King Abdullah University of Science and Technology KAUST

Lecture 1:

Introduction to low-rank tensor representation/approximation

Alexander Litvinenko

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology
Center for Uncertainty Quantification

http://sri-uq.kaust.edu.sa/

KAUST

Figure : KAUST campus, 5 years old, approx. 7000 people (include 1400 kids), 100 nations.

Stochastic Numerics Group at KAUST

Motivation
- Examples

Post-processing in low-rank data format

Lecture 2: Higher order tensors

Example: low-rank approx. with FFT-techniques

Definitions of different tensor formats

Applications
- Computation of the characteristic

Aims of low-rank/sparse data approximation

1. Represent/approximate multidimensional operators and functions in data sparse format
2. Perform linear algebra algorithms in tensor format (truncation is an issue)
3. Extract needed information from the data-sparse solution (e.g. mean, variance, quantiles etc)

- Book of W. Hackbusch 2012,
- Dissertations of I. Oseledets and M. Espig
- Articles of Tyrtyshnikov et al., De Lathauwer et al., L. Grasedyck, B. Khoromskij, M. Espig
- Lecture courses and presentations of Boris and Venera Khoromskij
- Software T. Kolda et al.; M. Espig et al.; D. Kressner, K. Tobler; I. Oseledets et al.

- modelling of multi-particle interactions in large molecular systems such as proteins, biomolecules,
- modelling of large atomic (metallic) clusters,
- stochastic and parametric equations,
- machine learning, data mining and information technologies,
- multidimensional dynamical systems,
- data compression
- financial mathematics,
- analysis of multi-dimensional data.

$$-\nabla \cdot (\kappa(x,\omega)\nabla u(x,\omega)) = f(x,\omega), \quad x \in \mathcal{G} \subset \mathbb{R}^d$$

where $\omega \in \Omega$, and $\mathcal{U} = L_2(\mathcal{G})$.

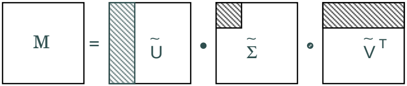
Further decomposition
$L_2(\Omega) = L_2(\times_j \Omega_j) \cong \bigotimes_j L_2(\Omega_j) \cong \bigotimes_j L_2(\mathbb{R}, \Gamma_j)$ results in
$u(x,\omega) \in \mathcal{U} \otimes \mathcal{S} = L_2(\mathcal{G}) \otimes \bigotimes_j L_2(\mathbb{R}, \Gamma_j)$.

Write first Karhunen-Loeve Expansion and then for uncorrelated random variables the Polynomial Chaos Expansion

$$u(x,\omega) = \sum_{i=1}^{K} \sqrt{\lambda_i}\varphi_i(x)\xi_i(\omega) = \sum_{i=1}^{K} \sqrt{\lambda_i}\varphi_i(x) \sum_{\alpha\in\mathcal{J}} \xi_i^{(\alpha)} H_\alpha(\boldsymbol{\theta}(\omega)) \quad (1)$$

where $\xi_i^{(\alpha)}$ is a tensor because $\alpha = (\alpha_1, \alpha_2, ..., \alpha_M, ...)$ is a multi-index.

Let $M := U\Sigma V^T \approx \tilde{U}\tilde{\Sigma}\tilde{V}^T = M_k$.
(Truncated Singular Value Decomposition).
Denote $A := \tilde{U}\tilde{\Sigma}$ and $B := \tilde{V}$, then $M_k = AB^T$.
Storage of A and B^T is $k(n+m)$ in contrast to nm for M.



Let $v \in \mathbb{R}^m$.
Suppose $M_k = AB^T \in \mathbb{R}^{n\times m}$, $A \in \mathbb{R}^{n\times k}$, $B \in \mathbb{R}^{m\times k}$ is given.
Property 1: $M_k v = AB^T v = (A(B^T v))$. Cost $\mathcal{O}(km + kn)$.

