A regularized least-squares method for sparse low-rank approximation of multivariate functions

Mathilde Chevreuil

joint work with Prashant Rai, Loic Giraldi, Anthony Nouy

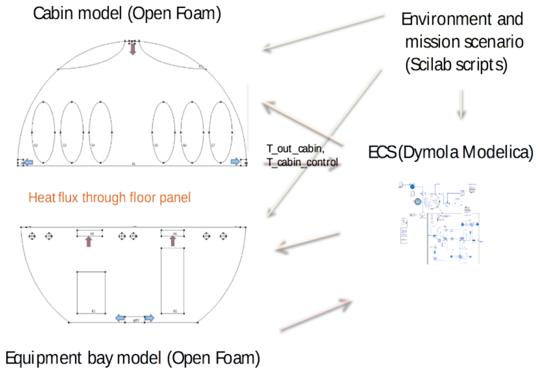
GeM – Institut de Recherche en Génie Civil et Mécanique
LUNAM Université
UMR CNRS 6183 / Université de Nantes / Centrale Nantes

Institut de Recherche en Génie Civil et Mécanique

Motivations

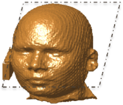
Chorus ANR project

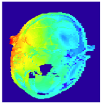
- Aero-thermal regulation in an aircraft cabin
  - 39 random parameters
  - Data basis of 2000 evaluations of the model



Motivations

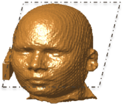
- In telecommunication: electromagnetic field and the Specific Absorption Rate (SAR) induced in the body
  - Over 4 random parameters
  - FDTD method: 2 days/run.
  - Few evaluations of the model available

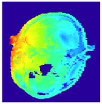




Motivations

- In telecommunication: electromagnetic field and the Specific Absorption Rate (SAR) induced in the body
  - Over 4 random parameters
  - FDTD method: 2 days/run.
  - Few evaluations of the model available





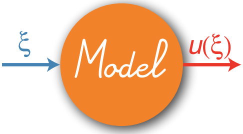
Aim

Construct a **surrogate model** of the true model **from a small collection of evaluations of the true model** that allows fast evaluations of output quantities of interest, observables or objective function.

- Propagation: estimation of quantiles, sensitivity analysis ...
- Optimization or identification
- Probabilistic inverse problem

- Stochastic/parametric models

Uncertainties represented by "simple" random variables $\xi = (\xi_1, \cdots, \xi_d) : \Theta \to \Xi$ defined on a probability space $(\Theta, \mathcal{B}, P)$.



Ideal approach

Compute an **accurate and explicit representation of** $u(\xi)$:

$$u(\xi) \approx \sum_{\alpha \in \mathcal{I}_P} u_\alpha \phi_\alpha(\xi), \quad \xi \in \Xi$$

where the $\phi_\alpha(\xi)$ constitute a suitable basis of multiparametric functions

- Polynomial chaos [Ghanem and Spanos 1991, Xiu and Karniadakis 2002, Soize and Ghanem 2004]
- Piecewise polynomials, wavelets [Deb 2001, Le Maître 2004, Wan 2005]

Motivations

- Aproximation spaces

$$\mathcal{S}_P = span\{\phi_\alpha(\xi) = \phi^{(1)}_{\alpha_1}(\xi_1)\dots\phi^{(d)}_{\alpha_d}(\xi_d); \alpha \in \mathcal{I}_P\}$$

with a pre-defined index set $\mathcal{I}_P$, *e.g.*

$$\left\{\alpha \in \mathbb{N}^d; |\alpha|_\infty \leq r\right\} \supset \left\{\alpha \in \mathbb{N}^d; |\alpha|_1 \leq r\right\} \supset \left\{\alpha \in \mathbb{N}^d; |\alpha|_q \leq r\right\},\ 0 < q < 1$$

Issue

- Approximation of a high dimensional function $u(\xi)$, $\xi \in \Xi \subset \mathbb{R}^d$

$$\boxed{\#(\mathcal{I}_P) \approx 10, 10^{10}, 10^{100}, 10^{1000}, \dots}$$

- Use of deterministic solvers in a black box manner

Numerous evaluations of possibly fine deterministic models

Motivations

Objective

Compute an approximation of $u \in \mathcal{S}_P$

$$u(\xi) \approx \sum_{\alpha \in \mathcal{I}_P} u_\alpha \phi_\alpha(\xi)$$

using few samples $\{u(y^q)\}_{q=1}^{Q}$

where $\{y^q\}_{q=1}^{Q}$ is a collection of sample points and the $u(y^q)$ are solutions of the deterministic problem

Exploit structures of $u(\xi)$

- $u(\xi)$ can be sparse on particular basis functions
- $u(\xi)$ can have suitable low rank representations

Can we exploit sparsity within low rank structure of u?

Outline

Outline

Low rank approximation

Approximation of function u using tensor approximation methods

- Exploit the tensor structure of function space

$$\mathcal{S}_P = \mathcal{S}^1_{P_1} \otimes \ldots \otimes \mathcal{S}^d_{P_d}; \qquad \mathcal{S}^k_{P_k} = \text{span}\left(\phi_i^{(k)}\right)_{i=1}^{P_k}$$

Low rank approximation

Approximation of function u using tensor approximation methods

- Exploit the tensor structure of function space

$$\mathcal{S}_P = \mathcal{S}^1_{P_1} \otimes \ldots \otimes \mathcal{S}^d_{P_d}; \qquad \mathcal{S}^k_{P_k} = \text{span}\left(\phi_i^{(k)}\right)_{i=1}^{P_k}$$

- Low rank tensor subsets $\mathcal{M}$

$$\mathcal{M} = \{v = F_{\mathcal{M}}(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n)\}$$

with $\boxed{dim(\mathcal{M}) = O(d)}$

[Nouy 2010, Khoromskij and Schwab 2010, Ballani 2010, Beylkin *et al* 2011, Matthies and Zander 2012, Doostan *et al* 2012, ...]

