# Low-rank and sparse methods for high-dimensional approximation and model order reduction

—–

## Lecture 8

Sampling methods for low-rank tensor approximation: a subspace point of view

# Introduction

Here, we consider the approximation of a function $u$ defined on a set $\mathcal{X}$ using samples of the function at some points $x^k$ in $\mathcal{X}$.

We first consider the case where

$$u : \mathcal{X} \to \mathcal{V}$$

is a function taking values in some vector space $\mathcal{V}$ (e.g. $\mathbb{R}^N$ or a function space).

Then we consider the approximation of a real-valued bivariate function defined on a product domain

$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2,$$

before considering the approximation of a real-valued multivariate function defined on a product domain

$$\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d.$$

# Outline


1. Low-rank approximation of a vector-valued function
2. Low-rank approximation of a bivariate function
3. Low-rank approximation of a multivariate function in Tucker format

# Outline


1. Low-rank approximation of a vector-valued function
2. Low-rank approximation of a bivariate function
3. Low-rank approximation of a multivariate function in Tucker format

# Vector-valued functions

Let $X$ be a random variable with values in a set $\mathcal{X}$ and let $\mu$ denote the probability law of $X$.

We here consider a function $u : \mathcal{X} \to \mathcal{V}$ defined on $\mathcal{X}$ with values in a Hilbert space $\mathcal{V}$.

We assume that $u$ is in the Hilbert tensor space

$$L^2_\mu(\mathcal{X};\mathcal{V}) = \overline{\mathcal{V} \otimes L^2_\mu(\mathcal{X})}$$

equipped with the canonical norm

$$\|u\| = \left(\int_{\mathcal{X}} \|u(x)\|^2_{\mathcal{V}} d\mu(x)\right)^{1/2} = \mathbb{E}(\|u(X)\|^2_{\mathcal{V}})^{1/2}.$$

# Best rank-$r$ approximation of vector-valued functions

We consider the problem of best approximation of $u$ by a rank-$r$ function

$$\min_{v \in \mathcal{R}_r} \|u - v\| = \min_{v \in \mathcal{R}_r} \left( \int_{\mathcal{X}} \|u(x) - v(x)\|_{\mathcal{V}}^2 d\mu(x) \right)^{1/2},$$

where

$$\mathcal{R}_r = \left\{ x \mapsto \sum_{i=1}^r v_i \lambda_i(x) : v_i \in \mathcal{V}, \lambda_i \in L^2_\mu(\mathcal{X}) \right\}.$$

# Best rank-$r$ approximation of vector-valued functions

A solution $u_r$ of the best rank-$r$ approximation problem is given by the truncated singular value decomposition of $u$:

$$u_r(x) = \sum_{i=1}^{r} \sigma_i v_i s_i(x),$$

where the $\sigma_i$ are the dominant singular values of $u$, the $v_i \in \mathcal{V}$ are the dominant left singular vectors and the $s_i \in L^2_\mu(\mathcal{X})$ are the dominant right singular vectors.

The $\{(v_i, \sigma_i^2)\}_{i=1}^r$ are the $r$ dominant eigenpairs of the correlation operator $C(u) : \mathcal{V} \to \mathcal{V}$ defined by

$$C(u)v = \int_{\mathcal{X}} u(x)(u(x), v)_{\mathcal{V}} d\mu(x) = \mathbb{E}(u(X)(u(X), v)_{\mathcal{V}}),$$

which is a compact and self-adjoint operator.

# Relation to principal component analysis

By introducing the subspace-based parametrization of $\mathcal{R}_r$,

$$\mathcal{R}_r = \left\{ v \in \mathcal{V}_r \otimes L^2_\mu(\mathcal{X}) : \mathcal{V}_r \subset \mathcal{V}, \dim(\mathcal{V}_r) = r \right\},$$

the best rank-$r$ approximation problem can be equivalently formulated as an optimization problem over $r$-dimensional spaces in $\mathcal{V}$:

$$\min_{\dim(\mathcal{V}_r)=r} \min_{v \in \mathcal{V}_r \otimes L^2_\mu(\mathcal{X})} \|u - v\|^2 = \min_{\dim(\mathcal{V}_r)=r} \int_{\mathcal{X}} \|u(x) - P_{\mathcal{V}_r} u(x)\|^2_{\mathcal{V}} d\mu(x),$$

where $P_{\mathcal{V}_r}$ is the orthogonal projection from $\mathcal{V}$ onto $\mathcal{V}_r$.