Suppose $M' = CD^T$, $C \in \mathbb{R}^{n\times k}$ and $D \in \mathbb{R}^{m\times k}$.

Property 2: $M_k + M' = A_{\text{new}} B_{\text{new}}^T$, $A_{\text{new}} := [A\ C] \in \mathbb{R}^{n\times 2k}$ and $B_{\text{new}} = [B\ D] \in \mathbb{R}^{m\times 2k}$.

Cost of rank truncation from $2k$ to k is $\mathcal{O}((n+m)k^2 + k^3)$.

In both cases below the matrix rank is full.
Gaussian kernel $exp^{-|x-y|^2}$ has the Kronecker rank 1.

The exponential kernel $e^{-|x-y|}$ can be approximated by a tensor with a low Kronecker rank

r	1	2	3	4	5	6	10
$\frac{\|C-C_r\|_\infty}{\|C\|_\infty}$	11.5	1.7	0.4	0.14	0.035	0.007	$2.8e-8$
$\frac{\|C-C_r\|_2}{\|C\|_2}$	6.7	0.52	0.1	0.03	0.008	0.001	$5.3e-9$

Let $\boldsymbol{W} := [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m]$, where $\mathbf{v}_i$ are snapshots.
Given tSVD $\boldsymbol{W}_k = \mathbf{A}\mathbf{B}^T \approx \boldsymbol{W} \in \mathbb{R}^{n\times m}$, $\mathbf{A} := U_k S_k$, $\mathbf{B} := V_k$.

$$\bar{\mathbf{v}} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{v}_i = \frac{1}{m}\sum_{i=1}^{m} \mathbf{A}\cdot\mathbf{b}_i = \mathbf{A}\bar{\mathbf{b}}, \tag{2}$$

$$\mathbf{C} = \frac{1}{m-1}\boldsymbol{W}_c\boldsymbol{W}_c^T = \frac{1}{m-1}\mathbf{A}\mathbf{B}^T\mathbf{B}\mathbf{A}^T = \frac{1}{m-1}\mathbf{A}\mathbf{A}^T. \tag{3}$$

Diagonal of $\mathbf{C}$ can be computed with the complexity $\mathcal{O}(k^2(m+n))$.

Lemma

Let $\|\boldsymbol{W} - \boldsymbol{W}_k\| \leq \varepsilon$ *and* $\overline{\mathbf{v}}_k$ *be a rank-k approximation of the mean* $\overline{\mathbf{v}}$*, then*

a) $\|\overline{\mathbf{v}} - \overline{\mathbf{v}}_k\| \leq \frac{1}{\sqrt{n}}\varepsilon$*,*

b) $\|\mathbf{C} - \mathbf{C}_k\| \leq \frac{1}{m-1}\varepsilon^2$*.*

Lecture 2: Higher order tensors

Definition of tensor of order d

Tensor of order d is a multidimensional array over a d-tuple index set $\mathcal{I} = I_1 \times \cdots \times I_d$,

$$A = [a_{i_1 \ldots i_d} : i_\ell \in I_\ell] \in \mathbb{R}^{\mathcal{I}}, \quad I_\ell = \{1, ..., n_\ell\}, \ell = 1, .., d.$$

A is an element of the linear space

$$\mathbb{V}_n = \bigotimes_{\ell=1}^{d} \mathbb{V}_\ell, \quad \mathbb{V}_\ell = \mathbb{R}^{I_\ell}$$

equipped with the Euclidean scalar product $\langle \cdot, \cdot \rangle : \mathbb{V}_n \times \mathbb{V}_n \to \mathbb{R}$, defined as

$$\langle A, B \rangle := \sum_{(i_1 \ldots i_d) \in \mathcal{I}} a_{i_1 \ldots i_d} b_{i_1 \ldots i_d}, \quad \text{for } A,\ B \in \mathbb{V}_n.$$

Tensor product of vectors $u^{(\ell)} = \{u^{(\ell)}_{i_\ell}\}^{n}_{i_\ell=1} \in \mathbb{R}^{I_\ell}$ forms the canonical rank-1 tensor

$$A_{(1)} \equiv [u_{\mathbf{i}}]_{\mathbf{i}\in\mathcal{I}} = u^{(1)} \otimes \ldots \otimes u^{(d)},$$

with entries $u_{\mathbf{i}} = u^{(1)}_{i_1} \otimes \ldots \otimes u^{(d)}_{i_d}$.

