15 Lectures on Tensor Numerical Methods for Multi-dimensional PDEs

Lect. 1-2. Approximation of functions by separation of variables. Matrix Case

Boris Khoromskij & Venera Khoromskaia,
Shanghai, Jiao Tong University, April 2017

Max-Planck-Institute for Mathematics in the Sciences, Leipzig

Outline of the Lecture Course

Lectures 1 - 3. Approximation of bivariate functions by separation of variables. Matrix case. Matrix SVD, EVD, Cholesky decompositions. FFT transform. Kronecker product. Matrix factorizations, examples and exercises in Matlab (SVD, Cholesky, QR, Kron, inv).

Lectures 4 - 6. Basic rank-structured tensor formats: canonical (CP), Tucker decompositions. Approximation of multivariate functions by the canonical and Tucker tensors. Algorithms for Tucker decomposition with Matlab examples. Canonical-to-Tucker algorithm.

Lectures 7 - 9. Polynomial and Sinc approximation of multivariate functions. Separable approximation in analytic form via the Laplace transform. Matrix product states (MPS, TT) decomposition. Quantized tensor approximation in log-complexity (QCan, QTT). Super-fast integration. Exercises in Matlab on *sinc*- and QTT approximation of discretized functions.

Lectures 10 - 12. Operators in many dimensions. Superfast Laplacian inverse, QTT-FFT and convolution. Multi-parametric (stochastic) PDEs. Preconditioning in many dimensions. Calculation of 3D integral operators in 1D complexity. Fast Hartree-Fock solver. Excitation energies by the Bethe-Salpeter equation.

Lectures 13 - 15. Time-dependent (parabolic) problems in $\mathbf{d}+\mathbf{1}$ formulation. Example of chemical master equation. Equations with oscillating coefficients. Asymptotic homogenization. The Hartree-Fock on lattice structured compounds. Super-fast summation of electrostatic potentials on finite lattices with defects. Range separated tensor formats and application to general many-particle systems.
Concluding discussion.

Outlook of Lectures 1 - 2.

- Motivations: Modern applications in higher dimensions. Principle of separation of variables in $\mathbb{R}^d$: low-parametric $O(d)$-approximation vs. "curse of dimensionality".
- Separability concept in computational quantum chemistry.
- Kolmogorow's paradigm.
- $d = 2$: Celebrated Schmidt's decomposition as a prototype of matrix SVD.
- Low-rank matrices, matrix SVD, QR, reduced and truncated SVD.
- Low-rank approximation via truncated Cholesky algorithm for spd matrices.
- Kronecker product of matrices: basic properties. Matrix prototype of a tensor product.
- From low to higher dimensions: what can be adopted from the traditional numerics?
- Based on hierarchical structures: $\mathcal{H}$ matrices, FFT, circulant convolution, FWT.
- Tensor methods win against supercomputing !

But tensor methods with supercomputing can be a great power ...

Modern high-dimensional problems:

- d-dim. operators: Green's functions, operator-valued functions, Fourier/convolution/wavelet transforms.
- Quantum chemistry (QC): electronic structure, quantum molecular dynamics, spin systems.
- PDEs in $\mathbb{R}^d$: stochastic/parametric PDEs, many-body systems, many-particles dynamics (Fokker-Planck, master eqn., etc.), homogenization.
- Data compression and classification: NMR, machine learning, data mining, image processing, principal component analysis (PCA).

Computational bottleneck – the "curse of dimensionality"

1961, R. Bellman: *In view of all that we have said ..., the many obstacles we appear to have surmounted, what casts the pall over our victory celebration? It is the curse of dimensionality, a malediction that has plagued the scientist from earliest days.*

Let $f(x_1, \ldots, x_d)$ be discretized in $[0,1]^d$ by sampling over $N \times N \times \ldots \times N$ grid, then we need N^d storage size.

Traditional $O(N^d)$-methods on $\underbrace{N \times N \times ... \times N}_{d \text{ times}}$ grids: exponential scaling in d.

Remedy: Principle of separation of variables

Example in QC: Schrödinger equation for N_e-electron system on 3D-grids: $d = 3N_e$, $N_e \sim 10^4$, $N \sim 10^3$. Number of atoms in Universe: 10^{60}.

Basic principle of separation of variables.

Let $f(x_1, \ldots, x_d) = f_1(x_1) \cdots f_d(x_d)$ be separable, then discretization in $[0,1]^d$ by sampling over $N \times N \times \ldots \times N$ grid reduces the storage size to $d\,N$.