A solution to this problem is given by the subspace

$$\mathcal{V}_r = \text{span}\{v_1, \ldots, v_r\}$$

which is generated by the dominant singular vectors of $u$, also called principal components of $u$.

The truncated singular value decomposition $u$ is such that

$$u_r(x) = P_{\mathcal{V}_r} u(x).$$

# Relation with optimal model order reduction

The best rank-$m$ approximation error

$$d_r^{(2)}(u) = \min_{\dim(\mathcal{V}_r)=r} \left( \int_{\mathcal{X}} \|u(x) - P_{\mathcal{V}_r} u(x)\|_{\mathcal{V}}^2 d\mu(x) \right)^{1/2}$$

measures how well the set

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

can be approximated by a $r$-dimensional space $\mathcal{V}_r$.

It quantifies the ideal performance of a reduced basis method with respect to the $L^2$-norm.

# Relation with optimal model order reduction

Since $\|u\| \leq \sup_{x\in\mathcal{X}} \|u(x)\|_{\mathcal{V}}$, we have

$$d_r^{(2)}(u) \leq \min_{\dim(\mathcal{V}_r)=r} \sup_{x\in\mathcal{X}} \|u(x) - P_{\mathcal{V}_r} u(x)\|_{\mathcal{V}} = d_r(u(\mathcal{X}))_{\mathcal{V}},$$

where

$$d_r(u(\mathcal{X}))_{\mathcal{V}} = \min_{\dim(\mathcal{V}_r)=r} \sup_{f\in u(\mathcal{X})} \|f - P_{\mathcal{V}_r} f\|_{\mathcal{V}}$$

corresponds to the Kolmogorov $r$-width of $u(\mathcal{X})$ in $\mathcal{V}$, which measures how well $u(\mathcal{X})$ can be approximated uniformly by a $r$-dimensional space.

Note that $d_r(u(\mathcal{X}))_{\mathcal{V}}$, contrary to $d_r^{(2)}(u)$, does not take into account the measure $\mu$.

# Sample-based estimation of principal components

Let us now assume that we have evaluations $u(x^k) \in \mathcal{V}$ of the function $u$ for $K$ random samples $x^k$, $1 \leq k \leq K$.

By introducing a statistical estimation of the expectation, an estimate of the optimal $r$-dimensional space is obtained by solving

$$\min_{\dim(\mathcal{V}_r)=r} \frac{1}{K} \sum_{k=1}^{K} \|u(x^k) - P_{\mathcal{V}_r} u(x^k)\|_{\mathcal{V}}^2. \quad (1)$$

The set of samples $\{u(x^1), \ldots, u(x^K)\} \in \mathcal{V}^K$ can be identified with a tensor

$$\mathbf{u} \in \mathcal{V} \otimes \mathbb{R}^K,$$

and a solution of (1) is given by the dominant left singular space of $\mathbf{u}$.

# Sample-based estimation of principal components

The optimal subspace is the dominant eigenspace of the empirical correlation operator

$$C_K(u) = \frac{1}{K} \sum_{k=1}^{K} u(x^k)(u(x^k), v)_{\mathcal{V}}.$$

**Remark.**

Note that other numerical integration methods using deterministic integration points could be considered for approximating the integral over $\mathcal{X}$.

# Outline

1. Low-rank approximation of a vector-valued function
2. Low-rank approximation of a bivariate function
3. Low-rank approximation of a multivariate function in Tucker format

# Bivariate functions

We consider a pair of independent random variables $X = (X_1, X_2)$ taking values in $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and with probability law $\mu = \mu_1 \otimes \mu_2$.

