Low-rank tensor methods for parametric and stochastic problems

Part 2: Sparse and low-rank tensor methods for high-dimensional approximation

Anthony Nouy

Ecole Centrale Nantes / GeM
Nantes, France

What we learned from Part 1

$$u \in X = L^p_\mu(\Xi; \mathcal{V})$$

- Subspace-based model reduction related with low-rank approximation

$$u(\xi) \approx \sum_{i=1}^{n} v_i s_i(\xi) \in \mathcal{V}_n \otimes L^p_\mu(\Xi)$$

- Optimal reduced spaces $\mathcal{V}_n = span\{v_1, \ldots, v_n\}$ can be defined w.r.t. L^p_μ-norm
- Algorithms for quasi-optimal constructions: greedy or not greedy.

Remaining issue: high-dimensional approximation

$$u(\xi) = u(\xi_1, \ldots, \xi_d)$$

- Different types of approximations, usually linear approximation

$$u_\Lambda(\xi) \in X_\Lambda = \left\{ \sum_{\alpha \in \Lambda} u_\alpha \psi_\alpha(\xi) : u_\alpha \in \mathcal{V} \right\}$$

with classical bases: polynomials, piecewise polynomials, wavelets...

- Different types of constructions depending on the expected accuracy:
  - uniform accuracy (for optimization, quantile estimation, ...) :

$$\|u(\xi) - u_\Lambda(\xi)\| \leq \epsilon \quad \forall \xi \in \Xi$$

  - mean-squared accuracy (for statistical moments, global sensitivity...) :

$$\int_\Xi \|u(\xi) - u_\Lambda(\xi)\|^2 \mu(d\xi) \leq \epsilon$$

  - ...

- Different types of constructions depending on the available information:
  - point evaluations $\Rightarrow$ interpolation, approximate projection using quadrature, discrete projection
  - model equations $\Rightarrow$ Galerkin projection
- Quasi-optimality under continuity and stability properties

$$\|u - u_\Lambda\|_X \leq C \inf_{v \in X_\Lambda} \|u - v\|_X$$

- The analysis of best approximation error $\inf_{v \in X_\Lambda} \|u - v\|$ requires extra-information on approximation spaces and the solution (regularity w.r.t. ξ)

Outline

1. Polynomial approximation spaces
2. High-dimensional approximation - tractability
3. Best n-term approximation and quasi-optimal approximation spaces
4. Higher-order tensors and low-rank tensor formats
5. Approximation in low-rank tensor formats
6. Algorithms for computing low-rank approximations
7. Higher-order low-rank methods for parametric and stochastic equations

Outline

1. Polynomial approximation spaces
2. High-dimensional approximation - tractability
3. Best n-term approximation and quasi-optimal approximation spaces
4. Higher-order tensors and low-rank tensor formats
5. Approximation in low-rank tensor formats
6. Algorithms for computing low-rank approximations
7. Higher-order low-rank methods for parametric and stochastic equations

Notations, definitions

For a multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$, we use the following notations.

- For $1 \leq p < \infty$,

$$|\alpha|_p = \left(\sum_{k=1}^{d} \alpha_k^p \right)^{1/p},$$

  and

$$|\alpha|_\infty = \max_{1 \leq k \leq d} \alpha_k$$

- $|\alpha| = |\alpha|_1$
- $\alpha! = \prod_{k=1}^d \alpha_k!$
- For $\omega = (\omega_1, \dots, \omega_d) \in \mathbb{R}^d$, $\alpha^\omega = \prod_{k=1}^d \alpha_k^{\omega_k}$
- For $u : \Xi \to X$,

$$D^\alpha u(\xi) = \partial_{\xi_1}^{\alpha_1} \dots \partial_{\xi_d}^{\alpha_d} u(\xi_1, \dots, \xi_d) = \frac{\partial^{|\alpha|}}{\partial \xi_1^{\alpha_1} \dots \partial \xi_d^{\alpha_d}} u(\xi_1, \dots, \xi_d)$$

Polynomial spaces

- For a set of multi indices $\Lambda \subset \mathbb{N}^d$, we define the polynomial space

$$\mathbb{P}_\Lambda(\Xi) = \mathrm{span}\left\{ \xi^\alpha = \prod_{k=1}^{d} \xi_k^{\alpha_k} : \alpha \in \Lambda \right\}$$

- Orthonormal basis $\{\psi_\alpha\}_{\alpha \in \Lambda}$ of $\mathbb{P}_\Lambda(\Xi) \subset L^2_\mu(\Xi)$ such that

$$\mathbb{P}_\Lambda(\Xi) = \mathrm{span}\left\{ \psi_\alpha(\xi) : \alpha \in \Lambda \right\}$$

- If product measure $\mu = \mu_1 \otimes \ldots \otimes \mu_d$ on $\Xi_1 \times \ldots \times \Xi_d$, then

$$\psi_\alpha(\xi) = \psi^{(1)}_{\alpha_1}(\xi_1) \ldots \psi^{(d)}_{\alpha_d}(\xi_d)$$

  with orthonormal polynomials $\{\psi^{(k)}_j\}_{j \geq 0}$ in $L^2_{\mu_k}(\Xi_k)$.

Polynomial approximation spaces

- Full tensor product (bounded partial degree)

$$\Lambda = \{\alpha : \alpha_k \leq p_k, 1 \leq k \leq d\}$$

$$\#\Lambda = \prod_{k=1}^{d} (p_k + 1)$$

$$\mathbb{P}_\Lambda(\Xi) = \mathbb{P}_{p_1}(\Xi_1) \otimes \ldots \otimes \mathbb{P}_{p_d}(\Xi_d)$$

Convergence of polynomial approximations

- For $X_{\Lambda_p} = \mathcal{V} \otimes \mathbb{P}_{\Lambda_p}(\Xi)$, let

$$\|u - u_{\Lambda_p}\|_{L^2_\mu(\Xi;\mathcal{V})} = \inf_{v \in X_{\Lambda_p}} \|u - v\|_{L^2_\mu(\Xi;\mathcal{V})}$$

- We have

$$\|u - u_{\Lambda_p}\|_{L^2_\mu(\Xi;\mathcal{V})} \leq \inf_{v \in X_{\Lambda_p}} \|u - v\|_{L^\infty_\mu(\Xi;\mathcal{V})} := \epsilon_p^{(\infty)}$$

so that results in L^2 norm can be deduced from stronger results in L^∞ norm.

Convergence of polynomial approximations

The following result can be found in [Chen-Quarteroni-Rozza 2014].

- Consider $\Xi = (-1,1)^d$ and $\Lambda_p = \{\alpha : \alpha_k \leq p_k\}$ and

$$\mathbb{P}_{\Lambda_p}(\Xi) = \otimes_{k=1}^d \mathbb{P}_{p_k}(\Xi_k) \quad \text{(full tensor product space)}$$

- Assume $u : \Xi \to \mathcal{V}$ is analytic and can be analytically extended to $\widehat{\Xi} = \left\{z \in \mathbb{C}^d : dist(z_k, \Xi_k) \leq \tau_k, 1 \leq k \leq d\right\}$. Then

$$\boxed{\epsilon_p^{(\infty)} \leq \sum_{k=1}^d C_{p_k} r_k^{-p_k}}$$

  where $r_k = \sqrt{1+\tau_k^2} + \tau_k > 1$ and $C_{p_k} \leq C \log(p_k + 1)$.

- In the case $p_k = p_\star$ for all k, then with $r_\star = \min_k r_k$,

$$\epsilon_p^{(\infty)} \leq C_{p_\star} d r_\star^{-p_\star}$$

  or with $n = \#\Lambda_p = (p_\star + 1)^d$,

$$\boxed{\epsilon_p^{(\infty)} \leq C_n r_\star^{1-n^{1/d}}}$$

  where $C_n \leq C \log(n)$.

- Convergence deteriorates with the dimension d: curse of dimensionality

Piecewise polynomial approximations

See e.g. [Deb-Babuska 2001, Babuska-Tempone-Zouraris 2004, Wan-Karniadakis 2005, Frauenfelder-Schwab-Todor 2005].

