Polynomial, sparse and low-rank approximations

Anthony Nouy

Centrale Nantes

Laboratoire de Mathématiques Jean Leray

RICAM Special Semester on
"Multivariate Algorithms and their Foundations in Number Theory",
Linz, December 14, 2018

Tutorial on Uncertainty Quantification -
Efficient Methods for PDEs with Random Coefficients

Uncertainty quantification

We consider a (numerical or experimental) model depending on a set of random parameters $X = (X_1, \ldots, X_d)$ that describe the uncertainties on the model, and some output variable of interest

$$Y = u(X).$$

- Forward problems: evaluation of statistics, probability of events, sensitivity indices...

$$\mathbb{E}(h(Y)) = \mathbb{E}(h \circ u(X)) = \int h(u(x_1, \ldots, x_d)) p(x_1, \ldots, x_d) dx_1 \ldots dx_d$$

- Inverse problems: from (partial) observations of Y, estimate the distribution μ of X

$$d\mu(x_1, \ldots, x_d)$$

Solving forward and inverse problems requires the evaluation of the model for many instances of X.

This is usually unaffordable when one evaluation requires a costly numerical simulation (or experiment).

Approximation for uncertainty quantification

In practice, we rely on approximations of the map

$$X \mapsto u(X)$$

used as predictive surrogate models (reduced order models, metamodels) which are easy to operate with (evaluation, integration, derivation...).

This requires

- approximation formats (model classes) that exploit some specific features of the functions (e.g. regularity, low effective dimension, sparsity, low rank...), possibly deduced from some knowledge on the model,
- algorithms for constructing approximations from available information: samples (black box), model's equations (white or grey box)...

Approximation for uncertainty quantification

An approximation $\tilde{Y} = \tilde{u}(X)$ of $Y = u(X)$ can be directly used for obtaining approximate solutions to forward and inverse problems, with a control of errors on quantities of interest, e.g.

$$|\mathbb{E}(Y) - \mathbb{E}(\tilde{Y})| \leq \int |u(x) - \tilde{u}(x)| d\mu(x) = \|u - \tilde{u}\|_{L^1_\mu},$$

but also to design variance reduction methods for Monte-Carlo methods, e.g. as a control variate

$$\mathbb{E}(Y) \approx \mathbb{E}(\tilde{Y}) + \frac{1}{N} \sum_{k=1}^{N} (u(X_k) - \tilde{u}(X_k)) := \hat{I}_N,$$

$$\mathbb{E}(|\hat{I}_N - \mathbb{E}(Y)|^2) = \mathbb{V}(\hat{I}_N) \leq \frac{1}{N} \|u - \tilde{u}\|^2_{L^2_\mu}.$$

Approximation

The goal is to approximate a function u from a space M by a function u_n from a subset M_n (model class) described by n (or $O(n)$) parameters.

We distinguish linear approximation, where M_n are linear spaces, from nonlinear approximation, where M_n are nonlinear sets.

The quality of an approximation u_n in M_n can be assessed by

$$d(u, u_n)$$

where d is a metric on M, and the quality of the model class is assessed by the best approximation error

$$e_n(u)_M = \inf_{v \in M_n} d(u, v)$$

Approximation

Given a function u, and given a family of model classes $(M_n)_{n\geq 1}$, fundamental problems are to determine if and how fast $e_n(u)_M$ tends to 0, and to provide algorithms which produce approximations $u_n \in M_n$ such that

$$d(u, u_n) \leq Ce_n(u)_M$$

with C independent of n or $C(n)e_n(u)_M \to 0$ as $n \to \infty$.

Worst-case and mean squared errors

For functions defined on a parameter space $\mathcal{X}$ (equipped with a measure μ) and with values in some Banach space V, a classical setting is to consider functions from the Bochner space

$$M = L^p_\mu(\mathcal{X}; V) = V \otimes L^p_\mu(\mathcal{X})$$

equipped with the metric

$$d(u, v) = \|u - v\|_{L^p_\mu(\mathcal{X};V)}.$$

Two typical cases are $p = \infty$ (worst-case setting),

$$\|u - v\|_{L^\infty_\mu(\mathcal{X};V)} = \operatorname*{ess\,sup}_{x \in \mathcal{X}} \|u(x) - v(x)\|_V$$

and $p = 2$ (mean-squared setting),

$$\|u - v\|^2_{L^2_\mu(\mathcal{X};V)} = \int_{\mathcal{X}} \|u(x) - v(x)\|^2_V d\mu(x) = \mathbb{E}(\|u(X) - v(X)\|^2_V)$$

where $X \sim \mu$.

Noting that $\|u - v\|_{L^2_\mu(\mathcal{X};V)} \leq \|u - v\|_{L^\infty_\mu(\mathcal{X};V)}$, approximation results in L^2 can be deduced from stronger results in L^∞.

Model classes for vector-valued functions

For the approximation of a function $u \in L^p_\mu(\mathcal{X}; V)$, typical model classes are

- $M_n = V \otimes S_n$, where S_n is a subspace of $L^p_\mu(\mathcal{X})$ (e.g. polynomials, wavelets...), which results in an approximation

$$u_n(x) = \sum_{i=1}^{n} v_i \varphi_i(x)$$

with an explicit expression as a function of x.

- $M_n = L^p_\mu(\mathcal{X}; V_n) = V_n \otimes L^p_\mu(\mathcal{X})$, where V_n is a low-dimensional subspace of V, which results in an approximation

$$u_n(x) = \sum_{i=1}^{n} v_i \varphi_i(x)$$

which is not explicit in terms of x.

When $u(x)$ is solution of a parameter-dependent equation, the approximation $u_n(x) \in V_n$ is obtained by some projection of $u(x)$ on V_n that exploits the equations. This corresponds to projection-based model order reduction methods.

Computing an approximation

An approximation u_n in a certain model class M_n can be obtained by

- an interpolation of u at a set of points Γ_n.

  For a vector space $M_n = V \otimes S_n$ and a set of points $\Gamma_n \subset \mathcal{X}$ unisolvent for S_n, the interpolation u_n is such that

  $$u_n(x) = u(x) \quad \forall x \in \Gamma_n,$$

  and

  $$\|u - u_n\|_{L^p} \leq (1 + L_n^{(p)}) e_n(u)_{L^p}$$

  where $L_n^{(p)}$ is the norm of the interpolation operator from $L^p_\mu(\mathcal{X})$ to S_n, which depends on the quality of the set of points Γ_n for S_n.

  For $p = \infty$, $L_n^{(\infty)}$ is the Lebesgue constant $L_n^{(\infty)} = \sup_{x \in \mathcal{X}} \sum_{i=1}^n |\ell_i(x)|$ where $\{\ell_i\}$ is a basis of S_n with the interpolation property.

