid Increase of Dimensionality, called Curse of Dimensionality ($d > 3$).

Consider $-\Delta u = f$ in $[0, 1] \times [0, 1] \subset \mathbb{R}^2$, uniformly discretized as

$$(I \otimes A + A \otimes I) x =: Ax = b \iff AX + XA^T = B$$

with $x = \text{vec} (X)$ and $b = \text{vec} (B)$ with low-rank right hand side $B \approx b_1 b_2^T$.

- Low-rankness of $\tilde{X} := VW^T \approx X$ follows from properties of $A$ and $B$, e.g., [Penzl ’00, Grasedyck ’04].

- We solve this using low-rank Krylov subspace solvers$^1$. These essentially require matrix-vector multiplication and vector computations.

---

$^1$Recent work by H. Elman analyzes multigrid solver in this context.
Motivation I: Low-Rank Solvers

Curse of Dimensionality

Increase matrix size of discretized differential operator for $h \rightarrow \frac{h}{2}$ by factor $2^d$. 

$\leadsto$ Rapid Increase of Dimensionality, called Curse of Dimensionality ($d > 3$).

Consider $-\Delta u = f$ in $[0, 1] \times [0, 1] \subset \mathbb{R}^2$, uniformly discretized as 

$$(I \otimes A + A \otimes I) x =: Ax = b \iff AX + XA^T = B$$

with $x = \text{vec} (X)$ and $b = \text{vec} (B)$ with low-rank right hand side $B \approx b_1 b_2^T$.

- Low-rankness of $\tilde{X} := VW^T \approx X$ follows from properties of $A$ and $B$, e.g.,

$\text{[Penzl '00, Grasedyck '04]}$.

- We solve this using low-rank Krylov subspace solvers$^1$. These essentially require matrix-vector multiplication and vector computations.

- Hence, $\text{Avec} (X_k) = \text{Avec} (V_k W_k^T) = \text{vec} \left( [AV_k, V_k ] [W_k, AW_k ]^T \right)$

$^1$Recent work by H. Elman analyzes multigrid solver in this context.
Curse of Dimensionality

Increase matrix size of discretized differential operator for \( h \to \frac{h}{2} \) by factor \( 2^d \).

\[ \approx \text{Rapid Increase of Dimensionality}, \text{ called } \textbf{Curse of Dimensionality } (d > 3). \]

Consider \(-\Delta u = f\) in \([0, 1] \times [0, 1] \subset \mathbb{R}^2\), uniformly discretized as

\[
(I \otimes A + A \otimes I)x =: Ax = b \iff AX + XA^T = B
\]

with \( x = \text{vec} (X) \) and \( b = \text{vec} (B) \) with low-rank right hand side \( B \approx b_1 b_2^T \).

- Low-rankness of \( \tilde{X} := VW^T \approx X \) follows from properties of \( A \) and \( B \), e.g., \([\text{Penzl '00, Grasedyck '04}]\).

- We solve this using low-rank Krylov subspace solvers\(^1\). These essentially require matrix-vector multiplication and vector computations.

- Hence,

\[
\text{Avec} (X_k) = \text{Avec} (V_k W_k^T) = \text{vec} \left( [AV_k, V_k] [W_k, AW_k]^T \right)
\]

- The rank of \([AV_k, V_k] \in \mathbb{R}^{n,2r}, [W_k, AW_k] \in \mathbb{R}^{n_t,2r}\) increases but can be controlled using truncation. \( \approx \text{Low-rank Krylov subspace solvers.} \)

\(^1\)Recent work by H. Elman analyzes multigrid solver in this context.
We consider the problem:

$$\min_{y \in \mathcal{Y}, u \in \mathcal{U}} \mathcal{J}(y, u) \quad \text{subject to} \quad c(y, u) = 0,$$

where

- $c(y, u) = 0$ represents a nonlinear PDE with uncertain coefficient(s).
- The state $y$ and control $u$ are random fields.
- The cost functional $\mathcal{J}$ is a real-valued differentiable functional on $\mathcal{Y} \times \mathcal{U}$. 
Curse of Dimensionality

Increase matrix size of discretized differential operator for $h \rightarrow \frac{h}{2}$ by factor $2^d$.

$\leadsto$ Rapid Increase of Dimensionality, called **Curse of Dimensionality** ($d > 3$).

Goal of this talk

Apply low-rank (Krylov) solvers to discrete optimality systems resulting from **PDE-constrained optimization problems under uncertainty**, and go one step further applying low-rank tensor (instead of matrix) techniques.
This Talk

Curse of Dimensionality [Bellman ’57]

Increase matrix size of discretized differential operator for \( h \to \frac{h}{2} \) by factor \( 2^d \).

\( \Rightarrow \) Rapid Increase of Dimensionality, called **Curse of Dimensionality** \((d > 3)\).

Goal of this talk

Apply low-rank (Krylov) solvers to discrete optimality systems resulting from

**PDE-constrained optimization problems under uncertainty**, and go one step further applying low-rank tensor (instead of matrix) techniques.

Take home message

Biggest problem solved so far has \( n = 1.29 \cdot 10^{15} \) unknowns (KKT system for unsteady incompressible Navier-Stokes control problem with uncertain viscosity).
This Talk

Curse of Dimensionality

Increase matrix size of discretized differential operator for $h \rightarrow \frac{h}{2}$ by factor $2^d$.

