Everything should be made as simple
as possible, but not simpler.
A. Einstein (1879-1955)

Introduction to Tensor Numerical Methods in Scientific Computing

(Bridging MLA with modern high-dimensional applications)

Boris N. Khoromskij

http://personal-homepages.mis.mpg.de/bokh

Summer School on Matrix Methods and Applied Linear Algebra, Rome, September 18-30, 2011

Lecture 1. Separable approximation in higher dimensions.

Lecture 2. From low to higher dimensions. Basic tensor formats: explicit representation and nonlinear approximation.

Lecture 3. Computations in the tensor train (TT) and quantized-TT low-parametric formats.

Lecture 4. Solving equations in tensor formats (BVPs, EVPs, transient problems). Numerical illustrations for some high-dim. applications.

Lecture notes: see Literature.

MATLAB Tensor Toolbox:

– http://csmr.ca.sandia.gov/~tgkolda/Tensor Toolbox/

– http://spring.inm.ras.ru/osel

(Group by E. Tyrtyshnikov: I. Oseledets/D. Savostianov/S. Dolgov/V. Kazeev)

Outlook of Lecture 1.

- Motivations: Moden applications in higher dimensions.
- From low to higher dimensions: what can be adopted from the traditional numerical methods?
- Rank structured separable representations of multi-variate functions in $\mathbb{R}^d$. Basic dimension splitting formats.
- Indispensable rank-structured tensor/matrix multilinear algebra (MLA).
- Kolmogorow's paradigm and "curse of dimensionality".
- $d = 2$: Celebrated Schmidt's decomposition (cf. SVD).
- Greedy Algorithms: simple but slowly convergent.
- Other model reduction approaches.

1929, Dirac:

The fundamental laws necessary for the mathematical treatment of large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.

1998, W. Kohn, A. Pople:

Nobel Prize in Chemistry for development of DFT, based on use of problem adapted (separable) GTO basis sets.

Nowadays: Spreading of tensor methods in multi-dimensional numerical modeling:

- ∘ MLA with linear complexity scaling in dimension d,
- ∘ Effective nonlinear approximation of functions/operators in $\mathbb{R}^d$,
- ∘ Initial applications in quantum chemistry, sPDEs, stochastic models.

Multi-dimensional equations in wide range applications

Basic physical models include (nonlocal) **multivariate transforms**.

Examples of high dimensional problems.

1. Multi-dimensional integral operators in $\mathbb{R}^d$ (Green's functions, convolution, Fourier and Laplace transforms).
2. Elliptic/parabolic/hyperbolic solution operators, preconditioning.
3. Schrödinger eq. for many-particle systems. Density matrix calculation in $\mathbb{R}^3 \times \mathbb{R}^3$ (DFT, Hartree-Fock/Kohn-Sham eqs.), quantum molecular dynamics, DMRG and quantum computing.
4. Stochastic/parametric PDEs, Kolmogorow forward/Fokker-Planck and master eqs.
5. Financial math. (Kolmogorow backward, Black-Scholes eqs).
6. Collision integrals in the deterministic Boltzmann eq. in $\mathbb{R}^3$ (dilute gas).
7. Multi-dimensional data in chemometrics, psychometrics, higher-order statistics, data mining, ...

Examples of the operator calculus

Tensor structured vectors and matrices in $\mathbb{R}^{n^{\otimes d}} \rightleftarrows \mathbb{R}^{n^d}$:

$$x \in \mathbb{R}^{n^d} \rightleftarrows \mathbb{R}^n \otimes ... \otimes \mathbb{R}^n, \quad A \in \mathbb{R}^{m^d \times n^d} \rightleftarrows \mathbb{R}^{m\times n} \otimes ... \otimes \mathbb{R}^{m\times n}.$$