Low rank approximation

Approximation of function u using tensor approximation methods

- Exploit the tensor structure of function space

$$\mathcal{S}_P = \mathcal{S}^1_{P_1} \otimes \ldots \otimes \mathcal{S}^d_{P_d}; \qquad \mathcal{S}^k_{P_k} = \text{span}\left(\phi_i^{(k)}\right)_{i=1}^{P_k}$$

- Low rank tensor subsets $\mathcal{M}$

$$\mathcal{M} = \{v = F_{\mathcal{M}}(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n)\}$$

with $\boxed{dim(\mathcal{M}) = O(d)}$

[Nouy 2010, Khoromskij and Schwab 2010, Ballani 2010, Beylkin *et al* 2011, Matthies and Zander 2012, Doostan *et al* 2012, ...]

- Sparse low rank tensor subsets $\mathcal{M}^{m\text{-sparse}}$, ideally

$$\mathcal{M}^{m\text{-sparse}} = \{v = F_{\mathcal{M}}(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n);\ \|\mathbf{p}_i\|_0 \leq m_i;\ 1 \leq i \leq n\}$$

with $\boxed{dim(\mathcal{M}^{m\text{-sparse}}) \ll dim(\mathcal{M})}$.

Low rank approximation

Least-squares in low rank subsets

- Approximation of $v(\xi) \in \mathcal{M}$ defined by

$$\min_{v \in \mathcal{M}} \|u - v\|_Q^2 \quad \text{with} \quad \|u - v\|_Q^2 = \sum_{k=1}^{Q} |u(y^k) - v(y^k)|^2$$

[Beylkin *et al* 2011, Doostan *et al* 2012]

Low rank approximation

Least-squares in low rank subsets

- Approximation of $v(\xi) \in \mathcal{M}$ defined by

$$\min_{v \in \mathcal{M}} \|u - v\|_Q^2 \quad \text{with} \quad \|u - v\|_Q^2 = \sum_{k=1}^{Q} |u(y^k) - v(y^k)|^2$$

[Beylkin *et al* 2011, Doostan *et al* 2012]

- Approximation of $v(\xi) \in \mathcal{M}^{m-sparse}$ defined by

$$\min_{v \in \mathcal{M}} \|u - v\|_Q^2 \;\; \text{s.t.} \;\; \|\mathbf{p}_i\|_0 \leq m_i$$

Low rank approximation

Least-squares in low rank subsets

- Approximation of $v(\xi) \in \mathcal{M}$ defined by

$$\min_{v \in \mathcal{M}} \|u - v\|_Q^2 \quad \text{with} \quad \|u - v\|_Q^2 = \sum_{k=1}^{Q} |u(y^k) - v(y^k)|^2$$

[Beylkin *et al* 2011, Doostan *et al* 2012]

- Approximation of $v(\xi) \in \mathcal{M}^{m-sparse}$ defined by

$$\min_{v \in \mathcal{M}} \|u - v\|_Q^2 \ \text{ s.t. } \ \|\mathbf{p}_i\|_0 \leq m_i \rightarrow \boxed{\min_{v \in \mathcal{M}} \|u - v\|_Q^2 + \sum_{i=1}^{n} \lambda_i \|\mathbf{p}_i\|_1} \ \text{(Lasso)}$$

Low rank approximation

Least-squares in low rank subsets

- Approximation of $v(\xi) \in \mathcal{M}$ defined by

$$\min_{v \in \mathcal{M}} \|u - v\|_Q^2 \quad \text{with} \quad \|u - v\|_Q^2 = \sum_{k=1}^{Q} |u(y^k) - v(y^k)|^2$$

[Beylkin *et al* 2011, Doostan *et al* 2012]

- Approximation of $v(\xi) \in \mathcal{M}^{m-sparse}$ defined by

$$\min_{v \in \mathcal{M}} \|u - v\|_Q^2 \ \text{ s.t. } \ \|\mathbf{p}_i\|_0 \leq m_i \rightarrow \boxed{\min_{v \in \mathcal{M}} \|u - v\|_Q^2 + \sum_{i=1}^{n} \lambda_i \|\mathbf{p}_i\|_1} \ \text{(Lasso)}$$

Alternating least-squares with sparse regularization

For $1 \leq i \leq n$ and for fixed $\mathbf{p}_j$ with $j \neq i$

$$\min_{\mathbf{p}_i} \|u - F_{\mathcal{M}}(\mathbf{p}_1, \ldots, \mathbf{p}_i, \ldots, \mathbf{p}_n)\|_Q^2 + \lambda_i \|\mathbf{p}_i\|_1$$

Outline

Approximation in canonical tensor subset

Rank-one canonical tensor subset

$$\mathcal{R}_1 = \left\{ w = w^{(1)} \otimes \ldots \otimes w^{(d)} \; ; \; w^{(k)} \in \mathcal{S}^k_{P_k} \text{ s.t. } w^{(k)}(\xi_k) = \phi^{(k)}(\xi_k)^T \mathbf{w}^{(k)} \right\}$$

Approximation in canonical tensor subset

Rank-one canonical tensor subset

$$\mathcal{R}_1 = \left\{ w = \langle \boldsymbol{\phi}, \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(d)} \rangle ; \mathbf{w}^{(k)} \in \mathbb{R}^{P_k} \right\}$$

where $\boldsymbol{\phi} = \left(\boldsymbol{\phi}^{(1)} \otimes \ldots \otimes \boldsymbol{\phi}^{(d)} \right)(\xi)$ and with $dim(\mathcal{R}_1) = \sum_{k=1}^{d} P_k$

Approximation in canonical tensor subset

Rank-one canonical tensor subset

$$\mathcal{R}_1^{\gamma} = \left\{ w = \langle \boldsymbol{\phi}, \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(d)} \rangle ; \mathbf{w}^{(k)} \in \mathbb{R}^{P_k}, \|\mathbf{w}^{(k)}\|_1 \leq \gamma_k \right\}$$

where $\boldsymbol{\phi} = \left(\boldsymbol{\phi}^{(1)} \otimes \ldots \otimes \boldsymbol{\phi}^{(d)} \right)(\xi)$ and with $dim(\mathcal{R}_1^{\gamma}) = \sum_{k=1}^{d} P_k$