We consider a bivariate function

$$u : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$$

and assume that $u$ is in the Hilbert tensor space

$$L^2_\mu(\mathcal{X}) = \overline{L^2_{\mu_1}(\mathcal{X}_1) \otimes L^2_{\mu_2}(\mathcal{X}_2)}$$

equipped with the canonical norm

$$\|u\| = \left( \int_{\mathcal{X}_1 \times \mathcal{X}_2} u(x_1, x_2)^2 d\mu_1(x_1) d\mu_2(x_2) \right)^{1/2} = \mathbb{E}(u(X_1, X_2)^2)^{1/2}.$$

A function $v$ with rank $r$ has a representation of the form

$$v(x_1, x_2) = \sum_{i=1}^{r} v_i^{(1)}(x_1) v_i^{(2)}(x_2).$$

# Low-rank approximation of bivariate functions

The problem of best rank-$r$ approximation of the function $u$ is defined by

$$\min_{v \in \mathcal{R}_r} \|u - v\|^2.$$

A solution $u_r$ is given by the truncated singular value decomposition of $u$

$$u_r(x_1, x_2) = \sum_{i=1}^{r} \sigma_i v_i^{(1)}(x_1) v_i^{(2)}(x_2)$$

where the $\sigma_i$ are the $r$ dominant singular values of $u$, and the $v_i^{(1)}$ and $v_i^{(2)}$ are the corresponding left and right singular vectors.

The dominant left and right singular spaces are respectively defined by

$$\mathcal{V}_r^1 = \text{span}\{v_1^{(1)}, \dots, v_r^{(1)}\} \quad \text{and} \quad \mathcal{V}_r^2 = \text{span}\{v_1^{(2)}, \dots, v_r^{(2)}\}.$$

# Low-rank approximation of bivariate functions

The dominant left singular space $\mathcal{V}_r^1$ is a solution of the optimization problem over $r$-dimensional spaces in $L^2_\mu(\mathcal{X}_1)$

$$\min_{\dim(\mathcal{V}_r^1)=r} \|u - (P_{\mathcal{V}_r^1} \otimes I)u\|.$$

Also, the dominant right singular space $\mathcal{V}_r^2$ is a solution of the optimization problem over $r$-dimensional spaces in $L^2_\mu(\mathcal{X}_2)$

$$\min_{\dim(\mathcal{V}_r^2)=r} \|u - (I \otimes P_{\mathcal{V}_r^2})u\|.$$

Then, given these optimal spaces $\mathcal{V}_r^1$ and $\mathcal{V}_r^2$, a best rank-$r$ approximation is given by

$$u_r = \arg \min_{v \in \mathcal{V}_r^1 \otimes \mathcal{V}_r^2} \|u - v\|$$

such that

$$u_r = (P_{\mathcal{V}_r^1} \otimes I)u = (I \otimes P_{\mathcal{V}_r^2})u = (P_{\mathcal{V}_r^1} \otimes P_{\mathcal{V}_r^2})u.$$

# Sample based estimation of optimal spaces

Given $K_1$ samples $\{x_2^k\}_{k=1}^{K_1}$ of $X_2$ and corresponding partial evaluations $\{u(\cdot, x_2^k)\}_{k=1}^{K_1}$ of $u$, an estimation $\mathcal{V}_r^1$ of the optimal left space can be obtained by solving

$$\min_{\dim(\mathcal{V}_r^1)=r} \frac{1}{K_1} \sum_{k=1}^{K_1} \| u(\cdot, x_2^k) - P_{\mathcal{V}_r^1} u(\cdot, x_2^k) \|^2_{L^2_{\mu_1}(\mathcal{X}_1)},$$

whose solution is given by the dominant eigenspace of the empirical correlation operator $C^1_{K_1}(u)$ defined for $v \in L^2_{\mu_1}(\mathcal{X}_1)$ by

$$C^1_{K_1}(u)v = \frac{1}{K_1} \sum_{k=1}^{K_1} u(\cdot, x_2^k) \int_{\mathcal{X}_1} u(x_1, x_2^k) v(x_1) \mu_1(dx_1).$$

Similarly, given $K_2$ samples $\{x_1^k\}_{k=1}^{K_2}$ of $X_1$ and corresponding partial evaluations $\{u(x_1^k, \cdot)\}_{k=1}^{K_2}$ of $u$, we obtain an estimation $\mathcal{V}_r^2$ of the optimal right space.