Computing an approximation

- A minimization of an empirical risk functional

$$\min_{v \in M_n} \frac{1}{m} \sum_{k=1}^{m} \ell(u(x_k), v(x_k)) \approx \min_{v \in M_n} \mathbb{E}(\ell(u(X), v(X)))$$

where the x_k are samples of X and the risk $\mathbb{E}(\ell(u(X), v(X)))$ provides some "distance" $d(u, v)$ between u and v.

A better performance can be obtained by solving

$$\min_{v \in M_n} \frac{1}{m} \sum_{k=1}^{m} w_k \ell(u(x_k), v(x_k))$$

where the x_k are samples in $\mathcal{X}$ drawn from a suitable distribution $d\nu(x) = \rho(x) d\mu(x)$ on $\mathcal{X}$, and the weights $w_k = \rho(x_k)^{-1}$.

Computing an approximation

- a (weighted) least-squares projection of $u \in L^2_\mu(\mathcal{X}; V)$, which is solution of

$$\min_{v \in M_n} \frac{1}{m} \sum_{k=1}^{m} \rho(x_k)^{-1} \|u(x_k) - v(x_k)\|_V^2$$

where the x_k are samples in $\mathcal{X}$ drawn from a certain distribution $d\nu(x) = \rho(x) d\mu(x)$ on $\mathcal{X}$.

For $M_n = V \otimes S_n$ with S_n a n-dimensional subspace of $L^2_\mu(\mathcal{X})$ with orthonormal basis $\{\varphi_i\}_{i=1}^n$, the quality of the least-squares projection depends on how far the empirical Gram matrix

$$G_{ij} = \frac{1}{m} \sum_{k=1}^{m} w_k \varphi_i(x_k) \varphi_j(x_k)$$

is from identity.

An optimal weighted least-squares method [Cohen and Migliorati 2017] is obtained with $\rho(x) = \frac{1}{n} \sum_{i=1}^n \varphi_i(x)^2$. Then for $m \geq n\epsilon^{-2} \log(2n\eta^{-1})$, this ensures that $\mathbb{P}(\|G - I\| > \epsilon) \leq \eta$ and (in particular)

$$\mathbb{E}(\|u - u_n\|_{L^2}^2) \leq C e_n(u)_{L^2}^2 + \|u\|^2 \eta, \quad \text{with} \quad C = 1 + \frac{1}{1-\epsilon} \frac{n}{m}.$$

Computing an approximation

- Given the model's equations

$$A(x)u(x) = f(x), \quad \text{with } A(x) : V \to W,\ f(x) \in W$$

an approximation u_n can be obtained through a Galerkin projection[1] of u, e.g. defined by

$$\min_{v \in M_n} \int_{\mathcal{X}} \|A(x)v(x) - f(x)\|_W^2 d\mu(x) \quad \text{or} \quad \min_{v \in M_n} \sup_{x \in \mathcal{X}} \|A(x)v(x) - f(x)\|_W$$

If $A(x)$ is a linear operator such that $\alpha\|v\|_V \leq \|A(x)v\|_W \leq \beta\|v\|_V$, then

$$\|u - u_n\|_{L^p_\mu(\mathcal{X};V)} \leq \frac{\beta}{\alpha} \inf_{v \in M_n} \|u - v\|_{L^p_\mu(\mathcal{X};V)}$$

[1] coined stochastic Galerkin projection

Outline

Outline

Polynomial spaces

Let $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d \subset \mathbb{R}^d$.

For each dimension k, we consider a family of univariate polynomials $\{\psi_n^k\}_{n \geq 0}$ with $\psi_n^k \in \mathbb{P}_n(\mathcal{X}_k)$.

Then we define the tensorised basis

$$\psi_\alpha(x) = \psi_{\alpha_1}^1(x_1) \ldots \psi_{\alpha_d}^d(x_d)$$

where α is a multi-index in $\mathbb{N}^d$.

For a set $\Lambda \subset \mathbb{N}^d$, we consider the space of polynomials

$$\mathbb{P}_\Lambda(\mathcal{X}) = \operatorname{span}\{\psi_\alpha(x) : \alpha \in \Lambda\}$$

In general, the polynomial space $\mathbb{P}_\Lambda(\mathcal{X})$ depends on the chosen univariate polynomial bases, except for downward closed sets Λ such that

$$\alpha \in \Lambda \text{ and } \beta \leq \alpha \quad \Rightarrow \quad \beta \in \Lambda$$

Polynomial interpolation

Let $\Gamma^k = (t_i^k)_{i \geq 0}$ be a sequence of points in $\mathcal{X}_k$ such that the set $(t_i^k)_{i=0}^n$ is unisolvent for $\mathbb{P}_n(\mathcal{X}_k)$, which means that for any $a \in \mathbb{R}^{n+1}$, there exists a unique polynomial $v \in \mathbb{P}_n(\mathcal{X}_k)$ such that

$$v(t_i^k) = a_i \quad \text{for all } 0 \leq i \leq n,$$

therefore allowing to define the interpolation operator $\mathcal{I}_n^k : \mathbb{R}^{\mathcal{X}_k} \to \mathbb{P}_n(\mathcal{X}_k)$.

Then for any downward closed set $\Lambda \subset \mathbb{N}^d$, the set

$$\Gamma_\Lambda = \{t_\alpha = (t_{\alpha_1}^1, \ldots, t_{\alpha_d}^d) : \alpha \in \Lambda\}$$

is unisolvent for $\mathbb{P}_\Lambda(\mathcal{X})$, that uniquely defines an interpolation operator (oblique projection)

$$\mathcal{I}_\Lambda : \mathbb{R}^{\mathcal{X}} \to \mathbb{P}_\Lambda(\mathcal{X})$$

whose norm can be bounded using upper bounds of the norm of one-dimensional interpolation operators.

Orthogonal polynomials

When using least-squares or Galerkin projections methods in $L^2_\mu(\mathcal{X})$, the use of orthonormal bases improves properties of numerical methods.

Let consider a product measure $\mu = \mu_1 \otimes \ldots \otimes \mu_d$ with support $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$. Let $\{\psi_n^k\}_{n \geq 0}$ be an orthonormal polynomial basis in $L^2_{\mu_k}(\mathcal{X}_k)$, with

$$\psi_n^k \in \mathbb{P}_n(\mathcal{X}_k)$$

such that

$$\int_{\mathcal{X}_k} \psi_n^k(x_k)\psi_m^k(x_k)d\mu_k(x_k) = \delta_{nm}$$

Then the tensorized polynomial basis $\{\psi_\alpha(x) = \psi^1_{\alpha_1}(x_1)\ldots\psi^d_{\alpha_d}(x_d)\}_{\alpha \in \mathbb{N}^d}$ constitutes an orthonormal basis of $L^2_\mu(\mathcal{X})$.