- For a given partition $(K^t)_{t\in T}$ of Ξ, let

$$\mathcal{S}_{\Xi,\Lambda} = \left\{ v : \Xi \to \mathbb{R} : v_{|K^t} \in \mathbb{P}_{\Lambda_t}(K^t) \text{ for all } t \in T \right\}$$

- As a standard case, assume $\Xi = \times_k \Xi_k$ with Ξ_k a bounded interval. Introduce a partition made of boxes $K^t = \times_{k=1}^d K_k^t$ with $K_k^t = (a_k^t, b_k^t) \subset \Xi_k$ with maximal sizes $h_k = \max_t |b_k^t - a_k^t|$ in dimension k. Then introduce

$$\mathcal{S}_{h,p} = \left\{ v : \Xi \to \mathbb{R} : v_{|K^t} \in \mathbb{P}_p(K^t) = \otimes_{k=1}^d \mathbb{P}_{p_k}(K_k^t) \right\}$$

Piecewise polynomial approximations

- If $u \in C^{p+1}(\Xi; \mathcal{V})$, it holds

$$\epsilon_{h,p}^{(2)} := \inf_{v \in \mathcal{V} \otimes \mathcal{S}_{h,p}} \|u - v\|_{L^2_\mu(\Xi;\mathcal{V})} \leq C \sum_{k=1}^{d} \left(\frac{h_k}{2}\right)^{p_k+1} \frac{\|\partial_{\xi_k}^{p_k+1} u\|_{L^2_\mu(\Xi;\mathcal{V})}}{(p_k+1)!}$$

$$\leq \sum_{k=1}^{d} C_{p_k} h_k^{p_k+1} \|u\|_{C^{p+1}(\Xi;\mathcal{V})}$$

  with C_{p_k} depending on p_k.

- If $p_k = p_\star$ and $h_k = h$ for all k, then

$$\epsilon_{h,p}^{(2)} \leq C_{p_\star} d h^{p_\star+1} \|u\|_{C^{p_\star+1}(\Xi;\mathcal{V})}$$

  With $n = \dim(\mathcal{S}_{h,p}) \sim (h^{-1}(p_\star+1))^d$,

$$\epsilon_{h,p}^{(2)} \lesssim D_{p_\star} d n^{-\frac{p_\star+1}{d}}$$

- Convergence deteriorates with the dimension d: curse of dimensionality $\Rightarrow$ requires adaptivity.

Outline

1. Polynomial approximation spaces
2. High-dimensional approximation - tractability
3. Best n-term approximation and quasi-optimal approximation spaces
4. Higher-order tensors and low-rank tensor formats
5. Approximation in low-rank tensor formats
6. Algorithms for computing low-rank approximations
7. Higher-order low-rank methods for parametric and stochastic equations

High-dimensional approximation

- Consider the approximation of the multivariate function

$$u(x_1, \dots, x_d)$$

- Tensorization of bases can be extended to other types of bases:

$$\psi_\alpha(x_1, \dots, x_d) = \psi^{(1)}_{\alpha_1}(x_1) \dots \psi^{(d)}_{\alpha_d}(x_d)$$

- Tensor product discretization yields high-dimensional parametrizations

$$u(x_1, \dots, x_d) \approx \sum_{\alpha_1=1}^{n} \dots \sum_{\alpha_d=1}^{n} a_{\alpha_1 \dots \alpha_d} \psi^{(1)}_{\alpha_1}(x_1) \dots \psi^{(d)}_{\alpha_d}(x_d), \quad a \in \mathbb{R}^{n^d}$$

  - Number of parameters for full tensor product approximation

$$N = n^d$$

  - Assuming $u \in C^s((0,1)^d)$, accuracy

$$\epsilon = O(n^{-s}) = O(N^{-s/d})$$

    and the number of parameters to achieve accuracy ϵ is

$$N = O(\epsilon^{-d/s}) \quad \text{(Curse of dimensionality)}$$

Curse of dimensionality - tractability (Novak,Wozniakowski)

- Curse of dimensionality related to computational intractability in high dimension.
  - Quantitive definition through information-based complexity analysis
  - Tractability depends on the measure of precision and the available information.

- $N(\epsilon, d)$ being the number of linear informations to obtain a precision ϵ, intractability (curse of dimensionality) when

$$\lim_{\epsilon^{-1}+d\to\infty} \frac{\log N(\epsilon, d)}{\epsilon^{-1}+d} > 0$$

Curse of dimensionality - tractability (Novak,Wozniakowski)

- Weak tractability

$$\lim_{\epsilon^{-1}+d\to\infty} \frac{\log N(\epsilon, d)}{\epsilon^{-1} + d} = 0$$

- Polynomial tractability

$$N(\epsilon, d) \leq C\epsilon^{-p} d^q \quad \text{for all } \epsilon \text{ and } d$$

- Strong polynomial tractability

$$N(\epsilon, d) \leq C\epsilon^{-p} \quad \text{for all } \epsilon$$

Intractability for the approximation of smooth functions

Consider the set of functions

$$F_d = \left\{ u \in C^\infty((0,1)^d) : \sup_\alpha \|D^\alpha u\|_\infty < \infty \right\}$$

- Minimal approximation error using linear informations:

$$\epsilon(N, d) = \inf_{A_N} \sup_{u \in F_d} \|u - A_N(u)\|_\infty$$

  where the infimum is taken over all algorithms A_N using N linear informations on u and providing an approximation $A_N(u)$.

- Optimal rate of convergence is infinite: for arbitrary large s

$$\epsilon(N, d) = O(N^{-s}) \quad \text{as } N \to \infty$$

$$N(\epsilon, d) = O(\epsilon^{-1/s}) \quad \text{as } \epsilon \to 0$$

- But what about the constants in O ? [Novak 2009] proves

$$N(\epsilon, d) \geq 2^{\lfloor \frac{d}{2} \rfloor}$$

Curse of dimensionality !

Remedies: structured approximations

Order reduction methods must exploit specific structures (application dependent)

- Anisotropic smoothness
- Low effective dimensionality, e.g.

$$u(x_1, \dots, x_d) \approx g(x_1, x_2)$$

- Low-order interactions, e.g.

$$u(x_1, \dots, x_d) \approx u_0 + \sum_i u_i(x_i) + \sum_{i \neq j} u_{i,j}(x_i, x_j)$$

- Sparsity (relatively to a basis or frame)

$$u(x_1, \dots, x_d) = \sum_{\alpha \in \mathbb{N}^d} a_\alpha \psi_\alpha(x_1, \dots, x_d) \approx \sum_{\alpha \in \Lambda} a_\alpha \psi_\alpha(x_1, \dots, x_d)$$

Structured (sparse) tensorization

$$\mathcal{S}_\Lambda = span\left\{\psi_\alpha(x) = \psi^{(1)}_{\alpha_1}(x_1)\dots\psi^{(d)}_{\alpha_d}(x_d) : \alpha \in \Lambda\right\}$$

- Total degree: isotropic (left), anisotropic (right)

$$\Lambda = \{\alpha : \sum_k \alpha_k \leq p\} \qquad \Lambda = \{\alpha : \sum_k \omega_k \alpha_k \leq p\}$$

- Hyperbolic Cross: isotropic (left), anisotropic (right)

$$\Lambda = \{\alpha : \prod_k (\alpha_k + 1) \leq p + 1\} \qquad \Lambda = \{\alpha : \prod_k (\alpha_k + 1)^{\omega_k} \leq p + 1\}$$

And now the question is ....

Can we define optimal or quasi-optimal approximation spaces ?