- Linear elliptic systems and spectral problems ($A = A(y)$)

$$Au = f, \quad Au = \lambda u \quad \Rightarrow \quad B \approx A^{-1}.$$

- Volume/interface preconditioning $\Rightarrow \Delta^{-\alpha}$, $\alpha = 1, \pm 1/2$.
- Parabolic equations
  $\frac{\partial u}{\partial t} + Au = f \quad \Rightarrow \quad \exp(-tA),\ (A + \frac{1}{\tau}I)^{-1},\ (\frac{\partial}{\partial t} + A)^{-1}.$
- Control theory: Matrix Lyapunov equation on $\mathbb{R}^{n\times n}$,
  $AX + XB = G \quad \Rightarrow \quad X = \int_0^\infty e^{-tA} G e^{-tB} dt, \quad \mathrm{sign}(A).$
- Convolution, FFT, QTT in $\mathbb{R}^{n^{\otimes d}}$.

1. Motivating applications:

Molecular systems: quantum molecular dynamics, DMRG in quant. chem.
FEM/BEM in $\mathbb{R}^d$: stochastic PDEs, atmospheric model., financial math.
Data mining: quantum computing, machine learning, image processing.

2. "Curse of dimensionality": (R. Bellman, Princeton UP, NJ, 1961).

$O(n^d)$-methods using $N_{vol} = \underbrace{n \times n \times ... \times n}_{d}$ grids (linear in volume size).

3. $O(dn)$-Methods via separation of variables:

Tensor-formatted representation of d-variate functions, operators, and solving equations on rank-structured tensor manifolds in $\mathbb{R}^d$, $d \geq 3$.

4. $\log$-volume super-compressed tensor representation:

Quantized-TT (QTT) approximation of n-d tensors, $n^d \to O(d \log n)$.

In *low dimensions* ($d = 1, 2, 3$) *the goal* is $O(N_{vol})$-methods.

Main principles: making use of *hierarchical* structures, *low-rank* pattern, *recursive* algorithms and *parallelization*.

Based on recursions via hierarchical structures:

Classical Fourier (1768-1830) methods, FFT in $O(N_{vol} \log N_{vol})$ op.
FFT-based circulant convolution, Toeplitz, Hankel matrices.

Multiresolution representation via wavelets, $O(N_{vol})$-FWT.

Multigrid methods: $O(N_{vol})$ - elliptic problem solvers.

Fast multipole, panel clustering, $\mathcal{H}$-matrix: $O(c^d N_{vol} \log N_{vol})$.
Well suited for integral (nonlocal) operators in FEM/BEM.

Parallelization:

Domain decomposition: $O(N_{vol}/p)$ - parallel algorithms.

Traditional numerical tools of reduced complexity

- High order methods: hp-FEM/BEM, spectral methods, bcFEM, Richardson extrapolation.
- Adaptive mesh refinement: a priori/a posteriori strateg.
- Dimension reduction: boundary/interface equations, Schur complement/domain decomposition methods.
- Combination of tensor-product basis with anisotropic adaptivity: hyperbolic cross approximation by FEM/wavelet (sparse grids).
- Model reduction: multi-scale, homogenization, neural networks, proper orthogonal decomposition (POD), etc.
- (Q)Monte-Carlo methods (e.g., for stochastic PDEs).

Separabe representation of functions in TPHS

Let H_ℓ $(\ell = 1, ..., d)$ be a real, separable Hilbert space of functions. **M. Reed, B. Simon, Functional analysis, AP, 1972**.

Def. 1.1. A tensor-product of Hilbert spaces H_ℓ (TPHS), $\mathbb{H} = H_1 \otimes ... \otimes H_d$, is defined as the closure of a set of finite sums, $\sum_k \bigotimes_{\ell=1}^d w_k^{(\ell)}$, of dual multilinear forms (linear functionals) on $H_1 \times ... \times H_d$. A single form is defined by

$$\bigotimes_{\ell=1}^{d} w^{(\ell)} \left(v^{(1)}, ..., v^{(d)}\right) := \prod_{\ell=1}^{d} \langle w^{(\ell)}, v^{(\ell)} \rangle_{H_\ell}.$$

The scalar product of rank-1 (separable) elements (tensors) in $\mathbb{H}$ is defined by

$$\langle w^{(1)} \otimes \ldots \otimes w^{(d)}, v^{(1)} \otimes \ldots \otimes v^{(d)} \rangle = \prod_{\ell=1}^{d} \langle w^{(\ell)}, v^{(\ell)} \rangle,$$

and it is extended by linearity.