Example: The Gaussian in $\mathbb{R}^d$ is most commonly used in physics and chemistry separable function

$$g(x) = e^{-\lambda \|x\|^2} = e^{-\lambda x_1^2} e^{-\lambda x_2^2} \cdots e^{-\lambda x_d^2}.$$

Goal: $N^d \to O(dN)$ complexity by tensor numerical methods for

- representation of d-variate functions (tensors), operators (matrices),
- solving equations in $\mathbb{R}^d$ on rank-structured "low-parametric manifolds".

Super-compressed representation $N^d \to O(d \log N)$ (log-volume scaling):

- Quantized tensor approximation (QTT) of functional vectors/matrices.

Separability concept in computational quantum chemistry and beyond

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.

Recent decade: Spreading of tensor methods in multi-dimensional numerical modeling:

◦ Multilinear algebra (MLA) with linear complexity scaling in dimension d.

◦ Effective nonlinear tensor approximation of functions and operators in $\mathbb{R}^d$.

◦ Slightly entangled systems: MPS/DMRG in quantum chemistry, spin systems.

◦ New approach (since 2006): grid-based tensor numerical methods in electronic structure calculations (Hartree-Fock, MP2, Bether-Salpeter eqn., density of states).

◦ Multi-dimensional stochastic dynamical systems, molecular dynamics (since 1993).

◦ Stochastic PDEs, uncertainty quantification (UQ), parametric systems.

A piece of history on dimensionality reduction: Kolmogorow's paradigm

▶ 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. K. (A. Kolmogorow, 1957) For $d \geq 2$, any function $f \in \mathbb{C}([0,1]^d)$ 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 $\phi_{i\ell} : [0,1] \to \mathbb{R}$ do not depend on f and belong to the class $Lip1$, while $g_i : \mathbb{R} \to \mathbb{R}$ are continuous functions.

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

▶ The opposite direction: quantized tensor method (QTT) approximates univariate functions by mapping to higher dimensions !

A piece of history on SVD: $d = 2$, Schmidt decomposition

▶ 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 non-increasing sequence of singular values of the integral operator,

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

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

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

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

$d = 2$: Best R-term approximation and SVD

- The kernel function of A (iterated kernel) is defined by

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

- The Schmidt decomposition (SD) is given by

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

- The best R-term 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 R-term bilinear approximation can be realized by the so-called Pure Greedy Algorithm (PGA).

- Best rank-R degenerate approximation of J_f: $J_f g \approx \sum\limits_{k=1}^{R} \sigma_k \varphi_k(x) \langle \psi_k, g \rangle$.

Applies to the Nyström/Galerkin discretization via matrix SVD.

Examples of the operator calculus on large functional vectors (d-dimensional arrays)

Tensorize long vectors (functions) and matrices (operators) in $\mathbb{R}^{n^d} \rightleftarrows \mathbb{R}^{n^{\otimes d}}$:

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

- Linear elliptic systems, spectral problems for PDEs (parameter depend. coeff)

$$Au = f, \quad Au = \lambda u \qquad \text{where} \quad A = A(y), \quad y \in \mathbb{R}^M.$$

- Preconditioning in FEM/BEM $\Rightarrow B \approx A^{-1}$, for example $B = \Delta^{-1}$.
- Parabolic equations (matrix exponential, matrix resolvent)

$$u(0) = u_0 : \ \frac{\partial u}{\partial t} + Au = f,\ f = 0, \quad u(t) = e^{-tA} u_0 \quad \Rightarrow \quad \exp(-tA),\ (A + \frac{1}{\tau} I)^{-1}.$$

- Control theory: matrix Lyapunov equation on $\mathbb{R}^{n^d \times n^d}$, A, B - elliptic operators

$$AX + XB = G \quad \Rightarrow \quad X = \int_0^{\infty} e^{-tA} G e^{-tB} dt.$$

- Signal/image processing: convolution, FFT, FWT, QTT, ODEs, PDEs in $\mathbb{R}^{n^{\otimes d}}$.

Low-rank matrices. I. Favorable features

The class of rank $\leq k$ matrices in $\mathbb{R}^{m\times n}$ will be called by $\mathcal{R}_k$-matrices, i.e.

$$rank(M) \leq k \quad \text{for} \quad M \in \mathcal{R}_k.$$

$rank(M)$ = number of linear independent columns/rows in M.
Each $M \in \mathcal{R}_k$ can be represented in the form

$$M = A \cdot B^T = \sum_{i=1}^{k} a_i b_i^T, \qquad A \in \mathbb{R}^{m\times k}, \quad B \in \mathbb{R}^{n\times k}.$$

Lem. 1.1. (Favorable features of structured $\mathcal{R}_k$-matrices).