Approximation in canonical tensor subset

Rank-one canonical tensor subset

$$\mathcal{R}_1^{\gamma} = \left\{ w = \langle \boldsymbol{\phi}, \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(d)} \rangle ; \mathbf{w}^{(k)} \in \mathbb{R}^{P_k}, \|\mathbf{w}^{(k)}\|_1 \leq \gamma_k \right\}$$

where $\boldsymbol{\phi} = \left(\boldsymbol{\phi}^{(1)} \otimes \ldots \otimes \boldsymbol{\phi}^{(d)} \right)(\xi)$ and with $dim(\mathcal{R}_1^{\gamma}) = \sum_{k=1}^{d} P_k$

Rank-m tensor subsets

$$\mathcal{R}_m^{\gamma^1,\ldots,\gamma^m} = \{ v = \sum_{i=1}^{m} w_i \; ; w_i \in \mathcal{R}_1^{\gamma^i} \}$$

Approximation in canonical tensor subset

Rank-one canonical tensor subset

$$\mathcal{R}_1^{\gamma} = \left\{ w = \langle \boldsymbol{\phi}, \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(d)} \rangle ; \mathbf{w}^{(k)} \in \mathbb{R}^{P_k}, \|\mathbf{w}^{(k)}\|_1 \leq \gamma_k \right\}$$

where $\boldsymbol{\phi} = \left(\boldsymbol{\phi}^{(1)} \otimes \ldots \otimes \boldsymbol{\phi}^{(d)} \right)(\xi)$ and with $dim(\mathcal{R}_1^{\gamma}) = \sum_{k=1}^{d} P_k$

Rank-m tensor subsets

$$\begin{aligned} \mathcal{R}_m^{\gamma^1, \ldots, \gamma^m} &= \{ v = \sum_{i=1}^{m} w_i \; ; w_i \in \mathcal{R}_1^{\gamma^i} \} \\ &= \left\{ v = \langle \boldsymbol{\phi}, \sum_{i=1}^{m} \mathbf{w}_i^{(1)} \otimes \ldots \otimes \mathbf{w}_i^{(d)} \rangle ; \|\mathbf{w}_i^{(k)}\|_1 \leq \gamma_k^i \right\} \end{aligned}$$

Algorithm for adaptive sparse tensor approximation

- Algorithms
  - Progressive construction based on corrections in $\mathcal{R}_1^\gamma$
    - greedy construction of a basis $\{w_i\}_{i=1}^m$ selected in a tensor subset $\mathcal{R}_1^{\gamma^i}$
    - Compute $u_m = \sum_{i=1}^m \alpha_i w_i$ using regularized least-squares

  [MC, R. Lebrun, A. Nouy, P. Ray, *A least-squares method for sparse low rank approximation of multivariate functions*, arXiv:1305.0030, 2013]

  - Direct approximation in $\mathcal{R}_m^{\gamma^1,\ldots,\gamma^m}$

Algorithm for adaptive sparse tensor approximation

Algorithm for progressive construction

Let $u_0 = 0$. For $m \geq 1$,

- Compute a sparse rank-one correction $w_m \in \mathcal{R}_1^\gamma$ by solving

$$\min_{w \in \mathcal{R}_1^\gamma} \|u - u_{m-1} - w\|_Q^2$$

  Computed using alternating minimization on the parameters of $\mathcal{R}_1^\gamma$.

Algorithm for adaptive sparse tensor approximation

Algorithm for progressive construction

Let $u_0 = 0$. For $m \geq 1$,

- Compute a sparse rank-one correction $w_m \in \mathcal{R}_1^{\gamma}$ by solving

$$\min_{\mathbf{w}^{(1)} \in \mathbb{R}^{P_1}, \dots, \mathbf{w}^{(d)} \in \mathbb{R}^{P_d}} \| u - u_{m-1} - \langle \phi, \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(d)} \rangle \|_Q^2 + \sum_{k=1}^{d} \lambda_k \| \mathbf{w}^{(k)} \|_1$$

  Computed using Alternating regularized Least Squares

Algorithm for adaptive sparse tensor approximation

Algorithm for progressive construction

Let $u_0 = 0$. For $m \geq 1$,

- Compute a sparse rank-one correction $w_m \in \mathcal{R}_1^\gamma$ by solving

$$\min_{\mathbf{w}^{(1)} \in \mathbb{R}^{P_1}, \ldots, \mathbf{w}^{(d)} \in \mathbb{R}^{P_d}} \|u - u_{m-1} - \langle \phi, \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(d)} \rangle\|_Q^2 + \sum_{k=1}^{d} \lambda_k \|\mathbf{w}^{(k)}\|_1$$

  Computed using Alternating regularized Least Squares: for $1 \leq j \leq d$

$$\min_{\mathbf{w}^{(j)} \in \mathbb{R}^{n_j}} \|\mathbf{z} - \mathbf{\Phi}^{(j)} \mathbf{w}^{(j)}\|_2^2 + \lambda_j \|\mathbf{w}^{(j)}\|_1 \qquad \text{where } (\mathbf{\Phi}^{(j)})_{qi} = \phi_i^{(j)}(y_j^q) \prod_{k \neq j} w^{(k)}(y_k^q)$$

  - Lasso problem computed with LARS algorithm
  - Optimal solution selected using the fast LOO CV error estimate [Blatman, Sudret 2011]

Algorithm for adaptive sparse tensor approximation

Algorithm for progressive construction

Let $u_0 = 0$. For $m \geq 1$,

- Compute a sparse rank-one correction $w_m \in \mathcal{R}_1^{\gamma}$ by solving

$$\min_{\mathbf{w}^{(1)} \in \mathbb{R}^{P_1}, \ldots, \mathbf{w}^{(d)} \in \mathbb{R}^{P_d}} \|u - u_{m-1} - \langle \phi, \mathbf{w}^{(1)} \otimes \ldots \otimes \mathbf{w}^{(d)} \rangle\|_Q^2 + \sum_{k=1}^{d} \lambda_k \|\mathbf{w}^{(k)}\|_1$$