$\Downarrow$ Rapid Increase of Dimensionality, called **Curse of Dimensionality** ($d > 3$).

Goal of this talk

Apply low-rank (Krylov) solvers to discrete optimality systems resulting from

**PDE-constrained optimization problems under uncertainty,**

and go one step further applying low-rank tensor (instead of matrix) techniques.

Take home message

Biggest problem solved so far has $n = 1.29 \cdot 10^{15}$ unknowns (KKT system for unsteady incompressible Navier-Stokes control problem with uncertain viscosity).

Would require $\approx 10$ **petabytes (PB) = 10,000 TB** to store the solution vector!
This Talk

Curse of Dimensionality

Increase matrix size of discretized differential operator for \( h \to \frac{h}{2} \) by factor \( 2^d \).

\[ \leadsto \text{Rapid Increase of Dimensionality, called Curse of Dimensionality } (d > 3). \]

Goal of this talk

Apply low-rank (Krylov) solvers to discrete optimality systems resulting from PDE-constrained optimization problems under uncertainty, and go one step further applying low-rank tensor (instead of matrix) techniques.

Take home message

Biggest problem solved so far has \( n = 1.29 \cdot 10^{15} \) unknowns (KKT system for unsteady incompressible Navier-Stokes control problem with uncertain viscosity).

Would require \( \approx 10 \text{ petabytes (PB)} = 10,000 \text{ TB} \) to store the solution vector!

Using low-rank tensor techniques, we need \( \approx 7 \cdot 10^7 \text{ bytes} = 70 \text{ GB} \) to solve the KKT system in MATLAB in less than one hour!
Consider the optimization problem

\[ J(t, y, u) = \frac{1}{2} \| y - \bar{y} \|_{L^2(0, T; \mathcal{D})}^2 + \frac{\alpha}{2} \| \text{std}(y) \|_{L^2(0, T; \mathcal{D})}^2 + \frac{\beta}{2} \| u \|_{L^2(0, T; \mathcal{D})}^2 \]

subject, \( \mathbb{P} \)-almost surely, to

\[
\begin{aligned}
\frac{\partial y(t, x, \omega)}{\partial t} - \nabla \cdot (a(x, \omega) \nabla y(t, x, \omega)) &= u(t, x, \omega), \quad \text{in } (0, T] \times \mathcal{D} \times \Omega, \\
y(t, x, \omega) &= 0, \quad \text{on } (0, T] \times \partial \mathcal{D} \times \Omega, \\
y(0, x, \omega) &= y_0, \quad \text{in } \mathcal{D} \times \Omega,
\end{aligned}
\]

where

- any \( z : \mathcal{D} \times \Omega \to \mathbb{R}, \ z(x, \cdot) \) is a random variable defined on the complete probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) for each \( x \in \mathcal{D} \),
- \( a(x, \omega) \) is assumed to be uniformly positive in \( \mathcal{D} \times \Omega \).
Discretization

We discretize and then optimize the stochastic control problem.

- Under finite noise assumption we can use $N$-term (truncated) Karhunen-Loève expansion (KLE)

$$a \equiv a(x, \omega) \approx a_N(x, \xi(\omega)) = a(x, \xi_1(\omega), \xi_2(\omega), \ldots, \xi_N(\omega)).$$

- Assuming a known continuous covariance $C_a(x, y)$, we get the KLE

$$a_N(x, \xi(\omega)) = \mathbb{E}[a](x) + \sigma_a \sum_{i=1}^{N} \sqrt{\lambda_i} \varphi_i(x) \xi_i(\omega),$$

where $(\lambda_i, \varphi_i)$ are the dominant eigenpairs of $C_a$.

- Doob-Dynkin Lemma admits same parametrization for solution $y$.

- Use linear finite elements for the spatial discretization and implicit Euler in time.

This is used within a stochastic Galerkin FEM (SGFEM) approach.
Overview of UQ techniques

**Monte Carlo Sampling**

Given a sample \( \{\omega_i\}_{i=1}^M \in \Omega \), we estimate desired statistical quantities using the law of large numbers.

- **Pros**: Simple, code reusability, etc.
- **Cons**: Slow convergence \( \mathcal{O}(1/\sqrt{M}) \).
Discretization

Overview of UQ techniques

- **Monte Carlo Sampling**
  Given a sample \( \{ \omega_i \}_{i=1}^{M} \in \Omega \), we estimate desired statistical quantities using the law of large numbers.
  - **Pros**: Simple, code reusability, etc.
  - **Cons**: Slow convergence \( O(1/\sqrt{M}) \).

- **Parametric**
  Expand \( y(x, \xi) = \sum_i y_i(x) H_i(\xi) \).
Overview of UQ techniques

- **Monte Carlo Sampling**
  Given a sample \( \{\omega_i\}_{i=1}^M \in \Omega \), we estimate desired statistical quantities using the law of large numbers.
  - **Pros**: Simple, code reusability, etc.
  - **Cons**: Slow convergence \( \mathcal{O}(1/\sqrt{M}) \).