1. Fixed k, the set $\{\mathcal{R}_p\}$, $p \leq k$, is closed (nontrivial result in numer. linear algebra)

2. Only $k(m+n)$ numbers are required to store an $\mathcal{R}_k$-matrix.

3. The matrix-vector multiplication

$$x \mapsto y := Mx, \quad x \in \mathbb{R}^n$$

can be implemented in two steps at the cost $2k(m+n)$:

$$y' := B^T x \in \mathbb{R}^k, \text{ and } y := Ay' \in \mathbb{R}^m.$$

Low rank matrices. Collecting properties.

4. The sum of two $\mathcal{R}_k$-matrices $R_1 = A_1 B_1^T$, $R_2 = A_2 B_2^T$ is an $\mathcal{R}_{2k}$-matrix,

$$R_1 + R_2 = [A_1|A_2][B_1|B_2]^T,$$

where by concatenation $[A_1|A_2] \in \mathbb{R}^{m\times 2k}$ and $[B_1|B_2] \in \mathbb{R}^{n\times 2k}$.

5. The product of $R = AB^T \in \mathcal{R}_k$ by an arbitrary matrix M gives an $\mathcal{R}_k$-matrix:

$$RM = AB^T M = A(M^T B)^T, \qquad MR = (MA)B^T.$$

6. (Discrete version of the Schmidt decomposition).
The best approximation of an arbitrary matrix $M \in \mathbb{R}^{m\times n}$ by an $\mathcal{R}_k$-matrix M_k, say in the Frobenius norm, that is

$$\|A\|_F^2 := \sum_{(i,j)\in\{m\}\times\{n\}} a_{ij}^2 = \sum_{p=1}^{\min\{m,n\}} \sigma_p^2,$$

can be calculated by the truncated SVD (see Lem. 1.2 and Alg. 1.1).

7. Smooth bivariate functions $f(x,y)$ sampled on a tensor $m \times n$ grid can be well approximated by low-rank matrices via polynomial interpolation $\to$ fast multi-pole, $\mathcal{H}$-matrices.

II. Best rank-R approximation: Matrix singular value decomposition (SVD)

Lem. 1.2. (Matrix SVD). Every real (complex) $m \times n$-matrix M can be represented as the (contracted) product

$$M = U\Sigma V^T := \Sigma \times_1 U \times_2 V \equiv \Sigma \times_1 U^{(1)} \times_2 U^{(2)},$$

where $\times_1, \times_2$ is the contracted product in the first and second index, respectively,

1. $U = U^{(1)} = [u_1, u_2, ..., u_m]$ is a unitary $m \times m$-matrix,
2. $V = U^{(2)} = [v_1, v_2, ..., v_n]$ is a unitary $n \times n$-matrix,
3. Σ is an $m \times n$-matrix (core tensor) with the properties of

$$(i) \; pseudodiagonality: \quad \Sigma = diag\{\sigma_1, \sigma_2, ..., \sigma_{\min(m,n)}\},$$

$$(ii) \; ordering: \quad \sigma_1 \geq \sigma_2 \geq ... \geq \sigma_{\min(m,n)} \geq 0.$$

The σ_i are singular values of M, and the vectors u_i and v_j are, respectively, an ith left and jth right singular vectors: $MM^T U = U\Sigma$, $M^T MV = V\Sigma$.

Remark. SVD is the basic algebraic tool in the rank-structured matrix/tensor calculus.

Best rank-R approximation: Truncated SVD

Alg. 1.1. (Truncated SVD). Given $k \leq n$, let $M = U\Sigma V^T$ be the SVD of M, i.e.,

$$\Sigma = diag\{\sigma_1, \ldots \sigma_k, ..., \sigma_n\}, \quad n = \min(m,n), \quad \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0,$$

and $U = [u_1, ..., u_k, u_{k+1}, ..., u_m]$, $V = [v_1, ..., v_k, v_{k+1}, ..., v_n]$.

Set $\Sigma_k := diag\{\sigma_1, \ldots, \sigma_k, 0, \ldots, 0\}$, $\overline{\Sigma}_k = diag\{\sigma_1, \ldots, \sigma_k\}$, same for $\overline{U}_k$, $\overline{V}_k$. Define

$$M_k := U\Sigma_k V^T \equiv \overline{U}_k \overline{\Sigma}_k \overline{V}_k^T \approx M.$$

- Error bound

$$\|M_k - M\|_F \leq \sqrt{\sum_{j=k+1}^{n} \sigma_j^2}.$$

- The complexity of the truncated SVD: $\mathcal{O}(mn^2)$ with $m \geq n$.

Too expensive for large m and n.

How to compute almost the best rank-k matrix approximation getting rid of full SVD ?

▶ Given $M \in \mathcal{R}_p$, then its best approximation $M_k \in \mathcal{R}_k$, $k < p$, can be computed by the following QR-SVD scheme (Alg. 1.2).