$\langle \cdot, \cdot \rangle$ is called the induced scalar product.

Lem. 1.1 $\langle\cdot,\cdot\rangle$ is well defined and it is positive definite.

Lem. 1.2 If $\{\phi^{(\ell)}_{k_\ell}\}$ is an orthonormal basis in H_ℓ, then $\{\Phi_{\mathbf{k}}\} = \{\bigotimes_{\ell=1}^{d} \phi^{(\ell)}_{k_\ell}\}$, $\mathbf{k} = (k_1, ..., k_d) \in \mathbb{N}^d$, is the orthonormal basis in $\mathbb{H}$.

Exer. 1.1. Prove Lem. 1.1 - 1.2.

The tensor product of univariate functions $f^{(\ell)}(x_\ell)$, $x_\ell \in I_\ell = [a_\ell, b_\ell]$, is a d-variate function (called as separable or rank-1) defined as follows

$$f := \bigotimes_{\ell=1}^{d} f^{(\ell)}, \quad \text{where} \quad f(x_1, ..., x_d) = \prod_{\ell=1}^{d} f^{(\ell)}(x_\ell).$$

Exer. 1.2. Prove $L^2(I_1 \times ... \times I_d) = \bigotimes_{\ell=1}^{d} L^2(I_\ell)$.

Ex. 1.2. Denote by $H^{\otimes d}$ the d-fold tensor product of spaces H. If $H = L^2(\mathbb{R})$, then an element $\psi \in \mathcal{F}(H) := \oplus_{d=0}^{\infty} H^{\otimes d}$, of the so-called Fock space over H, $\mathcal{F}(H)$, is a sequence of functions

$$\psi = \{\psi_0, \psi_1(x_1), \psi_2(x_1, x_2), \psi_3(x_1, x_2, x_3), \ldots\},$$

such that

$$|\psi_0|^2 + \sum_{d=1}^{\infty} \int_{\mathbb{R}^d} |\psi_n(x_1, \ldots, x_d)|^2 dx_1 \ldots dx_d < \infty.$$

The finite expansion in $\mathcal{F}(H)$ as above is known as ANOVA representation.

In the physical literature, the subspaces of $\mathcal{F}(H)$ consisting of symmetric/antisymmetric functions w.r.t. permutation of two arguments are called the boson and fermion Fock spaces, respectively.

Def. 1.2 d-th order tensor is a function of d discrete arguments, $f : I_1 \times ... \times I_d \to \mathbb{R}$, (multi-dimensional array over $I_1 \times ... \times I_d$). The respective TPHS $\mathbb{H}$ is equipped with Euclidean scalar product and Frobenius norm (More details in Lect. 2).

Ex. 1.3. $\mathbb{H} = \mathbb{R}^{I_1 \times \ldots \times I_d} = \bigotimes_{\ell=1}^{d} \mathbb{R}^{I_\ell}$, with $I_\ell = \{1, ..., n_\ell\}$ is the space of real valued tensors of order d.

Def. 1.3. (Canonical format). Call by $\mathcal{C}_R$ a subset of elements in $\mathbb{H}$, requiring at most R terms (rank-R functions),

$$\mathcal{C}_R = \left\{ w \in \mathbb{H} : \; w = \sum_{k=1}^{R} w_k^{(1)} \otimes w_k^{(2)} \otimes \ldots \otimes w_k^{(d)}, \; w_k^{(\ell)} \in H_\ell \right\}.$$

$w \in \mathcal{C}_R$ can be represented by the description of Rd elements $w_k^{(\ell)} \in H_\ell$. Storage on n^d-grid: dRn (linear in d).

