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

Lecture 7

Higher-order singular value decompositions and related tensor truncation schemes

Introduction

In this lecture, we consider a tensor u in a tensor product of Hilbert spaces $V = V^1 \otimes \ldots \otimes V^d$ and we assume that u is given as a full tensor or in a certain low-rank format.

We present truncation schemes for finding a low-rank approximation of u with reduced complexity, relying on the standard singular value decomposition of order-two tensors.

Although most of the ideas naturally extend to the case of infinite-dimensional tensors, we consider that the tensor space V in finite dimensional.

Outline

Outline

Singular value decomposition of matrices

A matrix $u \in \mathbb{R}^{n\times m}$ admits a singular value decomposition (SVD)

$$u = \sum_{k=1}^{N} \sigma_k v_k w_k^T,$$

where $N = \min\{n, m\}$, $\sigma_k \in \mathbb{R}^+$ are the singular values, and v_k and w_k are the associated singular vectors, which are orthonormal vectors. It can be equivalently written

$$u = VSW^T$$

where $V \in \mathbb{R}^{n\times N}$ and $W = \mathbb{R}^{m\times N}$ are orthogonal matrices, and $S = \mathrm{diag}(\sigma_1, \dots, \sigma_N) \in \mathbb{R}^{N\times N}$ is a diagonal matrix.

The set of singular values of u is denoted $\sigma(u) = \{\sigma_k(u)\}_{k=1}^N$.

Singular values of matrices and related matrix norms

The rank of u is the number of non-zero singular values,

$$\text{rank}(u) = \|\sigma(u)\|_0 = \#\{k : \sigma_k(u) \neq 0\}.$$

The canonical norm of u is

$$\|u\| = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} u_{i,j}^2} = \sqrt{\sum_{k=1}^{N} \sigma_k^2} = \|\sigma(u)\|_2,$$

which corresponds to the Frobenius norm.

It is a particular case of Schatten p-norms which are defined for $1 \leq p \leq \infty$ by

$$\|u\|_{\sigma_p} = \|\sigma(u)\|_p.$$

Singular value decomposition of order-two tensors

The notion of singular value decomposition can be extended to the case of Hilbert tensor spaces $\overline{V \otimes W}^{\|\cdot\|}$, where V and W are Hilbert spaces (e.g. spaces of functions), even infinite-dimensional, and where $\|\cdot\|$ denote the canonical norm on $V \otimes W$ (the Frobenius norm for u a matrix).

A tensor $u \in \overline{V \otimes W}^{\|\cdot\|}$ admits a singular value decomposition

$$u = \sum_{k=1}^{N} \sigma_k v_k \otimes w_k,$$

with $N = \min\{dim(V), dim(W)\} \in \mathbb{N} \cup \{\infty\}$, where v_k and w_k are orthonormal vectors.

The canonical norm

$$\|u\| = \|\sigma(u)\|_2$$

is also called the Hilbert-Schmidt norm.

Singular value decomposition of order-two tensors

Example 1 (Proper Orthogonal Decomposition)

For $\Omega \times I$ a space-time domain and V a Hilbert space of functions defined on Ω, a function $u \in L^2(I; V)$ admits a singular value decomposition

$$u(t) = \sum_{k=1}^{\infty} \sigma_k v_k w_k(t)$$

which is known as the Proper Orthogonal Decomposition (POD).

Example 2 (Karhunen-Loeve decomposition)

For a probability space (Ω, μ), an element $u \in L^2_\mu(\Omega; V)$ is a second-order V-valued random variable. If u is zero-mean, the singular value decomposition of u is known as the Karhunen-Loeve decomposition

$$u(\omega) = \sum_{k=1}^{\infty} \sigma_k v_k w_k(\omega)$$

where $w_k : \Omega \to \mathbb{R}$ are uncorrelated (orthogonal) random variables.

Outline

Truncated singular value decomposition for order-two tensors

Let u be an order-two tensor in the Hilbert space $V \otimes W$, where V and W are Hilbert spaces, and let $\|\cdot\|$ denote the canonical norm on $V \otimes W$ (the Frobenius norm for u a matrix).

Let consider a tensor u in $V \otimes W$ with singular value decomposition

$$u = \sum_{k=1}^{N} \sigma_k v_k \otimes w_k,$$

where the singular values $\sigma(u) = \{\sigma_k\}_{k=1}^{N}$ are sorted by decreasing order.

