Generating Sparse Representations by Adaptive Multiscale Approximations

Angela Kunoth

University of Cologne, Germany

Generating Sparse Representations by Adaptive Multiscale Approximations

Angela Kunoth

University of Cologne, Germany

Central topic: Efficient extraction, representation and analysis of information:
sparse representations

Goal: Maximal gain of knowledge with (ideally) provable
minimal amount of degrees of freedom and work

Essential ingredients: adaptive multiscale/wavelet representations

Generating Sparse Representations by Adaptive Multiscale Approximations

Angela Kunoth

University of Cologne, Germany

Central topic: Efficient extraction, representation and analysis of information:
sparse representations

Goal: Maximal gain of knowledge with (ideally) provable
minimal amount of degrees of freedom and work

Essential ingredients: adaptive multiscale/wavelet representations

Problem classes:

- Explicitly given information:
  fit and/or analysis of (multivariate, nonlinear) data on nonuniform grids
- Implicitly given information:
  - solution of (elliptic or parabolic) partial differential equations (PDEs)
  - PDE-constrained control problems

Problem:

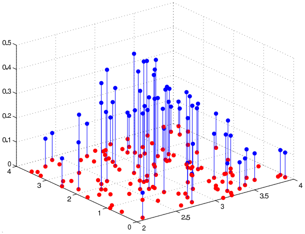
Given $P = \{(x_1, z_1), \ldots, (x_N, z_N)\}$ not uniformly distributed points

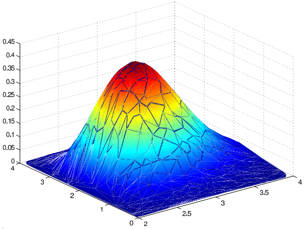
$X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n \qquad n \in \{1, 2, 3\}$

$Z = \{z_1, \ldots, z_N\} \subset \mathbb{R}$

Goal: Construct function $f : \mathbb{R}^n \to \mathbb{R}$ representing P

Example for $n = 2$:



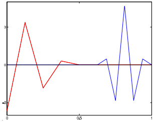


Solution method: Adaptive coarse–to–fine construction with thresholding

An Adaptive Coarse–to–Fine Method with Thresholding

[Castaño, Kunoth '03–'06]

Ansatz: $f(x) = \sum_{\lambda \in \Lambda} d_\lambda \psi_\lambda(x)$ $\quad \Lambda$ appropriate set of indices $\lambda = (j, k, \mathbf{e})$



Multiscale basis functions:
$\{\psi_\lambda\}_{\lambda \in \Lambda}$ preorthogonal (boundary–adapted) B–spline wavelets

Fitting of $\{d_\lambda\}_{\lambda \in \Lambda}$ using

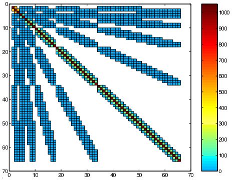
Approximation: $$\min \sum_{i=1}^{N} (z_i - f(x_i))^2$$

Approximation with regularization: $$\min \sum_{i=1}^{N} (z_i - f(x_i))^2 + \nu \|f\|^2_{H^\alpha}$$

Approximation (with regularization) $\rightsquigarrow$ normal equations
$(A^T A(+\nu D))d = A^T z \quad :\Longleftrightarrow: \quad (M(+\nu D))d = b$

$$M \in \mathbb{R}^{(\#\Lambda)\times(\#\Lambda)} \qquad M_{\lambda,\lambda'} = \sum_{i=1}^{N} \psi_\lambda(x_i)\psi_{\lambda'}(x_i)$$

$$b \in \mathbb{R}^{(\#\Lambda)} \qquad b_\lambda = \sum_{i=1}^{N} z_i \psi_\lambda(x_i)$$

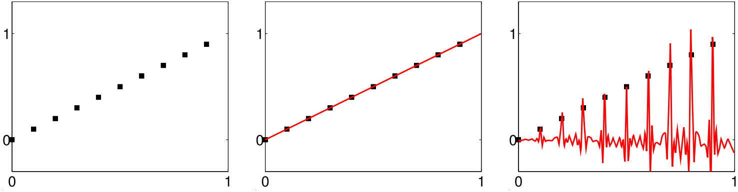


Typical structure of M

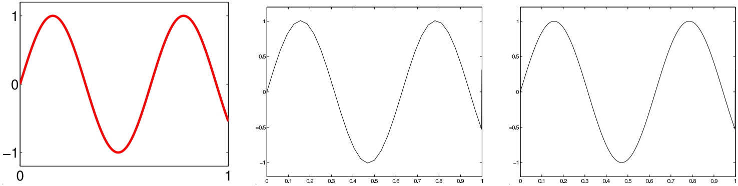
Amount of data $N \gg \#\Lambda$ degrees of freedom

Choice of index set Λ in $f = \sum_{\lambda \in \Lambda} d_\lambda \psi_\lambda$ in order to . . .

. . . get a reasonable reconstruction



. . . avoid processing redundant information



$\#P = 10.000$ points

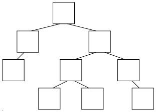
$\#\Lambda_a = 33$

$\#\Lambda_b = 8614$

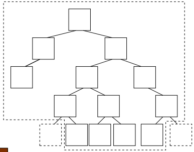
Data Driven Coarse–To–Fine Construction of Λ

1. Tree at (coarsest) level $j = 3$

Λ_3

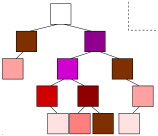


2. Discard children containing less than q data points in support



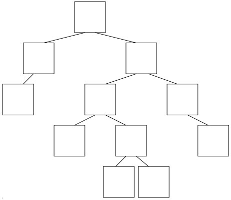
$\tilde{\Lambda}_3$

3. Compute approximation $f_3 := \sum_{\lambda \in \tilde{\Lambda}_3} d_\lambda \psi_\lambda$



4. Threshold small coefficients

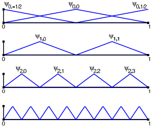
. . . and get tree for level $j = 4$



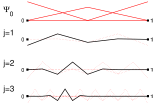
Λ_4

. . . . . . repeat 2.–4.