- **Parametric**
  Expand \( y(x, \xi) = \sum_i y_i(x)H_i(\xi) \).
  - **Stochastic collocation**.
    Compute \( y_i \) for a set of interpolation points \( \xi_i \), then connect the realizations with Lagrangian basis functions \( H_i := L_i \).
Discretization

Overview of UQ techniques

- **Monte Carlo Sampling**
  Given a sample \( \{ \omega_i \}_{i=1}^M \in \Omega \), we estimate desired statistical quantities using the law of large numbers.
  - **Pros**: Simple, code reusability, etc.
  - **Cons**: Slow convergence \( \mathcal{O}(1/\sqrt{M}) \).

- **Parametric**
  Expand \( y(x, \xi) = \sum_i y_i(x) H_i(\xi) \).
  - **Stochastic collocation**.
    Compute \( y_i \) for a set of interpolation points \( \xi_i \), then connect the realizations with Lagrangian basis functions \( H_i := L_i \).
  - **Stochastic Galerkin (Generalized Polynomial Chaos)**.
    Compute \( y_i \) projecting the equation onto a subspace spanned by orthogonal polynomials \( H_i := \psi_i \).
    - \( \xi \) are uniform random variables \( \rightarrow \psi_i \) Legendre polynomials.
    - \( \xi \) are Gaussian random variables \( \rightarrow \psi_i \) Hermite polynomials.
First order conditions of the discrete heat control problem are given by the KKT system

\[
\begin{bmatrix}
\tau M_1 & 0 & -K_t^T \\
0 & \beta \tau M_2 & \tau N^T \\
-K_t & \tau N & 0
\end{bmatrix}
\begin{bmatrix}
y \\
u \\
f
\end{bmatrix}
=
\begin{bmatrix}
\tau M_a y \\
0 \\
d
\end{bmatrix},
\]

\[M_1 = D \otimes G_\alpha \otimes M = D \otimes M_\alpha,\]

\[K_t = (I_{n_t} \otimes L) + (C \otimes M) = I_{n_t} \otimes \left[ \sum_{i=0}^{N} G_i \otimes \hat{K}_i \right] + (C \otimes G_0 \otimes M),\]

\[N = I_{n_t} \otimes G_0 \otimes M, \quad M_2 = D \otimes G_0 \otimes M\]

and

\[
\begin{cases}
G_0 = \text{diag} \left( \langle \psi_0^2 \rangle, \langle \psi_1^2 \rangle, \ldots, \langle \psi_{P-1}^2 \rangle \right), \\
G_i(j, k) = \langle \xi_i \psi_j \psi_k \rangle, \quad i = 1, \ldots, N,
\end{cases}
\]

with \( \psi \) the orthogonal (Legendre, Hermite, \ldots) polynomials and \( K_i \) are stiffness matrices involving terms from the KLE.
Solving the KKT System

This system is a saddle point system

\[
\begin{bmatrix}
A & B^T \\
B & 0
\end{bmatrix}
\] with preconditioner

\[
\begin{bmatrix}
\hat{A} & 0 \\
0 & \hat{S}
\end{bmatrix}.
\]

Lots of pioneering work by Elman, Ernst, Ullmann, Powell, Silvester, ...

\[\text{Theorem (Onwunta/Stoll '16)}\]

Let \( \alpha \in [0, +\infty) \).

Then, the eigenvalues of \( S^{-\frac{1}{2}} S \) satisfy

\[
\lambda(S^{-\frac{1}{2}} S) \subset \left[ \frac{1}{2}(1 + \alpha \beta N), 1 \right],
\]

\( \forall \alpha < \left( \sqrt{\kappa(K)} + 1 \right) \sqrt{\kappa(K)} - 1 \)²⁻¹,

where

\[
K = \sum_{i=0}^{N} G_i \otimes K_i
\]

and

\[
S = \tau(Kt + \tau \sqrt{1 + \alpha \beta N}) M^{-1} \left( Kt + \tau \sqrt{1 + \alpha \beta N} \right)^T.
\]
Solving the KKT System

This system is a saddle point system

\[
\begin{bmatrix}
A & B^T \\
B & 0
\end{bmatrix}
\]

with preconditioner \[
\begin{bmatrix}
\hat{A} & 0 \\
0 & \hat{S}
\end{bmatrix}.
\]

Lots of pioneering work by Elman, Ernst, Ullmann, Powell, Silvester, ...

Theorem ([B./Onwunta/Stoll ’16])

Let \( \alpha \in [0, +\infty) \). Then, the eigenvalues of \( S_2^{-1}S \) satisfy

\[
\lambda(S_2^{-1}S) \subset \left[ \frac{1}{2(1 + \alpha)}, 1 \right], \quad \forall \alpha < \left( \frac{\sqrt{\kappa(K)} + 1}{\sqrt{\kappa(K)} - 1} \right)^2 - 1,
\]

where \( \kappa = \sum_{i=0}^{N} G_i \otimes K_i \) and

\[
S_2 = \frac{1}{\tau} \left( \kappa_t + \tau \sqrt{\frac{1 + \alpha}{\beta} N} \right) M_1^{-1} \left( \kappa_t + \tau \sqrt{\frac{1 + \alpha}{\beta} N} \right)^T.
\]
The dimensionality of the saddle point system is vast ⇒ use tensor structure and low tensor ranks.