Tensor formats: CP, Tucker, TT

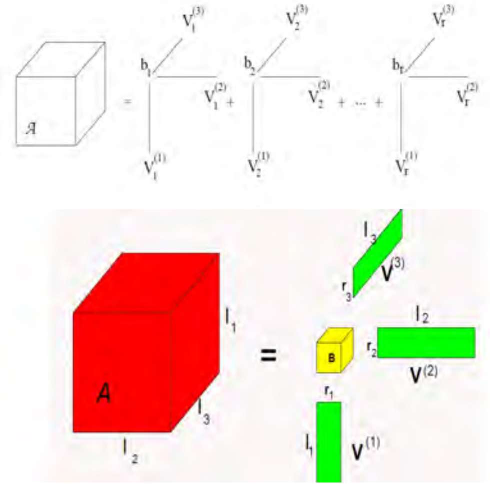
$$A(i_1, i_2, i_3) \approx \sum_{\alpha=1}^{r} u_1(i_1, \alpha) u_2(i_2, \alpha) u_3(i_3, \alpha)$$

$$A(i_1, i_2, i_3) \approx \sum_{\alpha_1, \alpha_2, \alpha_3} c(\alpha_1, \alpha_2, \alpha_3) u_1(i_1, \alpha_1) u_2(i_2, \alpha_2) u_3(i_3, \alpha_3)$$

$$A(i_1, ..., i_d) \approx \sum_{\alpha_1, ..., \alpha_{d-1}} G_1(i_1, \alpha_1) G_2(\alpha_1, i_2, \alpha_2) ... G_{d-1}(\alpha_{d-1}, i_d)$$

Discrete: $G_k(i_k)$ is a $r_{k-1} \times r_k$ matrix, $r_1 = r_d = 1$.

Canonical and Tucker tensor formats



(Taken from Boris Khoromskij)

Tensor $(A_\ell)_{\ell\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$
where $\mathcal{I} = I_1 \times I_2 \times \ldots \times I_d$, $\#I_\mu = n_\mu$, $\#\mathcal{I} = \prod_{\mu=1}^{d} n_\mu$.

Rank-1 tensor

$$A = u_1 \otimes u_2 \otimes \ldots \otimes u_d =: \bigotimes_{\mu=1}^{d} u_\mu$$

$$A_{i_1,\ldots,i_d} = (u_1)_{i_1} \cdot \ldots \cdot (u_d)_{i_d}$$

Rank-k tensor $A = \sum_{i=1}^{k} u_i \otimes v_i$, matrix $A = \sum_{i=1}^{k} u_i v_i^T$.
Kronecker product $A \otimes B \in \mathbb{R}^{nm\times nm}$ is a block matrix whose ij-th block is $[A_{ij}B]$.

Rank-1: $f = \exp(f_1(x_1) + ... + f_d(x_d)) = \prod_{j=1}^{d} \exp(f_j(x_j))$

Rank-2: $f = \sin(\sum_{j=1}^{d} x_j)$, since
$2i \cdot \sin(\sum_{j=1}^{d} x_j) = e^{i \sum_{j=1}^{d} x_j} - e^{-i \sum_{j=1}^{d} x_j}$