Advantage: Tremendous reduction of representation cost, removing d from the exponential, $n^d \to d\,R\,n$.

Limitations: Applies to special class of functions given analytically, nonrobust algebraic decomposition.

Probl. 1. Best rank-R approximation of a multi-variate function $f = f(x_1, ..., x_d) \in \mathbb{H}$ in the set $\mathcal{C}_R$.

Given a tuple of dimensions, $\mathbf{r} = (r_1, \ldots, r_d) \in \mathbb{N}^d$, choose $V_\ell = span\left\{\phi_k^{(\ell)}\right\}_{k=1}^{r_\ell} \subset H_\ell$, $r_\ell := \dim V_\ell < \infty$ $(1 \leq \ell \leq d)$ with orthogonal basis and build the tensor subspace,
$\mathbb{V} = V_1 \otimes V_2 \otimes \ldots \otimes V_d \subset \mathbb{H}$. Each $v \in \mathbb{V}$ can be represented by

$$v = \sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{r}} b_{\mathbf{k}} \phi_{k_1}^{(1)} \otimes \phi_{k_2}^{(2)} \otimes \ldots \otimes \phi_{k_d}^{(d)}, \quad b_{\mathbf{k}} \in \mathbb{R}. \tag{1}$$

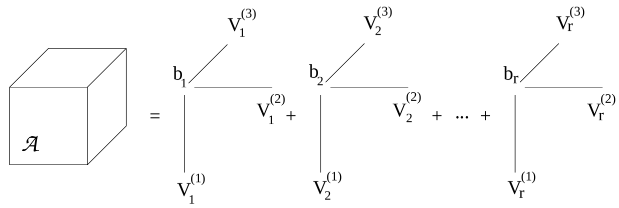
Def 1.4. (Tucker format) Given $\mathbf{r}$, define

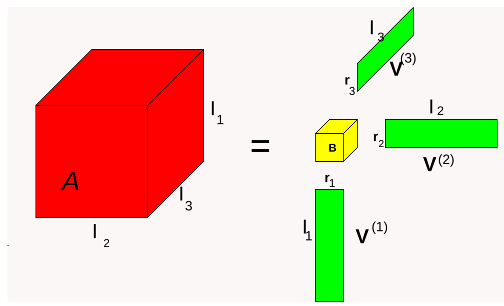
$$\mathcal{T}_{\mathbf{r}} := \{v \in \mathbb{V} \subset \mathbb{H} : \quad \forall\, V_\ell \text{ s.t. } \dim V_\ell = r_\ell \text{ with } b_{\mathbf{k}} \in \mathbb{R}\}.$$

Representing $w \in \mathcal{T}_{\mathbf{r}}$: $\prod_{\ell=1}^{d} r_\ell$ reals and the sampling of $\sum_{\ell=1}^{d} r_\ell$ functions $\phi_k^{(\ell)}$.
Robust but storage on n^d-grid: $r^d + rdn \ll n^d$, $r = \max r_\ell$.

Orthogonal separabe representation

Visualization of the canonical and Tucker models for $d=3$.





Probl. 2. Best rank-$\mathbf{r}$ orthogonal approx. of $f \in \mathbb{H}$ in $\mathcal{T}_{\mathbf{r}}$.

