Technical University of Denmark

Spectral Tensor-Train Decomposition for low-rank surrogate models

Bigoni, Daniele; Engsig-Karup, Allan Peter; Marzouk, Youssef M.

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Bigoni, D., Engsig-Karup, A. P., & Marzouk, Y. M. (2014). Spectral Tensor-Train Decomposition for low-rank surrogate models. Poster session presented at Spatial Statistics and Uncertainty Quantification on Supercomputers, Bath, United Kingdom.

Spectral tensor-train decomposition for low-rank surrogate models

Daniele Bigoni*[1], Allan P. Engsig-Karup[1], Youssef M. Marzouk[2]

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark
[2] Department of Aeronautics and Astronautics, Massachusetts Institute of Technology
* **Corresponding author: dabi@dtu.dk**

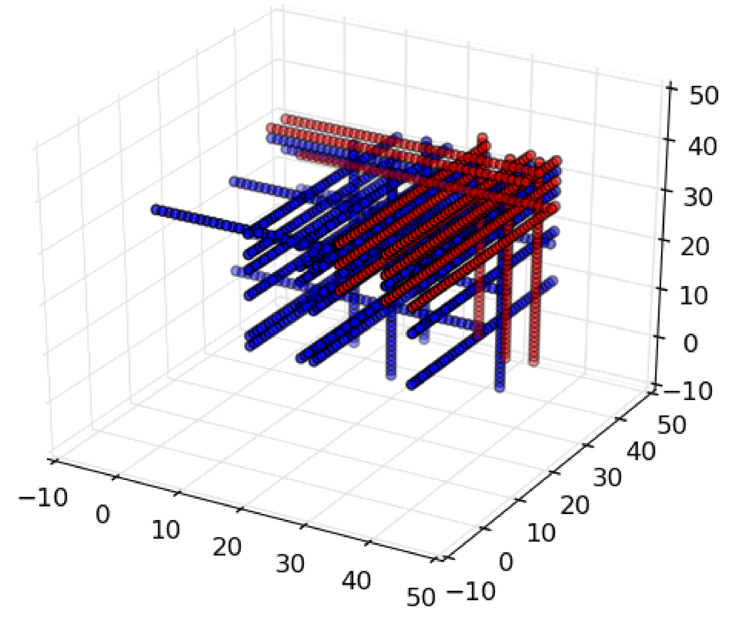


Figure 1: TT-cross

Software: http://www.compute.dtu.dk/∼dabi/
Python PyPi: TensorToolbox

Introduction

The construction of surrogate models is very important as a mean of acceleration in computational methods for uncertainty quantification (UQ). When the forward model is particularly expensive compared to the accuracy loss due to the use of a surrogate – as for example in computational fluid dynamics (CFD) – the latter can be used for the forward propagation of uncertainty [7] and the solution of inference problems [4].

Problem setting

We consider $f \in L^2([a,b]^d)$, where $d \gg 1$ and assume f is a computationally expensive function. Let $\boldsymbol{\xi} \in [a,b]^d$ be random variables entering the formulation of a parametric problem. In the context of UQ, we might want to:

- Compute relevant statistics
- Inquire the sensitivity of f to $\boldsymbol{\xi}$
- Infer the distribution of $\boldsymbol{\xi}$

In most real problems, these goals require an high number of evaluations of f. Often the construction of the surrogate and its evaluation in place of the original f provides a good payoff.

Tensor-train decomposition

Let f be evaluated at all points on a tensor grid $\boldsymbol{\mathcal{X}} = \otimes_{j=1}^d \mathbf{x}_j$, where $\mathbf{x}_j = (x_{i_j})_{i_j=1}^{n_j}$ for $j \in [1,d]$. Let $\boldsymbol{\mathcal{A}} = f(\boldsymbol{\mathcal{X}})$.

Discrete tensor-train approximation [5]

For $\mathbf{r} = (1, r_1, \ldots, r_{d-1}, 1)$, let $\boldsymbol{\mathcal{A}}_{TT}$ be s.t.

$$\boldsymbol{\mathcal{A}}(i_1,\ldots,i_d) = \boldsymbol{\mathcal{A}}_{TT}(i_1,\ldots,i_d) + \boldsymbol{\mathcal{E}}_{TT}(i_1,\ldots,i_d)$$

$$\boldsymbol{\mathcal{A}}_{TT} = \sum_{\alpha_0,\ldots,\alpha_d=1}^{\mathbf{r}} G_1(\alpha_0,i_1,\alpha_1)\ldots G_d(\alpha_{d-1},i_d,\alpha_d)$$