Outline

1. Polynomial approximation spaces
2. High-dimensional approximation - tractability
3. Best n-term approximation and quasi-optimal approximation spaces
4. Higher-order tensors and low-rank tensor formats
5. Approximation in low-rank tensor formats
6. Algorithms for computing low-rank approximations
7. Higher-order low-rank methods for parametric and stochastic equations

Best n-term approximation

- Let $\{\psi_\alpha\}_{\alpha \in \mathbb{N}^d}$ be an orthonormal basis of $L^2_\mu(\Xi)$.
- Let $u \in X = L^2_\mu(\Xi; \mathcal{V})$ such that

$$u = \sum_\alpha u_\alpha \psi_\alpha, \quad u_\alpha = \mathbb{E}_\mu(u(\xi)\psi_\alpha(\xi))$$

- For a index set Λ, let

$$X_\Lambda = \left\{ u(\xi) = \sum_{\alpha \in \Lambda} u_\alpha \psi_\alpha(\xi) : u_\alpha \in \mathcal{V} \right\} \subset X$$

- The best approximation u_Λ of u in X_Λ is defined by

$$\|u - u_\Lambda\|_{L^2_\mu(\Xi;\mathcal{V})} = \min_{v \in X_\Lambda} \|u - v\|_{L^2_\mu(\Xi;\mathcal{V})}$$

  and such that

$$\|u - u_\Lambda\|^2_{L^2_\mu(\Xi;\mathcal{V})} = \| \sum_{\alpha \notin \Lambda} u_\alpha \psi_\alpha \|^2_{L^2_\mu(\Xi;\mathcal{V})} = \sum_{\alpha \notin \Lambda} \|u_\alpha\|^2_{\mathcal{V}}$$

Best n-term approximation

- Best n-term approximation

$$\sigma_n^{(2)} = \min_{\#\Lambda_n = n} \min_{v \in X_{\Lambda_n}} \|u - v\|_{L^2_\mu(\Xi;\mathcal{V})}$$

where the minimum is taken over all subsets Λ_n with cardinal n. Optimal Λ_n obtained by retaining the n largest coefficients $\|u_\alpha\|_{\mathcal{V}}$.

- First question: How fast $\sigma_n^{(2)}$ converges with n ?
- Second question : How to construct Λ_n in practice (since u_α are not available) and how $\|u - u_{\Lambda_n}\|_X$ does it compare with $\sigma_n^{(2)}$?

Convergence of best n-term approximation: some results

- If the sequence $(\|u_\alpha\|_{\mathcal{V}})_\alpha \in \ell^p$ with $p < 1$, then there exists Λ_n with $\#\Lambda_n = n$ such that
$$\|u - u_{\Lambda_n}\|_{L^2_\mu(\Xi;\mathcal{V})} \leq C n^{-s}, \quad C = \|(\|u_\alpha\|_{\mathcal{V}})_\alpha\|_{\ell^p}, \quad s = \frac{1}{p} - \frac{1}{2}$$
so that
$$\sigma_n^{(2)} \leq C n^{-s}$$
- See [Cohen-DeVore-Schwab 2010,Chkifa-Cohen-Schwab 2014] for a proof of $(\|u_\alpha\|_{\mathcal{V}})_\alpha \in \ell^p$ for a large class of parametric problems. Results are working for infinitely many random variables.

Remark on L^∞ case

- For $u \in L^\infty(\Xi; \mathcal{V})$, one can be interested in controlling the norm in $L^\infty(\Xi; \mathcal{V})$

$$\|u - u_\Lambda\|_{L^\infty(\Xi;\mathcal{V})} = \sup_{\xi \in \Xi} \|u(\xi) - u_\Lambda(\xi)\|_{\mathcal{V}}$$

  Let $u = \sum_\alpha v_\alpha \varphi_\alpha$ with $\varphi_\alpha = \psi_\alpha / \|\psi_\alpha\|_{L^\infty}$, then

$$\|u - u_\Lambda\|_{L^\infty(\Xi;\mathcal{V})} \leq \sum_{\alpha \notin \Lambda} \|v_\alpha\|_{\mathcal{V}}$$

- Best n-term approximation can be defined in the L^∞ norm

$$\sigma_n^{(\infty)} = \min_{\#\Lambda_n = n} \min_{v \in X_{\Lambda_n}} \|u - v\|_{L^\infty(\Xi;\mathcal{V})}$$

- If $(\|v_\alpha\|_{\mathcal{V}})_\alpha \in \ell^p$ with $p < 1$, then there exists Λ_n with $\#\Lambda_n = n$ such that

$$\|u - u_{\Lambda_n}\|_{L^\infty(\Xi;\mathcal{V})} \leq C n^{-s}, \quad C = \|(\|v_\alpha\|_{\mathcal{V}})_\alpha\|_{\ell^p}, \quad s = \frac{1}{p} - 1$$

  so that

$$\sigma_n^{(\infty)} \leq C n^{-s}$$

- See [Cohen-DeVore-Schwab 2010,Chkifa-Cohen-Schwab 2014] for a proof of $(\|v_\alpha\|_{\mathcal{V}})_\alpha \in \ell^p$ for a large class of problems.

Quasi-optimal index sets

- Assume that we know a bound $\delta(\alpha)$ such that

$$\|u_\alpha\|_{\mathcal{V}} \leq \delta(\alpha) \quad (1)$$

- Quasi-optimal index set Λ_n^δ obtained by retaining the n largest values $\delta(\alpha)$. Close to the optimal if the bound (1) is sharp.
- Estimation of $\delta(\alpha)$ based on a priori analysis (a priori definition of the sequence Λ_n^δ) or based on a posteriori analysis (adaptive construction).
- Suppose that there exists $\gamma \geq 1$ such that

$$\gamma^{-1}\delta(\alpha) \leq \|u_\alpha\|_{\mathcal{V}} \leq \delta(\alpha)$$

Then

$$\|u - u_{\Lambda_n^\delta}\|^2_{L^2_\mu(\Xi;\mathcal{V})} = \sum_{\alpha \notin \Lambda_n^\delta} \|u_\alpha\|^2_{\mathcal{V}} \leq \sum_{\alpha \notin \Lambda_n^\delta} \delta(\alpha)^2 = \min_{\#\Lambda_n = n} \sum_{\alpha \notin \Lambda_n} \delta(\alpha)^2 \leq \gamma^2 \min_{\#\Lambda_n = n} \sum_{\alpha \notin \Lambda_n} \|u_\alpha\|^2_{\mathcal{V}}$$

and therefore

$$\|u - u_{\Lambda_n^\delta}\|_{L^2_\mu(\Xi;\mathcal{V})} \leq \gamma \sigma_n^{(2)} \quad \text{(Quasi-optimality)}$$

Quasi-optimal index sets based on a priori analysis

- In practice, define a sequence of subsets

$$\Lambda_p = \{\alpha : \delta(\alpha) \geq \epsilon(p)\}$$

with $(\epsilon(p))_{p\geq 0}$ a decreasing sequence.

- Assume

$$\|u_\alpha\|_{\mathcal{V}} \leq Ce^{-\sum_k \omega_k \alpha_k} := \delta(\alpha) \qquad (\star)$$

If C independent of α, take $\epsilon(p) = Ce^{-p}$, so that

$$\Lambda_p = \left\{\alpha : \sum_k \omega_k \alpha_k \leq p\right\} \quad \text{(Anisotropic total degree)}$$

If $C = C(\alpha)$, take $\epsilon(p) = e^{-p}$, so that

$$\Lambda_p = \left\{\alpha : \sum_k \omega_k \alpha_k - \log(C(\alpha)) \leq p\right\} \quad \text{(modified Anisotropic total degree)}$$

- See [Back-Nobile-Tamellini-Tempone 2011, 2012, 2014] for a proof of $(\star)$ for some classes of parametric problems.

Outline

1. Polynomial approximation spaces
2. High-dimensional approximation - tractability
3. Best n-term approximation and quasi-optimal approximation spaces
4. Higher-order tensors and low-rank tensor formats
5. Approximation in low-rank tensor formats
6. Algorithms for computing low-rank approximations
7. Higher-order low-rank methods for parametric and stochastic equations

Tensor spaces

An algebraic tensor space $V = V_1 \otimes \ldots \otimes V_d$ is the set of elements of the form

$$u = \sum_{i=1}^{m} v_i^1 \otimes \ldots \otimes v_i^d$$

or for multivariate functions

$$u(x_1, \ldots, x_d) = \sum_{i=1}^{m} v_i^1(x_1) \ldots v_i^d(x_d).$$

A tensor Banach space $V_{\|\cdot\|}$ is obtained by the completion of the algebraic tensor space V with respect to a norm $\|\cdot\|$:

$$V_{\|\cdot\|} = \overline{V_1 \otimes \ldots \otimes V_d}^{\|\cdot\|}.$$

W. Hackbusch.
Tensor Spaces and Numerical Tensor Calculus,
Springer, 2012.

Examples of (Banach) tensor spaces

- Multidimensional array

$$a \in \mathbb{R}^{n_1 \times \ldots \times n_d} = \mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_d}$$

$$a = \sum_{i_1=1}^{n_1} \ldots \sum_{i_d=1}^{n_d} a_{i_1,\ldots,i_d} e^1_{i_1} \otimes \ldots \otimes e^d_{i_d}$$