# Projection on reduced spaces

Once subspaces $\mathcal{V}_r^1$ and $\mathcal{V}_r^2$ have been obtained, a rank-$r$ approximation can be obtained by solving approximately the best approximation problem

$$u_r = \arg \min_{v \in \mathcal{V}_r^1 \otimes \mathcal{V}_r^2} \|u - v\|.$$

This can be achieved by a least-squares approach using $K$ random samples $\{x^k\}_{k=1}^K$ of $X$ (possibly the available samples) or any other approximate projection method.

# Outline

1. Low-rank approximation of a vector-valued function
2. Low-rank approximation of a bivariate function
3. Low-rank approximation of a multivariate function in Tucker format

# Multivariate functions

We consider a set of independent random variables $X = (X_1, \ldots, X_d)$ taking values in $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$ and with probability law $\mu = \mu_1 \otimes \ldots \otimes \mu_d$.

We consider a multivariate function

$$u : \mathcal{X}_1 \times \ldots \times \mathcal{X}_d \to \mathbb{R}$$

and assume that $u$ is in the Hilbert tensor space

$$L^2_\mu(\mathcal{X}) = \overline{L^2_{\mu_1}(\mathcal{X}_1) \otimes \ldots \otimes L^2_{\mu_d}(\mathcal{X}_d)}$$

equipped with the canonical norm

$$\|u\| = \left( \int_{\mathcal{X}_1 \times \ldots \times \mathcal{X}_d} u(x_1, \ldots, x_d)^2 d\mu_1(x_1) \ldots d\mu_d(x_d) \right)^{1/2} = \mathbb{E}(u(X_1, \ldots, X_d)^2)^{1/2}.$$

# Best approximation in Tucker format

A best approximation of $u$ with Tucker rank bounded by $r = (r_1, \ldots, r_d)$ is solution of

$$\min_{\mathcal{T}_r} \|u - v\|^2.$$

Noting that

$$\mathcal{T}_r = \left\{ v \in \mathcal{V}^1_{r_1} \otimes \ldots \otimes \mathcal{V}^d_{r_d} : \mathcal{V}^\nu_{r_\nu} \subset L^2_{\mu_\nu}(\mathcal{X}_\nu),\ \dim(\mathcal{V}^\nu_{r_\nu}) = r_\nu,\ 1 \leq \nu \leq d \right\},$$

the best approximation problem can be rewritten as an optimization problem over $r_\nu$-dimensional spaces $\mathcal{V}^\nu_{r_\nu}$:

$$\min_{\dim(\mathcal{V}^1_{r_1})=r_1} \cdots \min_{\dim(\mathcal{V}^d_{r_d})=r_d} \|u - (P_{\mathcal{V}^1_{r_1}} \otimes \ldots \otimes P_{\mathcal{V}^d_{r_d}})u\|^2$$

# Quasi-best approximation in Tucker format

We consider the $d$ independent optimization problems

$$\min_{\dim(\mathcal{V}^{\nu}_{r_\nu})=r_\nu} \|u - (I \otimes \ldots \otimes P_{\mathcal{V}^{\nu}_{r_\nu}} \otimes \ldots \otimes I)u\|^2, \quad 1 \leq \nu \leq d,$$

whose solutions are such that

$$u_r = (P_{\mathcal{V}^{1}_{r_1}} \otimes \ldots \otimes P_{\mathcal{V}^{d}_{r_d}})u \in \mathcal{T}_r$$

satisfies the quasi-optimality property

$$\|u - u_r\| \leq \sqrt{d} \min_{v \in \mathcal{T}_r} \|u - v\|.$$

$u_r$ is the truncated higher-order singular value decomposition of $u$.

# Estimation of sub-optimal subspaces

For a given $\nu \in D = \{1, \dots, d\}$, the function $u(X)$ can be seen as a bivariate function $u(X_\nu, X_{\nu^c})$, where $X_{\nu^c}$ denotes the group of $d-1$ variables $(X_\eta)_{\eta \in \nu^c}$.