Rank-d function $f(x_1, ..., x_d) = x_1 + x_2 + ... + x_d$ can be approximated by rank-2: with any prescribed accuracy:

$$f \approx \frac{\prod_{j=1}^{d}(1 + \varepsilon x_j)}{\varepsilon} - \frac{\prod_{j=1}^{d} 1}{\varepsilon} + \mathcal{O}(\varepsilon), \quad \text{as} \quad \varepsilon \to 0$$

$$\frac{e^{i\kappa\|x-y\|}}{\|x-y\|} \quad \text{Helmholtz kernel} \tag{4}$$

$$f(x) = \int_{[-a,a]^3} \frac{e^{-i\mu\|x-y\|}}{\|x-y\|} u(y)dy \quad \text{Yukawa potential} \tag{5}$$

Classical Green kernels, $x, y \in \mathbb{R}^d$

$$\log(\|x-y\|), \quad \frac{1}{\|x-y\|} \tag{6}$$

Other multivariate functions

$$(a)\ \frac{1}{x_1^2 + .. + x_d^2}, \quad (b)\ \frac{1}{\sqrt{x_1^2 + .. + x_d^2}}, \quad (c)\ \frac{e^{-\lambda\|x\|}}{\|x\|}. \tag{7}$$

For (a) use ($\rho = x_1^2 + ... + x_d^2 \in [1, R]$, $R > 1$)

$$\frac{1}{\rho} = \int_{\mathbb{R}_+} e^{-\rho t} dt. \tag{8}$$

Canonical rank d, TT rank 2.

$$f(x_1,...,x_d) = w_1(x_1) + w_2(x_2) + ... + w_d(x_d)$$

$$= (w_1(x_1), 1) \begin{pmatrix} 1 & 0 \\ w_2(x_2) & 1 \end{pmatrix} ... \begin{pmatrix} 1 & 0 \\ w_{d-1}(x_{d-1}) & 1 \end{pmatrix} \begin{pmatrix} 1 \\ w_d(x_d) \end{pmatrix}$$

Examples:

rank(f)=2

$$f = \sin(x_1 + x_2 + ... + x_d)$$
$$= (\sin x_1, \cos x_1) \begin{pmatrix} \cos x_2 & -\sin x_2 \\ \sin x_2 & \cos x_2 \end{pmatrix} ... \begin{pmatrix} \cos x_{d-1} & -\sin x_{d-1} \\ \sin x_{d-1} & \cos x_{d-1} \end{pmatrix} \begin{pmatrix} \cos x_d \\ \sin x_{d-1} \end{pmatrix}$$

Then the d-dimensional Fourier transformation $\tilde{\mathbf{u}} = \mathcal{F}^{[d]}(\mathbf{u})$ is given by

$$\tilde{\mathbf{u}} = \mathcal{F}^{[d]}(\mathbf{u}) = \left(\bigotimes_{i=1}^{d} \mathcal{F}_i\right) \sum_{j=1}^{k_u} \left(\bigotimes_{i=1}^{d} \mathbf{u}_{ji}\right) = \sum_{j=1}^{k_u} \bigotimes_{i=1}^{d} \left(\mathcal{F}_i\left(\mathbf{u}_{ji}\right)\right) \quad (9)$$

Lemma

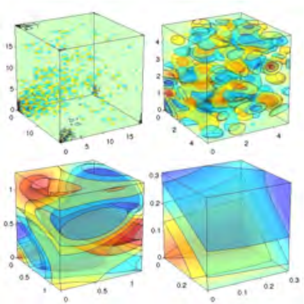
If

$$\mathbf{u} \circ \mathbf{q} = \left(\sum_{j=1}^{k_u} \bigotimes_{i=1}^{d} \mathbf{u}_{ji} \right) \circ \left(\sum_{\ell=1}^{k_q} \bigotimes_{i=1}^{d} \mathbf{q}_{\ell i} \right) = \sum_{j=1}^{k_u} \sum_{\ell=1}^{k_q} \bigotimes_{i=1}^{d} (\mathbf{u}_{ji} \circ \mathbf{q}_{\ell i}) .$$