An element of best approximation of u in the set of tensors with rank bounded by r is provided by the truncated singular value decomposition

$$u_r = \sum_{k=1}^{r} \sigma_k v_k \otimes w_k,$$

such that

$$\|u - u_r\|^2 = \min_{\mathrm{rank}(v) \leq r} \|u - v\|^2 = \sum_{k=r+1}^{N} \sigma_k^2.$$

Truncated singular value decomposition for order-two tensors

An approximation u_r with relative precision ϵ, such that

$$\|u - u_r\| \leq \epsilon \|u\|,$$

can be obtained by choosing a rank r such that

$$\sum_{k=r+1}^{N} \sigma_k^2 \leq \epsilon \sum_{k=1}^{N} \sigma_k^2.$$

Remark.

The complexity of computing the singular value decomposition of a tensor u is $O(n^3)$ if $\dim(V) = \dim(W) = O(n)$. If u is given in low-rank format $u = \sum_{k=1}^{R} a_k \otimes b_k$, with a rank $R < n$, the complexity breaks down to $O(R^3 + 2Rn^2)$.

Truncated singular value decomposition for order-two tensors

The truncated singular value can be interpreted as an orthogonal projection onto linear spaces generated by singular vectors.

Denoting by $V_r = \text{span}\{v_k\}_{k=1}^r$ and $W_r = \text{span}\{w_k\}_{k=1}^r$ the dominant singular spaces of u, and by $P_{V_r} : V_r \to V$ and $P_{W_r} : W \to W_r$ the orthogonal projections onto V_r and W_r respectively, we have

$$u_r = (P_{V_r} \otimes P_{W_r})u = (P_{V_r} \otimes I)u = (I \otimes P_{W_r})u.$$

Outline

Higher-order tensors as order-two tensors...

For a non-empty subset α in $D = \{1, \ldots, d\}$, a tensor $u \in V^1 \otimes \ldots \otimes V^d$ can be identified with its matricisation

$$\mathcal{M}_\alpha(u) \in V^\alpha \otimes V^{\alpha^c},$$

an order-two tensor which admits a singular value decomposition

$$\mathcal{M}_\alpha(u) = \sum_{k \geq 1} \sigma_k^\alpha v_k^\alpha \otimes w_k^{\alpha^c}.$$

The set $\sigma^\alpha(u) := \{\sigma_k^\alpha\}_{k \geq 1}$ is called the set of α-singular values of u. The α-rank of u is the number of non-zero α-singular values

$$\text{rank}_\alpha(u) = \|\sigma^\alpha(u)\|_0.$$

Higher-order tensors as order-two tensors...

By sorting the α-singular values by decreasing order, an approximation u_r with α-rank r can be obtained by retaining the r largest singular values, i.e.

$$u_r \quad \text{such that} \quad \mathcal{M}_\alpha(u_r) = \sum_{k=1}^{r} \sigma_k^\alpha v_k^\alpha \otimes w_k^{\alpha^c},$$

which satisfies

$$\|u - u_r\|^2 = \min_{\text{rank}_\alpha(v) \leq r} \|u - v\|^2 = \sum_{k>r} (\sigma_k^\alpha)^2.$$

There are 2^{d-1} different binary partitions $\alpha \cup \alpha^c$ of D, to each of which corresponds a singular value decomposition and a way to truncate a higher-order tensor !

Truncation scheme for the approximation in Tucker format

For each $\nu \in \{1, \dots, d\}$, we consider the singular value decomposition of the matricisation $\mathcal{M}_\nu(u)$ of a tensor u

$$\mathcal{M}_\nu(u) = \sum_{k \geq 1} \sigma_k^\nu v_k^\nu \otimes w_k^\nu .$$

Let $U_{r_\nu}^\nu = \mathrm{span}\{v_k^\nu\}_{k=1}^{r_\nu}$ be the subspace of V^ν generated by the r_ν dominant left singular vectors of $\mathcal{M}_\nu(u)$, and by $P_{U_{r_\nu}}$ the orthogonal projection from V^ν to $U_{r_\nu}^\nu$.

The tensor

$$u_r = (P_{U_{r_1}^1} \otimes \dots \otimes P_{U_{r_d}^d})u$$

is a projection of u onto the reduced tensor space

$$U_{r_1}^1 \otimes \dots \otimes U_{r_d}^d$$

and therefore, u_r is an element of the subset $\mathcal{T}_r$ of tensors with Tucker rank bounded by $r = (r_1, \dots, r_d)$,

$$\mathcal{T}_r = \{v \in U^1 \otimes \dots \otimes U^d : U^\nu \subset V^\nu, \dim(U^\nu) = r_\nu, 1 \leq \nu \leq d\}.$$

The sequence of approximations u_r is called a higher-order singular value decomposition (for Tucker format).