Then, the optimization problem which defines a quasi-optimal subspace $\mathcal{V}^\nu_{r_\nu}$ can be written

$$\min_{\dim(\mathcal{V}^\nu_{r_\nu}) = r_\nu} \|u - (P_{\mathcal{V}^\nu_{r_\nu}} \otimes I)u\|^2.$$

This problem can be treated as for the case of bivariate functions.

# Estimation of quasi-optimal subspaces

Given $K_\nu$ samples $\{x^k_{\nu^c}\}_{k=1}^{K_\nu}$ of the group of variables $X_{\nu^c}$, an estimation $\mathcal{V}^\nu_{r_\nu}$ of the quasi-optimal subspace can be obtained by solving

$$\min_{\dim(\mathcal{V}^\nu_{r_\nu})=r_\nu} \frac{1}{K_\nu}\sum_{k=1}^{K_\nu} \|u(\cdot,x^k_{\nu^c}) - P_{\mathcal{V}^\nu_{r_\nu}} u(\cdot,x^k_{\nu^c})\|^2_{L^2_{\mu_\nu}(\mathcal{X}_\nu)},$$

whose solution is given by the dominant eigenspace of the empirical correlation operator $C^\nu_{K_\nu}(u)$ defined for $v \in L^2_{\mu_\nu}(\mathcal{X}_\nu)$ by

$$C^\nu_{K_\nu}(u)v = \frac{1}{K_\nu}\sum_{k=1}^{K_\nu} u(\cdot,x^k_{\nu^c})\int_{\mathcal{X}_\nu} u(x_\nu,x^k_{\nu^c})v(x_\nu)\mu_\nu(dx_\nu).$$

# Estimation of quasi-optimal subspaces

In practice, the univariate functions $x_\nu \mapsto u(x_\nu, x^k_{\nu^c})$ can not be computed exactly and needs to be approximated.
For example, this can be done interpolation on a grid $\{t^i_\nu\}_{i=1}^{N_\nu}$ in $\mathcal{X}_\nu$.

The computation of $\mathcal{V}^\nu_{r_\nu}$ therefore requires the evaluation of $u$ at the $N_\nu K_\nu$ points of the tensorized grid

$$\{(t^i_\nu, x^k_{\nu^c}) : 1 \leq i \leq N_\nu, 1 \leq k \leq K_\nu\} \subset \mathcal{X}_\nu \times \mathcal{X}_{\nu^c}.$$

Then, the estimation of the quasi-optimal reduced space

$$\mathcal{V}^1_{r_1} \otimes \ldots \otimes \mathcal{V}^d_{r_d}$$

requires $\sum_{\nu=1}^d N_\nu K_\nu$ evaluations of the function, which scales linearly with the dimension $d$.

# Estimation of the projection on the reduced tensor product space

Once subspaces $\mathcal{V}^{\nu}_{r_\nu}$ have been computed, an optimal approximation $u_r$ in $\mathcal{V}^{1}_{r_1} \otimes \ldots \otimes \mathcal{V}^{d}_{r_d}$ is then defined by

$$u_r = \arg \min_{v \in \mathcal{V}^{1}_{r_1} \otimes \ldots \otimes \mathcal{V}^{d}_{r_d}} \|u - v\| = (P_{\mathcal{V}^{1}_{r_1}} \otimes \ldots \otimes P_{\mathcal{V}^{d}_{r_d}})u.$$

Of course, when the dimension $d$ is high, low-dimensional structures of $u_r$ have to be exploited. Sparse or low-rank tensor methods can be applied here.

Note that for each dimension $\nu$, there exists a natural hierarchy of subspaces

$$\mathcal{V}^{\nu}_{1} \subset \ldots \subset \mathcal{V}^{\nu}_{r_\nu}$$

given by dominant singular spaces of increasing dimension. This yields a straightforward strategy for applying working set algorithms for sparse tensor methods, either for computing a sparse approximation of the tensor $u_r$ or a low-rank approximation of $u_r$ with sparse parameters.