Use tensor train format and represent the tensor objects as

\[
y(i_1, \ldots, i_d) = \sum_{\alpha_1 \ldots \alpha_{d-1} = 1} y^{(1)}(i_1) y^{(2)}_{\alpha_1,\alpha_2}(i_2) \cdots y^{(d-1)}_{\alpha_{d-2},\alpha_{d-1}}(i_{d-1}) y^{(d)}_{\alpha_{d-1}}(i_d),
\]

and

\[
A(i_1 \cdots i_d, j_1 \cdots j_d) \approx \sum_{\beta_1 \ldots \beta_{d-1} = 1} A^{(1)}_{\beta_1}(i_1,j_1) A^{(2)}_{\beta_1,\beta_2}(i_2,j_2) \cdots A^{(d)}_{\beta_{d-1}}(i_d,j_d).
\]
### Numerical Results

#### Mean-Based Preconditioned TT-MinRes

<table>
<thead>
<tr>
<th>TT-MINRES</th>
<th># iter (t)</th>
<th># iter (t)</th>
<th># iter (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_t$</td>
<td>$2^5$</td>
<td>$2^6$</td>
<td>$2^8$</td>
</tr>
<tr>
<td>$\dim(\mathcal{A}) = 3JPn_t$</td>
<td>10,671,360</td>
<td>21,342,720</td>
<td>85,370,880</td>
</tr>
<tr>
<td>$\alpha = 1$, $\text{tol} = 10^{-3}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta = 10^{-5}$</td>
<td>6 (285.5)</td>
<td>6 (300.0)</td>
<td>8 (372.2)</td>
</tr>
<tr>
<td>$\beta = 10^{-6}$</td>
<td>4 (77.6)</td>
<td>4 (130.9)</td>
<td>4 (126.7)</td>
</tr>
<tr>
<td>$\beta = 10^{-8}$</td>
<td>4 (56.7)</td>
<td>4 (59.4)</td>
<td>4 (64.9)</td>
</tr>
<tr>
<td>$\alpha = 0$, $\text{tol} = 10^{-3}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta = 10^{-5}$</td>
<td>4 (207.3)</td>
<td>6 (366.5)</td>
<td>6 (229.5)</td>
</tr>
<tr>
<td>$\beta = 10^{-6}$</td>
<td>4 (153.9)</td>
<td>4 (158.3)</td>
<td>4 (172.0)</td>
</tr>
<tr>
<td>$\beta = 10^{-8}$</td>
<td>2 (35.2)</td>
<td>2 (37.8)</td>
<td>2 (40.0)</td>
</tr>
</tbody>
</table>
We model this as a **boundary control problem**.

Our constraint $c(y, u) = 0$ is given by the unsteady incompressible Navier-Stokes equations with **uncertain viscosity** $\nu := \nu(\omega)$. 
Minimize:

$$\mathcal{J}(v, u) = \frac{1}{2} \| \text{curl} v \|^2_{L^2(0, T; \mathcal{D}) \otimes L^2(\Omega)} + \frac{\beta}{2} \| u \|^2_{L^2(0, T; \mathcal{D}) \otimes L^2(\Omega)}$$  \hspace{1cm} (1)

subject to

$$\frac{\partial v}{\partial t} - \nu \Delta v + (v \cdot \nabla) v + \nabla p = 0, \quad \text{in} \quad \mathcal{D},$$

$$-\nabla \cdot v = 0, \quad \text{in} \quad \mathcal{D},$$

$$v = \theta, \quad \text{on} \quad \Gamma_{in},$$

$$v = 0, \quad \text{on} \quad \Gamma_{wall},$$

$$\frac{\partial v}{\partial n} = u, \quad \text{on} \quad \Gamma_{c},$$

$$\frac{\partial v}{\partial n} = 0, \quad \text{on} \quad \Gamma_{out},$$

$$v(\cdot, 0, \cdot) = v_0, \quad \text{in} \quad \mathcal{D}.$$  \hspace{1cm} (2)
We assume

- $\nu(\omega) = \nu_0 + \nu_1 \xi(\omega)$, $\nu_0, \nu_1 \in \mathbb{R}^+$, $\xi \sim \mathcal{U}(-1, 1)$.
- $\mathbb{P}(\omega \in \Omega : \nu(\omega) \in [\nu_{\text{min}}, \nu_{\text{max}}]) = 1$, for some $0 < \nu_{\text{min}} < \nu_{\text{max}} < +\infty$.
- $\Rightarrow$ velocity $\nu$, control $u$ and pressure $p$ are random fields on $L^2(\Omega)$.
- $L^2(\Omega) := L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space.
- $L^2(0, T; \mathcal{D}) := L^2(\mathcal{D}) \times L^2(\mathcal{T})$. 
We assume

- \( \nu(\omega) = \nu_0 + \nu_1 \xi(\omega) \), \( \nu_0, \nu_1 \in \mathbb{R}^+ \), \( \xi \sim \mathcal{U}(-1, 1) \).
- \( \mathbb{P}(\omega \in \Omega : \nu(\omega) \in [\nu_{\min}, \nu_{\max}]) = 1 \), for some \( 0 < \nu_{\min} < \nu_{\max} < +\infty \).
- \( \Rightarrow \) velocity \( \nu \), control \( u \) and pressure \( p \) are random fields on \( L^2(\Omega) \).
- \( L^2(\Omega) := L^2(\Omega, \mathcal{F}, \mathbb{P}) \) is a complete probability space.
- \( L^2(0, T; \mathcal{D}) := L^2(\mathcal{D}) \times L^2(\mathcal{T}) \).