Higher-order singular value decomposition for Tucker format

The operator

$$\mathcal{P}^{\nu}_{r_\nu} = \mathcal{M}_\nu^{-1} P_{U^{\nu}_{r_\nu}} \mathcal{M}_\nu = I \otimes \ldots \otimes P_{U^{\nu}_{r_\nu}} \otimes \ldots \otimes I$$

is the orthogonal projection from V onto

$$V^1 \otimes \ldots \otimes U^{\nu}_{r_\nu} \otimes \ldots \otimes V^d,$$

which is such that

$$\|u - \mathcal{P}^{\nu}_{r_\nu} u\| = \min_{\text{rank}_\nu(v) \le r_\nu} \|u - v\| = \sum_{k \ge r_\nu + 1} (\sigma_k^\nu)^2.$$

The approximation u_r can then be written

$$u_r = \mathcal{P}^{1}_{r_1} \ldots \mathcal{P}^{d}_{r_d} u,$$

and satisfies

$$\|u - u_r\|^2 = \sum_{\nu=1}^{d} \|u - \mathcal{P}^{\nu}_{r_\nu} u\|^2 = \sum_{\nu=1}^{d} \min_{\text{rank}_\nu(v) \le r_\nu} \|u - v\|^2,$$

from which we deduce the quasi-optimality property

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

Truncation scheme for the approximation in Tucker format

Also, from

$$\|u - u_r\|^2 = \sum_{\nu=1}^{d} \|u - \mathcal{P}_{r_\nu}^{\nu} u\|^2 = \sum_{\nu=1}^{d} \sum_{k_\nu > r_\nu} (\sigma_{k_\nu}^{\nu})^2,$$

we deduce that if we select the ranks $(r_1, \dots, r_d)$ such that for each ν

$$\sum_{k_\nu > r_\nu} (\sigma_{k_\nu}^{\nu})^2 \leq \frac{\epsilon^2}{d} \sum_{k_\nu \geq 1} (\sigma_{k_\nu}^{\nu})^2 = \frac{\epsilon^2}{d} \|u\|^2,$$

then the truncated singular value decomposition $\mathcal{P}_{r_\nu}^{\nu} u$ has a relative precision $\epsilon/\sqrt{d}$ and we finally obtain an approximation u_r with relative precision ϵ,

$$\|u - u_r\| \leq \epsilon \|u\|.$$

Note that the definition of u_r is independent on the order of the projections $\mathcal{P}_{r_\nu}^{\nu}$.

Truncation scheme for tree-based tensor formats

For tree-based (hierarchical) low-rank tensor formats

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

where T is a dimension partition tree over $D = \{1, \dots, d\}$, a higher order singular value decomposition (also called hierarchical singular value decomposition) can also be defined from singular value decompositions of matricisations $\mathcal{M}_\alpha(u)$ of a tensor u.

Truncation scheme for tree-based tensor formats

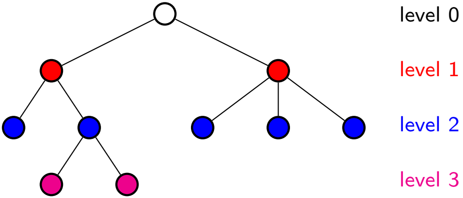
Letting $U^{\alpha}_{r_\alpha}$ be the subspace generated by the r_α dominant left singular vectors of $\mathcal{M}_\alpha(u)$, and letting $P_{U^{\alpha}_{r_\alpha}}$ be the orthogonal projector from V^α to $U^{\alpha}_{r_\alpha}$, we define the orthogonal projection

$$\mathcal{P}^{\alpha}_{r_\alpha} = \mathcal{M}_\alpha^{-1} P_{U^{\alpha}_{r_\alpha}} \mathcal{M}_\alpha.$$

Then, an approximation with tree-based rank $r = (r_\alpha)_{\alpha \in T}$ can be defined by

$$u_r = \mathcal{P}_r^{T,(L)} \mathcal{P}_r^{T,(L-1)} \dots \mathcal{P}_r^{T,(1)} u \quad \text{with} \quad \mathcal{P}^{T,(\ell)} = \prod_{\substack{\alpha \in T \\ \text{level}(\alpha)=\ell}} \mathcal{P}^{\alpha}_{r_\alpha}$$

where we apply to u a sequence of projections ordered by increasing level in the tree (from the root to the leaves). Here $L = \max_{\alpha \in T} \text{level}(\alpha)$.



Truncation scheme for tree-based tensor formats

Remark.