Examples on rank-R and Tucker formats

Ex. 1.4. $\mathbb{H} = L^2(I^d)$. Rank-1 elements, $f = f_1(x_1)...f_d(x_d)$, e.g. $f = \exp(g_1(x_1) + ... + g_d(x_d)) = \prod_{\ell=1}^d \exp(g_\ell(x_\ell))$. For the function $f = \sin\left(\sum_{j=1}^d x_j\right)$, $rank(f) = 2$ holds over the field $\mathbb{C}$,

$$2i \sin\left(\sum_{j=1}^d x_j\right) = e^{i\sum_{j=1}^d x_j} - e^{-i\sum_{j=1}^d x_j}.$$

Rank-d function $f(x) = x_1 + x_2 + \ldots + x_d$, can be approximated by a rank-2 expansion with any prescribed accuracy,

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

Ex. 1.5. The Tucker approximation in $\mathbb{H} = L^2(I^d)$ can be made by the tensor product polynomial interp. of order $\mathbf{r}$,

$$f(x_1, ..., x_d) \approx \sum_{\mathbf{j}=1}^{\mathbf{r}} f(\nu_{j_1}, ..., \nu_{j_d}) \prod_{\ell=1}^{d} L_{j_\ell}(x_\ell).$$

L_{j_ℓ} is a set of the Lagrange polynomials on $[-1,1]$ at, say, Chebyshev-Gauss-Lobatto grid, $\nu_{j_\ell} = \cos\frac{\pi j_\ell}{N}$, $j_\ell = 0, ..., r_\ell$.

Function product decomposition (Tensor train/chain)

Given $\mathcal{J} := \times_{\ell=1}^{d} J_\ell$, $J_\ell = \{1, ..., r_\ell\}$, and $J_0 = J_d$.

Def. 1.5 The rank-$\mathbf{r}$ functional tensor train/chain (FTT/FTC) format: product of functional tri-tensors over $\mathcal{J}$,

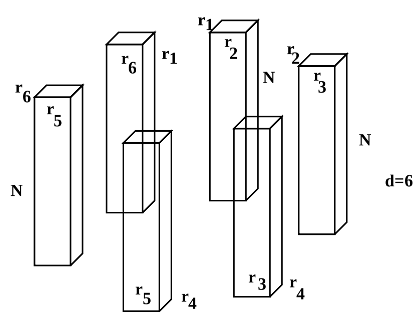
$$f(x_1, ..., x_d) = \sum_{\boldsymbol{\alpha} \in \mathcal{J}} f_1(\alpha_d, x_1, \alpha_1) f_2(\alpha_1, x_2, \alpha_2) \cdots f_d(\alpha_{d-1}, x_d, \alpha_d)$$
$$\equiv F_1(x_1) F_2(x_2) ... F_d(x_d),$$

If $J_0 = \{1\}$, we have the FTT decomp. Here $F_1(x_1)$ is a row $1 \times r_1$-vector function depending on x_1, $F_\ell(x_\ell)$ is a matrix of size $r_{\ell-1} \times r_\ell$ with functional elements depending on x_ℓ, $F_d(x_d)$ is a column vector of size $r_{d-1} \times 1$, depending on x_d.
Sampling on a $n^{\otimes d}$-grid: $O(dr^2 n)$-storage.

$f \in \mathbb{H}$ is represented by a product of matrices (matrix product functions), each depending on a single mode variable.

Examples on functional TT decomposition

Ex. 1.6. d-fold product of functional tensors over $J_1, ..., J_d$ $(d = 6)$.



Special case $r_6 = 1$: FTT[$\mathbf{r}$] = FTC[$\mathbf{r}$].

Exer. 1.3. In some cases the function product decomp. allows an explicit form, e.g. FTT-rank of $f(x) = x_1 + x_2 + \ldots + x_d$ is 2.

$$f(x) = \begin{pmatrix} x_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ x_2 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ x_{d-1} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ x_d \end{pmatrix}.$$

Since $\mathcal{T}_{\mathbf{r}}$, $\mathcal{C}_R$ and FTT$[\mathbf{r}]$ are not linear spaces, we obtain a nontrivial nonlinear approximation problem on estimation

$$f \in \mathbb{H}: \quad \sigma(f, \mathcal{S}) := \inf_{s \in \mathcal{S}} \|f - s\|, \tag{2}$$

where $\mathcal{S} = \{\mathcal{T}_{\mathbf{r}}, \mathcal{C}_R, FTT[\mathbf{r}]\}$.