Then

$$\mathcal{F}^{[d]}(\mathbf{u} \circ \mathbf{q}) = \sum_{\ell=1}^{k_u} \sum_{j=1}^{k_q} \bigotimes_{i=1}^{d} (\mathcal{F}_i (\mathbf{u}_{ji} \circ \mathbf{q}_{\ell i})) \quad .$$

Domain: $20m \times 20m \times 20m$, $25,000 \times 25,000 \times 25,000$ dofs.
4,000 measurements randomly distributed within the volume, with increasing data density towards the lower left back corner of the domain.
The covariance model is anisotropic Gaussian with unit variance and with 32 correlation lengths fitting into the domain in the horizontal directions, and 64 correlation lengths fitting into the vertical direction.



The top left figure shows the entire domain at a sampling rate of 1:64 per direction, and then a series of zooms into the respective lower left back corner with zoom factors (sampling rates) of 4 (1:16), 16 (1:4), 64 (1:1) for the top right, bottom left and bottom right plots, respectively. Color scale: showing the 95% confidence interval $[\mu - 2\sigma, \mu + 2\sigma]$.

MATHEMATICAL DEFINITIONS, PROPERTIES AND APPLICATIONS

Let $\mathcal{T} := \bigotimes_{\mu=1}^{d} \mathbb{R}^{n_\mu}$ be the tensor product constructed from vector spaces $(\mathbb{R}^{n_\mu}, \langle , \rangle_{\mathbb{R}^{n_\mu}})$ $(d \geq 3)$.
Tensor representation U is a multilinear map $U : P \to \mathcal{T}$, where parametric space $P = \times_{\nu=1}^{D} P_\nu$ $(d \leq D)$.
Further, P_ν depends on some representation rank parameter $r_\nu \in \mathbb{N}$.
A standard example of a tensor representation is the canonical tensor format.

(!!!)We distinguish between a tensor $v \in \mathcal{T}$ and its tensor format representation $p \in P$, where $v = U(p)$.

The set $\mathcal{R}_r$ of tensors which can be represented in $\mathcal{T}$ with r-terms is defined as

$$\mathcal{R}_r(\mathcal{T}) := \mathcal{R}_r := \left\{ \sum_{i=1}^{r} \bigotimes_{\mu=1}^{d} v_{i\mu} \in \mathcal{T} : v_{i\mu} \in \mathbb{R}^{n_\mu} \right\}. \qquad (10)$$

Let $v \in \mathcal{T}$. The tensor rank of v in $\mathcal{T}$ is

$$\operatorname{rank}(v) := \min \left\{ r \in \mathbb{N}_0 : v \in \mathcal{R}_r \right\}. \qquad (11)$$

Example: The Laplace operator in 3d:

$$\Delta^3 = \Delta^1 \otimes I \otimes I + I \otimes \Delta^1 \otimes I + I \otimes I \otimes \Delta^1$$

The canonical tensor format is defined by the mapping

$$U_{cp} : \bigtimes_{\mu=1}^{d} \mathbb{R}^{n_\mu \times r} \rightarrow \mathcal{R}_r, \tag{12}$$

$$\hat{v} := (v_{i\mu} : 1 \leq i \leq r,\ 1 \leq \mu \leq d) \mapsto U_{cp}(\hat{v}) := \sum_{i=1}^{r} \bigotimes_{\mu=1}^{d} v_{i\mu}.$$

Lemma

Let $r_1, r_2 \in \mathbb{N}$, $u \in \mathcal{R}_{r_1}$ and $v \in \mathcal{R}_{r_2}$. We have

(i) $\langle u, v\rangle_{\mathcal{T}} = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \prod_{\mu=1}^{d} \langle u_{j_1\mu}, v_{j_2\mu}\rangle_{\mathbb{R}^{n_\mu}}$. *The computational cost of* $\langle u, v\rangle_{\mathcal{T}}$ *is* $\mathcal{O}\left(r_1 r_2 \sum_{\mu=1}^{d} n_\mu\right)$.