5. Stop at highest level J only determined by data

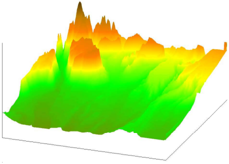


Hierarchical basis

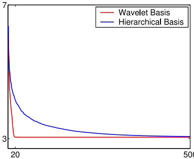


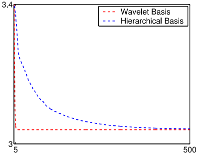
B–spline wavelet basis

Solution of $A^T A d = A^T z$



160.000 gridded data points from GTOPO30 Digital Elevation Model

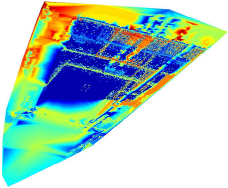


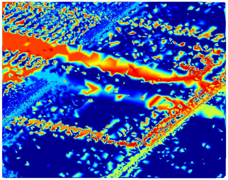


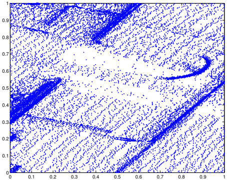
Error decay at highest level $J = 7$: log(error)/CG iterations: no nesting nested iterations

Further issues:
Regularization: Construction of surfaces with smoothness constraints
Robust regression: Handling of outliers [Castaño, Kunoth, IEEE Trans. Image Proc., 2006]

Example from Photogrammetric Application with fixed α, ν

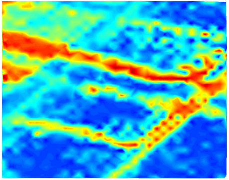


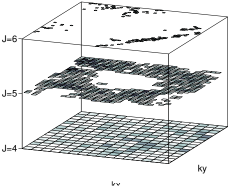


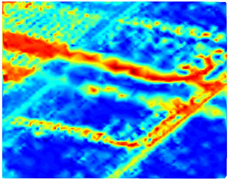


Original data (section) — vertical view of original data (section) — sampling geometry

3D point set (330.000 points) of industrial site taken by Leica Cyrax 2500, Prof. Staiger, GH Essen







reconstruction for $J = 4$ — coefficients of wavelets of type $(1, 1)$ — reconstruction for $J = 6$

wavelet reconstruction with regularization: $\nu = 0.01$, $\alpha = 4$
thresholding parameter $\varepsilon = 1e - 3$ $\qquad \#\Lambda_6 = 2623$ $\qquad$ full grid: 16384 coefficients

Part II: Implicitly Given Data: Optimal Control Problems Constrained by a Parabolic PDE

Given $\quad y_*(t,\cdot) \qquad f \qquad \omega > 0 \qquad$ end time $T > 0 \qquad$ initial condition y_0

$$
\begin{array}{lrcll}
\text{minimize} & \mathcal{J}(y,u) & = & \frac{1}{2}\int_0^T \|y(t,\cdot) - y_*(t,\cdot)\|_Z^2 \, dt + \frac{\omega}{2}\int_0^T \|u(t,\cdot)\|_U^2 \, dt & \\
\text{subject to} & y'(t) + A(t)y(t) & = & f(t) + u(t) \qquad \text{a.e. } t \in (0,T) =: I & \text{(PDE)} \\
 & y(0) & = & y_0 &
\end{array}
$$

$y' := \frac{\partial}{\partial t} y \qquad y = y(t,x)$ state $\qquad u = u(t,x)$ control

$Y = H_0^1(\Omega)$ state space $\quad Z = Y = H_0^1(\Omega)$ observation space $\quad U = Y' = H^{-1}(\Omega)$ control space

$A(t) : Y \to Y' \qquad \langle A(t)v(t,\cdot), w(t,\cdot)\rangle := \int_\Omega [\nabla v(t,x) \cdot \nabla w(t,x) + v(t,x)w(t,x)] \, dx$

$A(t)$ 2nd order linear selfadjoint coercive & continuous operator on $Y \qquad \Omega \subset \mathbb{R}^d$

Part II: Implicitly Given Data: Optimal Control Problems Constrained by a Parabolic PDE

Given $\quad y_*(t,\cdot) \qquad f \qquad \omega > 0 \qquad$ end time $T > 0 \qquad$ initial condition y_0

$$\begin{array}{lrcll} \text{minimize} & \mathcal{J}(y,u) & = & \frac{1}{2}\int_0^T \|y(t,\cdot) - y_*(t,\cdot)\|_Z^2 \, dt + \frac{\omega}{2}\int_0^T \|u(t,\cdot)\|_U^2 \, dt & \\ \text{subject to} & y'(t) + A(t)y(t) & = & f(t) + u(t) \qquad \text{a.e. } t \in (0,T) =: I & \text{(PDE)} \\ & y(0) & = & y_0 & \end{array}$$

$y' := \frac{\partial}{\partial t} y \qquad y = y(t,x)$ state $\qquad u = u(t,x)$ control

$Y = H_0^1(\Omega)$ state space $\quad Z = Y = H_0^1(\Omega)$ observation space $\quad U = Y' = H^{-1}(\Omega)$ control space

$A(t) : Y \to Y' \qquad \langle A(t)v(t,\cdot), w(t,\cdot)\rangle := \int_\Omega [\nabla v(t,x)\cdot\nabla w(t,x) + v(t,x)w(t,x)]\, dx$

$A(t)$ 2nd order linear selfadjoint coercive & continuous operator on $Y \qquad \Omega \subset \mathbb{R}^d$

PDE-constrained control problem $\rightsquigarrow$ requires repeated solution of (PDE)

$$\begin{array}{rcl} y'(t) + A(t)y(t) & = & f(t) + u(t) \\ y(0) & = & y_0 \end{array}$$

$\rightsquigarrow$ requires fast solver as core ingredient

Conventional time discretizations (e.g., Crank-Nicolson method) $\rightsquigarrow$
requires fast solver for elliptic PDE in each time step

Numerical Solution of a Single Elliptic PDE

Elliptic PDE $Ay = f$ s.th. $\|Av\|_{Y'} \sim \|v\|_Y \quad \rightsquigarrow$ find $y \in Y$: $\langle v, Ay\rangle = \langle v, f\rangle$ for all $v \in Y$

Conventional finite element discretization on a uniform grid: $Y_h \subset Y \quad \dim Y_h < \infty$

$$\rightsquigarrow \quad \boxed{A_h\, y_h = f_h}$$

Obstructions:

- Large linear systems of equations $\rightsquigarrow$ iterative solver
- High desired accuracy $\rightsquigarrow$ small h $\rightsquigarrow$ larger problem $\rightsquigarrow$ worse condition $\mathrm{cond}_2(A_h) \sim h^{-2}$
- Resolution of singularities in data and/or geometry $\rightsquigarrow$ small h

Numerical Solution of a Single Elliptic PDE

Elliptic PDE $Ay = f$ s.th. $\|Av\|_{Y'} \sim \|v\|_Y$ $\rightsquigarrow$ find $y \in Y$: $\langle v, Ay\rangle = \langle v, f\rangle$ for all $v \in Y$

Conventional finite element discretization on a uniform grid: $Y_h \subset Y$ $\dim Y_h < \infty$

$$\rightsquigarrow \quad \boxed{A_h\, y_h = f_h}$$

Obstructions:

- Large linear systems of equations $\rightsquigarrow$ iterative solver
- High desired accuracy $\rightsquigarrow$ small h $\rightsquigarrow$ larger problem $\rightsquigarrow$ worse condition $\mathrm{cond}_2(A_h) \sim h^{-2}$
- Resolution of singularities in data and/or geometry $\rightsquigarrow$ small h

Ingredients for Efficient Numerical Solution:

(i) Multilevel preconditioner C_h
multigrid methods, BPX preconditioner, wavelet discretization $\rightsquigarrow$ $\mathrm{cond}_2(C_h A_h) \sim 1$

Proofs: [Braess, Hackbusch '80s], [Dahmen, Kunoth '92], [Oswald '92]

(ii) Nested iteration

(iii) Additionally: adaptive refinement (for nonsmooth solutions)
a–posteriori error estimation $\rightsquigarrow$ local grid refinement $\rightsquigarrow$ convergence/convergence rates?