- Finite dimensional tensor spaces:

$$V = V_1 \otimes \ldots \otimes V_d = V_{\|\cdot\|}$$

Denoting $\{\phi_i^k\}_{i=1}^{n_k}$ a basis of the n_k-dimensional space V_k, $u \in V$ can be written

$$u = \sum_{i_1=1}^{n_1} \ldots \sum_{i_d=1}^{n_d} a_{i_1 \ldots i_d} \phi^1_{i_1} \otimes \ldots \otimes \phi^d_{i_d},$$

and identified with

$$a \in \mathbb{R}^{n_1 \times \ldots \times n_d} = \mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_d}$$

Examples of (Banach) tensor spaces

- Lebesgue space $L^p_\mu(\Xi)$ with product measure $\mu = \mu_1 \otimes \ldots \otimes \mu_d$ on $\Xi = \Xi_1 \times \ldots \times \Xi_d$:

$$L^p_\mu(\Xi_1 \times \ldots \times \Xi_d) = \overline{L^p_{\mu_1}(\Xi_1) \otimes \ldots \otimes L^p_{\mu_1}(\Xi_d)}^{\|\cdot\|_p} \quad (1 \leq p < \infty)$$

An element $u \in L^p_{\mu_1}(\Xi_1) \otimes \ldots \otimes L^p_{\mu_1}(\Xi_d)$ is of the form

$$u(\xi_1, \ldots, \xi_d) = \sum_{i=1}^m u_i^1(\xi_1) \ldots u_i^d(\xi_d), \quad (\xi_1, \ldots, \xi_d) \in \Xi.$$

- Sobolev space $W^{s,p}(I)$ on $I = I_1 \times \ldots \times I_d$, the set of measurable functions $u : I \to \mathbb{R}$ with bounded norm

$$\|u\|_{s,p} = \sum_{|\alpha| \leq s} \|\partial^\alpha u\|_p, \quad \partial^\alpha = \partial_1^{\alpha_1} \ldots \partial_d^{\alpha_d}.$$

$$W^{s,p}(I) = \overline{W^{s,p}(I_1) \otimes \ldots \otimes W^{s,p}(I_d)}^{\|\cdot\|_{s,p}} \quad (1 \leq p < \infty)$$

$W^{s,p}(I)$ is an intersection tensor space:

$$W^{s,p}(I) = \bigcap_{\alpha \in \Lambda_s} \overline{W^{\alpha_1,p} \otimes \ldots \otimes W^{\alpha_d,p}}^{\|\cdot\|_\alpha}$$

$$\Lambda_s = \{(0, \ldots, 0), (s, 0, \ldots, 0), (0, \ldots, 0, s)\}$$

Tensors in stochastic and parametric problems

- Stochastic/Parametric equations (PDEs, ODEs...):

$$\mathcal{F}(u(\xi);\xi) = 0, \quad u(\xi) \in \mathcal{V}$$

$$\xi \sim \mu, \quad supp(\mu) = \Xi.$$

$$u \in L^p_\mu(\Xi;\mathcal{V}) = \overline{\mathcal{V} \otimes L^p_\mu(\Xi)}$$

- Functions of independent random variables:

$$u(\xi_1, \xi_2, \dots, \xi_d)$$

$$\xi_k \sim \mu_k, \quad supp(\mu_k) = \Xi_k$$

$$u \in \overline{\mathcal{V} \otimes L^p_\mu(\Xi)} = \overline{\mathcal{V} \otimes L^p_{\mu_1}(\Xi_1) \otimes \dots \otimes L^p_{\mu_d}(\Xi_d)}$$

- Parametrized functions of random variables (robust optimization and control, statistical inverse problems):

$$u(\xi, \eta)$$

$$\xi \sim \mu, \quad supp(\mu) = \Xi, \quad \eta \in A$$

$$u \in \overline{\mathcal{V} \otimes L^p_\mu(\Xi) \otimes L^q_\nu(A)}$$

- Stochastic differential equations:

$$\begin{cases} dX_t = a(X_t, t)dt + \sigma(X_t, t)dW_t \\ X_0 = x_0 \end{cases} \qquad X_t = (X_t^1 \dots X_t^n)$$

  - The probability density function $u(\cdot, t)$ of X_t verifies a n-dimensional PDE (Kolmogorov forward equation)

$$u(\cdot, t) \in \overline{H^1_{\mu_1}(\mathbb{R}) \otimes \dots \otimes H^1_{\mu_n}(\mathbb{R})}$$

  - Approximation of Wiener process $W_t \approx \sum_{k=1}^d \varphi_k(t)\xi_k$, and

$$X_t \approx X_t(\xi_1, \dots, \xi_d) \in \overline{L^2_\Gamma(\mathbb{R}) \otimes \dots \otimes L^2_\Gamma(\mathbb{R})}$$

Notions of ranks for higher-order tensors

- For order-two tensors, a single notion of rank.

$$rank(v) \leq r \iff v = \sum_{i=1}^{r} v_i^1 \otimes v_i^2 \quad \left(v(x_1, x_2) = \sum_{i=1}^{r} v_i^1(x_1) v_i^2(x_2) \right)$$

- For higher-order tensors, different notions of rank.

- Canonical rank:

$$rank(v) \leq r \iff v = \sum_{i=1}^{r} v_i^1 \otimes \ldots \otimes v_i^d \quad \text{or} \quad v(x) = \sum_{i=1}^{r} v_i^1(x_1) \ldots v_i^d(x_d)$$

Parametrization with $\sum_{\nu=1}^{d} \dim(V_\nu) = O(d)$ parameters.

Example

Ishigami function

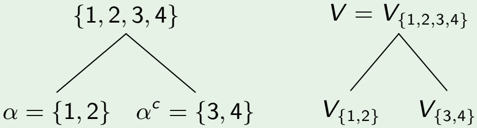
$$v(x_1, x_2, x_3) = \sin(x_1) + a\sin(x_2)^2 + bx_3^4 \sin(x_1) = \sin(x_1)(1 + bx_3^4) + a\sin(x_2)^2$$

has a canonical rank 2.

- α-rank:
  for $\alpha \subset \{1,\dots,d\}$, $V = V_\alpha \otimes V_{\alpha^c}$, with $V_\alpha = \bigotimes_{\mu\in\alpha} V_\mu$ and define the α-rank:

$$rank_\alpha(v) \le r_\alpha \quad \Longleftrightarrow \quad v = \sum_{i=1}^{r_\alpha} v_i^\alpha \otimes v_i^{\alpha^c}, \quad v_i^\alpha \in V_\alpha, \quad v_i^{\alpha^c} \in V_{\alpha^c}$$

Example



$u(x_1,\dots,x_4) = f(x_1,x_2)g(x_3,x_4)$ is such that $rank_{(1,2)}(u) = rank_{(3,4)}(u) = 1$.

- Relation with minimal subspaces:
  The minimal subspace $U_\alpha^{min}(v)$ of $v \in V$ is the smallest subspace in V_α such that

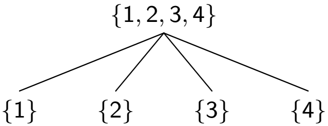
$$v \in U_\alpha^{min}(v) \otimes V_{\alpha^c}$$

$$\boxed{rank_\alpha(v) = dim(U_\alpha^{min}(v))}$$

For $v = \sum_{i=1}^{r_\alpha} v_i^\alpha \otimes v_i^{\alpha^c}$, $U_\alpha^{min}(v) = span\{v_i^\alpha : 1 \le i \le r_\alpha\}$.