Classical examples of univariate orthonormal polynomials are

- Legendre polynomials for $\mu_k \sim U(-1, 1)$,
- Hermite polynomials for $\mu_k \sim \mathcal{N}(0, 1)$

Polynomial approximations

Consider $\mathcal{X} = [-1,1]^d \subset \mathbb{R}^d$ and the space $\mathbb{P}_\Lambda(\mathcal{X})$ of polynomials with partial degree bounded by p, where

$$\Lambda = \{\alpha : \max_k \alpha_k \leq p\}.$$

with dimension $n = \#\Lambda = (p+1)^d$.

Assume that $u : \mathcal{X} \to V$ is analytic and can be analytically extended to $\{z \in \mathbb{C}^d : |z_k| \leq \tau\} \supset \mathcal{X}$, then

$$e_n(u)_{L^\infty(\mathcal{X})} \lesssim e^{-c_\tau n^{\mathbf{1}/d}}$$

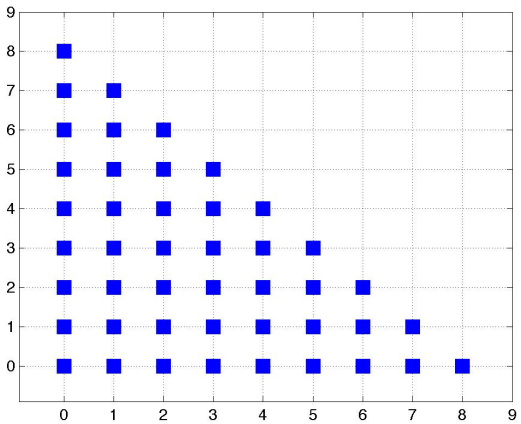
The convergence rate deteriorates with the dimension d (curse of dimensionality).

The key for circumventing the curse of dimensionality is to exploit some sparsity.

Sparse polynomial spaces

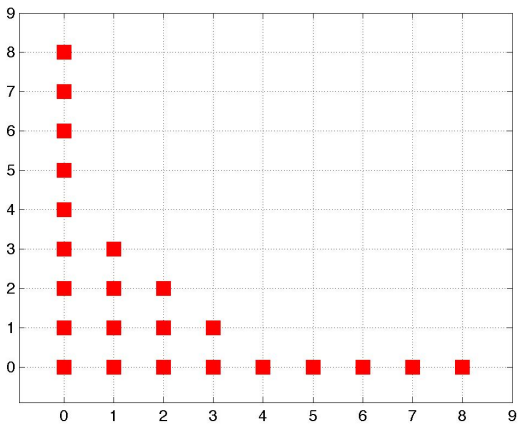
- Polynomials with bounded total degree

$$\Lambda = \{\alpha : \sum_k \alpha_k \leq p\} \text{ with } \#\Lambda = \frac{(d+p)!}{d!p!}$$



- Hyperbolic cross sets

$$\Lambda = \{\alpha : \prod_k (\alpha_k + 1) \leq p\} \text{ with } \#\Lambda \approx p \log(1+p)^d$$



Sparse polynomial spaces

- Additive polynomial functions: for

$$\Lambda = \{\alpha : \max_k \alpha_k \leq p \text{ and } \#\{k : \alpha_k \neq 0\} \leq 1\}$$

the space $\mathbb{P}_\Lambda(\mathcal{X})$ corresponds to additive functions

$$\sum_{i=1}^{d} u_i(x_i)$$

with univariate polynomial functions u_i of degree p.

- Polynomial functions with low-order interactions: for

$$\Lambda = \{\alpha : \max_k \alpha_k \leq p \text{ and } \#\{k : \alpha_k \neq 0\} \leq m\}$$

the space $\mathbb{P}_\Lambda(\mathcal{X})$ corresponds to functions with interactions of order m

$$\sum_{i_1,\ldots,i_m}^{d} u_{i_1,\ldots,i_m}(x_{i_1},\ldots,x_{i_m})$$

with m-variate polynomial functions $u_{i_1,\ldots,i_m}$ of degree p.

Outline

Best n-term approximation

Let $u \in M = L^p_\mu(\mathcal{X}; V)$ and let $\{\psi_\alpha\}_{\alpha \in \mathcal{F}}$ be a basis of $L^p_\mu(\mathcal{X})$, such that

$$u(x) = \sum_{\alpha \in \mathcal{F}} u_\alpha \psi_\alpha(x).$$

For a subset $\Lambda \subset \mathcal{F}$, let

$$M_\Lambda = \left\{ v(x) = \sum_{\alpha \in \Lambda} v_\alpha \psi_\alpha(x) : v_\alpha \in V \right\}.$$

Then we consider the nonlinear model class

$$M_n = \{v \in M_\Lambda : \Lambda \subset \mathcal{F}, \#\Lambda = n\} = \bigcup_{\#\Lambda = n} M_\Lambda$$

of functions that admit a representation with at most n non zero coefficients in the basis $\{\psi_\alpha\}_{\alpha \in \mathcal{F}}$.

Best n-term approximation

A best approximation of u in M_n is called a best n-term approximation of u relatively to the given basis.

A best n-term approximation u_n is solution of

$$\min_{v \in M_n} \|u - v\|_{L^p_\mu(\mathcal{X};V)} = \min_{\#\Lambda = n} \min_{v \in M_\Lambda} \|u - v\|_{L^p_\mu(\mathcal{X};V)} := e_n(u)_{L^p}$$

where the minimum is taken over all subsets Λ with cardinal n.

This notion can be extended to more general dictionaries of functions.

Best n-term approximation

Assuming that the functions ψ_α are normalized in $L^p_\mu(\mathcal{X})$,

$$\min_{v \in M_\Lambda} \|u - v\|_{L^p_\mu(\mathcal{X};V)} \leq \| \sum_{\alpha \notin \Lambda} u_\alpha \psi_\alpha \|_{L^p_\mu(\mathcal{X};V)} \leq \sum_{\alpha \notin \Lambda} \|u_\alpha\|_V .$$

Therefore, by choosing a set Λ_n corresponding to the n-largest terms $\|u_\alpha\|_V$, we obtain a bound of the best n-term approximation error

$$e_n(u)_{L^p} \leq \sum_{\alpha \notin \Lambda_n} \|u_\alpha\|_V$$

If the sequence $c = (\|u_\alpha\|_V)_\alpha \in \ell^r$ with $r < 1$, Stechkin's lemma yields

$$e_n(u)_{L^p} \leq C n^{-s}, \quad s = \frac{1}{r} - 1$$

with $C = \|c\|_{\ell^r} = (\sum_\alpha |c_\alpha|^r)^{1/r}$.