Computational challenges

- Nonlinearity (due to the nonlinear convection term \( (\nu \cdot \nabla)\nu \)).
- Uncertainty (due to random \( \nu(\omega) \)).
- High dimensionality (of the resulting linear/optimality systems).
OTD Strategy and Picard (Oseen) Iteration

state equation
\[ \nu_t - \nu \Delta v + (\bar{v} \cdot \nabla) v + \nabla p = 0 \]
\[ \nabla \cdot v = 0 + \text{boundary conditions} \]

adjoint equation
\[ -\chi_t - \Delta \chi - (\bar{v} \cdot \nabla) \chi + (\nabla \bar{v})^T \chi + \nabla \mu = -\text{curl}^2 v \]
\[ \nabla \cdot \chi = 0 \]
\[ \text{on } \Gamma_{\text{wall}} \cup \Gamma_{\text{in}} : \quad \chi = 0 \]
\[ \text{on } \Gamma_{\text{out}} \cup \Gamma_{\text{c}} : \quad \frac{\partial \chi}{\partial n} = 0 \]
\[ \chi(\cdot, T, \cdot) = 0 \]

gradient equation
\[ \beta u + \chi|_{\Gamma_c} = 0. \]
Optimality System in Function Space: Optimize-then-Discretize (OTD)

OTD Strategy and Picard (Oseen) Iteration

state equation
\[ v_t - \nu \Delta v + (\overline{v} \cdot \nabla) v + \nabla p = 0 \]
\[ \nabla \cdot v = 0 + \text{boundary conditions} \]

adjoint equation
\[ -\chi_t - \Delta \chi - (\overline{v} \cdot \nabla) \chi + (\nabla \overline{v})^T \chi + \nabla \mu = -\text{curl}^2 v \]
\[ \nabla \cdot \chi = 0 \]
on \( \Gamma_{\text{wall}} \cup \Gamma_{\text{in}} \) : \( \chi = 0 \)

on \( \Gamma_{\text{out}} \cup \Gamma_{\text{c}} \) :
\[ \frac{\partial \chi}{\partial n} = 0 \]
\[ \chi(\cdot, T, \cdot) = 0 \]

gradient equation
\[ \beta u + \chi|_{\Gamma_{\text{c}}} = 0. \]

- \( \overline{v} \) denotes the velocity from the previous Oseen iteration.
- Having solved this system, we update \( \overline{v} = v \) until convergence.
Velocity $v$ and control $u$ are of the form

$$z(t, x, \omega) = \sum_{k=0}^{P-1} \sum_{j=1}^{J_v} z_{jk}(t) \phi_j(x) \psi_k(\xi) = \sum_{k=0}^{P-1} z_k(t, x) \psi_k(\xi).$$

Pressure $p$ is of the form

$$p(t, x, \omega) = \sum_{k=0}^{P-1} \sum_{j=1}^{J_p} p_{jk}(t) \tilde{\phi}_j(x) \psi_k(\xi) = \sum_{k=0}^{P-1} p_k(t, x) \psi_k(\xi).$$

Here,

- $\{\phi_j\}_{j=1}^{J_v}$ and $\{\tilde{\phi}_j\}_{j=1}^{J_p}$ are Q2–Q1 finite elements;
- $\{\psi_k\}_{k=0}^{P-1}$ are Legendre polynomials.

Implicit Euler/dG(0) used for temporal discretization.
Linearization and SGFEM discretization yields the following saddle point system

\[
\begin{bmatrix}
M_y & 0 & L^* \\
0 & M_u & N^T \\
L & N & 0
\end{bmatrix}
\begin{bmatrix}
y \\
u \\
\lambda
\end{bmatrix}
= \begin{bmatrix}
f \\
0 \\
g
\end{bmatrix}.
\]

Each of the block matrices in \( A \) is of the form

\[
\sum_{\alpha=1}^{R} X_\alpha \otimes Y_\alpha \otimes Z_\alpha,
\]

corresponding to temporal, stochastic, and spatial discretizations.
Linearization and SGFEM discretization yields the following saddle point system

$$
A \begin{bmatrix} y \\ u \\ \lambda \end{bmatrix} = \begin{bmatrix} f \\ 0 \\ b \end{bmatrix},
$$

where

$$
A = \begin{bmatrix} M_y & 0 & L^* \\ 0 & M_u & N^T \\ L & N & 0 \end{bmatrix}.
$$

Each of the block matrices in $A$ is of the form

$$
\sum_{\alpha=1}^{R} X_\alpha \otimes Y_\alpha \otimes Z_\alpha,
$$

corresponding to temporal, stochastic, and spatial discretizations.

Size: $\sim 3n_t P (J_v + J_p)$, e.g., for $P = 10$, $n_t = 2^{10}$, $J \approx 10^5 \rightarrow \approx 10^9$ unknowns!
Tensor Techniques

Separation of variables and low-rank approximation

Approximate: \( \mathbf{x}(i_1, \ldots, i_d) \approx \sum_{\alpha} \mathbf{x}_\alpha^{(1)}(i_1) \mathbf{x}_\alpha^{(2)}(i_2) \cdots \mathbf{x}_\alpha^{(d)}(i_d) \).