First goal: Realize discretization error accuracy ε with minimal amount of work $\mathcal{O}(N(\varepsilon))$

A-priori Estimates for Finite Elements:

Quality measure: Approximation in norm $\|y - y_h\|_{L_2(\Omega)} \leq \varepsilon$

A–priori error estimates: $\Omega \subset \mathbb{R}^d$ $\dim Y_h = N \sim h^{-d}$ uniform grid

$$\|y - y_h\|_{L_2(\Omega)} \lesssim h^r \|y\|_{H^r(\Omega)} \qquad y_h \in Y_h \qquad 0 \leq r \leq r_{\max}$$

$$\iff \quad \|y - y_N\|_{L_2(\Omega)} \lesssim N^{-r/d} \|y\|_{H^r(\Omega)}$$

N degrees of freedom $\longleftrightarrow$ accuracy $\mathcal{O}(N^{-r/d})$

Approximation rate determined by

(i) approximation order $r_{\max}$ of Y_h

(ii) space dimension d

(iii) amount of smoothness of y in L_2

Target: Realize discretization error accuracy $\varepsilon \sim h^2 \sim 2^{-2J}$ for grid with spacing $h \sim 2^{-J}$

Problem complexity: For $h \sim 2^{-J}$ a total of $N \sim 2^{Jd}$ unknowns

Optimal complexity for iterative solver: Minimal amount of work is $\mathcal{O}(N)$

Theorem: For smooth solution y:
Optimal multilevel preconditioner & nested iteration yields method of optimal complexity $\mathcal{O}(N_J)$
to reach discretization error accuracy on finest level J

Proofs: [Braess, Hackbusch '80s], [Dahmen, Kunoth '92], [Oswald '92]

First goal: Realize discretization error accuracy ε with minimal amount of work $\mathcal{O}(N(\varepsilon))$

A-priori Estimates for Finite Elements:

Quality measure: Approximation in norm $\|y - y_h\|_{L_2(\Omega)} \leq \varepsilon$

A–priori error estimates: $\Omega \subset \mathbb{R}^d$ dim $Y_h = N \sim h^{-d}$ uniform grid

$$\|y - y_h\|_{L_2(\Omega)} \lesssim h^r \|y\|_{H^r(\Omega)} \qquad y_h \in Y_h \qquad 0 \leq r \leq r_{\max}$$

$$\iff \qquad \|y - y_N\|_{L_2(\Omega)} \lesssim N^{-r/d} \|y\|_{H^r(\Omega)}$$

N degrees of freedom $\longleftrightarrow$ accuracy $\mathcal{O}(N^{-r/d})$

Approximation rate determined by

(i) approximation order $r_{\max}$ of Y_h

(ii) space dimension d

(iii) amount of smoothness of y in L_2

Target: Realize discretization error accuracy $\varepsilon \sim h^2 \sim 2^{-2J}$ for grid with spacing $h \sim 2^{-J}$

Problem complexity: For $h \sim 2^{-J}$ a total of $N \sim 2^{Jd}$ unknowns

Optimal complexity for iterative solver: Minimal amount of work is $\mathcal{O}(N)$

Theorem: For smooth solution y:
Optimal multilevel preconditioner & nested iteration yields method of optimal complexity $\mathcal{O}(N_J)$ to reach discretization error accuracy on finest level J

Proofs: [Braess, Hackbusch '80s], [Dahmen, Kunoth '92], [Oswald '92]

. . . back to PDE-constrained control problem . . .

Optimal Control Problems Constrained by a Parabolic PDE

Given $y_*(t,\cdot)$ $\quad f \quad$ $\omega > 0$ $\quad$ end time $T > 0$ $\quad$ initial condition y_0

$$
\boxed{
\begin{array}{rrcll}
\text{minimize} & \mathcal{J}(y,u) & = & \frac{1}{2}\int_0^T \|y(t,\cdot) - y_*(t,\cdot)\|_Z^2\,dt + \frac{\omega}{2}\int_0^T \|u(t,\cdot)\|_U^2\,dt & \\
\text{subject to} & y'(t) + A(t)y(t) & = & f(t) + u(t) \qquad \text{a.e. } t \in (0,T) =: I & \text{(PDE)} \\
 & y(0) & = & y_0 &
\end{array}
}
$$

$y' := \frac{\partial}{\partial t} y$ $\qquad$ $y = y(t,x)$ state $\qquad$ $u = u(t,x)$ control

$Y = H_0^1(\Omega)$ state space $\quad$ $Z = Y = H_0^1(\Omega)$ observation space $\quad$ $U = Y' = H^{-1}(\Omega)$ control space

$A(t) : Y \to Y'$ $\qquad$ $\langle A(t)v(t,\cdot), w(t,\cdot)\rangle := \int_\Omega [\nabla v(t,x) \cdot \nabla w(t,x) + v(t,x)w(t,x)]\,dx$

$A(t)$ 2nd order linear selfadjoint coercive & continuous operator on Y $\qquad$ $\Omega \subset \mathbb{R}^d$

Optimal Control Problems Constrained by a Parabolic PDE

Given $y_*(t,\cdot)$ $\quad f \quad$ $\omega > 0$ $\quad$ end time $T > 0$ $\quad$ initial condition y_0

$$
\begin{array}{lrcll}
\text{minimize} & \mathcal{J}(y,u) & = & \frac{1}{2}\int_0^T \|y(t,\cdot) - y_*(t,\cdot)\|_Z^2 \, dt + \frac{\omega}{2}\int_0^T \|u(t,\cdot)\|_U^2 \, dt & \\
\text{subject to} & y'(t) + A(t)y(t) & = & f(t) + u(t) \qquad \text{a.e. } t \in (0,T) =: I & \text{(PDE)} \\
 & y(0) & = & y_0 &
\end{array}
$$

$y' := \frac{\partial}{\partial t} y$ $\quad$ $y = y(t,x)$ state $\quad$ $u = u(t,x)$ control

$Y = H_0^1(\Omega)$ state space $\quad$ $Z = Y = H_0^1(\Omega)$ observation space $\quad$ $U = Y' = H^{-1}(\Omega)$ control space

$A(t) : Y \to Y'$ $\qquad$ $\langle A(t)v(t,\cdot), w(t,\cdot)\rangle := \int_\Omega [\nabla v(t,x) \cdot \nabla w(t,x) + v(t,x)w(t,x)] \, dx$

$A(t)$ 2nd order linear selfadjoint coercive & continuous operator on Y $\qquad$ $\Omega \subset \mathbb{R}^d$

PDE-constrained control problem $\quad \rightsquigarrow \quad$ requires repeated solution of (PDE)

$$
\begin{aligned}
y'(t) + A(t)y(t) &= f(t) + u(t) \\
y(0) &= y_0
\end{aligned}
$$