Best n-term approximation

Assuming that $\{\psi_\alpha\}$ is an orthonormal basis in $L^2_\mu(\mathcal{X})$,

$$\min_{v \in M_\Lambda} \|u - v\|^2_{L^2_\mu(\mathcal{X};V)} = \|\sum_{\alpha \notin \Lambda} u_\alpha \psi_\alpha\|^2_{L^2_\mu(\mathcal{X};V)} = \sum_{\alpha \notin \Lambda} \|u_\alpha\|^2_V.$$

Therefore, by choosing a set Λ_n corresponding to the n-largest terms $\|u_\alpha\|_V$, we obtain the best n-term approximation error

$$e_n(u)^2_{L^2} = \sum_{\alpha \notin \Lambda_n} \|u_\alpha\|^2_V$$

If the sequence $c = (\|u_\alpha\|_V)_\alpha \in \ell^r$ with $r < 1$, Stechkin's lemma yields

$$e_n(u)_{L^2} \leq C n^{-s}, \quad s = \frac{1}{r} - \frac{1}{2}$$

with $C = \|c\|^{1/2}_{\ell^{r/2}}$.

Parameter-dependent equations

Consider the parameter-dependent equation

$$-\nabla \cdot (a(x)\nabla u(x)) = f \quad \text{in } D \subset \mathbb{R}^m, \quad u(x) = 0 \quad \text{on } \partial D,$$

with the uniform ellipticity assumption $0 < \gamma \leq a(x) \leq \beta < \infty$, and a particular parametrization

$$a(x) = a_0 + \sum_{i=1}^{d} a_i x_i, \quad x \in \mathcal{X} = [-1,1]^d, \quad d \in \mathbb{N} \cup \{+\infty\}$$

Consider the Taylor expansion of u at 0

$$u(x) = \sum_{\alpha \in \mathcal{F}} u_\alpha x^\alpha, \quad u_\alpha = \frac{1}{\alpha!}\partial^\alpha u(0).$$

Parameter-dependent equations

Bounds of $\|u_\alpha\|_V$ can be obtained by complex analysis.

The solution admits an analytic extension to the complex domain (polydisc) $\{z \in \mathbb{C}^d : |z_k| \leq 1\}$.

If $\rho = (\rho_i)_{i \geq 1}$ is any sequence such that

$$\sum_{i \geq 1} \rho_i |a_i| \leq a_0 - \zeta$$

for some $0 < \zeta < \gamma$, the solution admits an analytic extension $u(z)$ to an even larger complex domain (polydisc)

$$\{z \in \mathbb{C}^d : |z_k| \leq \rho_k\}, \quad \rho_k > 1,$$

and

$$\|u_\alpha\|_V \leq \delta(\alpha), \quad \delta(\alpha) = C_\zeta \prod_{i \geq 1} \rho_i^{-\alpha_i}$$

Parameter-dependent equations

Assuming that $(\|a_i\|_{L^\infty(D)})_{i\geq 1} \in \ell^r$, we can design a sequence ρ such that $(\delta(\alpha))_{\alpha\in\mathcal{F}} \in \ell^r$.

Therefore if $(\|a_i\|_{L^\infty(D)})_{i\geq 1} \in \ell^r$ for some $r < 1$, then $(\|u_\alpha\|_V)_{\alpha\in\mathcal{F}} \in \ell^r$ and the best n-term approximation in the canonical basis $\{x^\alpha\}_\alpha$ is such that

$$e_n(u)_{L^\infty} \leq Cn^{-s}, \quad s = \frac{1}{r} - 1$$

We observe an algebraic convergence rate independent of the number of parameters, possibly infinite !

This result is still valid in the more general case of parameter-dependent operator equations

$$A(x)u(x) = f$$

where $A(x) : V \to W$ is such that $A(x) = A_0 + \sum_{i=1}^m A_i x_i$ and $(\|A_i\|_{W\leftarrow V})_{i\geq 1} \in \ell^r$.

The same performances are obtained by imposing to the sets Λ to be downward closed.

More general parameter-dependent equations

For different types of models (different parametrizations, nonlinearity), the solution may not admit an analytic extension to a complex polydisc containing $\mathcal{X}$, so that Taylor expansion may not converge.

However, by using a Legendre polynomial basis (or rescaled Legendre basis), it is possible to exploit the fact that the solution admits an analytic extension on a smaller complex domain (contained in a polyellipse).

Index sets based on estimates of coefficients

Assuming that we know an upper bound of the coefficients,

$$\|u_\alpha\|_V \leq \delta(\alpha) \tag{1}$$

a subset Λ_n^δ is obtained by retaining the n largest values $\delta(\alpha)$. The resulting set is close to optimal if the bound (1) is sharp.

Upper bounds $\delta(\alpha)$ can be obtained based on a priori analysis (a priori definition of the sequence Λ_n^δ) or based on a posteriori analysis (adaptive construction).

Assuming that there exists $\gamma \geq 1$ such that

$$\gamma^{-1}\delta(\alpha) \leq \|u_\alpha\|_V \leq \delta(\alpha),$$

we have

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

and therefore

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

Index sets based on estimates of coefficients

In practice, we can define a sequence of subsets

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

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

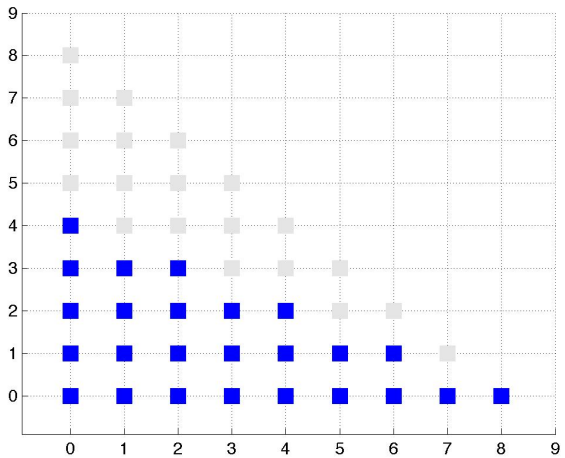
Assume that

$$\|u_\alpha\|_V \leq C \prod_k \rho_k{}^{-\alpha_k} = e^{-\sum_k \omega_k \alpha_k} := \delta(\alpha)$$

Taking $\epsilon(p) = Ce^{-p}$, we obtain

$$\Lambda_p = \left\{ \alpha : \sum_k \omega_k \alpha_k \leq p \right\}$$

which corresponds to polynomials with bounded weighted total degree.



Index sets based on estimates of coefficients

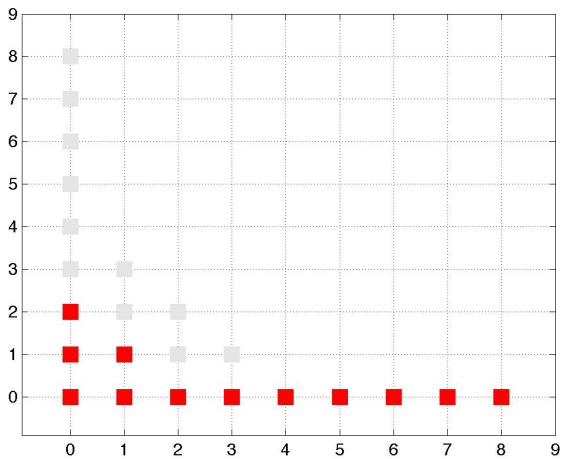
Assume that

$$\|u_\alpha\|_V \leq C \prod_k (1+\alpha_k)^{-\omega_k} := \delta(\alpha)$$

Taking $\epsilon(p) = Cp^{-1}$, we obtain

$$\Lambda_p = \left\{ \alpha : \prod_k (1+\alpha_k)^{\omega_k} \leq p \right\}$$

which is an anisotropic hyperbolic cross set.