(ii) $u + v \in \mathcal{R}_{r_1+r_2}$.

(iii) $u \odot v \in \mathcal{R}_{r_1 r_2}$, *where* $\odot$ *denotes the point wise Hadamard product. Further,* $u \odot v$ *can be computed in the canonical tensor format with* $r_1 r_2 \sum_{\mu=1}^{d} n_\mu$ *arithmetic operations.*

Let $R_1 = A_1 B_1^T$, $R_2 = A_2 B_2^T$ be rank-k matrices, then $R_1 + R_2 = [A_1 A_2][B_1 B_2]^T$ be rank-$2k$ matrix. Rank truncation!

Let $u = \sum_{j=1}^{k} \bigotimes_{i=1}^{d} u_{ji}$, $u_{ji} \in \mathbb{R}^n$.

$$\mathcal{F}^{[d]}(\tilde{u}) = \sum_{j=1}^{k} \bigotimes_{i=1}^{d} \left(\mathcal{F}_i \left(\tilde{u}_{ji} \right) \right), \quad \text{where} \quad \mathcal{F}^{[d]} = \bigotimes_{i=1}^{d} \mathcal{F}_i. \qquad (13)$$

Let $S = AB^T = \sum_{i=0}^{k_1} a_i b_i^T \in \mathbb{R}^{n \times m}$,
$T = CD^T = \sum_{j=0}^{k_2} c_i d_i^T \in \mathbb{R}^{n \times m}$ where $a_i,\, c_i \in \mathbb{R}^n$, $b_i,\, d_i \in \mathbb{R}^m$,
$k_1, k_2, n, m > 0$. Then

$$\mathcal{F}^{(2)}(S \circ T) = \sum_{i=0}^{k_1} \sum_{j=0}^{k_2} \mathcal{F}(a_i \circ c_j) \mathcal{F}(b_i^T \circ d_j^T).$$

Tucker Tensor Format

$$A = \sum_{i_1=1}^{k_1} \dots \sum_{i_d=1}^{k_d} c_{i_1,\dots,i_d} \cdot u_{i_1}^1 \otimes \dots \otimes u_{i_d}^d \qquad (14)$$

Core tensor $c \in \mathbb{R}^{k_1 \times \dots \times k_d}$, rank $(k_1, \dots, k_d)$.
Nonlinear fixed rank approximation problem:

$$X = \text{argmin}_X \quad \min_{\text{rank}(k_1,\dots,k_d)} \|A - X\| \qquad (15)$$

- Problem is well-posed but not solved
- There are many local minima
- HOSVD (Lathauwer et al.) yields rank $(k_1, \dots, k_d)$ $Y \ : \ \|A - Y\| \leq \sqrt{d}\|A - X\|$
- reliable arithmetic, exponential scaling ($c \in \mathbb{R}^{k_1 \times k_2 \times \dots \times k_d}$)

d-Laplacian over uniform tensor grid. It is known to have the Kronecker rank-d representation,

$$\Delta_d = A \otimes I_N \otimes \ldots \otimes I_N + I_N \otimes A \otimes \ldots \otimes I_N + \ldots + I_N \otimes I_N \otimes \ldots \otimes A \in \mathbb{R}^{I^{\otimes d} \otimes I^{\otimes d}} \tag{16}$$

with $A = \Delta_1 = \mathrm{tridiag}\{-1, 2, -1\} \in \mathbb{R}^{N \times N}$, and I_N being the $N \times N$ identity. Notice that for the canonical rank we have rank $k_C(\Delta_d) = d$, while TT-rank of Δ_d is equal to 2 for any dimension due to the explicit representation

$$\Delta_d = (\Delta_1 \ I) \times \begin{pmatrix} I & 0 \\ \Delta_1 & I \end{pmatrix} \times \ldots \times \begin{pmatrix} I & 0 \\ \Delta_1 & I \end{pmatrix} \times \begin{pmatrix} I \\ \Delta_1 \end{pmatrix} \tag{17}$$

where the rank product operation "$\times$" is defined as a regular matrix product of the two corresponding core matrices, their blocks being multiplied by means of tensor product. The similar bound is true for the Tucker rank $\mathrm{rank}_{Tuck}(\Delta_d) = 2$.