The definition of the projection $\mathcal{P}^{T,(\ell)}$ is independent of the ordering of the projections $\mathcal{P}^{\alpha}_{r_\alpha}$ of the same level ℓ. Subsets $\alpha \in T$ with level ℓ form a partition of D, and $\mathcal{P}^{T,(\ell)} = \bigotimes_{\substack{\alpha \in T \\ \text{level}(\alpha)=\ell}} P_{U^{\alpha}_{r_\alpha}}$ is the orthogonal projection from V onto $\bigotimes_{\substack{\alpha \in T \\ \text{level}(\alpha)=\ell}} U^{\alpha}_{r_\alpha}$.

Truncation scheme for tree-based tensor formats

The obtained approximation u_r is such that

$$\|u - u_r\|^2 = \sum_{\alpha \in T \setminus D} \min_{\text{rank}_\alpha(v) \leq r_\alpha} \|u - v\|^2 = \sum_{\alpha \in T \setminus D} \sum_{k_\alpha > r_\alpha} (\sigma^\alpha_{k_\alpha})^2,$$

from which we deduce that u_r is a quasi-optimal approximation of u in $\mathcal{T}^T_r$ such that

$$\|u - u_r\| \leq C(T) \min_{v \in \mathcal{T}^T_r} \|u - v\|,$$

where $C(T) = \sqrt{\#T - 1}$ is the square root of the number of projections applied to the tensor. The number of nodes of a dimension partition tree T being bounded by $2d - 1$,

$$C(T) \leq \sqrt{2d - 2}.$$

Also, if we select the ranks $(r_\alpha)_{\alpha \in T \setminus D}$ such that for all α

$$\sum_{k_\alpha > r_\alpha} (\sigma^\alpha_{k_\alpha})^2 \leq \frac{\epsilon^2}{C(T)^2} \sum_{k_\alpha \geq 1} (\sigma^\alpha_{k_\alpha})^2 = \frac{\epsilon^2}{C(T)^2} \|u\|^2,$$

we finally obtain an approximation u_r with relative precision ϵ,

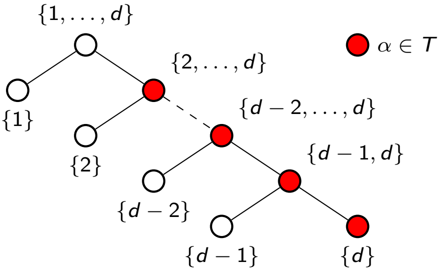
$$\|u - u_r\| \leq \epsilon \|u\|.$$

Truncation scheme for TT format

The procedure for general tree-based formats also applies to the particular case of the TT-format $\mathcal{TT}_r$, which corresponds to $\mathcal{T}_r^T$ for

$$T = \{\{\nu+1,\ldots,d\} : 1 \leq \nu \leq d-1\},$$

which is a subset of a linearly structured tree.



Here, since $\#T = d-1$, the resulting approximation satisfies the quasi-optimality property with

$$C(T) = \sqrt{d-1}.$$

Sequential higher-order singular value decomposition

For T (a subset of) a dimension partition tree, a sequential higher-order singular value decomposition can be defined recursively [Hackbusch 2012, section 11.4.2.3].

Letting $T \setminus D = \{\alpha_1, \ldots, \alpha_M\}$ with $\text{level}(\alpha_{k+1}) \leq \text{level}(\alpha_k)$, we start from $u^{(0)} = u$ and we define a sequence of approximations $u^{(k)}$

$$u^{(k)} = \mathcal{P}^{\alpha_k}_{r_{\alpha_k}} u^{(k-1)}$$

for $1 \leq k \leq M$, where $\mathcal{P}^{\alpha_k}_{r_{\alpha_k}}$ is the projection associated with the dominant α-singular vectors of $u^{(k-1)}$ (and not u).

Finally, we obtain an approximation $u_r := u^{(M)}$ in $\mathcal{T}_r^T$ which satisfies the quasi-optimality property

$$\|u - u_r\| \leq C(T) \min_{v \in \mathcal{T}_r^T} \|u - v\|.$$

where $C(T) = \sqrt{\#T - 1}$ is the square root of the number of projections applied.

For another version of a sequential higher-order singular value decomposition, see [Hackbusch 2012, section 11.4.2.2].