Adaptive constructions of index sets

Adaptive algorithms for sparse approximation construct an increasing sequence of subsets $(\Lambda_n)_{n\geq 1}$ in $\mathcal{F}$ and a sequence of approximations $u_n \in M_{\Lambda_n}$ computed through interpolation, regression or other projection methods.

The sequence of subsets is defined by

$$\Lambda_n = \Lambda_{n-1} \cup A_n$$

where A_n is a subset of a candidate set N_n.

The definition of N_n requires a strategy for the exploration of the set $\mathcal{F}$.

The definition of A_n requires a selection strategy, usually based on error estimates.

Adaptive constructions of index sets

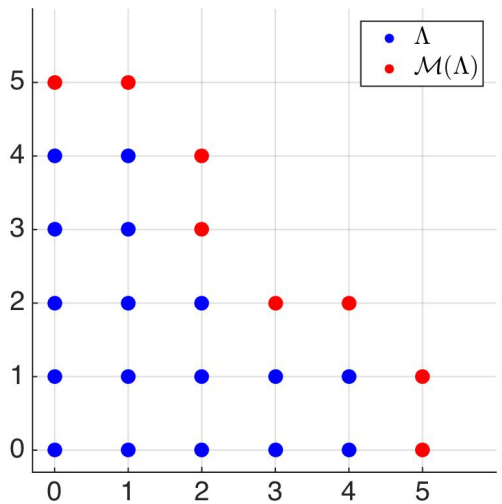
For a given downward closed set Λ, a natural neighborhood is given by the margin of Λ

$$\mathcal{M}(\Lambda) = \{\alpha \in \mathcal{F} \setminus \Lambda : \exists \beta \in \Lambda \text{ s.t. } \|\alpha - \beta\|_1 = 1\}$$

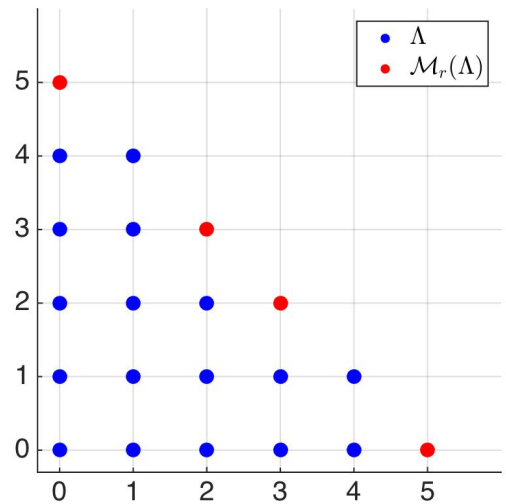
or the reduced margin of Λ

$$\mathcal{M}_r(\Lambda) = \{\alpha \in \mathcal{F} \setminus \Lambda : \alpha - e_k \in \Lambda \text{ for all } k \text{ s.t. } \alpha_k > 1\}$$

a set Λ and its margin $\mathcal{M}(\Lambda)$



a set Λ and its reduced margin $\mathcal{M}_r(\Lambda)$



For a downward closed set Λ, an interesting property of the reduced margin $\mathcal{M}_r(\Lambda)$ is that for any subset $A \subset \mathcal{M}_r(\Lambda)$, $\Lambda \cup A$ is downward closed.

Outline

Parameter-dependent equations

We consider the case of models described by parameter-dependent equations

$$\mathcal{F}(u(x); x) = 0, \quad x \in \mathcal{X},$$

where the solution $u(x)$ is in a high-dimensional space V (e.g. a finite element approximation space for PDEs).

The complexity limits the number of evaluations of $u(x)$.

However, for many problems, the solution manifold

$$\mathcal{M} = \{u(x) : x \in \mathcal{X}\}$$

has a low effective dimension, i.e. it can be well approximated by a low dimensional subspace V_n of V.

Parameter-dependent equations

This is exploited by projection-based model reduction methods that consist in projecting the solution $u(x)$ in a suitable subspace V_n, which results in an approximation

$$u_n(x) = \sum_{i=1}^{n} v_i \varphi_i(x)$$

where the $v_i \in V$ form a basis of V_n, and $\varphi_i : \mathcal{X} \to \mathbb{R}$.

This can be interpreted as a rank-n approximation of u, seen as an element of $V \otimes \mathbb{R}^{\mathcal{X}}$.

For $u \in L^p_\mu(\mathcal{X}; V)$, this is equivalent to consider model classes

$$M_n = L^p_\mu(\mathcal{X}; V_n) = V_n \otimes L^p_\mu(\mathcal{X}).$$

Measuring the quality of subspaces

Consider a Banach space V equipped with a norm $\|\cdot\|_V$.

For a given instance $x \in \mathcal{X}$, the quality of a subspace V_n is measured through the best approximation error

$$d(u(x), V_n) = \inf_{v \in V_n} \|u(x) - v\|_V$$

When we are interested in controlling the worst-case error, the map u is seen as an element of $L^\infty(\mathcal{X}; V)$ and the quality of V_n is measured by

$$\inf_{v \in L^\infty(\mathcal{X}; V_n)} \|u - v\|_{L^\infty(\mathcal{X}; V)} = \sup_{x \in \mathcal{X}} d(u(x), V_n) = \sup_{f \in \mathcal{M}} d(f, V_n)$$

When $\mathcal{X}$ is equipped with a measure and we are interesting in controlling a mean-squared error, the map is seen as an element of $L^2_\mu(\mathcal{X}; V)$ and the quality of V_n is measured by

$$\inf_{v \in L^2_\mu(\mathcal{X}; V_n)} \|u - v\|^2_{L^2(\mathcal{X}; V)} = \int_{\mathcal{X}} d(u(x), V_n)^2 d\mu(x) = \int_{\mathcal{M}} d(f, V_n)^2 d\nu(f)$$

where $\nu = u_\# \mu$ is the push-forward measure of μ through the solution map u.

Optimal subspaces in the worst case setting

Optimal spaces V_n for the worst-case error are solution of

$$\inf_{\dim(V_n)=n} \inf_{v \in L^\infty(\mathcal{X};V_n)} \|u - v\|_{L^\infty(\mathcal{X};V)} = \inf_{\dim(V_n)=n} \sup_{f \in \mathcal{M}} d(f, V_n) := d_n(\mathcal{M})_V$$

$d_n(\mathcal{M})_V$ is the Kolmogorov n-width of the set $\mathcal{M}$ in V which measures how well $\mathcal{M}$ can be approximated by n-dimensional subspaces.