Why the problem (2) might be difficult for $d \geq 3$?

Prop. 1.1 [Beylkin, Mohlenkamp] The trigonometric identity ($d \geq 2$)

$$f(x) := \sin(\sum_{j=1}^{d} x_j) = \sum_{j=1}^{d} \sin(x_j) \prod_{k \in \{1,\ldots,d\} \setminus \{j\}} \frac{\sin(x_k + \alpha_k - \alpha_j)}{\sin(\alpha_k - \alpha_j)} \tag{3}$$

holds for any $\alpha_k \in \mathbb{R}$, s.t. $\sin(\alpha_k - \alpha_j) \neq 0$ for all $j \neq k$.

For $d \geq 3$ it can be proven by induction (nontrivial exercise!).

Exer. 1.4. Prove that FTT-rank of f in (3) is 2. (Lem. 1.3)

Lem. 1.3 The function $f(x) := \sin(\sum_{j=1}^{d} x_j)$, $x \in \mathbb{R}^d$, admits the explicit rank-2 FTT decomposition

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

Proof. Induction, similar to Exer. 1.3,

$$\begin{aligned} f(x) &= \sin x_1 \cos(x_2 + \ldots + x_d) + \cos x_1 \sin(x_2 + \ldots + x_d) \\ &= \begin{pmatrix} \sin x_1 & \cos x_1 \end{pmatrix} \begin{pmatrix} \cos(x_2 + \ldots + x_d) \\ \sin(x_2 + \ldots + x_d) \end{pmatrix} \\ &= \begin{pmatrix} \sin x_1 & \cos x_1 \end{pmatrix} \begin{pmatrix} \cos x_2 & -\sin x_2 \\ \sin x_2 & \cos x_2 \end{pmatrix} \begin{pmatrix} \cos(x_3 + \ldots + x_d) \\ \sin(x_3 + \ldots + x_d) \end{pmatrix}. \end{aligned}$$

Cor. 1.1 For any $d \geq 3$, $\varepsilon > 0$, we have for the Helmholtz kernels, [7],

$$rank_{FTT,\varepsilon}(f_{1,\kappa}(\|x\|)) \leq C(|\log \varepsilon| + \kappa), \quad f_{1,\kappa}(\|x\|) := sin(\kappa\|x\|)/\|x\|,$$

$$rank_{FTT,\varepsilon}(f_{2,\kappa}(\|x\|)) \leq Crank_C(\|x\|)(|\log \varepsilon| + \kappa), \quad f_{2,\kappa}(\|x\|) := \frac{2sin^2(\frac{\kappa}{2}\|x\|)}{\|x\|}.$$

Expansion (3) shows the lack of uniqueness (ambiguity) of the best rank-d separable representation. The minimisation process in (2) might be non-robust (multiple local minima).

Principal questions (no ultimate answers):

- **Is the "curse of dimensionality" unremovable?**
- **How to solve (2) efficiently?** (Extend truncated SVD)
- **Can one expect the fast (exponential) convergence in the rank parameters R, $r = \max r_\ell$?**
- **Can one solve the physical equations on nonlinear tensor manifold $\mathcal{S}$ getting rid of "curse of dimension"?**

Our approach: Tensor-structured numerical methods based on the efficient *multilinear* algebra (MLA).

Hilbert 13th problem: A solution of the algebraic equation of degree 7 cannot be written as superposition of continuous bivariate functions.

Solved via celebrated theorem by Kolmogorow on the superposition of univariate functions.