Goals:
- Store and manipulate \( \mathbf{x} \) \( \mathcal{O}(dn) \) cost instead of \( \mathcal{O}(n^d) \).
- Solve equations \( A\mathbf{x} = b \) \( \mathcal{O}(dn^2) \) cost instead of \( \mathcal{O}(n^{2d}) \).
Discrete separation of variables:

\[
\begin{bmatrix}
  x_{1,1} & \cdots & x_{1,n} \\
  \vdots & \ddots & \vdots \\
  x_{n,1} & \cdots & x_{n,n}
\end{bmatrix} = \sum_{\alpha=1}^{r} \begin{bmatrix}
  v_{1,\alpha} \\
  \vdots \\
  v_{n,\alpha}
\end{bmatrix} \begin{bmatrix}
  w_{\alpha,1} & \cdots & w_{\alpha,n}
\end{bmatrix} + O(\varepsilon).
\]

Diagrams:

- Rank \( r \ll n \).
- \( \text{mem}(v) + \text{mem}(w) = 2nr \ll n^2 = \text{mem}(x) \).
- **Singular Value Decomposition (SVD)**
  \[\Rightarrow \varepsilon(r) \text{ optimal w.r.t. spectral/Frobenius norm.}\]
Matrix Product States/Tensor Train (TT) format

For indices

\[ i_p \ldots i_q = (i_p-1)n_{p+1} \cdots n_q + (i_{p+1}-1)n_{p+2} \cdots n_q + \cdots + (i_{q-1}-1)n_q + i_q, \]

the TT format can be expressed as

\[
\sum_{\alpha=1}^{r} \mathbf{x}^{(1)}_{\alpha_1}(i_1) \cdot \mathbf{x}^{(2)}_{\alpha_1,\alpha_2}(i_2) \cdot \mathbf{x}^{(3)}_{\alpha_2,\alpha_3}(i_3) \cdots \mathbf{x}^{(d)}_{\alpha_{d-1},\alpha_d}(i_d)
\]

or

\[
\mathbf{x}(i_1 \ldots i_d) = \mathbf{x}^{(1)}(i_1) \cdots \mathbf{x}^{(d)}(i_d), \quad \mathbf{x}^{(k)}(i_k) \in \mathbb{R}^{r_{k-1} \times r_k}.
\]

or

\[
\begin{array}{cccccccc}
\mathbf{x}^{(1)} & \alpha_1 & \mathbf{x}^{(2)} & \alpha_2 & \cdots & \alpha_{k-2} & \mathbf{x}^{(k-1)} & \alpha_{k-1} & \mathbf{x}^{(k)} & \alpha_k & \mathbf{x}^{(k+1)} & \alpha_{k+1} & \cdots & \alpha_{d-1} & \mathbf{x}^{(d)} \\
& i_1 & & i_2 & & \cdots & & i_{k-1} & & i_k & & i_{k+1} & & \cdots & & i_d \\
\end{array}
\]
Overloading Tensor Operations

Always work with factors $x^{(k)} \in \mathbb{R}^{r_{k-1} \times n_k \times r_k}$ instead of full tensors.

- Sum $z = x + y \implies$ increase of tensor rank $r_z = r_x + r_y$.
- TT format for a high-dimensional operator

$$A(i_1 \ldots i_d, j_1 \ldots j_d) = A^{(1)}(i_1, j_1) \cdots A^{(d)}(i_d, j_d)$$

- Matrix-vector multiplication $y = Ax; \implies$ tensor rank $r_y = r_A \cdot r_x$.
- Additions and multiplications increase TT ranks.
- Decrease ranks quasi-optimally via QR and SVD.
Central Question

How to solve $Ax = b$?

Data are given in TT format:

$A(i, j) = A(1)(i_1, j_1) \cdots A(d)(i_d, j_d)$.

$b(i) = b(1)(i_1) \cdots b(d)(i_d)$.

Seek the solution in the same format:

$x(i) = x(1)(i_1) \cdots x(d)(i_d)$.

Use a new block-variant of Alternating Least Squares in a new block TT format to overcome difficulties with indefiniteness of KKT system matrix.
Central Question

How to solve $Ax = b$?

Data are given in TT format:

- $A(i, j) = A^{(1)}(i_1, j_1) \cdots A^{(d)}(i_d, j_d)$.
- $b(i) = b^{(1)}(i_1) \cdots b^{(d)}(i_d)$.

Seek the solution in the same format:

- $x(i) = x^{(1)}(i_1) \cdots x^{(d)}(i_d)$.
Central Question

How to solve $Ax = b$?

Data are given in TT format:

- $A(i,j) = A^{(1)}(i_1,j_1) \cdots A^{(d)}(i_d,j_d)$.
- $b(i) = b^{(1)}(i_1) \cdots b^{(d)}(i_d)$.

Seek the solution in the same format:

- $x(i) = x^{(1)}(i_1) \cdots x^{(d)}(i_d)$.

Use a new block-variant of Alternating Least Squares in a new block TT format to overcome difficulties with indefiniteness of KKT system matrix.
Central Question

How to solve $Ax = b$?

Data are given in TT format:

- $A(i, j) = A^{(1)}(i_1, j_1) \cdots A^{(d)}(i_d, j_d)$.
- $b(i) = b^{(1)}(i_1) \cdots b^{(d)}(i_d)$.

Seek the solution in the same format:

- $x(i) = x^{(1)}(i_1) \cdots x^{(d)}(i_d)$.