  Computed using Alternating regularized Least Squares: for $1 \leq j \leq d$

$$\min_{\mathbf{w}^{(j)} \in \mathbb{R}^{n_j}} \|\mathbf{z} - \mathbf{\Phi}^{(j)} \mathbf{w}^{(j)}\|_2^2 + \lambda_j \|\mathbf{w}^{(j)}\|_1 \qquad \text{where } (\mathbf{\Phi}^{(j)})_{qi} = \phi_i^{(j)}(y_j^q) \prod_{k \neq j} w^{(k)}(y_k^q)$$

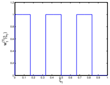
  - Lasso problem computed with LARS algorithm
  - Optimal solution selected using the fast LOO CV error estimate [Blatman, Sudret 2011]
- Set $U_m = span\{w_i\}_{i=1}^m$ (reduced approximation space)
- Compute $u_m = \sum_{i=1}^m \alpha_i w_i \in \mathcal{R}_m^{\gamma^1, \ldots, \gamma^m}$ using sparse regularization

$$\min_{\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m} \|u - \sum_{i=1}^{m} \alpha_i w_i\|_Q^2 + \lambda' \|\boldsymbol{\alpha}\|_1$$

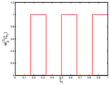
Best rank selected using cross validation method

Illustration: checker-board function

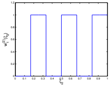
- Rank-2 function: $u(\xi_1, \xi_2) = \sum_{i=1}^{2} w_i^{(1)}(\xi_1) w_i^{(2)}(\xi_2)$



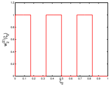
(a) $w_1^{(1)}(\xi_1)$



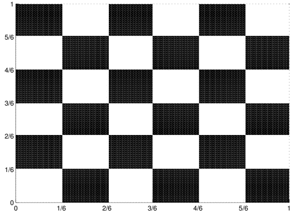
(b) $w_2^{(1)}(\xi_1)$



(c) $w_1^{(2)}(\xi_2)$



(d) $w_2^{(2)}(\xi_2)$



with dimension: $d = 2$
$\xi_i \in U(0,1)$. $\Xi = (0,1)^2$.

- Approximation of u in $\mathcal{S}^1_{P_1} \otimes \mathcal{S}^2_{P_2}$

Piecewise polynomials of degree p defined on a uniform partition of Ξ_k of s intervals:

$$\boxed{\mathcal{S}^k_{P_k} = \mathbb{P}_{p,s}}$$

Illustration: checker-board function

- Performance of the method for sparse low rank approximation
  - $Q = 200$ samples
  - Optimal rank m_{op} selected using 3-fold cross validation
  - Relative error ε estimated with Monte Carlo integration

Comparison of different regularizations within Alternated Least Squares

	OLS		ℓ_2		ℓ_1	
Approximation space	ε	m_{op}	ε	m_{op}	ε	m_{op}
$\mathcal{R}_m(\mathbb{P}_{2,3} \otimes \mathbb{P}_{2,3}), P = 9^2$	0.527	2	0.508	2	0.507	2
$\mathcal{R}_m(\mathbb{P}_{2,6} \otimes \mathbb{P}_{2,6}), P = 18^2$	0.664	2	0.061	8	$2.41\,10^{-13}$	2
$\mathcal{R}_m(\mathbb{P}_{2,12} \otimes \mathbb{P}_{2,12}), P = 36^2$	-	-	0.566	4	$1.50\,10^{-12}$	3
$\mathcal{R}_m(\mathbb{P}_{10,6} \otimes \mathbb{P}_{10,6}), P = 66^2$	-	-	0.855	10	$7.88\,10^{-13}$	2

With few samples:

- ℓ_1-regularization detects sparsity and gives accurate results
- Rank 2 is retrieved

Illustration: Friedman function

- Friedman function

$$f(\xi) = 10sin(\pi\xi_1\xi_2) + 20(\xi_3 - 0.5)^2 + 10\xi_4 + 5\xi_5$$

Dimensions: $d = 5$
$\xi_i, i = 1, \ldots, 5$ are uniform random variables over $[0, 1]$.

- Approximation in $\mathcal{S}_P = \bigotimes_{k=1}^{5} \mathcal{S}_{P_k}^k$

Polynomials of degree p: $\mathcal{S}_{P_k}^k = \mathbb{P}_p$

Illustration: Friedman function

- Friedman function

$$f(\xi) = 10sin(\pi\xi_1\xi_2) + 20(\xi_3 - 0.5)^2 + 10\xi_4 + 5\xi_5$$

Dimensions: $d = 5$
$\xi_i, i = 1, \ldots, 5$ are uniform random variables over $[0, 1]$.

- Approximation in $\mathcal{S}_P = \bigotimes_{k=1}^{5} \mathcal{S}_{P_k}^k$

Polynomials of degree p: $\mathcal{S}_{P_k}^k = \mathbb{P}_p$

What is the sufficient number of samples Q^* just needed given an *a priori* underlying approximation space ?

$$Q^* = f(p, d, m)$$

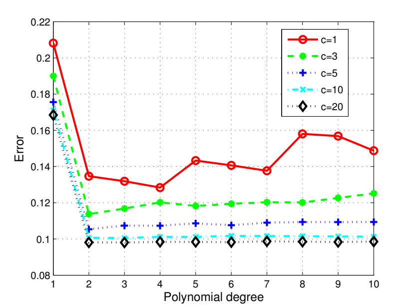
[G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone, 2011]: sufficient condition for a stable approximation of a multivariate function using OLS: $Q^* \sim (\#(\mathcal{I}_P))^2$