It quantifies the ideal performance of linear approximation methods since for any approximation of u of the form $u_n(x) = \sum_{i=1}^n v_i \varphi_i(x)$,

$$\|u - u_n\|_{L^\infty(\mathcal{X};V)} \geq d_n(\mathcal{M})_V.$$

Upper bounds for $d_n(\mathcal{M})_V$ can be obtained by constructing particular approximations $u_n(x)$ (e.g. polynomial approximations)

Optimal subspaces in the mean-squared setting

Optimal spaces V_n in the mean-squared sense are solution of

$$\inf_{\dim(V_n)=n} \inf_{v \in L^2_\mu(\mathcal{X};V_n)} \|u - v\|^2_{L^2_\mu(\mathcal{X};V)} = \inf_{\dim(V_n)=n} \int_{\mathcal{X}} d(u(x), V_n)^2 d\mu(x) := e_n(u)^2_{L^2}$$

$e_n(u)_{L^2}$ is another notion of linear n-width of the manifold $\mathcal{M}$ equipped with the measure $\nu = u_\# \mu$.

If V is a Hilbert space and μ is a probability measure,

$$e_n(u)^2_{L^2} = \inf_{\dim(V_n)=n} \int_{\mathcal{X}} \|u(x) - P_{V_n} u(x)\|^2_V d\mu(x) = \inf_{\dim(V_n)=n} \mathbb{E}(\|u(X) - P_{V_n} u(X)\|^2_V)$$

and optimal spaces V_n are the n-dimensional principal subspaces of the V-valued random variable $u(X)$.

This corresponds to principal component analysis and the optimal approximation $u_n(x) = P_{V_n} u(x)$ is the truncated Karhunen-Loeve decomposition of $u(X)$.

n-widths for parameter-dependent equations

Consider the parameter-dependent equation

$$-\nabla \cdot (a(x)\nabla u(x)) = f \quad \text{in } D \subset \mathbb{R}^m, \quad u(x) = 0 \quad \text{on } \partial D,$$

with the assumption $0 < \gamma \leq a(x) \leq \beta < \infty$, $\forall x \in \mathcal{X}$.

The problem admits a unique solution $u(x) \in H_0^1(D) = V$ and $\|u(x)\|_V \leq \frac{1}{\gamma}\|f\|_{H^{-1}(D)}$. Therefore the solution manifold $\mathcal{M}$ is a bounded subset of V. This says nothing about the convergence of $d_n(\mathcal{M})_V$.

If $f \in H^{s-1}(D)$, $a(x) \in C^s(D)$ and D is sufficiently regular, then $\mathcal{M}$ is a bounded subset of $H^{s+1}(D)$, therefore compact in V when $s \geq 1$, and

$$d_n(\mathcal{M})_V \lesssim n^{-s/m}.$$

This performance is achieved by generic approximation spaces V_n such as splines on uniform meshes.

Finer assumptions are required to reveal an interest of projection-based model reduction methods.

n-widths for parameter-dependent equations

Consider a particular parametrization

$$a(x) = a_0 + \sum_{i=1}^{d} a_i x_i, \quad x_i \in [-1, 1].$$

From results on best n-term approximations using polynomial bases, we obtain bounds on the n-widths of $\mathcal{M}$.

If $d < \infty$, we have an exponential convergence of $d_n(\mathcal{M})_V$, with a deterioration of the convergence rate when m increases.

If $d = \infty$ and $(\|a_i\|_\infty)_{i \geq 1} \in \ell^r$ for some $r < 1$, then

$$d_n(\mathcal{M})_V \lesssim n^{-s}, \quad s = \frac{1}{r} - 1.$$

n-widths for parameter-dependent equations

More general results have been obtained for parameter-dependent equations

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

where a belongs to some compact set $\mathcal{A}$ of a complex Banach space A (e.g. $L^\infty(D)$).

If $u : a \in \mathcal{A} \mapsto u(a) \in \mathcal{M}$ is holomorphic, then

$$d_n(\mathcal{A})_A \lesssim n^{-s} \quad \Rightarrow \quad d_n(\mathcal{M})_V \lesssim n^{-r} \text{ with } r < s - 1.$$

For details, see [Cohen & DeVore 2015].

Practical construction of subspaces in the mean-squared setting

Optimal subspaces V_n are usually out of reach but suboptimal constructions can be proposed.

In the mean-squared setting, Empirical Principal Component Analysis (or Proper Orthogonal Decomposition) defines subspaces V_n as solutions of

$$\min_{\dim(V_n)=n} \frac{1}{m} \sum_{i=1}^{m} \|u(x^i) - P_{V_n} u(x^i)\|_V^2$$

where $u(x^i)$ are samples of $u(X)$. The resulting spaces V_n are nested subspaces contained in $span\{u(x^1), \ldots, u(x^m)\}$.

Proper Generalized Decomposition (or Generalized Spectral Decomposition) defines spaces V_n solution of

$$\min_{\dim(V_n)=n} \inf_{v \in L^2_\mu(\mathcal{X}; V_n)} \int_{\mathcal{X}} \Delta(u(x), v(x)) \mu(dx).$$

Assuming that $\Delta(u, v) \sim \|u - v\|_V^2$, the resulting spaces V_n are such that

$$\mathbb{E}(\|u(X) - P_{V_n} u(X)\|_V^2) \lesssim e_n(u)_{L^2}^2.$$

Constructive algorithms are obtained by imposing a nestedness property $V_n \supset V_{n-1}$. See [Nouy 2017].

Practical construction of subspaces in the worst-case setting

In the worst-case setting, a greedy algorithm defines spaces

$$V_n = span\{u(x^1), \ldots, u(x^n)\}$$

with adaptively chosen samples

$$x^{n+1} = \arg\max_{x \in \mathcal{X}} \|u(x) - P_{V_n} u(x)\|_V.$$

The quality of V_n is assessed by

$$\sigma_n = \sup_{f \in \mathcal{M}} \|f - P_{V_n} f\|_V$$

- If $d_n(\mathcal{M})_V \lesssim n^{-s}$, then $\sigma_n \lesssim n^{-s}$.
- If $d_n(\mathcal{M})_V \lesssim e^{-an^\alpha}$, then $\sigma_n \lesssim e^{-bn^\alpha}$.

See [DeVore et al 2013]

Practical construction of subspaces in the worst-case setting

In practice, samples are chosen such that

$$x^{n+1} = \arg\max_{x \in \mathcal{X}_N} \Delta(u(x), u_n(x))$$

where $\mathcal{X}_N$ is a discrete (training) set in $\mathcal{X}$, $u_n(x)$ is some projection of $u(x)$ onto V_n (typically a Galerkin projection) and $\Delta(u(x), u_n(x))$ is an estimator of $\|u(x) - u_n(x)\|$. This is the basic idea of reduced basis methods.