Necessary and Sufficient Conditions for Optimality

Optimal control problem constrained by parabolic PDE

$\rightsquigarrow$ System of parabolic PDEs coupled globally in time (and space)

$$
\begin{aligned}
y'(t) + A(t)\, y(t) &= f(t) + u(t) && \text{a.e. } t \in I \\
y(0) &= y_0 \\
\omega\, \tilde{R}^{-1} u(t) + p(t) &= 0 && \text{a.e. } t \in I \\
-p'(t) + A(t)^T p(t) &= \tilde{R}\,(y_*(t) - y(t)) && \text{a.e. } t \in I \\
p(T) &= 0
\end{aligned}
$$

Riesz operator $\tilde{R}$ defined by $\langle v, \tilde{R}w \rangle_{Y \times Y'} := (v, w)_Y$ for all $v, w \in Y$

Necessary and Sufficient Conditions for Optimality

Optimal control problem constrained by parabolic PDE

$\rightsquigarrow$ System of parabolic PDEs coupled globally in time (and space)

$$\begin{aligned} y'(t) + A(t)\, y(t) &= f(t) + u(t) && \text{a.e. } t \in I \\ y(0) &= y_0 \\ \omega\, \tilde{R}^{-1} u(t) + p(t) &= 0 && \text{a.e. } t \in I \\ -p'(t) + A(t)^T p(t) &= \tilde{R}\,(y_*(t) - y(t)) && \text{a.e. } t \in I \\ p(T) &= 0 \end{aligned}$$

Riesz operator $\tilde{R}$ defined by $\langle v, \tilde{R}w\rangle_{Y\times Y'} := (v,w)_Y$ for all $v, w \in Y$

Obstructions for numerical solution:

- convential time discretizations: time-marching methods
  $\rightsquigarrow$ need storage of $y(t_i), u(t_i), p(t_i)$ for all discrete times $0 = t_0, \ldots, T = t_N$
- in each time step: solve elliptic PDE $\rightsquigarrow$ large linear system of equations
  $\rightsquigarrow$ iterative solver $\rightsquigarrow$ need preconditioning in (conjugate) gradient method
- singularities in data/domain: adaptive (FE) mesh(es) for $y(t_i), u(t_i), p(t_i)$ for all t_i
  one mesh for all variables, refinement/coarsening ? [Meidner, Vexler '07], ...
  convergence ? complexity ??

Necessary and Sufficient Conditions for Optimality

Optimal control problem constrained by parabolic PDE

$\rightsquigarrow$ System of parabolic PDEs coupled globally in time (and space)

$$
\begin{aligned}
y'(t) + A(t)\,y(t) &= f(t) + u(t) && \text{a.e. } t \in I \\
y(0) &= y_0 \\
\omega\,\tilde{R}^{-1}u(t) + p(t) &= 0 && \text{a.e. } t \in I \\
-p'(t) + A(t)^T p(t) &= \tilde{R}\,(y_*(t) - y(t)) && \text{a.e. } t \in I \\
p(T) &= 0
\end{aligned}
$$

Riesz operator $\tilde{R}$ defined by $\langle v, \tilde{R}w\rangle_{Y\times Y'} := (v,w)_Y$ for all $v, w \in Y$

Obstructions for numerical solution:

- convential time discretizations: time-marching methods
  $\rightsquigarrow$ need storage of $y(t_i), u(t_i), p(t_i)$ for all discrete times $0 = t_0, \ldots, T = t_N$
- in each time step: solve elliptic PDE $\rightsquigarrow$ large linear system of equations
  $\rightsquigarrow$ iterative solver $\rightsquigarrow$ need preconditioning in (conjugate) gradient method
- singularities in data/domain: adaptive (FE) mesh(es) for $y(t_i), u(t_i), p(t_i)$ for all t_i
  one mesh for all variables, refinement/coarsening ? [Meidner, Vexler '07], ...
  convergence ? complexity ??

Solution Ansatz here: full weak space-time form of parabolic PDE constraint

Variational Space-Time Form for a Single Parabolic Evolution PDE

[Ladyshenskaya et al. 1967], [Wloka '82], [Dautray, Lions '92], [Schwab, Stevenson '09], [Chegini, Stevenson '11], [Stapel '11]

(PDE) $$\begin{aligned} y'(t) + A(t)\,y(t) &= f(t) \qquad \text{a.e. } t \in I \\ y(0) &= y_0 \end{aligned}$$

solution space: Lebesgue-Bochner space $\mathcal{Y} := (L_2(I) \otimes Y) \cap (H^1(I) \otimes Y') \hookrightarrow \mathcal{C}^0(\overline{I}) \otimes L_2(\Omega)$

with norm $\|w\|_{\mathcal{Y}}^2 := \|w\|_{L_2(I)\otimes Y}^2 + \|w'\|_{H^1(I)\otimes Y'}^2$

test space: $\mathcal{Q} := (L_2(I) \otimes Y) \times L_2(\Omega)$ with norm $\|v\|_{\mathcal{Q}}^2 := \|v_1\|_{L_2(I)\otimes Y}^2 + \|v_2\|_{L_2(\Omega)}^2$

bilinear form $b(\cdot,\cdot) : \mathcal{Y} \times \mathcal{Q} \to \mathbb{R}$

$$b(w,(v_1,v_2)) := \int_I \left[\langle w'(t,\cdot), v_1(t,\cdot)\rangle + \langle A(t)w(t,\cdot), v_1(t,\cdot)\rangle\right] dt + \langle w(0,\cdot), v_2\rangle =: \langle Bw, v\rangle$$

right hand side

$$\langle f, v\rangle := \int_I \langle f(t,\cdot), v_1(t,\cdot)\rangle\, dt + \langle y_0, v_2\rangle$$

(PDE) $\rightsquigarrow$ given $f \in \mathcal{Q}'$, find $y \in \mathcal{Y}$: $\boxed{By = f}$

Existence and uniqueness of solution:

Theorem $\|Bw\|_{\mathcal{Q}'} \sim \|w\|_{\mathcal{Y}}$ for all $w \in \mathcal{Q}$ mapping property (MP)

Formulations with 1/2 time derivatives: [Fontes "99], [Larsson, Schwab '15]