The construction can be built through the evaluation of f on the most important *fibers* (Fig. 1), detected using the TT-cross algorithm [6].
For example, let $f(x,y) = \frac{1}{x+y+1}\sin(4\pi(x+y))$

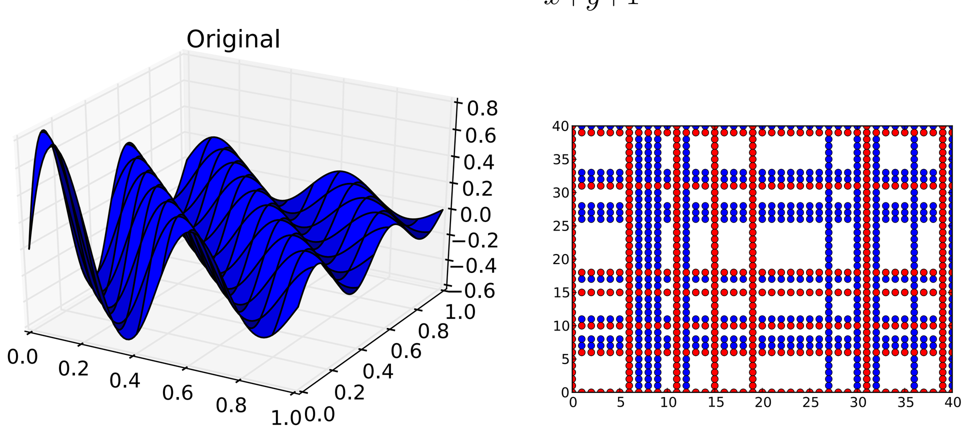


Figure 2: TT-cross: selection of fibers.

- Existence of low-rank best approximation
- Memory complexity: linear in d
- Computational complexity: linear in d

It tackles the **curse of dimensionality**.

Functional TT-decomposition

Using the spectral theory on (non-symmetric) Hilbert-Schmidt kernels, we can construct a functional counterpart of the discrete TT-approximation.

Functional tensor-train approximation [1]

For $\mathbf{r} = (1, r_1, \ldots, r_{d-1}, 1)$, let f_{TT} be s.t.

$$f(\mathbf{x}) = f_{TT}(\mathbf{x}) + R_{TT}(\mathbf{x})$$

$$f_{TT}(\mathbf{x}) = \sum_{\alpha_0,\ldots,\alpha_d=1}^{\mathbf{r}} \gamma_1(\alpha_0,x_1,\alpha_1)\cdots\gamma_d(\alpha_{d-1},x_d,\alpha_d)$$

where $\gamma_i(\alpha_{i-1},\cdot,\alpha_i)$ are orthogonal (see [1]).

f_{TT} is constructed through the eigenvalue decomposition of Hermitian integral operators defined in terms of f. It can be proved that [1]:

- for fixed $\mathbf{r}$, f_{TT} is optimal
- if $\frac{\partial^\beta f}{\partial x_1^{\beta_1}\cdots\partial x_d^{\beta_d}}$ exists and is continuous, then $\gamma_k(\alpha_{k-1},\cdot,\alpha_k) \in \mathcal{C}^{\beta_k}(I_k)$ for all k, α_{k-1} and α_k.

The latter statement can be relaxed:

FTT-decomposition and Sobolev spaces [1]

Let $\mathbf{I} \subset \mathbb{R}^d$ be closed and bounded, and $f \in L^2_\omega(\mathbf{I})$ be a Hölder continuous function with exponent $> 1/2$ such that $f \in \mathcal{H}^k_\omega(\mathbf{I})$. Then f_{TT} is such that $\gamma_j(\alpha_{j-1},\cdot,\alpha_j) \in \mathcal{H}^k_{\omega_j}(I_j)$ for all j, α_{j-1} and α_j.

Spectral TT-decomposition

Let $P_{\mathrm{N}} : L^2_\omega(\mathbf{I}) \to \mathrm{span}\left(\{\Phi_{\mathrm{i}}\}_{\mathrm{i}=0}^{\mathrm{N}}\right)$ where $\{\Phi_{\mathrm{i}}\}_{\mathrm{i}=0}^{\mathrm{N}}$ are orthogonal polynomials:

STT-Projection

$$P_{\mathrm{N}} f_{TT} = \sum_{\mathrm{i}=0}^{\mathrm{N}} \hat{c}_{\mathrm{i}}\Phi_{\mathrm{i}}$$

$$c_{\mathrm{i}} = \sum_{\alpha_0,\ldots,\alpha_d=1}^{\mathbf{r}} \beta_1(\alpha_0,i_1,\alpha_1)\ldots\beta_d(\alpha_{d-1},i_d,\alpha_d)$$

$$\beta_n(\alpha_{n-1},i_n,\alpha_n) = \int_{I_n} \gamma_n(\alpha_{n-1},x_n,\alpha_n)\phi_{i_n}(x_n)dx_n$$