▶ For symmetric positive definite (spd) matrices: truncated Cholesky algorithm.

Some definitions:

▷ QR factorization.

$$QR: \quad A = Q_A R_A \in \mathbb{R}^{q\times m} \quad \text{with the unitary matrix} \quad Q_A \in \mathbb{R}^{q\times m},$$

where R_A is a low triangular matrix.

▷ LU decomposition.

$$LU: \quad A = LU \in \mathbb{R}^{m\times m},$$

where L is a low triangular matrix, and U is an upper triangular matrix.

▷ In the particular case $A = A^T \geq 0$, we have the so-called Cholesky factorization,

$$A = L^T L \in \mathbb{R}^{m\times m}.$$

▷ Hadamard product of matrices with equal size is defined by

$$A \odot B = C, \quad \text{where} \quad c_{ij} = a_{ij} b_{ij}.$$

Rank reduction: Reduced truncated SVD

Alg. 1.2. (Reduced truncated SVD, RT-SVD). Given $M = AB^T \in \mathcal{R}_p$, $k < p$.

(i) Calculate the QR-decompositions $A = Q_A R_A$ and $B = Q_B R_B$, with the unitary matrices $Q_A \in \mathbb{R}^{m\times p}$, and $Q_B \in \mathbb{R}^{n\times p}$, and upper triangular matrices $R_A, R_B \in \mathbb{R}^{p\times p}$.

(ii) Calculate a SVD, $R_A R_B^T = U\Sigma V^T$ (with the cost $O(p^3)$).

(iii) Define $M_k = A_k B_k^T$ with $A_k := Q_A U_k \Sigma_k \in \mathbb{R}^{m\times k}$ and $B_k := Q_B V_k \in \mathbb{R}^{n\times k}$, where

$$U_k := [u_1, \ldots, u_k], \quad V_k := [v_1, \ldots, v_k]$$

(in both cases, first k columns of U and V, respectively).
(iv) The reduced matrix Σ_k of Σ is defined by truncated SVD of $R_A R_B^T = U\Sigma V^T$.

▶ Alg. 1.2. can be implemented in $\mathcal{O}(p^2(m+n)+p^3)$ operations.

Exer. 1.2. Look on the decay of singular values for the Hilbert matrix $A = \{a_{ij}\}$, $(i,j = 1,...,n)$ for $n = 10^3, 10^4$.
Compute the rank-R, $R = 2M+1$, $M = 64, 100$, *sinc*-quadrature approximation to A for $n = 10^4, 10^5$, (Exercise in Lec. 9)

$$a_{ij} = 1/(i+j) = \int_0^\infty e^{-(i+j)t}\,dt \approx \sum_{k=-M}^{M} c_k e^{-(i+j)t_k}.$$

Introduction to adaptive cross approximation of matrices (ACA)

Often, nearly best (suboptimal) rank-R approximation can be computed over partial data by the heuristic method called adaptive cross approximation (ACA).

Many matrix decomposition algorithms can be represented as a sequence of rank-one *Wedderburn updates.*

J. H. M. Wedderburn, Lectures on matrices, colloquium publications, vol. XVII, AMS, NY, 1934.

Main idea: For a given $m \times n$ matrix A and vectors x, y, s.t. $x^T Ay \neq 0$, matrix

$$B = A - \frac{Ayx^T A}{x^T Ay}, \quad \text{has} \quad rank(B) = rank(A) - 1.$$

For the rank-r matrix $A_0 = A$ after R updates (if do not fail) of form

$$A_k = A_{k-1} - \frac{A_{k-1} y_k x_k^T A_{k-1}}{x_k^T A_{k-1} y_k}, \quad \text{with} \quad x_k^T A_{k-1} y_k \neq 0,$$

the matrix A_R becomes zero leading to rank-R decomposition of A.

Example. Gaussian elimination (GE) with pivoting can be considered as a low rank matrix approximation. GE step is a rank-1 update (greedy algorithm):

$$A \longleftarrow A - A(j,:)A(:,k)/A(j,k).$$

ACA by truncated Cholesky factorization

Proposition. For symmetric positive definite (spd) matrix B the ACA can be computed by the truncated Pivoted Cholesky Factorization: $B \rightarrow B_\varepsilon = UU^T$, $U = [\mathbf{u}_1, \ldots, \mathbf{u}_r]$. The error control requires the diagonal elements only ! Book [N. Higham]