Advantages and disadvantages

Denote k - rank, d-dimension, $n = \#$ dofs in 1D:

1. CP: ill-posed approx. alg-m, $\mathcal{O}(dnk)$, hard to compute approx.
2. Tucker: reliable arithmetic based on SVD, $\mathcal{O}(dnk + k^d)$
3. Hierarchical Tucker: based on SVD, storage $\mathcal{O}(dnk + dk^3)$, truncation $\mathcal{O}(dnk^2 + dk^4)$
4. TT: based on SVD, $\mathcal{O}(dnk^2)$ or $\mathcal{O}(dnk^3)$, stable, linear in d, polynomial in r
5. Quantics-TT: $\mathcal{O}(n^d) \rightarrow \mathcal{O}(d\log^q n)$

Postprocessing in the canonical tensor format

Plan: Want to compute *frequency* and *level sets* (needed for the probability density function)

An intermediate step: compute the *sign* function and then the *characteristic* function $\chi_I(u)$

Definition

The *characteristic* $\chi_I(u) \in \mathcal{T}$ of $u \in \mathcal{T}$ in $I \subset \mathbb{R}$ is for every multi-index $\underline{i} \in \mathcal{I}$ pointwise defined as

$$(\chi_I(u))_{\underline{i}} := \begin{cases} 1, & u_{\underline{i}} \in I; \\ 0, & u_{\underline{i}} \notin I. \end{cases} \tag{18}$$

Furthermore, the $\mathrm{sign}(u) \in \mathcal{T}$ is for all $\underline{i} \in \mathcal{I}$ pointwise defined by

$$(\mathrm{sign}(u))_{\underline{i}} := \begin{cases} 1, & u_{\underline{i}} > 0; \\ -1, & u_{\underline{i}} < 0; \\ 0, & u_{\underline{i}} = 0. \end{cases} \tag{19}$$

Lemma: Let $u \in \mathcal{T}$, $a, b \in \mathbb{R}$, and $\mathbb{1} = \bigotimes_{\mu=1}^{d} \underline{1}_\mu$, where $\underline{1}_\mu := (1, \ldots, 1)^t \in \mathbb{R}^{n_\mu}$.

(i) If $I = \mathbb{R}_{<b}$, then we have $\chi_I(u) = \frac{1}{2}(\mathbb{1} + \operatorname{sign}(b\mathbb{1} - u))$.

(ii) If $I = \mathbb{R}_{>a}$, then we have $\chi_I(u) = \frac{1}{2}(\mathbb{1} - \operatorname{sign}(a\mathbb{1} - u))$.

(iii) If $I = (a, b)$, then we have $\chi_I(u) = \frac{1}{2}(\operatorname{sign}(b\mathbb{1} - u) - \operatorname{sign}(a\mathbb{1} - u))$.

Definition

Let $I \subset \mathbb{R}$ and $u \in \mathcal{T}$. The *level set* $\mathcal{L}_I(u) \in \mathcal{T}$ of u respect to I is pointwise defined by

$$(\mathcal{L}_I(u))_{\underline{i}} := \begin{cases} u_{\underline{i}}, u_{\underline{i}} \in I & ; \\ 0, u_{\underline{i}} \notin I & , \end{cases} \tag{20}$$

for all $\underline{i} \in \mathcal{I}$.