- Tucker rank:

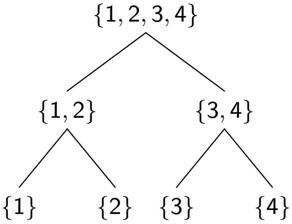
$$rank_T(v) = (rank_1(v), \dots, rank_d(v))$$



$$rank_T(v) \leq r = (r_1, \dots, r_d) \iff v = \sum_{i_1=1}^{r_1} \dots \sum_{i_d=1}^{r_d} a_{i_1 \dots i_d} v^1_{i_1} \otimes \dots \otimes v^d_{i_d}$$

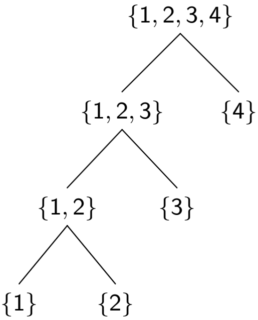
- Tree-based Tucker rank:

$$rank_T(v) = (rank_\alpha(v); \alpha \in T) \quad \text{with } T \text{ a dimension tree}$$



- TT-rank:

$$rank_{TT}(v) = (rank_{\{1\}}(v), rank_{\{1,2\}}(v) \dots, rank_{\{1,\dots,d-1\}}(v))$$



$$rank_{TT}(v) \leq r = (r_1, \dots, r_{d-1}) \iff v = \sum_{i_1=1}^{r_1} \dots \sum_{i_d=1}^{r_{d-1}} v^1_{1,i_1} \otimes v^2_{i_1,i_2} \otimes \dots \otimes v^d_{i_{d-1},1}$$

Example

$v(x) = v_1(x_1) + \dots + v_d(x_d)$ has a tree based rank $(2, \dots, 2)$.

Low-rank tensor subsets

- Different notions of rank yield different low-rank tensor subsets: Canonical, Tucker, Tree-based Tucker (HT, TT), ...

$$\mathcal{M}_{\leq r} = \{v \in V : rank(v) \leq r\}$$

- $\mathcal{M}_{\leq r}$ has a small dimension $n(r, d)$ (i.e. can be parameterized with a small number n of parameters), typically

$$n(r, d) = O(dr^s)$$

Geometry of low-rank tensors subsets

See [Holtz-Rohwedder-Schneider 11, Uschmajew-Vandereycken 13] for Hilbert setting. See [Falco-Hackbusch-Nouy 14] for a Banach setting and for general tree-based format.

- Subsets of tensors with fixed tree-based rank have a manifold structure :

$$\mathcal{M}_{\leq r} = \bigcup_{s \leq r} \mathcal{M}_{=s}$$

$$\mathcal{M}_{=s} = \{v \in V : rank_T(v) = s\} = \left\{ v = F_{\mathcal{M}}(p) \; ; \; p = (p_1, \ldots, p_L) \in \mathcal{P}^1 \times \ldots \times \mathcal{P}^L \right\}$$

  where $F_{\mathcal{M}}$ is a multilinear map and the $\mathcal{P}^l$ are low-dimensional vector spaces (or manifolds).

  - $\mathcal{M}_{=r}$ is an analytic Banach manifold, with explicit local charts.
  - Under the same assumptions as before on the norms $\{\| \cdot \|_\alpha \colon \alpha \in T\}$, $\mathcal{M}_{=r}$ is an embedded submanifold.

- Interesting consequences:
  - Optimization algorithms on manifolds
  - Dynamical systems on low-rank manifolds (Dirac-Frenkel can be extended to topological tensor spaces)

Outline

1. Polynomial approximation spaces
2. High-dimensional approximation - tractability
3. Best *n*-term approximation and quasi-optimal approximation spaces
4. Higher-order tensors and low-rank tensor formats
5. Approximation in low-rank tensor formats
6. Algorithms for computing low-rank approximations
7. Higher-order low-rank methods for parametric and stochastic equations

Approximation in low-rank subsets

- Approximation of a high order tensor

$$u \in \overline{V_1 \otimes \ldots \otimes V_d}^{\|\cdot\|}$$

in a subset of tensors with bounded rank

$$\mathcal{M}_{\leq r} = \{v \in V_1 \otimes \ldots \otimes V_d : rank(v) \leq r\}$$

- For all tensor formats, since $\cup_r \mathcal{M}_{\leq r}$ is dense in the tensor space, then

$$\inf_{\mathcal{M}_{\leq r}} \|u - v\|$$

converges to zero when $r \to \infty$.
- Questions :
  - How fast does it converge to zero ?
  - Do we beat the curse of dimensionality ?
  - Existence of best approximations ?

Approximation in low-rank tensor subsets (canonical format)

- Good approximation for smooth functions [Temlyakov, Uschmajew-Schneider]

Example (Sobolev regularity: approximation in canonical format)

$$\inf_{v \in \mathcal{R}_r} \|u - v\|_{L^2} \lesssim r^{-sd/(d-1)} \quad \forall u \in B^s_{mix} \subset L^2(\pi_d)$$

$$\text{with} \quad B^s_{mix} = \left\{ u \in L^2(\pi_d); \|u\|_{H^s_{mix}} \leq 1 \right\} \subset H^s_{mix}(\pi_d)$$

That means that for any $u \in B^s_{mix}$ and for $\epsilon > 0$, it could be possible to find an approximation $v(x_1, \ldots, x_d) = \sum_{i=1}^r \phi_i^1(x_1) \ldots \phi_i^d(x_d)$ such that

$$\|u - v\| \leq \epsilon \quad \text{with} \quad r \gtrsim \epsilon^{-\frac{d-1}{sd}}$$

- But low-rank approximations are expected to exploit additional features.

Sparse tensor approximation

Suppose $\{\phi_i^\nu(x_\nu) : i \in \Lambda_\nu\}$ is a basis of V_ν (e.g. polynomial spaces). Then

$$\{\phi_{i_1}^1(x_1)\dots\phi_{i_d}^d(x_d) : (i_1,\dots,i_d) \in \Lambda = \Lambda_1 \times \dots \times \Lambda_d\}$$

is a basis of $V_1 \otimes \dots \otimes V_d$. An element $v \in V_1 \otimes \dots \otimes V_d$ can be written

$$v(x_1,\dots,x_d) = \sum_{i\in\Lambda} a_i \phi_{i_1}^1(x_1)\dots\phi_{i_d}^d(x_d)$$

A sparse tensor approximation v_N of v is under the form

$$v_N(x_1,\dots,x_d) = \sum_{i\in\Lambda_N} a_i \phi_{i_1}^1(x_1)\dots\phi_{i_d}^d(x_d)$$

$$\Lambda_N \subset \Lambda, \quad \#\Lambda_N = N \ll \#\Lambda$$

and has a canonical rank

$$rank(v_N) \leq N$$

Consequence: if the best N-term approximation v_N (for a fixed basis) is such that $\|v - v_N\| = \sigma_N$, then

$$\inf_{v_N \in \mathcal{R}_N} \|v - v_N\| \leq \sigma_N$$

Approximation of rank one functions

Consider the function $u : [0,1]^d \to \mathbb{R}$

$$u(x_1, \dots, x_d) = u_1(x_1) \dots u_d(x_d), \quad u_k \in C^s(0,1)$$

- Approximation of the factors

$$u_k(x_k) \approx \sum_{i=1}^{n} c_{k,i} \varphi_i(x_k) \quad \text{with error} \sim n^{-s}$$

- Number of parameters

$$N = dn$$

- Number of parameters to achieve accuracy ϵ

$$N(\epsilon, d) = O(d^{1+1/s} \epsilon^{-1/s})$$

Complexity of approximating rank-one tensors

- Consider the set of functions [Bachmayr 2013,Novak 2014]

$$F_M = \left\{ u(x_1, \dots, x_d) = \prod_{i=1}^{d} u_i(x_i) : \|u_i\|_\infty \leq 1, \ \|u_i^{(s)}\|_\infty \leq M \right\}$$

  - If $M \geq 2^s s!$,

$$N(\epsilon, d) \geq 2^d \quad \text{for all } \epsilon < 1 \quad (\text{curse of dimensionality})$$

  - Polynomial tractability (algorithm available) for the sets of functions

$$F_M^{x^*} = \{u \in F_M : u(x^*) \neq 0 \quad \text{for a known } x^*\}$$

$$F_M^V = \{u \in F_M : u \neq 0 \text{ on a box of measure greater than } V\}$$

Approximation with low-rank tensors (canonical format)

- Consider that function $u : [0,1]^d \to \mathbb{R}$ can be approximated with accuracy ϵ by

$$u(x_1, \dots, x_d) \approx \sum_{i=1}^{r(\epsilon)} u_{1,i}(x_1) \dots u_{d,i}(x_d), \quad u_{k,i} \in C^s(0,1)$$

  - Approximation of each factor $u_{k,i}$ using n parameters with accuracy $\sim n^{-s}$
  - Number of parameters $N = r(\epsilon) nd$
  - Number of parameters to achieve accuracy ϵ

$$N(\epsilon, d) = O(r(\epsilon)^{1/s} d^{1+1/s} \epsilon^{-1/s})$$

- Do we beat the curse of dimensionality ?
  - What about $r(\epsilon)$ with respect to d ?
  - Which information on u and which algorithm ?