Reformulation of PDE-Constrained Optimal Control Problem

$$\text{minimize} \quad \mathcal{J}(y,u) = \tfrac{1}{2}\|y - y_*\|^2_{L_2(I)\otimes Y} + \tfrac{\omega}{2}\|u\|^2_{L_2(I)\otimes Y'}$$

$$\text{subject to} \quad B\,y = f + E\,u \qquad \text{(PDE)} \qquad B:\mathcal{Y}\to\mathcal{Q}' \text{ satisfies (MP)}$$

$$E := (\mathrm{Id}, 0) : L_2(I)\otimes Y' \to \mathcal{Q}'$$

Necessary and Sufficient Conditions — Karush-Kuhn-Tucker (KKT) system

$$\mathcal{L}(y,u,p) := \mathcal{J}(y,u) + \langle p, By - f - Eu\rangle$$

Riesz operator $\langle v, Rw\rangle_{(L_2(I)\otimes Y)\times(L_2(I)\otimes Y')} := (v,w)_{L_2(I)\otimes Y}$

$$\delta\mathcal{L} = 0 \rightsquigarrow \boxed{\begin{aligned} B^* p &= R(y_* - y)\\ \omega R^{-1} u &= E^* p\\ B\,y &= f + E\,u \end{aligned}}$$

$$\Longleftrightarrow \quad \begin{pmatrix} R & 0 & B^*\\ 0 & \omega R^{-1} & -E^*\\ B & -E & 0 \end{pmatrix}\begin{pmatrix} y\\ u\\ p\end{pmatrix} = \begin{pmatrix} Ry_*\\ 0\\ f\end{pmatrix} \qquad \text{(SPP)}$$

$\rightsquigarrow$ saddle point operator

$$\langle \mathcal{G}q, \tilde{q}\rangle := \left\langle \begin{pmatrix} R & 0 & B^*\\ 0 & \omega R^{-1} & -E^*\\ B & -E & 0 \end{pmatrix} q, \tilde{q}\right\rangle$$

$\mathcal{A} := \mathrm{diag}(R, \omega R^{-1})$ pos. def., $\mathcal{B} := (B, -E)$ full rank

symmetric, continuous, boundedly invertible on $\boldsymbol{\mathcal{X}} := \mathcal{Y}\times\mathcal{U}\times\mathcal{Q}$

$\Longrightarrow$ unique solution $\begin{pmatrix} y\\ u\\ p\end{pmatrix} =: q$ of system of PDEs (SPP)

Formulations with 1/2 time derivatives: [Langer, Wolfmayr '13], [Kunoth, Mollet '15, unpublished]

Reformulation of PDE-Constrained Optimal Control Problem

minimize	$\mathcal{J}(y,u) = \frac{1}{2}\|y - y_*\|^2_{L_2(I)\otimes Y} + \frac{\omega}{2}\|u\|^2_{L_2(I)\otimes Y'}$	
subject to	$B\,y = f + E\,u$ (PDE)	$B : \mathcal{Y} \to \mathcal{Q}'$ satisfies (MP) $E := (\mathrm{Id}, 0) : L_2(I)\otimes Y' \to \mathcal{Q}'$

Necessary and Sufficient Conditions — Karush-Kuhn-Tucker (KKT) system

$\mathcal{L}(y,u,p) := \mathcal{J}(y,u) + \langle p, By - f - Eu\rangle$

Riesz operator $\langle v, Rw\rangle_{(L_2(I)\otimes Y)\times(L_2(I)\otimes Y')} := (v,w)_{L_2(I)\otimes Y}$

$$\delta\mathcal{L} = 0 \rightsquigarrow \boxed{\begin{array}{rcl} B^* p &=& R(y_* - y) \\ \omega R^{-1} u &=& E^* p \\ B\,y &=& f + E\,u \end{array}} \iff \boxed{\begin{pmatrix} R & 0 & B^* \\ 0 & \omega R^{-1} & -E^* \\ B & -E & 0 \end{pmatrix}\begin{pmatrix} y \\ u \\ p \end{pmatrix} = \begin{pmatrix} Ry_* \\ 0 \\ f \end{pmatrix} \quad \text{(SPP)}}$$

$\rightsquigarrow$ saddle point operator

$$\langle \mathcal{G}q, \tilde{q}\rangle := \left\langle \begin{pmatrix} R & 0 & B^* \\ 0 & \omega R^{-1} & -E^* \\ B & -E & 0 \end{pmatrix} q, \tilde{q}\right\rangle$$

$\mathcal{A} := \mathrm{diag}(R, \omega R^{-1})$ pos. def., $\mathcal{B} := (B, -E)$ full rank

symmetric, continuous, boundedly invertible on $\boldsymbol{\mathcal{X}} := \mathcal{Y} \times \mathcal{U} \times \mathcal{Q}$

$\Longrightarrow$ unique solution $\begin{pmatrix} y \\ u \\ p \end{pmatrix} =: q$ of system of PDEs (SPP)

Formulations with 1/2 time derivatives: [Langer, Wolfmayr '13], [Kunoth, Mollet '15, unpublished]

Next: discretization in space and time variables by wavelets

Building Blocks: (Biorthogonal Spline–) Wavelets

H Hilbert space on domain $\Omega \subset \mathbb{R}^d$ with $\|\cdot\|_H$ $\qquad$ H' dual space for H with $\langle \cdot, \cdot \rangle$

$\Psi := \{\psi_\lambda : \lambda \in \mathbb{I}\} \subset H$ Wavelets $\qquad$ $\mathbb{I}$ (infinite) index set

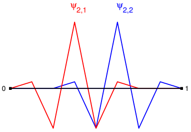
(NE) Ψ Riesz basis for H

$$v \in H: \quad v = \mathbf{v}^T \Psi := \sum_{\lambda \in \mathbb{I}} \langle v, \tilde{\psi}_\lambda \rangle \, \psi_\lambda \quad \text{such that} \quad \|v\|_H \sim \|\mathbf{v}\|_{\ell_2(\mathbb{I})}$$

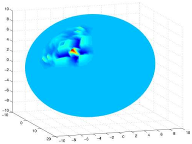
(L) Locality $\qquad \operatorname{diam}(\operatorname{supp} \psi_\lambda) \sim 2^{-|\lambda|}$ $\qquad$ $|\lambda|$ resolution

ψ_λ centered around $2^{-|\lambda|} k$

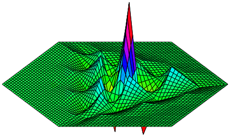
(CP) Vanishing moments $\quad \langle v, \psi_\lambda \rangle \lesssim 2^{-|\lambda|(\frac{d}{2} + \tilde{m})} \|v^{(\tilde{m})}\|_{L_\infty(\operatorname{supp} \psi_\lambda)}$ for some $\tilde{m}$



[Dahmen, Kunoth, Urban '99]



[Dahmen, Schneider '99], [Kunoth, Sahner '06]



[Harbrecht, Schneider '00]

Paradigm of Adaptive Wavelet Method for One Stationary PDE

[Cohen, Dahmen, DeVore '01/'02]

(i) Well–posed variational problem: given $f \in \mathcal{Q}'$, $B : \mathcal{Y} \to \mathcal{Q}'$, find $y \in \mathcal{Y}$ such that $\boxed{By = f}$

$$\boxed{\text{(MP)} \qquad \|Bw\|_{\mathcal{Q}'} \sim \|w\|_{\mathcal{Y}} \qquad \text{for all } w \in \mathcal{Y} \qquad \text{mapping property}}$$