Alg. TPCF (Truncated Pivoted Cholesky Factorization of an spd matrix)
Input: Routine to calculate the matrix entry (column) of $B \in \mathbb{R}^{N \times N}$. Error toler. $\varepsilon > 0$.
(1) Compute the diagonal $b = \text{diag}(B) = \{B(i,i)\}_{i=1}^N$.
(2) Set $r = 1$, err $= \|b\|_1$ and initialize $\pi = \{1, ..., N\}$;
While err $> \varepsilon$ perform (3) - (9)
(3) Find $m = \text{argmax}\{b(\pi_j) : j = r, r+1, ..., N\}$; update π by swapping π_r and π_m;
(4) Set $u_{r,\pi_r} = \sqrt{b(\pi_r)}$;
For $r + 1 \leq m \leq N$ perform (5) - (7)
(5) Compute the entire column of B via $B(:,r)$
(6) Compute the U-column $u_{r,\pi_m} = B(r,\pi_m) - \sum_{j=1}^{r-1} u_{j,\pi_r} u_{j,\pi_m}$;
(7) Update the stored diagonal $b(\pi_m) = b(\pi_m) - u_{r,\pi_m}^2$;
(8) Compute err $= \sum_{j=r+1}^N b(\pi_m)$;
(9) Increase $r = r + 1$;
Output: Low-rank decomposition of B, $B_\varepsilon = UU^T$, such that $tr(B - B_\varepsilon) \leq \varepsilon$.

Exer. Apply truncated Cholesky decomposition to the Hilbert matrix (Exer. in Lect. 3).

Kronecker product of matrices

Def. The Kronecker product (KP) operation $A \otimes B$ of two matrices $A = [a_{ij}] \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{h\times g}$ is an $mh \times ng$ matrix that has the block representation $A \otimes B = [a_{ij} B]$, $i = 1, ..., m$; $j = 1, ..., n$.

The equivalent representation via concatenation

$$A \otimes B = [a_1 \otimes b_1 \; a_1 \otimes b_2 \; \ldots \; a_1 \otimes b_g \otimes a_2 \otimes b_1 \ldots a_n \otimes b_{g-1} \; a_n \otimes b_g].$$

Def. The Kronecker sum of $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{n\times n}$ is defined by

$$\sum\nolimits_{\otimes}(A, B) = A \otimes I_n + I_m \otimes B \in \mathbb{R}^{mn\times mn},$$

where $I = I_n$ is the $n \times n$ identity.

The Kronecker sum can be extended by the associative law to the multiple sum

$$\sum\nolimits_{\otimes}(A, B, C) = \sum\nolimits_{\otimes}([\sum\nolimits_{\otimes}(A, B)], C) = A \otimes I_n \otimes I_k + I_m \otimes B \otimes I_k + I_m \otimes I_n \otimes C.$$

Exer. 2.1. The discrete FD Laplacian on $H_0^1([0, 1]^2)$, over $n \times n$ grid is defined by

$$\Delta^{(2)} := \Delta \otimes I + I \otimes \Delta \in \mathbb{R}^{n^2\times n^2},$$

where Δ is the 1D FD equi-spaced Laplacian in $H_0^1([0, 1])$,

$$\Delta = 1/(n + 1)^2 \text{tridiag}\{-1, 2, -1\} \in \mathbb{R}^{n\times n}.$$

$\Delta^{(2)}$ has Kronecker (tensor) rank $R = 2$.

Kronecker product of matrices: main properties

KP inherits many properties from matrices A and B.

(1) Let $C \in \mathbb{R}^{s\times t}$, then the KP satisfies the *associative law*,

$$(A \otimes B) \otimes C = A \otimes (B \otimes C) = A \otimes B \otimes C,$$

and therefore we do not use brackets above. The matrix $A \otimes B \otimes C := (A \otimes B) \otimes C$ has (mhs) rows and (ngt) columns.

(2) Let $C \in \mathbb{R}^{n\times r}$ and $D \in \mathbb{R}^{g\times s}$, then the standard matrix-matrix product in the Kronecker format takes the form

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

The corresponding extension to d-th order Kronecker product

$$(A_1 \otimes ... \otimes A_d)(B_1 \otimes ... \otimes B_d) = (A_1B_1) \otimes ... \otimes (A_dB_d).$$

(3) The *distributive law*

$$(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D.$$

(4) Rank relation: $rank(A \otimes B) = rank(A)rank(B)$.

Invariance of some matrix properties:

(5) If A and B are diagonal then $A \otimes B$ is diagonal, and conversely ($A \otimes B \neq 0$).

(6) The upper/lower triangular matrices are preserved.

(7) Let A, B be Hermitian/normal/orthogonal matrix ($A^* = A$, $A^{-1} = A$, $A^{-1} = A^T$). Then $A \otimes B$ is of the corresponding type.

(8) Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$. Then

$$det(A \otimes B) = (detA)^m (detB)^n.$$

Hint: Reduce to the case of diagonal matrices.