Best approximation in low-rank tensor subsets : canonical format

- If $\|\cdot\|$ is stronger than the injective norm, then the set of rank-one tensors $\mathcal{R}_1$ is weakly closed and best rank-one approximation problem is well-posed

$$\inf_{v \in \mathcal{R}_1} \|u - v\|$$

- For $d \geq 3$ and $r \geq 2$, the set $\mathcal{R}_r$ of tensors with canonical rank bounded by r is not closed, and then the best approximation problem in canonical format

$$\inf_{v \in \mathcal{R}_r} \|u - v\|$$

is ill posed.

Best approximation in low-rank tensor subsets: tree-based formats

- Best approximation problems in tree-based low-rank subsets $\mathcal{M}_{\leq r}$ are well-posed for any r, provided some conditions on tensor norms.

Theorem [Falco-Hackbusch-Nouy 14]

Assume that for all nodes $\alpha \in T$ with sons $S(\alpha) \neq \emptyset$,

- $V_\alpha = \bigotimes_{\beta \in S(\alpha)} V_\beta$ equipped with a norm $\|\cdot\|_\alpha$
- $\bigotimes : \times_{\beta \in S(\alpha)} (\overline{V_\beta}^{\|\cdot\|_\beta}, \|\cdot\|_\beta) \to \left(\bigotimes_{\beta \in S(\alpha)} \overline{V_\beta}^{\|\cdot\|_\beta}, \|\cdot\|_\alpha \right)$ is continuous
- $\|\cdot\|_\alpha$ is stronger than the injective norm $\|\cdot\|_{\vee(S(\alpha))}$ induced by the norms $\{\|\cdot\|_\beta : \beta \in S(\alpha)\}$.

Then $rank_T(\cdot) : \overline{V}^{\|\cdot\|_D} \to \mathbb{R}^{\#T}$ is weakly l.s.c.. Therefore $\mathcal{M}_{\leq r}$ is weakly closed.

- Theorem directly applies for L^p spaces.
- For Sobolev spaces $W^{s,p}$, the theorem can be used indirectly, by writing $W^{s,p}$ as an intersection of tensor spaces for which the conditions hold.

Outline

1 Polynomial approximation spaces

2 High-dimensional approximation - tractability

3 Best *n*-term approximation and quasi-optimal approximation spaces

4 Higher-order tensors and low-rank tensor formats

5 Approximation in low-rank tensor formats

6 Algorithms for computing low-rank approximations

7 Higher-order low-rank methods for parametric and stochastic equations

Alternating minimization algorithm

- Parametrization of low-rank subsets

$$\mathcal{M}_{\leq r} = \left\{ v = F(p) \ ; \ p = (p_1, \ldots, p_L) \in \mathcal{P}^1 \times \ldots \times \mathcal{P}^L \right\}$$

  with F a multilinear map.

- Best approximation problem

$$\inf_{\mathcal{M}_{\leq r}} J(v) = \inf_{p_1, \ldots, p_L} J(F(p_1, \ldots, p_L))$$

- Alternating minimization algorithm: solve successively

$$\min_{p_k} J(F(p_1, \ldots, p_k, \ldots, p_L)) \qquad (\star)$$

  - If J is convex, then the partial map $G_k : p_k \mapsto J(F(p_1, \ldots, p_k, \ldots, p_L))$ is also convex.
  - If J is quadratic, then G_k is quadratic and $(\star)$ is a quadratic optimization problem.
  - Under standard assumption on J, convergence to a stationary point can be proved but no guaranty for obtaining the global optimum (except for $d = 2$ and $\mathcal{J}(v) = \|u - v\|$ with $\| \cdot \|$ a canonical inner product norm).

Quasi-best approximations (canonical inner product norm and tree-based formats)

- Consider Hilbert tensor spaces with induced canonical norms.
- For tree-based tensor formats, algorithms based on higher-order versions of SVD for obtaining quasi-best approximations

$$u_r \in \mathcal{M}_{\leq r} \quad \text{such that} \quad \|u - u_r\| \leq \gamma(d) \inf_{v \in \mathcal{M}_{\leq r}} \|u - v\|$$

$$\gamma(d) = \begin{cases} \sqrt{d} & \text{for Tucker tensors} \\ \sqrt{2d-2} & \text{for tree-based Tucker tensors} \end{cases}$$

- Higher-order SVD for Tucker format [De Lathauwer et al 2000]
  - Truncated SVD of matricisation of $u \in V_k \otimes V_{[k]}$:

$$u \approx u^k_{r_k} = \sum_{i=1}^{r_k} v^k_i \otimes v^{[k]}_i$$

  - Subspaces $U^k_{r_k} = span\{v^k_1 \dots v^k_{r_k}\}$.
  - Approximation u_r defined by $U^1_{r_1} \otimes \dots \otimes U^d_{r_d}$

$$\|u - u_r\| = \min_{v \in U^1_{r_1} \otimes \dots \otimes U^d_{r_d}} \|u - v\|$$

  and such that

$$u_r = (P^1_{r_1} \otimes \dots \otimes P^d_{r_d})u$$

  where $P^k_{r_k}$ is the orthogonal projector from V^k onto $U^k_{r_k}$.

- Higher-order SVD for tree-based format [Grasedyck 2010]
  - Truncated SVD of matricisations of $u \in V_\alpha \otimes V_{\alpha^c}$
  - Subspaces $U^\alpha_{r_\alpha} \subset V_\alpha$.
  - Hierarchical composition of projections.

Approximation of higher-order tensors in canonical format

- Optimization in canonical format is ill-posed

$$\inf_{v \in \mathcal{R}_r} \mathcal{E}(u, v)$$

- A quasi-optimal approximation $u_r \in \mathcal{R}_r$ could (in principle) be obtained

$$\mathcal{E}(u, u_r) \leq (1 + \epsilon) \inf_{v \in \mathcal{R}_r} \mathcal{E}(u, v)$$

  but usually numerically unstable
- No notion of decomposition

$$u_r = \sum_{i=1}^{r} u_i^{(1,r)} \otimes \ldots \otimes u_i^{(d,r)}$$

Greedy approximation in canonical format (Proper Generalized Decomposition)

- Suboptimal construction of canonical representation using greedy algorithms.
- Starting from $u_0 = 0$, then
$$u_m = u_{m-1} + w_m$$
with $w_m = w_m^1 \otimes \ldots \otimes w_m^d$ defined by
$$\mathcal{E}(u, u_{m-1} + w_m) = \min_{w \in \mathcal{R}_1} \mathcal{E}(u, u_{m-1} + w)$$
- Notion of decomposition
$$u_m = \sum_{i=1}^{m} w_i^1 \otimes \ldots \otimes w_i^d$$
- Possible optimization of functions after each correction for improving convergence (but we loose the notion of decomposition).

Greedy approximation (Proper Generalized Decomposition)

- Possible corrections w_m in other low-rank subsets.
- Convergence results available [Cances & al 2011, Falco & N. 2012] (not really specific to tensor setting) but no a priori estimates (except for very general and pessimistic results).
- Convergence may be slow compared to $\inf_{w \in \mathcal{R}_r} \mathcal{E}(u, w)$
- The construction does not really exploit the tensor structure.

Greedy approximation in Tucker format: a subspace point of view

- Tucker tensors with bounded multilinear rank:

$$\mathcal{T}_r = \{v : rank_\mu(v) \leq r_\mu, \forall \mu\}$$
$$= \left\{ v \in \bigotimes_{\mu=1}^{d} U_\mu \;:\; dim(U_\mu) \leq r_\mu, \forall \mu \right\}$$

- Best approximation — a subspace point of view:

$$\min_{v \in \mathcal{T}_r} \mathcal{E}(u, v) = \min_{dim(U_1)=r_1} \dots \min_{dim(U_d)=r_d} \min_{v \in U_1 \otimes \dots \otimes U_d} \mathcal{E}(u, v)$$

  This yields sequences of optimal but non necessarily nested subspaces $\{U_\mu^{r_\mu} : r_\mu \geq 1\}$.