(ii) $\Psi^{\mathcal{Y}}, \Psi^{\mathcal{Q}}$ wavelet bases for $\mathcal{Y}, \mathcal{Q}$:

$$\boxed{\text{(NE)} \qquad \|\mathbf{w}^T \Psi^{\mathcal{Y}}\|_{\mathcal{Y}} \sim \|\mathbf{w}\|_{\ell_2} \qquad \text{for all } \mathbf{w} = (w_\lambda)_{\lambda \in \mathbb{I}} \in \ell_2}$$

$$\mathbf{Bw} := (\langle \psi_\lambda^{\mathcal{Y}}, Bw \rangle)_{\lambda \in \mathbb{I}} \qquad \mathbf{f} := (\langle \psi_\lambda^{\mathcal{Y}}, f \rangle)_{\lambda \in \mathbb{I}}$$

$\rightsquigarrow$ **Theorem** $\quad By = f \iff \mathbf{By} = \mathbf{f}$ well-posed in ℓ_2 $\quad (\mathbf{B} : \ell_2 \to \ell_2)$

$$\text{(MP)} + \text{(NE)} \iff \|\mathbf{Bw}\|_{\ell_2} \sim \|\mathbf{w}\|_{\ell_2} \quad \text{for all } \mathbf{w} \in \ell_2$$

(iii) Practical solution schemes for $\mathbf{By} = \mathbf{f}$:

(A) Perturbed Richardson iteration (for symmetric $\mathbf{B}$):

(A.1) $\mathbf{y}^{n+1} = \mathbf{y}^n + (\mathbf{f} - \mathbf{B}\mathbf{y}^n) \quad n = 0, 1, 2, \ldots \qquad \|\mathbf{y}^{n+1} - \mathbf{y}\|_{\ell_2} \le \rho \, \|\mathbf{y}^n - \mathbf{y}\|_{\ell_2} \quad \rho < 1$

(A.2) Approximate realization: adaptive evaluation of $\mathbf{B}\mathbf{y}^n$ in SOLVE $[\varepsilon, \mathbf{B}, \mathbf{f}] \to \mathbf{y}_\varepsilon$

(A.3) Coarsening (thresholding) of the iterands (for complexity)

(B) Adaptive wavelet Galerkin method and bulk chasing strategy

Extension to a Single Parabolic Evolution PDE

[Schwab, Stevenson '09]

(i) Variational space-time form of (PDE)
$$\begin{aligned} y'(t) + A(t)\,y(t) &= f(t) \qquad \text{a.e. } t \in I \\ y(0) &= y_0 \end{aligned}$$

solution space: Lebesgue-Bochner space $\mathcal{Y} := (L_2(I) \otimes Y) \cap (H^1(I) \otimes Y')$
with norm $\|w\|_{\mathcal{Y}}^2 := \|w\|_{L_2(I)\otimes Y}^2 + \|w'\|_{H^1(I)\otimes Y'}^2$

test space $\mathcal{Q} := L_2(I; Y) \times L_2(\Omega)$ with norm $\|v\|_{\mathcal{Q}}^2 := \|v_1\|_{L_2(I)\otimes Y}^2 + \|v_2\|_{L_2(\Omega)}^2$

bilinear form $b(\cdot,\cdot) : \mathcal{Y} \times \mathcal{Q} \to \mathbb{R}$

$$b(y,(v_1,v_2)) := \int_I \left[\langle y'(t,\cdot), v_1(t,\cdot)\rangle + \langle A(t)y(t,\cdot), v_1(t,\cdot)\rangle\right] dt + \langle y(0,\cdot), v_2\rangle =: \langle By, v\rangle$$

right hand side

$$\langle f, v\rangle := \int_I \langle f(t,\cdot), v_1(t,\cdot)\rangle\, dt + \langle y_0, v_2\rangle$$

(PDE) $\rightsquigarrow$ given $f \in \mathcal{Q}'$, find $y \in \mathcal{Y}$: $\boxed{By = f}$

Theorem	(MP)	$\|Bw\|_{\mathcal{Q}'} \sim \|w\|_{\mathcal{Y}}$	for all $w \in \mathcal{Y}$	mapping property

(ii) $\Psi^{\mathcal{Y}}, \Psi^{\mathcal{Q}}$ wavelet bases for $\mathcal{Y}, \mathcal{Q}$ $\rightsquigarrow$ $\mathbf{By} := (\langle \psi_\lambda^{\mathcal{Q}}, By\rangle)_{\lambda\in\mathbb{I}}$ $\quad \mathbf{f} := (\langle \psi_\lambda^{\mathcal{Q}}, f\rangle)_{\lambda\in\mathbb{I}}$

Theorem $By = f \iff \mathbf{By} = \mathbf{f}$ $\quad \mathbf{B} : \ell_2 \to \ell_2$ and $\mathbf{By} = \mathbf{f}$ well-posed in ℓ_2

(MP) + (NE) $\Longrightarrow$ $\|\mathbf{Bv}\|_{\ell_2} \sim \|\mathbf{v}\|_{\ell_2}$, $\mathbf{v} \in \ell_2$ $\quad$ $\mathbf{B}$ unsymmetric

Control problem in wavelet coordinates

minimize $\quad \mathbf{J}(\mathbf{y},\mathbf{u}) \;=\; \frac{1}{2}\,\|\mathbf{R}^{1/2}(\mathbf{y}-\mathbf{y}_*)\|^2 + \frac{\omega}{2}\,\|\mathbf{R}^{-1/2}\mathbf{u}\|^2$

subject to $\quad \mathbf{B}\mathbf{y} = \mathbf{f} + \mathbf{u} \qquad\qquad \mathbf{B} : \ell_2 \to \ell_2$ automorphism $\qquad \|\cdot\| := \|\cdot\|_{\ell_2}$

Necessary and Sufficient Conditions — Karush-Kuhn-Tucker (KKT) system

$$\mathbf{L}(\mathbf{y},\mathbf{u},\mathbf{p}) \;:=\; \mathbf{J}(\mathbf{y},\mathbf{u}) + \langle \mathbf{p},\, \mathbf{B}\mathbf{y} - (\mathbf{f}+\mathbf{u})\rangle$$

$$\delta\mathbf{L} = 0 \;\rightsquigarrow\; \boxed{\begin{aligned} \mathbf{B}\mathbf{y} &= \mathbf{f}+\mathbf{u} \\ \omega\mathbf{R}^{-1}\mathbf{u} &= \mathbf{p} \\ \mathbf{B}^*\mathbf{p} &= \mathbf{R}(\mathbf{y}_*-\mathbf{y}) \end{aligned}} \quad\Longleftrightarrow\quad \boxed{\mathbf{Q}\,\mathbf{u} = \mathbf{g}}$$

$$\Longleftrightarrow\quad \boxed{\begin{pmatrix} \mathbf{R} & \mathbf{0} & \mathbf{B}^* \\ \mathbf{0} & \omega\mathbf{R}^{-1} & -\mathbf{E} \\ \mathbf{B} & -\mathbf{E} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{R}\mathbf{y}_* \\ \mathbf{0} \\ \mathbf{f} \end{pmatrix} \qquad \text{(SPP)}} \qquad \mathbf{Q} : \ell_2 \to \ell_2 \;\text{ automorphism}$$

where
$$\begin{aligned} \mathbf{Q} &:= \mathbf{B}^{-*}\mathbf{R}\mathbf{B}^{-1} + \omega\mathbf{R}^{-1} \\ \mathbf{g} &:= \mathbf{B}^{-*}(\mathbf{R}\mathbf{y}_* - \mathbf{R}\mathbf{B}^{-1}\mathbf{f}) \end{aligned}$$