Use a new block-variant of *Alternating Least Squares* in a new block TT format to overcome difficulties with indefiniteness of KKT system matrix.
If $A = A^\top > 0$: minimize $J(x) = x^\top Ax - 2x^\top b$.

**Alternating Least Squares (ALS):**

- replace $\min_x J(x)$ by iteration
- for $k = 1, \ldots, d$, solve $\min_{x^{(k)}} J\left(x^{(1)}(i_1) \cdots x^{(k)}(i_k) \cdots x^{(d)}(i_d)\right)$. (all other blocks are fixed)
ALS for $d = 3$

1. $\hat{x}^{(1)} = \arg\min_{x^{(1)}} J(x^{(1)}(i_1)x^{(2)}(i_2)x^{(3)}(i_3))$
ALS for $d = 3$

1. $\hat{x}^{(1)} = \arg\min_{x^{(1)}} J(x^{(1)}(i_1)x^{(2)}(i_2)x^{(3)}(i_3))$
2. $\hat{x}^{(2)} = \arg\min_{x^{(2)}} J(\hat{x}^{(1)}(i_1)x^{(2)}(i_2)x^{(3)}(i_3))$
ALS for $d = 3$

1. $\hat{x}^{(1)} = \arg \min_{x^{(1)}} J(\hat{x}^{(1)}(i_1)x^{(2)}(i_2)x^{(3)}(i_3))$
2. $\hat{x}^{(2)} = \arg \min_{x^{(2)}} J(\hat{x}^{(1)}(i_1)x^{(2)}(i_2)x^{(3)}(i_3))$
3. $\hat{x}^{(3)} = \arg \min_{x^{(3)}} J(\hat{x}^{(1)}(i_1)\hat{x}^{(2)}(i_2)x^{(3)}(i_3))$

4. $x^{(2)} = \arg \min_{x^{(2)}} J(\hat{x}^{(1)}(i_1)\hat{x}^{(2)}(i_2)x^{(3)}(i_3))$

5. Repeat 1.–4. until convergence
ALS for $d = 3$

1. $\hat{x}^{(1)} = \arg\min_{x^{(1)}} J (x^{(1)}(i_1)x^{(2)}(i_2)x^{(3)}(i_3))$
2. $\hat{x}^{(2)} = \arg\min_{x^{(2)}} J (\hat{x}^{(1)}(i_1)x^{(2)}(i_2)x^{(3)}(i_3))$
3. $\hat{x}^{(3)} = \arg\min_{x^{(3)}} J (\hat{x}^{(1)}(i_1)\hat{x}^{(2)}(i_2)x^{(3)}(i_3))$
4. $x^{(2)} = \arg\min_{x^{(2)}} J (\hat{x}^{(1)}(i_1)x^{(2)}(i_2)\hat{x}^{(3)}(i_3))$

5. repeat 1.–4. until convergence
1. \( \hat{x}^{(1)} = \arg\min_{x^{(1)}} J (x^{(1)}(i_1)x^{(2)}(i_2)x^{(3)}(i_3)) \)
2. \( \hat{x}^{(2)} = \arg\min_{x^{(2)}} J (\hat{x}^{(1)}(i_1)x^{(2)}(i_2)x^{(3)}(i_3)) \)
3. \( \hat{x}^{(3)} = \arg\min_{x^{(3)}} J (\hat{x}^{(1)}(i_1)\hat{x}^{(2)}(i_2)x^{(3)}(i_3)) \)
4. \( x^{(2)} = \arg\min_{x^{(2)}} J (\hat{x}^{(1)}(i_1)x^{(2)}(i_2)\hat{x}^{(3)}(i_3)) \)
5. repeat 1.–4. until convergence
If we differentiate $J$ w.r.t. TT blocks, we see that...

- ...each step means solving a *Galerkin linear system*

\[
\left( X_{\neq k}^\top A X_{\neq k} \right) \hat{x}^{(k)} = \left( X_{\neq k}^\top b \right) \in \mathbb{R}^{nr^2}.
\]

- $X_{\neq k} = \text{TT} \left( \hat{x}^{(1)} \cdots \hat{x}^{(k-1)} \right) \otimes I_{n \times n} \otimes \text{TT} \left( x^{(k+1)} \cdots x^{(d)} \right)$.
If we differentiate $J$ w.r.t. TT blocks, we see that...

- ...each step means solving a *Galerkin linear system*

$$
\left( X_{\neq k}^{\top} A X_{\neq k} \right) \hat{x}^{(k)} = \left( X_{\neq k}^{\top} b \right) \in \mathbb{R}^{nr^2}.
$$

- $X_{\neq k} = \text{TT} \left( \hat{x}^{(1)} \cdots \hat{x}^{(k-1)} \right) \otimes I_{n \times n} \otimes \text{TT} \left( x^{(k+1)} \cdots x^{(d)} \right)$.  

**Properties of ALS include:**

+ Effectively 1D complexity in a prescribed format.
- Tensor format (ranks) is fixed and cannot be adapted.
- Convergence may be very slow, stagnation is likely.
ALS: Getting rid of “–”

- **Density Matrix Renormalization Group (DMRG)**  
  - updates *two* blocks $x^{(k)} x^{(k+1)}$ *simultaneously*.  
  [White ’92]

- **Alternating Minimal Energy (AMEn)**  
  - *augments* $x^{(k)}$ by a TT block of the *residual* $z^{(k)}$.  
  [Dolgov/Savostyanov ’13]
### ALS: Getting rid of “–”

- **Density Matrix Renormalization Group (DMRG)**  
  - updates two blocks $x^{(k)}x^{(k+1)}$ *simultaneously.*  
  
- **Alternating Minimal Energy (AMEn)**  
  - augments $x^{(k)}$ by a TT block of the residual $z^{(k)}$.