- Greedy construction of subspaces with nestedness property

$$\mathcal{E}(u, u_m) = \min_{U_1^m \supset U_1^{m-1}} \dots \min_{U_d^m \supset U_d^{m-1}} \min_{v \in U_1^m \otimes \dots \otimes U_d^m} \mathcal{E}(u, v)$$

  Constructive approach. Possible introduction of anisotropic enrichment.

- Possible (tricky!) extension to tree-based tensor formats.

Outline

1. Polynomial approximation spaces
2. High-dimensional approximation - tractability
3. Best n-term approximation and quasi-optimal approximation spaces
4. Higher-order tensors and low-rank tensor formats
5. Approximation in low-rank tensor formats
6. Algorithms for computing low-rank approximations
7. Higher-order low-rank methods for parametric and stochastic equations

Sample-based low-rank tensor approximation

- Computation of an output variable of interest

$$s(\xi) = \ell(u(\xi); \xi), \quad \xi \sim \mu,$$

$$s \in L^2_\mu(\mathbb{R}^d)$$

- Best L^2 approximation in a low-rank subset $\mathcal{M}$ (not computable)

$$\min_{v \in \mathcal{M}} \|s - v\|^2 = \min_{v \in \mathcal{M}} \mathbb{E}_\mu((s(\xi) - v(\xi))^2)$$

- Suppose we have evaluations $\{s(\xi^k)\}_{k=1}^K$ of s at sample points $\{\xi^k\}_{k=1}^K$ (Simulations of the full order model or of a reduced order model).
- Low rank approximation using least-square minimization:

$$\min_{v \in \mathcal{M}} \|s - v\|_K^2$$

$$\text{with} \quad \|s - v\|_K^2 = \frac{1}{K} \sum_{k=1}^{K} (s(\xi^k) - v(\xi^k))^2 \approx \mathbb{E}_\mu((s(\xi) - v(\xi))^2)$$

Sample-based low-rank tensor approximation

- Regularization could be required

$$\min_{v\in\mathcal{M}} \|s - v\|_K^2 + \text{“regularization”}$$

- For a given tensor format

$$\mathcal{M} = \{v = F_{\mathcal{M}}(p_1, \dots, p_L); p_k \in \mathbb{R}^{m_k}, 1 \le k \le L\}$$

solve

$$\min_{p_1,\dots p_L} \|s - F_{\mathcal{M}}(p_1, \dots, p_L)\|_K^2 + \sum_k \lambda_k \|p_k\|_s$$

that corresponds to a minimization in a subset of $\mathcal{M}$:

$$\mathcal{M}_\gamma = \{v = F_{\mathcal{M}}(p_1, \dots, p_L); p_k \in \mathbb{R}^{m_k}, \|p_k\|_s \le \gamma_k, 1 \le k \le L\}$$

-
- Sparsity-inducing regularization with $0 \le s \le 1$.
- Some references [Beylkin-Garcke-Mohlenkamp 2011; Doostan-Validi-Iaccarino 2013; Chevreuil-Lebrun-Nouy-Rai 2014]

Now, entering the model...

Higher order tensor structure

- Suppose that $\mu = \mu_1 \otimes \ldots \otimes \mu_d$, a product measure on $\Xi_1 \times \ldots \times \Xi_d$ (e.g. when $\xi = (\xi_1, \ldots, \xi_d)$ are independent random variables). Then
$$L^2_\mu(\Xi) = \overline{L^2_{\mu_1}(\Xi_1) \otimes \ldots \otimes L^2_{\mu_d}(\Xi_d)}$$
- Suppose that the approximation space $\mathcal{S} \subset L^2_\mu(\Xi)$ is a finite dimensional tensor space
$$\mathcal{S} = \mathcal{S}_1 \otimes \ldots \otimes \mathcal{S}_d, \quad \mathcal{S}_\nu \subset L^2_{\mu_\nu}(\Xi_\nu)$$
and the same for $\widetilde{\mathcal{S}} = \widetilde{\mathcal{S}}_1 \otimes \ldots \otimes \widetilde{\mathcal{S}}_d$.
- $\lambda^{(\nu)} : \Xi_\nu \to \mathbb{R}$ can be identified with an operator $\Lambda^{(\nu)} : \mathcal{S}_\nu \to \widetilde{\mathcal{S}}'_\nu$ such that
$$\langle \Lambda^{(\nu)} s, \tilde{s} \rangle = \mathbb{E}_{\mu_\nu}(\lambda^{(\nu)}(\xi_\nu) s(\xi_\nu) \tilde{s}(\xi_\nu))$$
- A function $\lambda : \Xi \to \mathbb{R}$ such that $\lambda(\xi) = \lambda^{(1)}(\xi_1) \ldots \lambda^{(d)}(\xi_d)$ can be identified with an operator $\Lambda : \mathcal{S} \to \widetilde{\mathcal{S}}'$ such that
$$\Lambda = \Lambda^{(1)} \otimes \ldots \otimes \Lambda^{(d)}$$

- Equation

$$A(\xi)u(\xi) = f(\xi)$$

- Suppose that

$$A(\xi) = \sum_{k=1}^{R} A_k \lambda_k(\xi), \quad \text{with} \quad \lambda_k(\xi) = \lambda_k^{(1)}(\xi_1)\dots\lambda_k^{(d)}(\xi_d)$$

and

$$f(\xi) = \sum_{k=1}^{L} f_k \eta_k(\xi), \quad \text{with} \quad \eta_k(\xi) = \eta_k^{(1)}(\xi_1)\dots\eta_k^{(d)}(\xi_d)$$

- Tensor structured equation for $u \in \mathcal{V} \otimes \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_d$

$$Bu = F \quad \Longleftrightarrow \quad \left(\sum_{k=1}^{R} B_k \otimes \Lambda_k^{(1)} \otimes \dots \otimes \Lambda_k^{(d)}\right) u = \sum_{k=1}^{L} f_k \otimes \eta_k^{(1)} \otimes \dots \otimes \eta_k^{(d)}$$

- Tensor structured equation in algebraic form for $\mathbf{u} \in \mathbb{R}^{\mathcal{N}} \otimes \mathbb{R}^{\mathcal{P}_1} \otimes \dots \otimes \mathbb{R}^{\mathcal{P}_d}$

$$\mathbf{Bu} = \mathbf{F} \quad \Longleftrightarrow \quad \left(\sum_{k=1}^{R} \mathbf{A}_k \otimes \boldsymbol{\Lambda}_k^{(1)} \otimes \dots \otimes \boldsymbol{\Lambda}_k^{(d)}\right) \mathbf{u} = \sum_{k=1}^{L} \mathbf{f}_k \otimes \boldsymbol{\eta}_k^{(1)} \otimes \dots \otimes \boldsymbol{\eta}_k^{(d)}$$

Classical iterative methods with low-rank truncations

- Equation in tensor format

$$Bu = F, \quad u \in \mathcal{V} \otimes \mathcal{S}_1 \otimes \ldots \otimes \mathcal{S}_d$$

- Iterative solver

$$u^{(k)} = T(u^{(k-1)}) \quad (T\text{: iteration map})$$

- Approximate iterations using low-rank truncations:

$$u^{(k)} \in \mathcal{M}_{r(\epsilon)} \quad \text{such that} \quad \|u^{(k)} - T(u^{(k-1)})\| \leq \epsilon$$

- For the canonical norm $\|\cdot\|$, truncation based on higher-order SVD
- Computational requirements: low-rank algebra and efficient truncation algorithms
- Analysis : perturbation of algorithms.

(see e.g. [Khoromskij-Schwab 2011, Kressner-Tobler 2011])

Minimal residual low-rank approximation

- Residual-based error

$$\mathcal{E}(u, w) = \|Bw - F\|_C = \|w - u\|_{B^*CB}$$

with a certain residual norm $\|\cdot\|_C^2 = \langle C\cdot, \cdot\rangle$.