Complexity Analysis

Based on benchmark:
decay rate s for (wavelet-)best N term approximation $\mathcal{A}^s := \{\mathbf{v} \in \ell_2 : \|\mathbf{v} - \mathbf{v}_N\| \lesssim N^{-s}\}$

Work/accuracy balance of best N term approximation:

$$\boxed{\text{Target accuracy } \varepsilon \;(\sim N^{-s}) \quad \longleftrightarrow \quad \text{Work } \varepsilon^{-1/s} \;(\sim N)}$$

Convergence and Complexity

For solution routine (A): (Idealized) iteration (for symmetric $\mathbf{B}$)
$\mathbf{v}^{n+1} = \mathbf{v}^n + (\mathbf{f} - \mathbf{B}\mathbf{v}^n)$ update via $\textsc{Res}\,[\eta, \mathbf{B}, \mathbf{f}, \mathbf{v}] \to \mathbf{r}_\eta$ $\rightsquigarrow$ $\textsc{Solve}\,[\varepsilon, \mathbf{B}, \mathbf{f}] \to \mathbf{v}_\varepsilon$

Benchmark Theorem [Cohen, Dahmen, DeVore '01/'02]

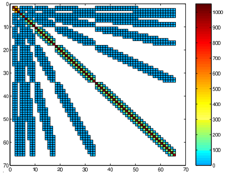
Vanishing moments (CP) for wavelets $\Longrightarrow$ $\mathbf{B}$ is s^*–compressible

$\Longrightarrow$ for variational problem satisfying (MP) scheme $\textsc{Solve}$ can be designed with properties:

(I) For every target accuracy $\varepsilon > 0$ $\textsc{Solve}$ produces after finitely many steps approximate solution $\mathbf{v}_\varepsilon$ such that $\|\mathbf{v} - \mathbf{v}_\varepsilon\| \leq \varepsilon$

(II) Exact solution $\mathbf{v} \in \mathcal{A}^s$ $\Longrightarrow$ $\operatorname{supp} \mathbf{v}_\varepsilon$, # flops $\sim \varepsilon^{-1/s} \sim N$

Core Ingredient of SOLVE: Compressible Operators



(CP) $\rightsquigarrow$ **B** is s^*–compressible:
for every $s \in (0, s^*)$ there exists $\mathbf{B}_j$
with $\leq \alpha_j 2^j$ nonzero entries per row and column s.th. for $j \in \mathbb{N}_0$

$$\|\mathbf{B}-\mathbf{B}_j\| \leq \alpha_j 2^{-sj} \qquad \sum_{j\in\mathbb{N}_0} \alpha_j < \infty \qquad (\mathbf{B} \text{ 'close to' sparse matrix})$$

Application of (Non)Linear Operators in Wavelet Bases

Theory: [Dahmen, Schneider, Xu '00], [Cohen, Dahmen, DeVore '03] ...
$d = 2$, isotropic tensor-product wavelets: [Vorloeper '10] general d: [Stapel '11], [Mollet, Pabel '12], [Pabel '15]

Input: finitely supported vector $\mathbf{v} = (v_\mu)_{\mu\in\Lambda}$ $\quad \Lambda \subset \mathbb{I}$ finite

Output: approximation of **Bv** with infinite-dimensional operator $\mathbf{B} : \ell_2(\mathbb{I}) \to \ell_2(\mathbb{I})$

$B : \mathcal{Y} \to \mathcal{Q}' \rightsquigarrow$ expand $Bv \in \mathcal{Q}'$ in dual wavelet basis for $\mathcal{Q}'$ and v in primal wavelet basis for $\mathcal{Y}$
$\rightsquigarrow$

$$Bv = (\mathbf{Bv})^T\tilde{\Psi} = \sum_{\lambda\in\mathbb{I}} \langle Bv, \psi_\lambda\rangle\,\tilde{\psi}_\lambda = \sum_{\lambda\in\mathbb{I}} \langle B(\sum_{\mu\in\Lambda} v_\mu\psi_\mu, \psi_\lambda\rangle)\,\tilde{\psi}_\lambda = \sum_{\lambda\in\mathbb{I}}\sum_{\mu\in\Lambda} v_\mu \langle B\psi_\mu, \psi_\lambda\rangle\,\tilde{\psi}_\lambda$$

$\rightsquigarrow$ compute $\langle B\psi_\mu, \psi_\lambda\rangle$ for given $\mu \in \Lambda$ (finite) and all $\lambda \in \mathbb{I}$

Compressibility of B: $\quad |\langle B\psi_\mu, \psi_\lambda\rangle| \leq C_{\|\mathbf{v}\|} \sup_{\mu:\ S_\lambda\cap S_\mu\neq\emptyset} 2^{-\gamma(|\lambda|-|\mu|)}\,|v_\mu| \qquad \gamma > \frac{d}{2}+1$

follows from wavelet property (CP)

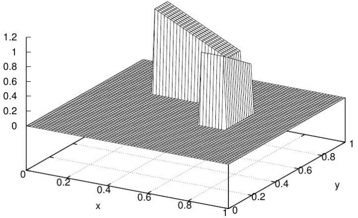
Essential data structure (for nonlinear operators): tree-type index sets
input $\mathbf{v} \rightsquigarrow$ prediction of tree index set based on supp $\mathbf{v}$ and properties of **B**
$\rightsquigarrow$ computation of $(\mathbf{Bv})_\lambda$ after transformation to piecewise polynomials
$\rightsquigarrow$ application of **B** in optimal linear complexity

Application of (Non)Linear Operators in Wavelet Bases: Numerical Example

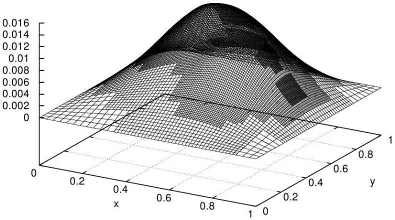
[Mollet, Pabel '12], [Pabel '15]

PDE with nonlinear term

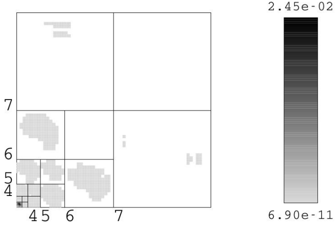
$$\begin{aligned} -\Delta y + y^3 &= f \qquad \text{in } \Omega := (0,1)^2 \\ y &= 0 \qquad \text{on } \partial\Omega \end{aligned}$$