Illustration: Friedman function

- Number of samples needed (no regularization): rank-1

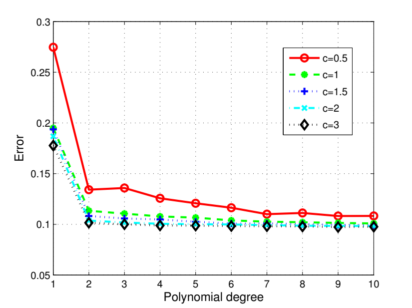
Number of samples: $Q = cd(p+1)$

$c \in \mathbb{R}_+^*$



Number of samples: $Q = cd(p+1)^2$

$c \in \mathbb{R}_+^*$



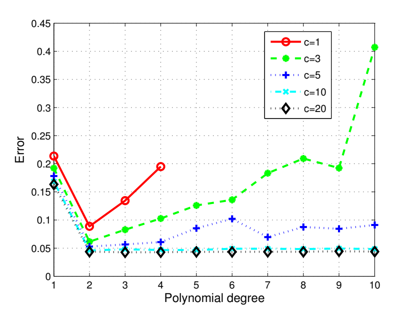
$$Q^* = d(p+1)^2$$

Illustration: Friedman function

- Number of samples needed (no regularization): rank-4

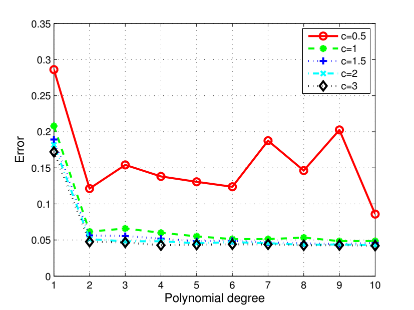
Number of samples: $Q = cmd(p+1)$

$c \in \mathbb{R}_+^*$

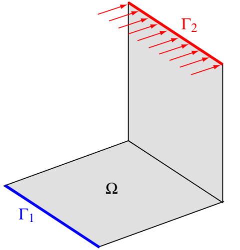


Number of samples: $Q = cmd(p+1)^2$

$c \in \mathbb{R}_+^*$



$$Q^* = md(p+1)^2$$



- Discrete problem

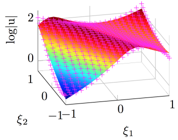
$$\mathbf{u} \in \mathbb{C}^N, \quad (-\omega^2 \mathbf{M} + i\omega \mathbf{C} + \mathbf{K})\mathbf{u} = \mathbf{f}$$

where $\mathbf{K} = E\tilde{\mathbf{K}}$ and $\mathbf{C} = i\omega\eta E\tilde{\mathbf{K}}$ with

$$E = \begin{cases} 0.975 + 0.025\xi_1 & \text{on horizontal plate,} \\ 0.975 + 0.025\xi_2 & \text{on vertical plate,} \end{cases}$$

$$\eta = \begin{cases} 0.0075 + 0.0025\xi_3 & \text{on horizontal plate,} \\ 0.0075 + 0.0025\xi_4 & \text{on vertical plate,} \end{cases}$$

where the $\xi_k \sim U(-1,1)$, $k = 1, \cdots, 4$. $\Xi = (-1,1)^8$.



- Approximation of a Variable of Interest $I(u)$ in $\mathcal{S}_P = \bigotimes_{k=1}^{5} \mathcal{S}_{P_k}^k$

$$I(u)(\xi) = \log \|u_c\|,$$

Polynomials of degree p: $\mathcal{S}_{P_k}^k = \mathbb{P}_p$

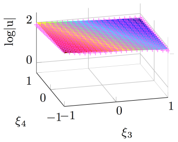
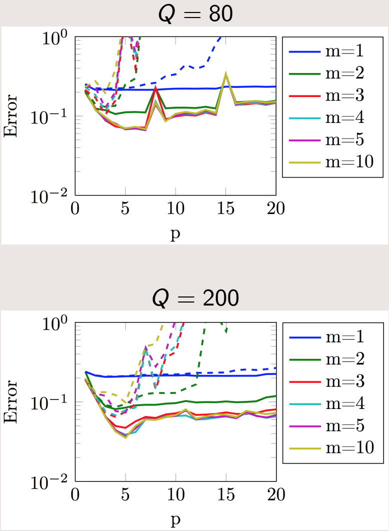
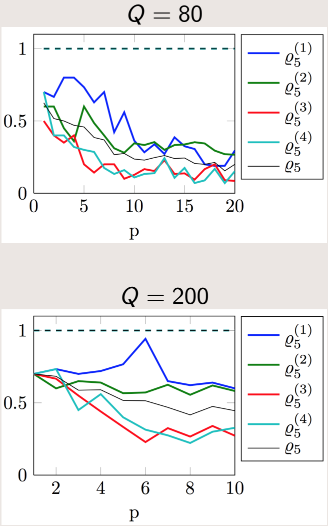


Illustration: vibration analysis

Evolution of error w.r.t. p



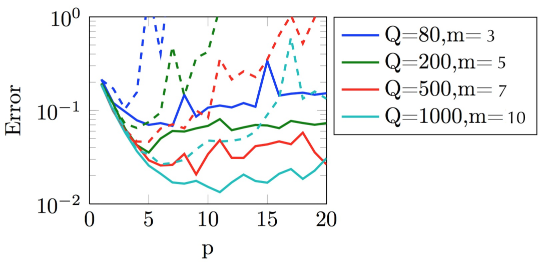
Sparsity ratio w.r.t. p



dashed lines: OLS, solid lines: with ℓ_1 regularization

Illustration: vibration analysis

Evolution of error w.r.t. p

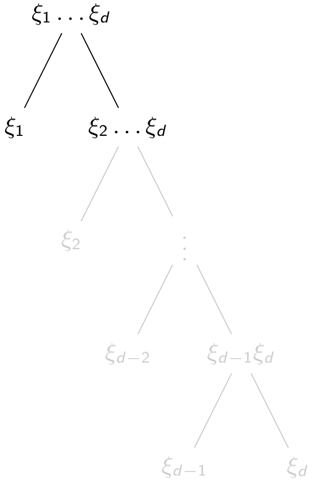


dashed lines: OLS, solid lines: with ℓ_1 regularization

Outline

1. Motivations and framework
2. Sparse low rank approximation
3. Tensor formats and algorithms
   - Canonical decomposition
   - Tensor Train format
4. Conclusion

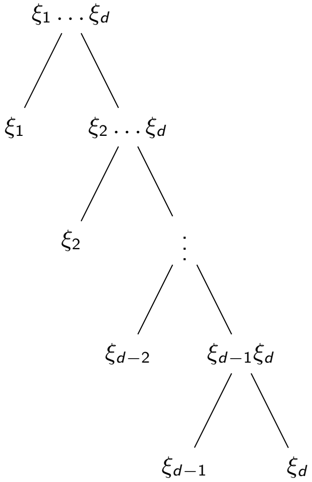
Tensor Train format



$$v = \sum_{i_1=1}^{r_1} v_{1,i_1}^{(1)} \otimes v_{i_1}^{(2,\dots,d)}$$

[Oseldets 2009,...]