**But... what about saddle point systems $A$?**

- Recall our KKT system:

$$
\begin{bmatrix}
M_y & 0 & L^* \\
0 & M_u & N^T \\
L & N & 0
\end{bmatrix}
\begin{bmatrix}
y \\
u \\
\lambda
\end{bmatrix}
=
\begin{bmatrix}
f \\
g
\end{bmatrix}.
$$

The whole matrix is *indefinite* ⇒ $X_{\neq k}^TAX_{\neq k}$ can be degenerate.
Work-around: Block TT representation

\[
\begin{bmatrix}
y \\ u \\ \lambda
\end{bmatrix} = \mathbf{x}_{\alpha_1} \otimes \cdots \otimes \begin{bmatrix}
y^{(k)}_{\alpha_{k-1},\alpha_k} \\ u^{(k)}_{\alpha_{k-1},\alpha_k} \\ \lambda^{(k)}_{\alpha_{k-1},\alpha_k}
\end{bmatrix} \otimes \cdots \otimes \mathbf{x}_{\alpha_{d-1}}.
\]

\(X_{\neq k}\) is the same for \(y, u, \lambda\).
Work-around: Block TT representation

\[
\begin{bmatrix}
y \\ u \\ \lambda
\end{bmatrix} = \mathbf{x}_{\alpha_1}^{(1)} \otimes \cdots \otimes \mathbf{x}_{\alpha_k}^{(k)} \otimes \cdots \otimes \mathbf{x}_{\alpha_{d-1}}^{(d)}.
\]

\(X_{\neq k}\) is the same for \(y, u, \lambda\).

Project each submatrix:

\[
\begin{bmatrix}
\hat{M}_y & 0 & \hat{L}^* \\
0 & \hat{M}_u & \hat{N}^T \\
\hat{L} & \hat{N} & 0
\end{bmatrix}
\begin{bmatrix}
y^{(k)} \\
u^{(k)} \\
\lambda^{(k)}
\end{bmatrix} =
\begin{bmatrix}
\hat{f} \\
0 \\
\hat{g}
\end{bmatrix},
\]

\(\mathbf{X}_{\neq k}^T(\cdot)\mathbf{X}_{\neq k}\)
Numerical experiments

Vary one of the default parameters:

- TT truncation tolerance $\varepsilon = 10^{-4}$,
- mean viscosity $\nu_0 = 1/20$,
- uncertainty $\nu_1 = 1/80$,
- regularization/penalty parameter $\beta = 10^{-1}$,
- number of time steps: $n_t = 2^{10}$,
- time horizon $T = 30$,
- spatial grid size $h = 1/4 \leadsto J = 2488$,
- max. degree of Legendre polynomials: $P = 8$.

Solve projected linear systems using block-preconditioned GMRES using efficient approximation of Schur complement [B/Dolgov/Onwunta/Stoll ’16a].
Varying regularization $\beta$ (left) and time $T$ (right)
Varying spatial $h$ (left) / temporal $n_t$ (right) mesh

![Graphs showing varying spatial $h$ and temporal $n_t$.]
Varying different viscosity parameters

![Graphs showing CPU time, TT rank, and memory consumption for varying viscosity parameters.]

**Figure:** Left: $\nu_0 = 1/10$, $\nu_1$ is varied. Right: $\nu_1$ and $\nu_0$ are varied together as $\nu_1 = 0.25\nu_0$. 
Cost functional, squared vorticity (top) and streamlines (bottom)
Conclusions & Outlook

- Low-rank tensor solver for unsteady heat and Navier-Stokes equations with uncertain viscosity.
- Similar techniques already used for 3D Stokes(-Brinkman) optimal control problems.
- Adapted AMEn (TT) solver to saddle point systems.
- With 1 stochastic parameter, the scheme reduces complexity by up to 2–3 orders of magnitude.
- To consider next:
Conclusions & Outlook

- Low-rank tensor solver for unsteady heat and Navier-Stokes equations with uncertain viscosity.
- Similar techniques already used for 3D Stokes(-Brinkman) optimal control problems.
- Adapted AMEn (TT) solver to saddle point systems.
- With 1 stochastic parameter, the scheme reduces complexity by up to 2–3 orders of magnitude.
- To consider next:
  - many parameters coming from uncertain geometry or Karhunen-Loève expansion of random fields;
    Initial results: the more parameters, the more significant is the complexity reduction w.r.t. memory — up to a factor of $10^9$ for the control problem for a backward facing step.
  - exploit multicore technology (need efficient parallelization of AMEn).
References


3D Stokes-Brinkman control problem

Mean

Standard deviation

State

Control

- Full size: $n_x n_\xi n_t \approx 3 \cdot 10^9$.
- Reduction: $\frac{\text{mem}(TT)}{\text{mem}(\text{full})} = 0.002$.  

© Peter Benner, benner@mpi-magdeburg.mpg.de  
PDE-constrained optimization under uncertainty  
37/37