Other simple properties:

$$(A \otimes B)^T = A^T \otimes B^T, \quad (A \otimes B)^* = A^* \otimes B^*.$$

Exer. 2.2. In general $A \otimes B \neq B \otimes A$. Formulate conditions on A and B that ensure $A \otimes B = B \otimes A$.
Hint: Check on diagonal matrices.

Thm. 2.1. Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$ be invertible matrices. Then

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

Proof. Since $det(A) \neq 0$, $det(B) \neq 0$ and the above property (**8**), we have $det(A \otimes B) \neq 0$. Thus $(A \otimes B)^{-1}$ exists and

$$(A^{-1} \otimes B^{-1})(A \otimes B) = (A^{-1}A) \otimes (B^{-1}B) = I_{nm}.$$

Lem. 2.2. Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$ be unitary matrices. Then $A \otimes B$ is a unitary matrix.
Proof. Since $A^* = A^{-1}$, $B^* = B^{-1}$ we have

$$(A \otimes B)^* = A^* \otimes B^* = A^{-1} \otimes B^{-1} = (A \otimes B)^{-1}.$$

Define the commutator $[A, B] := AB - BA$.
Lem. 2.3. Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$. Then

$$[A \otimes I_n, I_m \otimes B] = 0 \in \mathbb{R}^{m^2 \times n^2}.$$

Proof.

$$[A \otimes I_n, I_m \otimes B] = (A \otimes I_n)(I_m \otimes B) - (I_m \otimes B)(A \otimes I_n)$$
$$= A \otimes B - A \otimes B = 0.$$

II. Matrix operations

Rem. 2.2. Let $A, B \in \mathbb{R}^{n\times n}$, $C, D \in \mathbb{R}^{m\times m}$ and $[A, B] = 0$, $[C, D] = 0$. Then

$$[A \otimes C, B \otimes D] = 0.$$

Proof. Apply the identity $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.

Lem. 2.4. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times m}$. Then

$$tr(A \otimes B) = tr(A)tr(B).$$

Proof. Since $diag(a_{ii}B) = a_{ii}diag(B)$, we have

$$tr(A \otimes B) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ii}b_{jj} = \sum_{i=1}^{n} a_{ii} \sum_{j=1}^{m} b_{jj}.$$

Thm. 2.5. Let $A, B, I \in \mathbb{R}^{n\times n}$. Then

$$\exp(A \otimes I + I \otimes B) = (\exp A) \otimes (\exp B).$$

Proof. Since $[A \otimes I, I \otimes B] = 0$, we have

$$\exp(A \otimes I + I \otimes B) = \exp(A \otimes I) \exp(I \otimes B).$$

Proof of Thm. 2.5 continued

Furthermore, since

$$\exp(A \otimes I) = \sum_{k=0}^{\infty} \frac{(A \otimes I)^k}{k!}, \quad \exp(I \otimes B) = \sum_{m=0}^{\infty} \frac{(I \otimes B)^m}{m!}$$

the arbitrary term in $\exp(A \otimes I) \exp(I \otimes B)$ is given by

$$\frac{1}{k!}\frac{1}{m!}(A \otimes I)^k(I \otimes B)^m.$$

Imposing

$$(A \otimes I)^k(I \otimes B)^m = (A^k \otimes I^k)(I^m \otimes B^m) = (A^k \otimes I)(I \otimes B^m) \equiv A^k \otimes B^m,$$

we finally arrive at

$$\frac{1}{k!}\frac{1}{m!}(A \otimes I)^k(I \otimes B)^m = (\frac{1}{k!}A^k) \otimes (\frac{1}{m!}B^m).$$

Rem. Thm. 2.5 can be extended to the case of many-term sum

$$\exp(A_1 \otimes I \otimes ... \otimes I + I \otimes A_2 \otimes ... \otimes I + ... + I \otimes ... \otimes I \otimes A_d) = (e^{A_1}) \otimes ... \otimes (e^{A_d}).$$

Rem. Similar properties hold for other analytic functions, e.g. $\sin(A)$, $\cos(A)$, ...

III. Eigenvalue problem for Kronecker sums

Lem. 2.6. Let $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{n\times n}$ have the eigen-data $\lambda_1, ..., \lambda_m$, $u_1, ..., u_m$, and $\mu_1, ..., \mu_n$, $v_1, ..., v_n$, respectively. Then $A \otimes B$ has the eigenvalues $\lambda_j \mu_k$ with the corresponding eigenvectors $u_j \otimes v_k$, $1 \leq j \leq m$, $1 \leq k \leq n$.

Thm. 2.7. Under the conditions of Lem. 2.6 the eigenvalues/eigenfunctions of $A \otimes I_n + I_m \otimes B$ are given by $\lambda_j + \mu_k$ and $u_j \otimes v_k$, respectively.