Outline

Hard thresholding of singular values

The hard singular value thresholding operator $\mathcal{HT}_\tau$ is defined for an order-two tensor u with singular value decomposition $\sum_{k=1}^N \sigma_k v_k \otimes w_k$ by

$$\mathcal{HT}_\tau(u) = \sum_{k=1}^N HT_\tau(\sigma_k) v_k \otimes w_k,$$

where $HT_\tau(t) = t\,1_{|t|>\tau}$ is the hard thresholding function such that

$$HT_\tau(\sigma_k) = \begin{cases} \sigma_k & \text{if } \sigma_k > \tau \\ 0 & \text{if } \sigma_k \leq \tau \end{cases}.$$

The error after hard thresholding is

$$\|u - \mathcal{HT}_\tau(u)\|^2 = \sum_{k=1}^N \sigma_k^2\, 1_{\sigma_k \leq \tau}.$$

$\mathcal{HT}_\tau(u)$ is a solution of the problem

$$\min_v \|u - v\|^2 + \tau^2 \operatorname{rank}(v)$$

where $\operatorname{rank}(v) = \|\sigma(v)\|_0$.

Soft thresholding of singular values

The soft singular value thresholding operator $\mathcal{ST}_\tau$ is defined for a tensor u with singular value decomposition $\sum_{k=1}^N \sigma_k v_k \otimes w_k$ by

$$\mathcal{ST}_\tau(u) = \sum_{k=1}^N ST_\tau(\sigma_k) v_k \otimes w_k,$$

where $ST_\tau(t) = (|t| - \tau)_+ \operatorname{sign}(t)$ is the soft thresholding function, such that

$$ST_\tau(\sigma_k) = (\sigma_k - \tau)_+ = \begin{cases} \sigma_k - \tau & \text{if } \sigma_k \geq \tau \\ 0 & \text{if } \sigma_k < \tau \end{cases}.$$

The error after soft thresholding is

$$\|u - \mathcal{ST}_\tau(u)\|^2 = \sum_{k=1}^N (\sigma_k - (\sigma_k - \tau)_+)^2 = \sum_{\sigma_k \leq \tau} {\sigma_k}^2 + \sum_{\sigma_k > \tau} \tau^2.$$

Soft thresholding of singular values

$\mathcal{ST}_\tau(u)$ is a solution of the problem

$$\min_v \frac{1}{2}\|u-v\|^2 + \tau\|\sigma(v)\|_1$$

where $\|\sigma(v)\|_1$ is the nuclear norm of v, which is the convex regularization of the functional $v \mapsto \text{rank}(v)$.

In convex analysis, $\mathcal{ST}_\tau$ is known as the proximal operator of the convex function $v \mapsto \tau\|\sigma(v)\|_1$.

The operator $\mathcal{ST}_\tau$ is non-expansive, that means for all u, v,

$$\|\mathcal{ST}_\tau(u) - \mathcal{ST}_\tau(v)\| \leq \|u-v\|,$$

which is an important property for the analysis of algorithms with tensor truncations.

Hard and soft singular values thresholding for higher order tensors

For a higher order tensor u, we can naturally define hard and soft singular values thresholding operators $\mathcal{HS}_\tau^\alpha$ and $\mathcal{ST}_\tau^\alpha$ associated with the singular value decomposition of the matricisation $\mathcal{M}_\alpha(u)$ of u.

These operators are such that

$$\mathcal{HS}_\tau^\alpha(u) = \arg\min_v \|u - v\|^2 + \tau^2 \operatorname{rank}_\alpha(v),$$

and

$$\mathcal{ST}_\tau^\alpha(u) = \arg\min_v \frac{1}{2}\|u - v\|^2 + \tau\|\sigma^\alpha(u)\|_1.$$

Hard and soft singular values thresholding for higher order tensors

For a tree-based format associated with a dimension partition tree T (or a subset T of a dimension partition tree), hard and soft thresholding operators $\mathcal{HT}_\tau^T$ and $\mathcal{ST}_\tau^T$ can be defined as compositions of hard and soft thresholding operators,

$$\mathcal{HT}_\tau^T = \mathcal{HT}_\tau^{\alpha_M} \circ \ldots \circ \mathcal{HT}_\tau^{\alpha_1}$$

and

$$\mathcal{ST}_\tau^T = \mathcal{ST}_\tau^{\alpha_M} \circ \ldots \circ \mathcal{ST}_\tau^{\alpha_1}$$

where the set of nodes $\{\alpha_1, \ldots, \alpha_M\} = T \setminus D$ is sorted according to

$$\text{level}(\alpha_{k+1}) \geq \text{level}(\alpha_k).$$

The soft-thresholding operator $\mathcal{ST}_\tau^T$ is non-expansive, i.e.

$$\|\mathcal{ST}_\tau^T(u) - \mathcal{ST}_\tau^T(v)\| \leq \|u - v\|$$

for all tensors u, v.