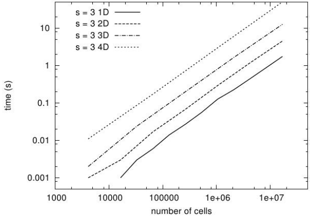
right hand side f



solution y (with Richardson scheme and residual error bound 10^{-3})



distribution of 7177 active wavelet coefficients



Runtime (seconds) for evaluating y^3 for $d \leq 4$

Essential ideas: $\textsc{Res}$ for $\textsc{Solve}\,[\ldots,\mathbf{Q},\ldots]$ reduced to $\textsc{Res}$ for $\textsc{Solve}\,[\ldots,\mathbf{B},\ldots]$

applied to normal equations

and KKT system $\longleftrightarrow$ condensed system $\mathbf{Qu} = \mathbf{g}$

Theorem [Dahmen, Kunoth, SICON '05], [Gunzburger, Kunoth, SICON '11]

For any target accuracy $\varepsilon > 0$ $\textsc{Solve}\,[\varepsilon, \mathbf{Q}, \mathbf{g}] \to \mathbf{u}_\varepsilon$ converges in finitely many steps

$\|\mathbf{u} - \mathbf{u}_\varepsilon\| \le \varepsilon$ $\|\mathbf{y} - \mathbf{y}_\varepsilon\| \lesssim \varepsilon$ $\|\mathbf{p} - \mathbf{p}_\varepsilon\| \lesssim \varepsilon$ $\mathbf{u}_\varepsilon, \mathbf{y}_\varepsilon, \mathbf{p}_\varepsilon$ finitely supported

$\mathbf{u}, \mathbf{y}, \mathbf{p} \in \mathcal{A}^s \Longrightarrow$

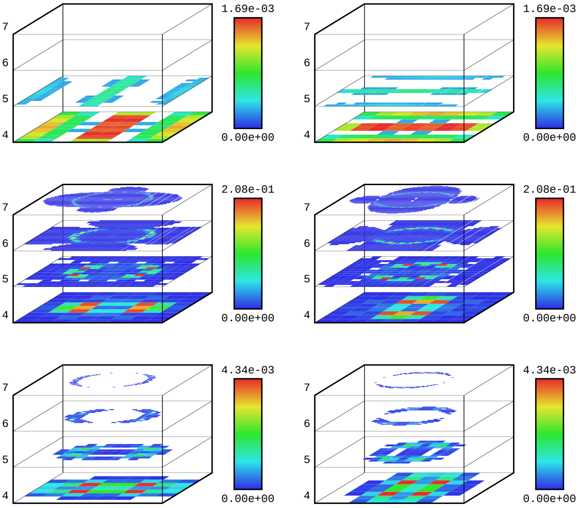
$$(\#\,\mathrm{supp}\,\mathbf{u}_\varepsilon) + (\#\,\mathrm{supp}\,\mathbf{y}_\varepsilon) + (\#\,\mathrm{supp}\,\mathbf{p}_\varepsilon) \lesssim \varepsilon^{-1/s}\left(\|\mathbf{u}\|_{\mathcal{A}^s}^{1/s} + \|\mathbf{y}\|_{\mathcal{A}^s}^{1/s} + \|\mathbf{p}\|_{\mathcal{A}^s}^{1/s}\right)$$

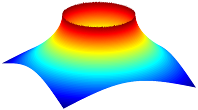
$$\|\mathbf{u}_\varepsilon\|_{\mathcal{A}^s} + \|\mathbf{y}_\varepsilon\|_{\mathcal{A}^s} + \|\mathbf{p}_\varepsilon\|_{\mathcal{A}^s} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s} + \|\mathbf{y}\|_{\mathcal{A}^s} + \|\mathbf{p}\|_{\mathcal{A}^s}$$

#flops $\sim \varepsilon^{-1/s}$

Numerical Example for Elliptic Control Problem (2D)

type $\mathbf{e} = (1,0)$ type $\mathbf{e} = (0,1)$





target state y_*

[Burstedde '05]

Numerical Example for One Parabolic PDE

[Chegini, Stevenson '11], [Stapel '11]

Compute $y = y(t,x)$ such that

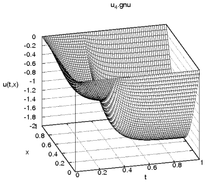
$$\begin{aligned} y_t(t,x) - y_{xx}(t,x) &= g(t) \otimes (-\pi^2)\sin(\pi x) && \text{in } I \times \Omega := (0,1)^2 \\ y(t,0) &= y(t,1) = 0 && \text{for } t \geq 0 \\ y(0,x) &= 0 && \text{for } x \in (0,1) \end{aligned}$$

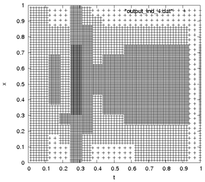
and $g(t) := \begin{cases} 1 & t \in [0, \frac{1}{3}) \\ 2 & t \in [\frac{1}{3}, 1] \end{cases}$

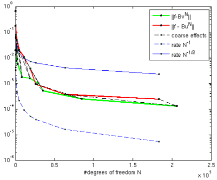
Problem formulation and implementation:

- Modified problem with zero initial conditions $\rightsquigarrow$
  solution space $\mathcal{Y} = (L_2(I) \otimes H^1(\Omega)) \cap (H^1_{(0}(I) \otimes H^{-1}(\Omega))$ and test space $\mathcal{Q} = L_2(I) \otimes Y$
- Inhomogeneous initial data: homogenization of initial conditions $\rightsquigarrow$ modification of r.h.s.
- Implementation based on AWM Toolbox by [Vorloeper '10]
  biorthogonal isotropic wavelets of order $m = 2, \tilde{m} = 4$
- Iterative solution by GMRES

Plot of Solution, Refined Grid and Residual Error Reduction







8526 degrees of freedom   Expected rate in H^1 (isotropic wavelets): $1/2$ red: after coarsening

Summary (Part II)

- Control problem constrained by parabolic PDE
- Full weak space-time formulation of evolution PDE

  $\rightsquigarrow$ saddle point system of PDEs coupled globally in time and space
- For smooth solutions: multilevel/wavelet preconditioners + nested iteration

  $\rightsquigarrow$ numerical solution scheme with optimal complexity
- For non-smooth solutions:
  proofs of convergence and optimal complexity based on adaptive wavelets

Extensions and Outlook

- Control problems constrained by elliptic or parabolic PDEs
  with stochastic coefficients (uncertainty quantification)
- Multilevel Monte-Carlo methods or stochastic Galerkin schemes (generalized polynomial chaos approximations) for PDE-constrained control problems (and finite-dimensional noise assumption)

  [Gunzburger, Lee, Lee '11], [Hou, Lee, Manouzi '11], [Chen, Quarteroni, Rozza '13], [Gunzburger, Webster, Zhang '14], ...
- Infinitely many stochastic parameters [Kunoth, Schwab, SICON '13], [Kunoth, Schwab, SIAM JUQ '16]
- Deterministic PDE-constrained control problems:
  Adaptive methods based on finite elements ? Convergence and optimal complexity ?

  [Becker, Mao '11], [Gong, Liu, Tan, Yan '18] ...

  One or different meshes for all variables ?