Proof. Due to Lem. 2.6 we have

$$\begin{aligned}(A \otimes I_n + I_m \otimes B)(u_j \otimes v_k) &= (A \otimes I_n)(u_j \otimes v_k) + (I_m \otimes B)(u_j \otimes v_k)\\ &= (Au_j) \otimes (I_n v_k) + (I_m u_j) \otimes (Bv_k)\\ &= (\lambda_j u_j) \otimes v_k + u_j \otimes (\mu_k v_k)\\ &= (\lambda_j + \mu_k)(u_j \otimes v_k).\end{aligned}$$

Thm. 2.7 can be extended to the case of many-term sum:

Cor. In the case $d \geq 3$ the eigenvalues/eigenfunctions of the Laplacian-type matrix

$$\Delta_d = A_1 \otimes I \otimes ... \otimes I + I \otimes A_2 \otimes ... \otimes I + ... + I \otimes ... \otimes I \otimes A_d$$

are given by $\Lambda_{j_1,...,j_d} = \sum_{\ell=1}^{d} \lambda_{j_\ell}^{(\ell)}$ and $U_{j_1,...,j_d} = \bigotimes_{\ell=1}^{d} u_{j_\ell}^{(\ell)}$, respectively.

Application to matrix Lyapunov/Sylvester equations (elliptic eqn., control problems)

- The matrix Sylvester equation for $X \in \mathbb{R}^{m\times n}$

$$AX + XB^T = G \in \mathbb{R}^{m\times m} \qquad (1)$$

with $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{n\times n}$, can be written in vector form

$$(I_n \otimes A + B \otimes I_m) vec(X) = vec(G).$$

Now the solvability conditions and certain solution methods can be derived. The matrix Sylvester equation (1) is uniquely solvable if (see Thm. 2.7)

$$\lambda_j(A) + \mu_k(B) \neq 0.$$

- In the case $A = B$ we arrive at the Lyapunov equation.

Rem. 2.4. Since $I_n \otimes A$ and $B \otimes I_m$ commute, we can apply methods based on *sinc*-quadratures, to represent the inverse

$$(I_n \otimes A + B \otimes I_m)^{-1} = \int_{\mathbb{R}_+} e^{-t(I_n \otimes A + B \otimes I_m)} dt \approx \sum_{k=-M}^{M} c_k e^{-t_k A} \otimes e^{-t_k B}.$$

If A and B represent the discrete elliptic operators in $\mathbb{R}^d$, we obtain the low-rank tensor-product approximation to the Sylvester solution operator.

Large problems in low dimensions: overview of traditional approaches

In *low dimensions* ($d = 1, 2, 3$)

the goal of traditional methods is $O(N_{vol})$-algorithms with $N_{vol} = n^d$.

Main principles: Making use of *hierarchical* structures, *low-rank* pattern, *recursive* algorithms, data-sparse (sparse grids) representations, multi-resolution techniques, reduced basis.

Based on recursions via hierarchical structures:

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

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

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

Fast multipole, $\mathcal{H}$-matrix, $\mathcal{H}$-LU: $O(c^d N_{vol} \log N_{vol})$.
Well suited for integral (nonlocal, say Δ^{-1}) operators in FEM/BEM.

Massive parallelization (p cores):

Domain decomposition: $O(N_{vol}/p)$ - algorithms. Challenging for temporal eqn.

$\mathcal{H}$-matrix format: brief survey

$\mathcal{H}$- and $\mathcal{H}^2$-matrix technique is the matrix analogue of *the fast multi-pole method* (FMM). V. Rochlin '85; Greengard, Rochlin '87.

This techniques allows data-sparse matrix-matrix operations.

$\mathcal{M}_{\mathcal{H},r}(T_{I\times I}, \mathcal{P})$, *the class of data-sparse hierarchical $\mathcal{H}$-matrices*
Hackbusch, Khoromskij, Börm, Grasedyck, Bebendorf, Sauter ('99 - '05).

Ingredients of the $\mathcal{H}$-matrix construction defined on the product index set $I \times I$:

- An $\mathcal{H}$-tree $T(I)$ of the index set I (hierarchical cluster tree).
- The q-admissible partitioning $\mathcal{P}_q$ of $I \times I$ based on a block cluster tree $T_{I\times I}$.
- Rank-r approximation of all large enough blocks in $\mathcal{P}_q$.