Tensor Train format



[Oseldets 2009,...]

$$v = \sum_{i_1=1}^{r_1} v^{(1)}_{1,i_1} \otimes v^{(2,\dots,d)}_{i_1}$$

$$v = \sum_{i_1=1}^{r_1} v^{(1)}_{1,i_1} \otimes \sum_{i_2=1}^{r_2} v^{(2)}_{i_1 i_2} \otimes \dots \otimes \sum_{i_{d-1}=1}^{r_{d-1}} v^{(d-1)}_{i_{d-2} i_{d-1}} \otimes v^{(d)}_{i_{d-1},1}$$

Tensor Train subsets $\mathcal{TT}_{(1,r_1,\dots,r_{d-1},1)} = \mathcal{TT}_r$

The set of tensors $\mathcal{TT}_r(\mathcal{S})$ is defined by

$$\mathcal{TT}_r = \left\{ v = \sum_{i \in \mathcal{I}} \bigotimes_k v^{(k)}_{i_{k-1} i_k};\ \ v^{(k)}_{i_{k-1} i_k} \in \mathcal{S}^k_{P_k} \right\}.$$

where $\mathcal{I} = \{ i = (i_0, i_1, \dots, i_{d-1}, i_d);\ \ i_k \in \{1, \dots, r_k\}\}$
with $r_0 = r_d = 1$

Parameterization

$$\mathcal{TT}_r = \left\{ v = F_r(\mathbf{v}_1, \dots, \mathbf{v}_d); \mathbf{v}_k \in \left(\mathbb{R}^{P_k}\right)^{r_{k-1} \times r_k} \right\}$$

Algorithms

Alternating least-squares in $\mathfrak{TT}_r$

- For a given rank vector r

$$\min_{v \in \mathfrak{TT}_r} \|u - v\|_Q^2 + \sum_{k=1}^{d} \lambda_k \|vec(\mathbf{v}_k)\|_1$$

- Question of selection of rank vector r

Algorithm for adaptive sparse tensor approximation: DMRG

Re-parameterization

- Consider the tensor $w^{(k)} \in (\mathcal{S}^{k}_{P_k})^{r_{k-1}} \otimes (\mathcal{S}^{k+1}_{P_{k+1}})^{r_{k+1}}$: $w^{(k)} = \sum_{i_k=1}^{r_k^*} v_{i_k}^{(k,*)} \otimes v_{i_k}^{(k+1,*)}$

$$\Rightarrow v = F_r^k(v, w^{(k)}) = \sum_{i_1=1}^{r_1} \dots \sum_{i_{k-1}=1}^{r_{k-1}} \sum_{i_{k+1}=1}^{r_{k+1}} \dots \sum_{i_{d-1}=1}^{r_{d-1}} v_{1 i_1}^{(1)} \otimes \dots \otimes w_{i_{k-1} i_{k+1}}^{(k)} \otimes \dots \otimes v_{i_{d-1} 1}^{(d)}$$

- Compute sparse low-rank $w^{(k)}$ with adaptive rank

Modified alternating least-squares algorithm

For $k \in \{1, \cdots, d-1\}$

- Compute $w^{(k)} \in (\mathcal{S}^{k}_{P_k})^{r_{k-1}} \otimes (\mathcal{S}^{k+1}_{P_{k+1}})^{r_{k+1}}$ by solving

$$\min_{\mathbf{w}^{(k)} \in \mathbb{R}^{(r_{k-1} P_k) \times (r_{k+1} P_{k+1})}} \left\| \mathrm{u} - \mathrm{F}^k(\mathrm{v}, \mathbf{w}^{(k)}) \right\|_Q^2$$

Algorithm for adaptive sparse tensor approximation: DMRG

Re-parameterization

- Consider the tensor $w^{(k)} \in (\mathcal{S}^k_{P_k})^{r_{k-1}} \otimes (\mathcal{S}^{k+1}_{P_{k+1}})^{r_{k+1}}$: $w^{(k)} = \sum_{i_k=1}^{r_k^*} v_{i_k}^{(k,*)} \otimes v_{i_k}^{(k+1,*)}$

$$\Rightarrow v = F_r^k(v, w^{(k)}) = \sum_{i_1=1}^{r_1} \dots \sum_{i_{k-1}=1}^{r_{k-1}} \sum_{i_{k+1}=1}^{r_{k+1}} \dots \sum_{i_{d-1}=1}^{r_{d-1}} v_{1 i_1}^{(1)} \otimes \dots \otimes w_{i_{k-1} i_{k+1}}^{(k)} \otimes \dots \otimes v_{i_{d-1} 1}^{(d)}$$

- Compute sparse low-rank $w^{(k)}$ with adaptive rank

Modified alternating least-squares algorithm

For $k \in \{1, \cdots, d-1\}$

- Compute sparse $w^{(k)} \in (\mathcal{S}^k_{P_k})^{r_{k-1}} \otimes (\mathcal{S}^{k+1}_{P_{k+1}})^{r_{k+1}}$ by solving

$$\min_{\mathbf{w}^{(k)} \in \mathbb{R}^{(r_{k-1} P_k) \times (r_{k+1} P_{k+1})}} \left\| \mathrm{u} - \mathrm{F}^k(\mathrm{v}, \mathbf{w}^{(k)}) \right\|_Q^2 + \lambda_k \| vec(\mathbf{w}^{(k)}) \|_1$$