Let $\Pi_N : L^2_\omega(\mathbf{I}) \to \mathrm{span}\left(\{l_i\}_{i=0}^N\right)$, $\{l_i\}_{i=0}^N$ being the Lagrange polynomials:

STT-Interpolation

$$\Pi_{\mathbf{N}} f_{TT} = \sum_{\alpha_0,\ldots,\alpha_d=1}^{\mathbf{r}} \beta_1(\alpha_0,\hat{x}_1,\alpha_1)\cdots\beta_d(\alpha_{d-1},\hat{x}_d,\alpha_d)$$

$$\beta_n(\alpha_{n-1},\hat{x}_n,\alpha_n) = L^{(n)}\gamma_n(\alpha_{n-1},x_n,\alpha_n)$$

where $L^{(n)}$ is the Lagrange interpolation matrix.

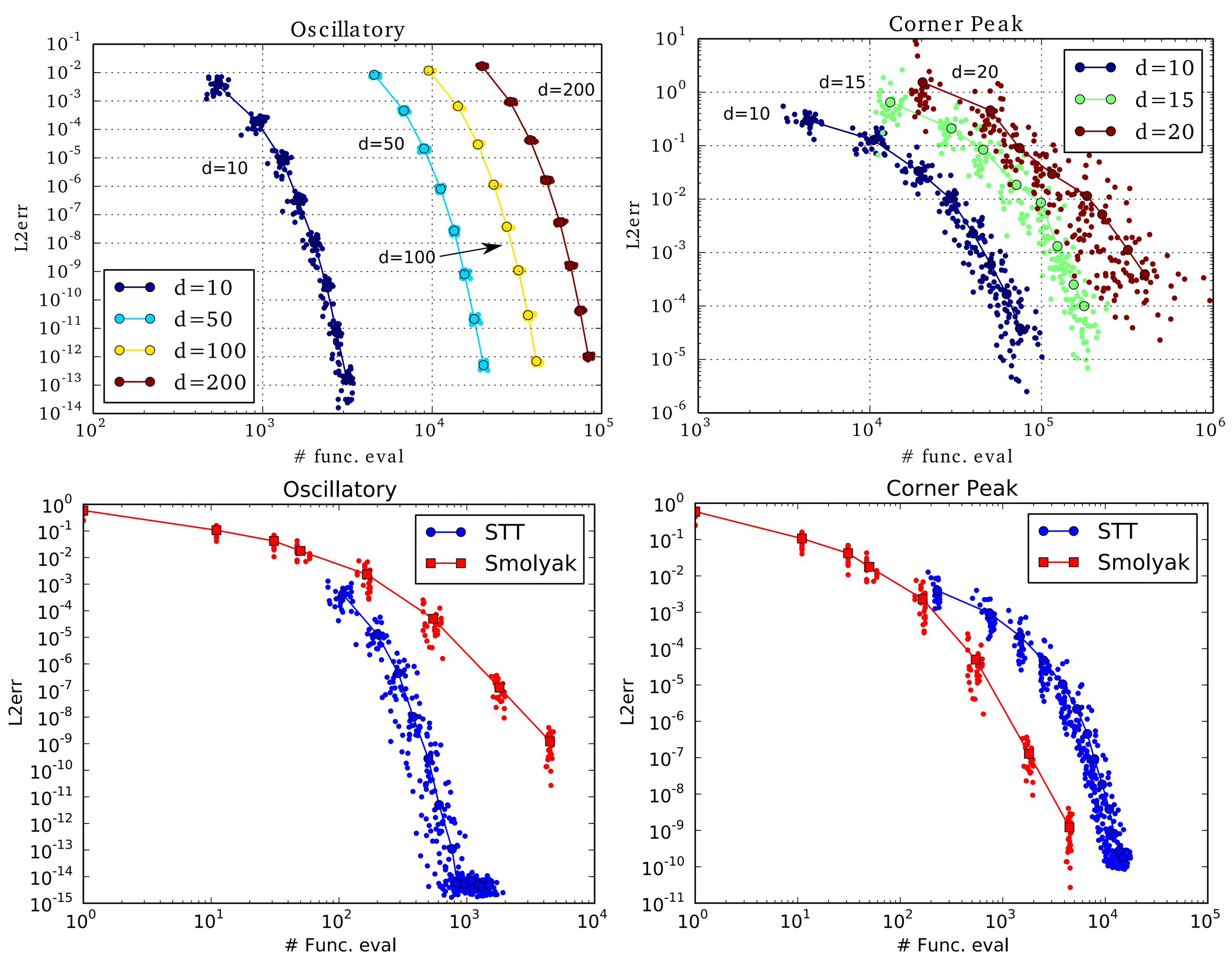
Conclusions

- Tackles the **curse of dimensionality**.
- **Spectral convergence** on smooth functions.