Thm. 1. (A. Kolmogorow, 1957) Let $I = [0,1]$. For $d \geq 2$, any function $f \in \mathbb{C}[I^d]$ can be represented in the form

$$f(x_1, ..., x_d) = \sum_{i=1}^{2d+1} g_i \left(\sum_{\ell=1}^{d} \phi_{i\ell}(x_\ell) \right),$$

where functions $\phi_{i\ell} : I \to \mathbb{R}$ do not depend on f and belong to the class $Lip1$, while $g_i : \mathbb{R} \to \mathbb{R}$ are continuous functions.

Thm. 1. is not constructive, but in our context it says that in the discrete setting, any function f can be represented by $O(2dN + (2d+1)dN)$ reals, N corresponds to the size of the interpolating table for g_i **[Griebel '08]**.

The approximation of functions $f(x,y)$ by bilinear forms

$$f \approx \sum_{k=1}^{R} u_k(x) v_k(y) \quad \text{in} \quad L^2([0,1]^2),$$

is due to E. Schmidt, 1907 (celebrated theorem). The result is a continuous analogue to SVD of matrices.

Let $\{\sigma_k(J_f)\}$, $\sigma_1 \geq \sigma_2 \geq ... \geq 0$, be a nonincreasing sequence of singular values of the IO,

$$J_f(g) := \int_0^1 f(x,y) g(y) dy,$$

$$\sigma_k(J_f) := \lambda_k[(A)^{1/2}], \quad A = J_f^* J_f, \quad J_f^* \text{ adjoint to } J_f$$

with orthonormal sequences $\{\varphi_k(x)\}$, $\{\psi_k(y)\}$,

$$A\psi_k(y) = \lambda_k \psi_k(y); \quad A^*\varphi_k(x) = \lambda_k \varphi_k(x), \quad k = 1, 2, ...$$

The kernel function of A is given by

$$f_A(x,y) := \int_0^1 f(x,z) f(z,y) dz.$$

The Schmidt decomposition (SD) is given by

$$f(x,y) = \sum_{k=1}^{\infty} \sigma_k(J_f) \varphi_k(x) \psi_k(y).$$

The best bilinear approximation property reads as,

$$\left\| f(x,y) - \sum_{k=1}^{R} \sigma_k \varphi_k(x) \psi_k(y) \right\|_{L^2} = \inf_{u_k, v_k \in L^2,\ k=1,\ldots,R} \left\| f(x,y) - \sum_{k=1}^{R} u_k(x) v_k(y) \right\|_{L^2}.$$

SD ensures that for $d=2$ the best bilinear approximation can be realised by the so-called Pure Greedy Algorithm (PGA).

For Nyström's approximation the problem is reduced to SVD.

Computing canonical decomposition: simple method

For $\mathcal{S} = \mathcal{C}_R$, the canonical decomposition can be considered in the framework of the best R-term approximation with regard to a redundant dictionary of rank-1 functions.

Def. 1.6 A system $\mathcal{D}$ of functions from $\mathbb{H}$ is called a dictionary if each $g \in \mathcal{D}$ has norm one and its linear span is dense in $\mathbb{H}$.

Denote by $\Sigma_R(\mathcal{D})$ the collection of $s \in \mathbb{H}$ which can be written in the form (cardinality bounded by R)

$$s = \sum_{g \in \Lambda} c_g g, \quad \Lambda \subset \mathcal{D}: \; \#\Lambda \leq R \in \mathbb{N} \quad \text{with } c_g \in \mathbb{R}.$$

For $f \in \mathbb{H}$, the best R-term approximation error is defined by

$$\sigma_R(f, \mathcal{D}) := \inf_{s \in \Sigma_R(\mathcal{D})} \|f - s\|.$$

Pure Greedy Algorithm

The *Pure Greedy Algorithm* (PGA) inductively computes an estimate to the best R-term approximation.