The *frequency* $\mathcal{F}_I(u) \in \mathbb{N}$ of u respect to I is defined as

$$\mathcal{F}_I(u) := \# \operatorname{supp} \chi_I(u). \tag{21}$$

Proposition

Let $I \subset \mathbb{R}$, $u \in \mathcal{T}$, and $\chi_I(u)$ its characteristic. We have

$$\mathcal{L}_I(u) = \chi_I(u) \odot u$$

and $\text{rank}(\mathcal{L}_I(u)) \leq \text{rank}(\chi_I(u))\text{rank}(u)$.
The frequency $\mathcal{F}_I(u) \in \mathbb{N}$ of u respect to I is

$$\mathcal{F}_I(u) = \langle \chi_I(u), \mathbb{1} \rangle ,$$

where $\mathbb{1} = \bigotimes_{\mu=1}^{d} \tilde{1}_\mu$, $\tilde{1}_\mu := (1, \ldots, 1)^T \in \mathbb{R}^{n_\mu}$.

Let $u = \sum_{j=1}^{r} \bigotimes_{\mu=1}^{d} u_{j\mu} \in \mathcal{R}_r$, then the mean value $\overline{u}$ can be computed as a scalar product

$$\overline{u} = \left\langle \left(\sum_{j=1}^{r} \bigotimes_{\mu=1}^{d} u_{j\mu} \right), \left(\bigotimes_{\mu=1}^{d} \frac{1}{n_\mu} \tilde{1}_\mu \right) \right\rangle = \sum_{j=1}^{r} \bigotimes_{\mu=1}^{d} \frac{\langle u_{j\mu}, \tilde{1}_\mu \rangle}{n_\mu} = \tag{22}$$

$$= \sum_{j=1}^{r} \prod_{\mu=1}^{d} \frac{1}{n_\mu} \left(\sum_{k=1}^{n_\mu} u_{j\mu} \right), \tag{23}$$

where $\tilde{1}_\mu := (1, \ldots, 1)^T \in \mathbb{R}^{n_\mu}$.
Numerical cost is $\mathcal{O}\left(r \cdot \sum_{\mu=1}^{d} n_\mu \right)$.

Let $u \in \mathcal{R}_r$ and

$$\tilde{u} := u - \overline{u} \bigotimes_{\mu=1}^{d} \frac{1}{n_\mu} \underline{1} = \sum_{j=1}^{r+1} \bigotimes_{\mu=1}^{d} \tilde{u}_{j\mu} \in \mathcal{R}_{r+1}, \tag{24}$$

then the variance $\mathrm{var}(u)$ of u can be computed as follows

$$\begin{aligned} \mathrm{var}(u) &= \frac{\langle \tilde{u}, \tilde{u} \rangle}{\prod_{\mu=1}^{d} n_\mu} = \frac{1}{\prod_{\mu=1}^{d} n_\mu} \left\langle \left(\sum_{i=1}^{r+1} \bigotimes_{\mu=1}^{d} \tilde{u}_{i\mu} \right), \left(\sum_{j=1}^{r+1} \bigotimes_{\nu=1}^{d} \tilde{u}_{j\nu} \right) \right\rangle \\ &= \sum_{i=1}^{r+1} \sum_{j=1}^{r+1} \prod_{\mu=1}^{d} \frac{1}{n_\mu} \langle \tilde{u}_{i\mu}, \tilde{u}_{j\mu} \rangle . \end{aligned}$$

Numerical cost is $\mathcal{O}\left((r+1)^2 \cdot \sum_{\mu=1}^{d} n_\mu \right)$.

Conclusion

Today we discussed:

- Motivation: why do we need low-rank tensors
- Tensors of the second order (matrices)
- Post processing: Computation of mean and variance
- CP, Tucker and tensor train tensor formats
- Many classical kernels have (or can be approximated in) low-rank tensor format
- An example from kriging

Ivan Oseledets et al., Tensor Train toolbox (Matlab),
http://spring.inm.ras.ru/osel

D.Kressner, C. Tobler, Hierarchical Tucker Toolbox (Matlab),
http://www.sam.math.ethz.ch/NLAgroup/htucker_toolbox.html

M. Espig, et al
Tensor Calculus library (C): http://gitorious.org/tensorcalculus