An algorithm using a random selection of training sets $\mathcal{X}_N$ is analyzed in [Cohen et al 2018].

Any projection $u_n(x)$ of $u(x)$ onto $V_n = span\{u(x^1), \ldots, u(x^n)\}$ interpolates the solution map u at points $\{x^1, \ldots, x^n\}$.

For parameter-dependent equations $A(x)u(x) = f(x)$ with $A(x) : V \to W$, a Galerkin projection can be defined by

$$u_n(x) = \arg\min_{v \in V_n} \|A(x)v - f(x)\|_W.$$

If $A(x)$ is linear and $A(x)$ and $f(x)$ depend polynomially in x, $u_n(x)$ is a rational interpolation of $u(x)$.

Outline

Model classes for high-dimensional approximation

Standard model classes include

- Linear models

$$a_1 x_1 + \ldots + a_d x_d$$

- Polynomial models

$$\sum_{\alpha \in \Lambda} a_\alpha x^\alpha$$

where $\Lambda \subset \mathbb{N}^d$ is a set of multi-indices, either fixed (linear approximation) or free (nonlinear approximation).

Other model classes include

- More general expansions

$$\sum_{i=1}^{n} a_i \psi_i(x)$$

where the ψ_i are either fixed (linear approximation) or freely selected in a dictionary of functions (nonlinear approximation).

Model classes for high-dimensional approximation

- Additive models

$$u_1(x_1) + \ldots + u_d(x_d)$$

or more generally

$$\sum_{\alpha \subset T} u_\alpha(x_\alpha)$$

where $T \subset 2^{\{1,\ldots,d\}}$ is either fixed (linear approximation) or a free parameter (nonlinear approximation).

- Multiplicative models

$$u_1(x_1) \ldots u_d(x_d)$$

or more generally

$$\prod_{\alpha \in T} u_\alpha(x_\alpha)$$

where $T \subset 2^{\{1,\ldots,d\}}$ is either a fixed or a free parameter.

Composition of functions

$$f(g(x)) = f(g_1(x), \ldots, g_m(x))$$

with g is a map from $\mathbb{R}^d$ to $\mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}$ has a low-dimensional parametrization.

- Linear transformations (ridge functions)

  $$f(Wx), \quad W \in \mathbb{R}^{m \times d}$$

  A typical example is the perceptron

  $$f(y) = a\sigma(w^T x + b)$$

- For large m, requires specific models for f, e.g.

  $$f(g_1(x), \ldots, g_m(x)) = f_1(g_1(x)) + \ldots + f_m(g_m(x))$$

  A sum of m perceptrons is a shallow neural network (with one hidden layer of width m)

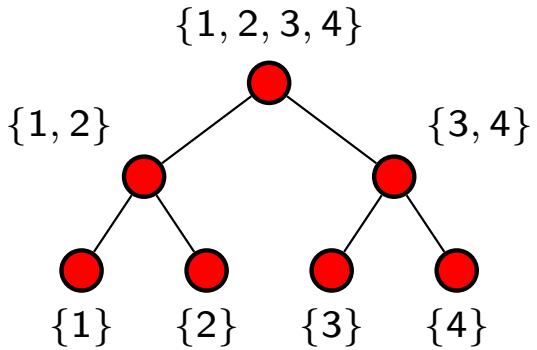
  $$\sum_{i=1}^{m} a_i \sigma(w_i{}^T x + b_i)$$

More compositions... deep neural networks

$$g_L \circ g_{L-1} \circ \ldots \circ g_2 \circ g_1(x)$$

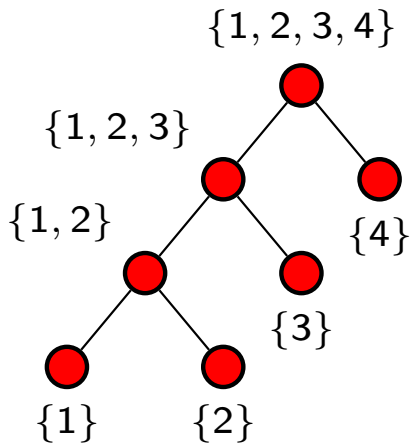
- Deep convolutional networks

$$f_{1,2,3,4}\left(f_{1,2}\left(f_1(x_1), f_2(x_2)\right), f_{3,4}\left(f_3(x_3), f_4(x_4)\right)\right)$$



- Deep recurrent networks

$$f_{1,2,3,4}\left(f_{1,2,3}\left(f_{1,2}\left(f_1(x_1), f_2(x_2)\right), f_3(x_3)\right), f_4(x_4)\right)$$



Low rank tensor formats

A multivariate function $v(x_1, \ldots, x_d)$ is identified with an an element of a tensor product space

$$\mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_d$$

where $\mathcal{H}_\nu$ is a vector space of functions of the variable x_ν.

- Function with rank one (elementary tensor)

$$v(x) = u_1(x_1) \ldots u_d(x_d)$$

- Function with canonical rank r

$$v(x) = \sum_{k=1}^{r} u_1^k(x_1) \ldots u_d^k(x_d)$$

Low rank tensor formats

- For a subset of variables $\alpha \subset \{1, \ldots, d\} := D$, $v(x)$ can be identified with a bivariate function

$$v(x_\alpha, x_{\alpha^c}),$$

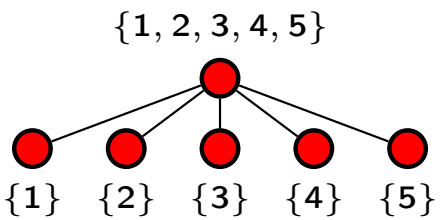
where x_α and x_{α^c} are complementary groups of variables. The canonical rank of this bivariate function is called the α-rank of v, denoted $\text{rank}_\alpha(v)$, which is the minimal integer r_α such that

$$v(x) = \sum_{k=1}^{r_\alpha} v_k^\alpha(x_\alpha) w_k^{\alpha^c}(x_{\alpha^c})$$

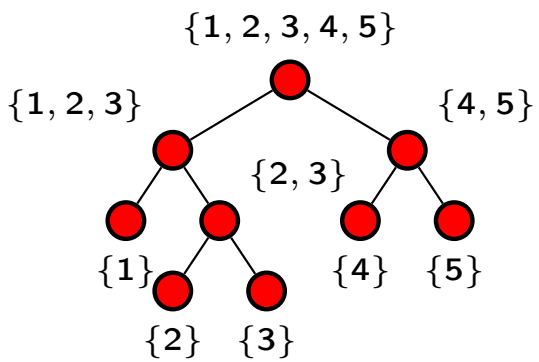
- For $T \subset 2^D$ a collection of subsets of D, a tensor format is defined by

$$\mathcal{T}_r^T = \{v : \text{rank}_\alpha(v) \leq r_\alpha, \alpha \in T\}$$

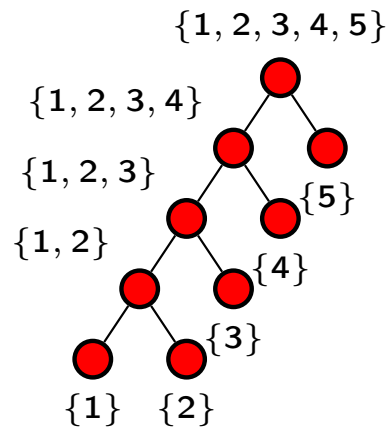
- Tree-based formats correspond to a tree-structured T.