Algorithm for adaptive sparse tensor approximation: DMRG

Re-parameterization

- Consider the tensor $w^{(k)} \in (\mathcal{S}^k_{P_k})^{r_{k-1}} \otimes (\mathcal{S}^{k+1}_{P_{k+1}})^{r_{k+1}}$: $w^{(k)} = \sum_{i_k=1}^{r_k^*} v^{(k,*)}_{i_k} \otimes v^{(k+1,*)}_{i_k}$

$$\Rightarrow v = F^k_r(v, w^{(k)}) = \sum_{i_1=1}^{r_1} \dots \sum_{i_{k-1}=1}^{r_{k-1}} \sum_{i_{k+1}=1}^{r_{k+1}} \dots \sum_{i_{d-1}=1}^{r_{d-1}} v^{(1)}_{1 i_1} \otimes \dots \otimes w^{(k)}_{i_{k-1} i_{k+1}} \otimes \dots \otimes v^{(d)}_{i_{d-1} 1}$$

- Compute sparse low-rank $w^{(k)}$ with adaptive rank

Modified alternating least-squares algorithm

For $k \in \{1, \cdots, d-1\}$

- Compute sparse $w^{(k)} \in (\mathcal{S}^k_{P_k})^{r_{k-1}} \otimes (\mathcal{S}^{k+1}_{P_{k+1}})^{r_{k+1}}$ by solving

$$\min_{\mathbf{w}^{(k)} \in \mathbb{R}^{(r_{k-1}P_k) \times (r_{k+1}P_{k+1})}} \left\| \mathrm{u} - \mathrm{F}^k(\mathrm{v}, \mathbf{w}^{(k)}) \right\|_Q^2 + \lambda_k \| vec(\mathbf{w}^{(k)}) \|_1$$

- Compute best low-rank approximation in $(\mathcal{S}^k_{P_k})^{r_{k-1}} \otimes (\mathcal{S}^{k+1}_{P_{k+1}})^{r_{k+1}}$ using SVD $\rightarrow$ adaptive rank r_k^*

$$v = \sum_{i_1=1}^{r_1} \dots \sum_{i_k=1}^{r_k^*} \dots \sum_{i_{d-1}=1}^{r_{d-1}} v^{(1)}_{1 i_1} \otimes \dots \otimes v^{(k,*)}_{i_{k-1} i_k} \otimes v^{(k+1,*)}_{i_k i_{k+1}} \otimes \dots \otimes v^{(d)}_{i_{d-1} 1}$$

Illustration: sine of a sum

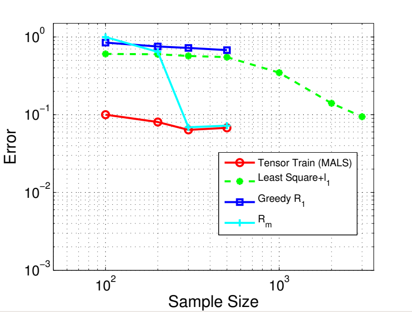
- Sine function:

$$u(\xi) = sin(\xi_1 + \xi_2 + \ldots + \xi_6)$$

with $\xi_i \in U(-1,1)$. $\Xi = (-1,1)^6$.

- Evolution of error with respect to sample size Q

Approx. in $\mathcal{S}_P = \bigotimes_{k=1}^{6} \mathcal{S}_{P_k}^k ; \mathcal{S}_{P_k}^k = \mathbb{P}_2$



Approx. in $\mathcal{S}_P = \bigotimes_{k=1}^{6} \mathcal{S}_{P_k}^k ; \mathcal{S}_{P_k}^k = \mathbb{P}_4$

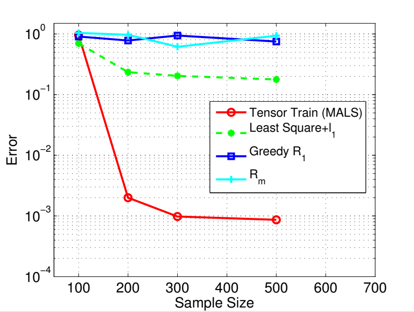


Illustration: borehole function

- The Borehole function models water flow through a borehole:

$$f(\xi) = \frac{2\pi T_u(H_u - H_l)}{\ln(r/r_w)\left(1 + \frac{2LT_u}{\ln(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l}\right)}$$

Dimension: $d = 8$

r_w	radius of borehole (m)	$\mathcal{N}(\mu = 0.10, \sigma = 0.0161812)$
r	radius of influence (m)	$\mathcal{LN}(\mu = 7.71, \sigma = 1.0056)$
T_u	transmissivity of upper aquifer (m^2/yr)	$\mathcal{U}[63070, 115600]$
H_u	potentiometric head of upper aquifer (m)	$\mathcal{U}[990, 1110]$
T_l	transmissivity of lower aquifer (m^2/yr)	$\mathcal{U}[63.1, 116]$
H_l	potentiometric head of lower aquifer (m)	$\mathcal{U}[700, 820]$
L	length of borehole (m)	$\mathcal{U}[1120, 1680]$
K_w	hydraulic conductivity of borehole (m/yr)	$\mathcal{U}[9855, 12045]$

- Approximation in $\mathcal{S}_P = \bigotimes_{k=1}^{8} \mathcal{S}_{P_k}^k$

Polynomials of degree p: $\mathcal{S}_{P_k}^k = \mathbb{P}_p$

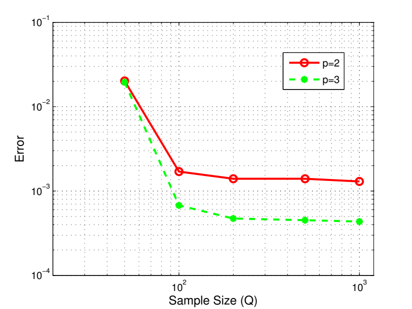
$p = 2$: $P = 6561$

$p = 3$: $P = 65536$

Illustration: borehole function

- Behavior of the algorithm

Evolution of error w.r.t. Q



TT ranks (Q=200, p=3)