Ongoing works

- Anisotropic heterogeneous adaptivity.
- Ordering problem.
- Application in the fields of coastal engineering [2, 3] and geoscience.

Numerical Examples



Genz functions:

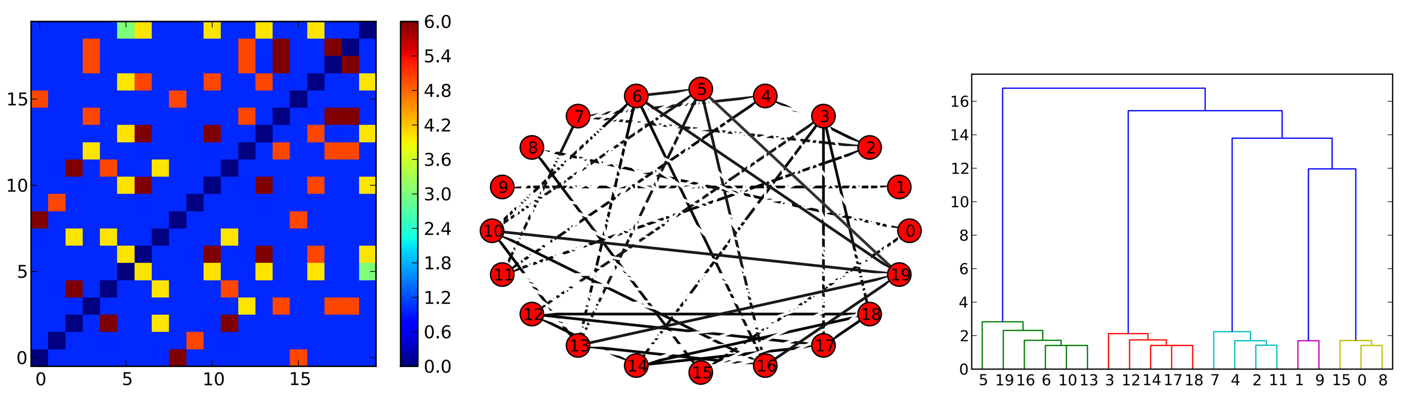
$$f_1(\mathbf{x}) = \cos\left(2\pi w_1 + \sum_{i=1}^d c_i x_i\right)$$

$$f_2(\mathbf{x}) = \left(1 + \sum_{i=1}^d c_i x_i\right)^{-(d+1)}$$

The method shows spectral convergence on both the tests, even on f_2, when there is no analytical low-rank representation.
For $d = 5$, we compare the non-adaptive STT-Projection with the anisotropically adaptive Smolyak Sparse Grid.

Ordering problem

TT and STT are negatively affected by the wrong ordering of the dimensions, leading to an increased computational cost and severe loss of accuracy.
We propose a strategy to find a good ordering.



(a) Vicinity matrix (b) Undirected graph (c) Hierarchical clustering

We construct a vicinity matrix based on the 2nd order ranks of the tensor. We then need to solve the Traveling Salesman Problem.

References

[1] Bigoni, D., Marzouk, Y. M., and Engsig-Karup, A. P. Spectral tensor-train decomposition.

[2] Engsig-Karup, A., Madsen, M. G., and Glimberg, S. L. A massively parallel GPU-accelerated model for analysis of fully nonlinear free surface waves. *International Journal for Numerical Methods in Fluids 70*, 1 (2011), 20–36.

[3] Engsig-Karup, A. P. Analysis of efficient preconditioned defect correction methods for nonlinear water waves. *International Journal for Numerical Methods in Fluids*, January (2014), 749–773.

[4] Marzouk, Y. M., and Najm, H. N. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics 228*, 6 (Apr. 2009), 1862–1902.

[5] Oseledets, I. Tensor-train decomposition. *SIAM Journal on Scientific Computing 33*, 5 (2011), 2295–2317.

[6] Oseledets, I., and Tyrtyshnikov, E. TT-cross approximation for multidimensional arrays. *Linear Algebra and its Applications 432*, 1 (Jan. 2010), 70–88.

[7] Xiu, D., and Karniadakis, G. E. The Wiener-Askey polynomial chaos for stochastic differential equations. Tech. rep., DTIC Document, 2003.