Tucker format



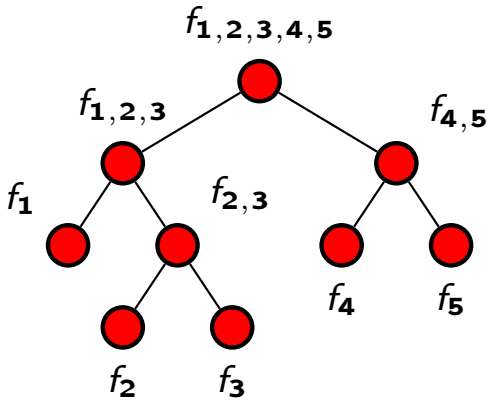
Hierarchical Tucker



Tensor Train format

Tree-based tensor formats as deep networks

- A tensor v in $\mathcal{T}_r^T$ admits a parametrization with parameters $\{f_\alpha\}_{\alpha \in T}$ forming a tree network of low dimensional multilinear functions (tensors).



$$v(x) = f_{1,2,3,4,5}\left(f_{1,2,3}\left(f_1(x_1), f_{2,3}(f_2(x_2), f_3(x_3)\right), f_{4,5}\left(f_4(x_4), f_5(x_5)\right)\right)$$

where for $1 \leq \nu \leq d$, $f_\nu : \mathcal{X}_\nu \to \mathbb{R}^{r_\nu}$, and for any node α with children $\beta_1 ... \beta_s$,

$$f_\alpha : \mathbb{R}^{r_{\beta_1}} \times ... \times \mathbb{R}^{r_{\beta_s}} \to \mathbb{R}^{r_\alpha}$$

is a multilinear function, which is identified with a tensor in $\mathbb{R}^{r_\alpha \times r_{\beta_1} \cdots \times r_{\beta_s}}$.

- Corresponds to a deep network with particular architecture and multilinear functions.
- Very specific structure allowing the design of stable algorithms for constructing approximations in this format.

Conclusions

A lot remains to be done for nonlinear approximation tools:

- characterize classes of functions for which these approximation tools achieve a certain performance (e.g. algebraic or exponential rates of convergence).
- find problems sthat involve these classes of functions,
- provide algorithms (interpolation, regression, Galerkin...) that achieve (almost) the ideal performance.

References I

- **Polynomial and Sparse approximation**

Ivo Babuska, Fabio Nobile, and Raul Tempone.
A stochastic collocation method for elliptic partial differential equations with random input data.
SIAM Review, 52(2):317–355, 2010.

G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone.
Analysis of discrete L^2 projection on polynomial spaces with random evaluations.
Foundations of Computational Mathematics, 14(3):419–456, 2014.

A. Chkifa, A. Cohen, and C. Schwab.
High-dimensional adaptive sparse polynomial interpolation and applications to parametric pdes.
Foundations of Computational Mathematics, 14(4):601–633, 2014.

Albert Cohen and Ronald DeVore.
Approximation of high-dimensional parametric pdes.
Acta Numerica, 24:1–159, 5 2015.

A. Cohen and G. Migliorati.
Optimal weighted least-squares methods.
SMAI Journal of Computational Mathematics, 3:181–203, 2017.

References II


Y. Maday, N. C. Nguyen, A. T. Patera, and G. S. H. Pau.
A general multipurpose interpolation procedure: the magic points.
Communications On Pure and Applied Analysis, 8(1):383–404, 2009.

- **Projection-based model order reduction - Reduced Basis methods**


A. Nouy.
Low-Rank Tensor Methods for Model Order Reduction, pages 857–882.
Springer International Publishing, Cham, 2017.


Alfio Quarteroni, Andrea Manzoni, and Federico Negri.
Reduced Basis Methods for Partial Differential Equations: An Introduction, volume 92.
Springer, 2015.


Jan S. Hesthaven, Gianluigi Rozza, and Benjamin Stamm.
Certified Reduced Basis Methods for Parametrized Partial Differential Equations.
Springer Briefs in Mathematics. Springer, Switzerland, 1 edition, 2015.


R. DeVore, G. Petrova, and P. Wojtaszczyk.
Greedy algorithms for reduced bases in Banach spaces.
Constructive Approximation, 37(3):455–466, 2013.


Albert Cohen and Ronald DeVore.
Kolmogorov widths under holomorphic mappings.
IMA Journal of Numerical Analysis, page dru066, 2015.

References III

Albert Cohen, Wolfgang Dahmen, and Ronald DeVore.
Reduced Basis Greedy Selection Using Random Training Sets.
arXiv e-prints, page arXiv:1810.09344, October 2018.

Y. Maday and O. Mula.
A generalized empirical interpolation method: Application of reduced basis techniques to data assimilation.
In Franco Brezzi, Piero Colli Franzone, Ugo Gianazza, and Gianni Gilardi, editors, *Analysis and Numerics of Partial Differential Equations*, volume 4 of *Springer INdAM Series*, pages 221–235. Springer Milan, 2013.

- **Low-rank tensors**

W. Hackbusch.
Tensor spaces and numerical tensor calculus, volume 42 of *Springer series in computational mathematics*.
Springer, Heidelberg, 2012.

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 methods for high-dimensional approximation and model order reduction.
In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model Reduction and Approximation: Theory and Algorithms*. SIAM, Philadelphia, PA, 2017.

References IV

Markus Bachmayr, Reinhold Schneider, and André Uschmajew.

Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations.

Foundations of Computational Mathematics, pages 1–50, 2016.

B. B. Khoromskij and C. Schwab.

Tensor-structured galerkin approximation of parametric and stochastic elliptic pdes.

SIAM Journal on Scientific Computing, 33(1):364–385, 2011.

A. Nouy.

Higher-order principal component analysis for the approximation of tensors in tree-based low-rank formats.

Numerische Mathematik, 2019. Arxiv eprint available.

E. Grelier, A. Nouy, and M. Chevreuil.

Learning with tree-based tensor formats.

arXiv e-prints, page arXiv:1811.04455, November 2018.