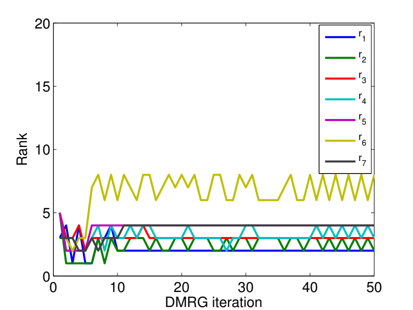
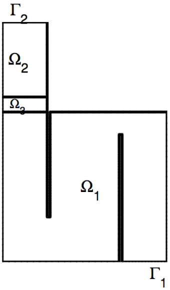


Illustration: Canister

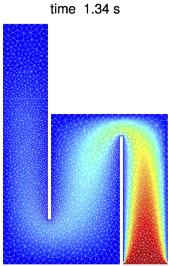
- Stochastic PDE

$$\begin{cases} \frac{\partial u}{\partial t} - \nabla(\kappa \nabla u) + c(D \cdot \nabla u) = \sigma u \quad \text{on} \quad \Omega_1 \cup \Omega_2 \\ u = \xi_1 \text{ on } \Gamma_1 \times \Omega_t \\ u = 0 \text{ on } \Gamma_2 \times \Omega_t \\ u_{,n} = 0 \text{ on } (\partial\Omega \backslash (\Gamma_1 \cup \Gamma_2)) \times \Omega_t \end{cases}$$



with

ξ_1	$u(t=0)$	$U[0.8, 1.2]$ on Ω
ξ_2	σ	$U[8, 12]$ on Ω_2
ξ_3	σ	$U[0.8, 1]$ on Ω_1
ξ_4	c	$U[1, 5]$
ξ_5	κ	$U[0.02, 0.03]$



- Approximation of a Variable of Interest $I(u)$ in $\mathcal{S}_P = \bigotimes_{k=1}^{5} \mathcal{S}_{P_k}^k$

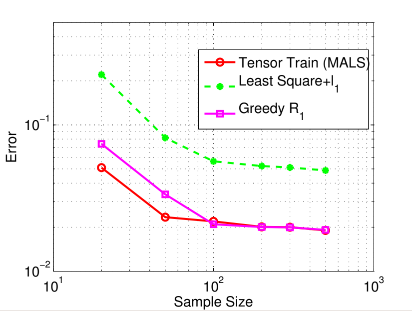
$$I(u) = \int_T \int_{\Omega_3} u(x,t) dxdt$$

Polynomials of degree p: $\boxed{\mathcal{S}_{P_k}^k = \mathbb{P}_p}$

Illustration: Canister

- Evolution of error with respect to sample size Q

Approx. in $\mathcal{S}_P = \bigotimes_{k=1}^{5} \mathcal{S}_{P_k}^k ; \mathcal{S}_{P_k}^k = \mathbb{P}_2$



Approx. in $\mathcal{S}_P = \bigotimes_{k=1}^{5} \mathcal{S}_{P_k}^k ; \mathcal{S}_{P_k}^k = \mathbb{P}_3$

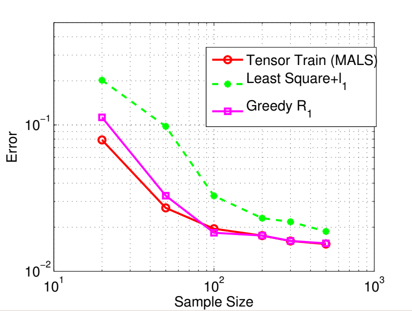
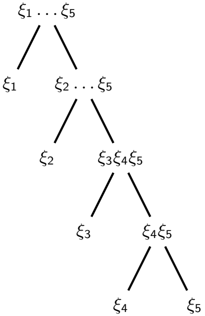


Illustration: Canister

- Order of separation of variables



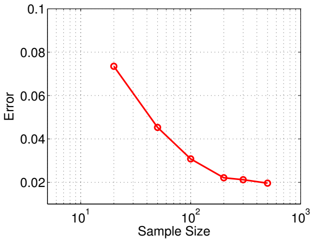
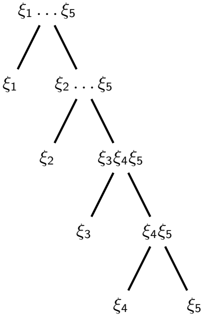
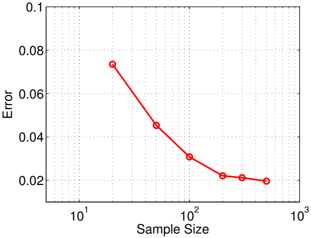


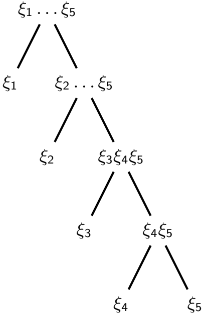
Illustration: Canister

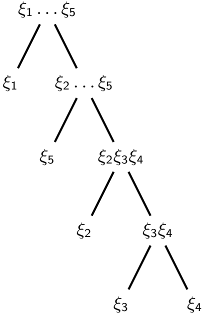
- Order of separation of variables

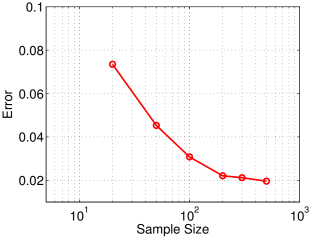


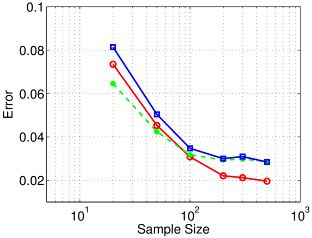


- Order of separation of variables









Outline

Conclusion

Least-squares method for sparse low rank approximation of high dimensional functions

- A non intrusive method
- Detects and exploits low-rank and sparsity
- Adaptive rank

Outlook

- More analyses on the suffisant number of samples to find an approximation in a tensor subset
- Include adaptivity with respect to polynomial degree for underlying approximation spaces
- Strategies for optimal separation of variables (choice of tree)

This research was supported by EADS Innovation Works
and by the French National Research Agency (ANR) (CHORUS project)