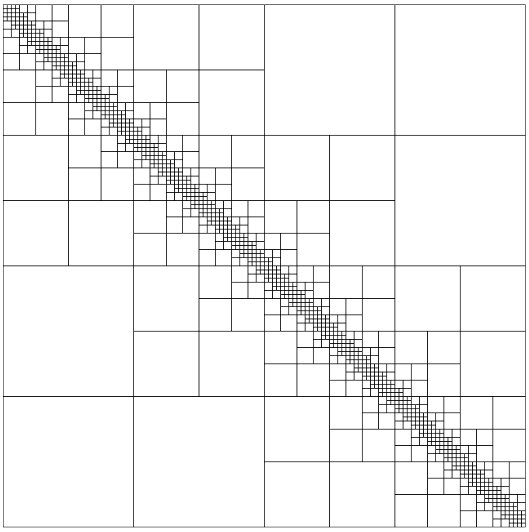
$\mathcal{H}$-matrices apply to approximation of boundary/volume integral operators for $d \leq 3$ with asymptotically smooth kernels, providing log-linear complexity in the volume size.

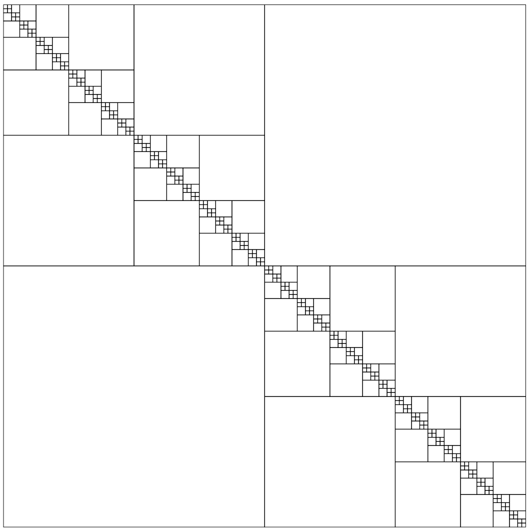
Example. The Slater, Newton, Yukawa, Helmholtz kernels ($\mathbf{r}$: the Euclidean distance):

$$\exp(-\lambda \mathbf{r}), \quad \frac{1}{\mathbf{r}}, \quad \frac{\exp(-\lambda \mathbf{r})}{\mathbf{r}}, \quad \frac{\exp(i\lambda \mathbf{r})}{\mathbf{r}} \quad \text{for} \quad d = 3.$$

Standard and weak-admissible hierarchical partitioning

Hierarchical Partitioning $\mathcal{P}_{1/2}(I \times I)$ and $\mathcal{P}_W(I \times I)$





Standard- (left) and Weak-admissible $\mathcal{H}$-partitioning for a matrix on "$1D$ index set" I.

Data-intensive applications: tensor methods beat super-computers ?

▶ **Prerequisits:**

▷ *Algebraic operations on high-dimensional data require large computer resources.*

▷ *Linear cost $O(N)$, $N = n^d$, is satisfactory only for small $d \leq 3$ and moderate n.*

▷ For large d *traditional "asymptotically optimal" methods suffer from*

"the curse of dimensionality" (R. Bellman 1961)

▷ *Complexity of $N \times N$ matrix operations in full arithmetics: $O(N^3)$.*

It is too large already for $d = 3$, i.e., $N = n^3 \Rightarrow N^3 = n^9$, say for $n \sim 10^3$.

▶ *A paradigm* of up-to-date numerical simulations:

– Higher computer capacities *only slightly reduce the curse of dimension.*, $N^d \to N^{d-1}$.

– Bridging tensor methods with supercomputing may lead to a great power !

▶ **Remedy:**

– The identification and efficient use of low-rank tensor structured representations providing linear complexity scaling in d.

– Development of tensor numerical methods with validation on the real-life problems.

Big picture. Most commonly used tensor formats in $\mathbb{R}^{n_1\times\cdots\times n_d}$.

- $d = 2$: rank-R $m\times n$ matrices admit three different representations (formats)

$$V = \sum_{k=1}^{R} u_k v_k^T \equiv UDV^T \equiv UV^T, \quad U \in \mathbb{R}^{m\times R}, \quad V \in \mathbb{R}^{n\times R}, \quad D \in \mathbb{R}^{R\times R}.$$

- $d \geq 3$: Rank-structured tensor formats in $\mathbb{R}^{n_1\times\cdots\times n_d}$.

$\Rightarrow$ **Canonical (CP)** tensor format (d-dimensional rank-R additive representation).
[Hitchcock '27]

$\Rightarrow$ **The orthogonal Tucker** decomposition (multidimensional SVD).
[Tucker '66]

$\Rightarrow$ **Matrix Product States (MPS)** (d-dimensional rank-R factorization).
[White, in physics & chemistry '90].
Special case of MPS: Tensor Train (**TT**) format.

$\Rightarrow$ **Range separated (RS) tensor format** (separate treatment of local and global features in many-particle interactions). [Benner, Khoromskaia, BNK '17].