- Best approximation in $\mathcal{M}_{\leq r}$

$$\mathcal{E}(u, u_r) = \min_{w \in \mathcal{M}_{\leq r}} \mathcal{E}(u, w)$$

- If

$$\alpha\|u - w\| \leq \mathcal{E}(u, w) \leq \beta\|u - w\|$$

then

$$\|u - u_r\| \leq \frac{\beta}{\alpha} \min_{w \in \mathcal{M}_{\leq r}} \|u - w\|$$

Illustration : stationary advection-diffusion-reaction equation

$$-\nabla \cdot (\kappa \nabla u) + c \cdot \nabla u + \gamma u = \delta I_{\Omega_1}(x) \quad on \quad \Omega$$

Random field

$$\kappa(x, \boldsymbol{\xi}) = \mu_\kappa + \sum_{i=1}^{40} \sqrt{\sigma_i} \kappa_i(x) \xi_i, \quad \xi_i \in U(-1, 1)$$

Spatial modes $\kappa_i(x)$

Amplitudes σ_i

Stochastic approximation

$$\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{40}), \quad \boldsymbol{\Xi} = (-1,1)^{40} = \Xi_1 \times \ldots \times \Xi_{40}$$

$$\mathcal{S} = \mathbb{P}_4(\Xi_1) \otimes \ldots \otimes \mathbb{P}_4(\Xi_{40})$$

$$\boxed{dim(\mathcal{S}) = 5^{40} \approx 10^{28}}$$

Finite element mesh

$$\boxed{dim(\mathcal{V}_N) = 4435}$$

Solution $u(\cdot, \boldsymbol{\mu}_{\boldsymbol{\xi}})$ for mean parameters

A basic hierarchical format

Deterministic/stochastic separation

$$u(\xi) \approx u_m(\xi) = \sum_{i=1}^{m} v_i s_i(\xi)$$

$$\hookrightarrow \quad \boxed{\mathcal{V}_m = span\{v_i\}_{i=1}^m}$$

low-rank approximation of parametric functions

$$\mathbf{s}(\xi) := (\mathbf{s}_i)_{i=1}^m \approx \mathbf{s}_Z(\xi) = \sum_{k=1}^{Z} \phi_k^0 \prod_{j=1}^{d} \phi_k^j(\xi_j)$$

$$\hookrightarrow \quad \boxed{\mathcal{S}_Z = span\{\prod_{j=1}^d \phi_k^j(\xi_j)\}_{k=1}^Z}$$

For a precision $\|u - u_{M,Z}\|_{L^2} \leqslant 10^{-2}$

- $\boxed{dim(\mathcal{V}_m) \approx 15} \ll 4435 = dim(\mathcal{V})$
- $\boxed{dim(\mathcal{S}_Z) \approx 10} \ll 10^{28} = dim(\mathcal{S})$
- 15 classical deterministic problems in order to build $\mathcal{V}_M \subset \mathcal{V}_N$
- about 1 minute computation on a laptop with matlab

Convergence properties of quantities of interest

Probability of events

Quantity of interest

$$Q(\boldsymbol{\xi}) = \int_{\Omega_2} u(x, \boldsymbol{\xi})\, dx$$

$$Q_M(\boldsymbol{\xi}) = \int_{\Omega_2} u_M(x, \boldsymbol{\xi})\, dx$$

$$P(Q > q), \quad q \in (3.5, 5.4)$$

Convergence properties of quantities of interest

Sensitivity analysis

$$Q(\boldsymbol{\xi}) \approx Q_M(\boldsymbol{\xi}) \approx Q_{M,Z}(\boldsymbol{\xi}) = \sum_{k=1}^{Z} q_k \Psi_k(\boldsymbol{\xi}), \quad \Psi_k(\boldsymbol{\xi}) = \prod_{i=1}^{40} \phi_k^i(\xi_i)$$

First order Sobol sensitivity index with respect to parameter ξ_i

$$\boxed{S_i = \frac{Var(E(Q|\xi_i))}{Var(Q)}} \quad E(Q|\xi_i) = \sum_{k=1}^{Z} \alpha_k^i \phi_k^i(\xi_i), \quad \alpha_k^i = q_k \prod_{\substack{j=1 \\ j\neq i}}^{40} E(\phi_k^i(\xi_i))$$

First order Sobol sensivity indices S_i

M=1 M=5 M=15

Challenging issues

- Classify applications and dedicated reduced order methods
    - Quantum physics : a long history for the construction of tensor formats
    - Machine learning and statistical learning: a huge literature on reduced order models for high dimensional functions.
- A priori estimates
- Automatic selection of reduced order formats (bases or frames for sparse approximation, tensor formats for low-rank approximation).
- Well-conditioned formulations for (quasi-)optimal model reduction
- Samples-based constructions: How to sample given an approximation format ? How many samples ?
- Software engineering. Minimize interactions with existing codes.
- Goal-oriented model order reduction: variable of interest $s = \ell(u)$, rare event computation, sensitivity analysis, optimization, inverse problems, ...

References I

- **SPECTRAL METHODS, GALERKIN METHODS**

M. Deb, I. Babuska, and J. T. Oden. Solution of stochastic partial differential equations using galerkin finite element techniques. *Computer Methods in Applied Mechanics and Engineering*, 190:6359–6372, 2001.

I. Babuska, R.Tempone, and G. E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numerical Analysis*, pages 800–825, 2004.

P. Frauenfelder, C. Schwab, and R. A. Todor. Finite elements for elliptic problems with stochastic coefficients. *Computer Methods in Applied Mechanics and Engineering*, 194(2-5):205–228, 2005.

- **SPARSE APPROXIMATION, BEST N-TERM APPROXIMATION**

F. Nobile, R. Tempone, and C. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2411–2442, 2008.

A. Cohen, R. DeVore, and C. Schwab. Convergence rates of best n-term galerkin approximations for a class of elliptic spdes. *Foundations of Computational Mathematics*, 10(6):615–646, 2010.

A. Chkifa, A. Cohen, and C. Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric {PDEs}. *Journal de Mathématiques Pures et Appliquées*, (0):–, 2014.

J. Beck, R. Tempone, F. Nobile, And L. Tamellini. On the optimal polynomial approximation of stochastic pdes by galerkin and collocation methods. *Mathematical Models and Methods in Applied Sciences*, 22(09):1250023, 2012.

References II

J. Beck, F. Nobile, L. Tamellini, and R. Tempone. Convergence of quasi-optimal stochastic galerkin methods for a class of PDES with random coefficients. *Computers & Mathematics with Applications*, 67(4):732 – 751, 2014.

- **LOW-RANK TENSOR METHODS**

W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*, Springer, 2012.

L. Grasedyck and D. Kressner and C. Tobler. A literature survey of low-rank tensor approximation techniques. 2013

F. Chinesta, P. Ladeveze, and E. Cueto. A short review on model order reduction based on proper generalized decomposition. *Archives of Computational Methods in Engineering*, 18(4):395–404, 2011.

B. Khoromskij. Tensors-structured numerical methods in scientific computing: Survey on recent advances. *Chemometrics and Intelligent Laboratory Systems*, 110(1):1 – 19, 2012.

A. Nouy. Low-rank tensor approximation methods and Proper Generalized Decompositions. Book Chapter. Preprint.

A. Falco, W. Hackbusch and A. Nouy. Geometric structures in tensor representations. Preprint.

A. Falco and A. Nouy. Proper generalized decomposition for nonlinear convex problems in tensor Banach spaces. Numerische Mathematik, 2012.

References III

- L. Giraldi, A. Nouy and G. Legrain. Low-rank approximate inverse for preconditioning tensor-structured linear systems. *SIAM Journal on Scientific Computing*, 36(4):A1850–A1870, 2014.
- M. Billaud-Friess, A. Nouy, and O. Zahm. A tensor approximation method based on ideal minimal residual formulations for the solution of high-dimensional problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, eFirst, 4 2014.

LOW-RANK FOR UQ AND PARAMETRIC MODELS

- A. Nouy. Proper Generalized Decompositions and separated representations for the numerical solution of high dimensional stochastic problems. Archives Comp. Meth. Eng., 2010.
- B. N. Khoromskij and C. Schwab. Tensor-structured galerkin approximation of parametric and stochastic elliptic pdes. *SIAM Journal on Scientific Computing*, 33(1):364–385, 2011.
- M. Chevreuil, R. Lebrun, A. Nouy, and P. Rai. A least-squares method for sparse low rank approximation of multivariate functions. *ArXiv e-prints*, April 2013.
- L. Tamellini, O. Le Maitre, and A. Nouy. Model reduction based on proper generalized decomposition for the stochastic steady incompressible navier-stokes equations. *MOX–Report No. 26/2012*, 2012.
- H. G. Matthies and E. Zander. Solving stochastic systems with low-rank tensor compression. *Linear Algebra and its Applications*, 436(10), 2012.