Let $g = g(f) \in \mathcal{D}$ ($\|g\| = 1$), maximise $|\langle f, g\rangle|$ (best rank-1 nonlinear approximation !). Define

$$G(f) := \langle f, g\rangle g, \quad R(f) := f - G(f).$$

The PGA reads as: Given $f \in \mathbb{H}$, introduce

$$R_0(f) := f \quad \text{and} \quad G_0(f) := 0.$$

Then, for all $1 \leq m \leq R$, we inductively define

$$G_m(f) := G_{m-1}(f) + G(R_{m-1}(f)),$$

$$R_m(f) := f - G_m(f) = R(R_{m-1}(f)).$$

Pure Greedy Algorithm

Applying PGA to functions characterised via the approximation property (low order approximation)

$$\sigma_R(f, \mathcal{D}) \le R^{-q}, \quad R = 1, 2, ...,$$

with some $q \in (0, 1/2]$, leads to the error bound (Temlyakov)

$$\|f - G_R(f, \mathcal{D})\| \le C(q, \mathcal{D}) R^{-q}, \quad R = 1, 2, ...,$$

which is "too pessimistic" for applications (cf. Monte-Carlo).

Our goal: The constructive R-term approximation on a class of analytic functions (possibly with point singularities), providing exponential convergence in $R = 1, 2, ...,$

$$\sigma_R(f, \mathcal{D}) \le C \exp(-R^q), \quad q = 1 \text{ or } q = 1/2.$$

Methods of choice: Quadrature/interpolation-based sinc approximation by exponential sums with rank recompression.

Greedy completely orthogonal decomposition

The decomposition in $\mathcal{C}_R$,

$$f = \sum_{k=1}^{R} a_k v_k, \quad v_k = \phi_k^{(1)}(x_1) \otimes ... \otimes \phi_k^{(d)}(x_d) \in \mathcal{C}_1,$$

is called *completely orthogonal* if

$$\langle \phi_k^{(\ell)}, \phi_m^{(\ell)} \rangle = \delta_{k,m} \quad \forall \ell = 1, ..., d, \Leftrightarrow \Phi^{(\ell)} = [\phi_1^{(\ell)}, ..., \phi_R^{(\ell)}] - orthog.$$

Greedy completely orthogonal decomposition (GCOD) is defined as PGA with the orthogonality constraint on $\Phi^{(\ell)}$.

Lem. 1.4 (Tucker format with the diagonal core.) Let $f \in \mathbb{H}$ allow a rank-R *completely orthogonal decomposition*. Then the GCOD algorithm correctly computes it. If $a_1 > a_2 > ... > a_R > 0$, then the decomp. is unique.

Exer. 1.5 Prove Lem. 1.3. **[Golub, Zhang 2001].**

Limitations. Poor approximation properties of the COD !

Literature to Lecture 1

1. G. Beylkin, M. Mohlenkamp: *Numerical operator calculus in higher dimensions*, Proc. Natl. Acad. Sci. USA, **99** (2002), 10246–10251.

2. M. Reed, and B. Simon: *Methods of Modern Mathematical Physics, I. Functional Analysis*, AP, NY, 1972.

3. B.N. Khoromskij: *An introduction to Structured Tensor-product representation of Discrete Nonlocal Operators.* Lecture notes 27, MPI MIS, Leipzig 2005.

4. I.V. Oseledets: *Constructive representation of functions in tensor formats.* Preprint INM Moscow, 2010.

5. V.N. Temlyakov: *Greedy Algorithms and M-Term Approximation with Regard to Redundant Dictionaries.* J. of Approx. Theory **98** (1999), 117-145.

6. T. Zhang, and G.H. Golub: *Rank-one approximation to high order tensors.* SIAM J. Matrix Anal. Appl. **23** (2001), 534-550.

7. B.N. Khoromskij: *Introduction to Tensor Numerical Methods in Scientific Computing.* Lecture Notes, University/ETH Zuerich, Preprint 06-2011, Uni. Zuerich 2011, pp. 1-238. http://www.math.uzh.ch/fileadmin/math/preprints/06_11.pdf

http://personal-homepages.mis.mpg.de/bokh