LOW-RANK MATRICES IN MATHEMATICS AND APPLICATIONS

Eugene Tyrtyshnikov

Institute of Numerical Mathematics of Russian Academy of Sciences

eugene.tyrtyshnikov@gmail.com

90 anniversary of Guri Marchuk

8 June, 2015

USEFUL AND USELESS THINGS

- Practical issues for big data in numerical problems:
  - Tensor Trains (TT) and Hierarchical Tucker (HT)
  - Tensor-based optimization techniques
  - Low-rank approximations using frames
- Nice mathematical questions roaming a bit far from practice:
  - Rank-one conjecture for trilinear decompositions
  - Generic rank conjectures

WE NEED STRUCTURES WHEN DATA IS REALLY BIG

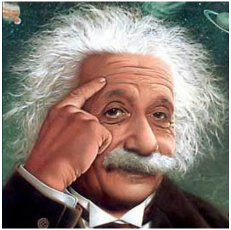
Arrays with d indices of size $n \times \ldots \times n$:

$$a(i_1, \ldots, i_d), \quad 1 \leqslant i_1, \ldots, i_d \leqslant n$$

$n = 2,\ d = 300 \quad \Rightarrow \quad$ the entries $2^{300} \gg 10^{83}$
more than atoms in the universe.

Curse of dimensionality!

BIG DATA SUGGESTS THAT NOT ALL IS IMPORTANT



Everything that can be counted
does not necessarily counts,
and everything that counts
cannot be necessarily counted.

LOW-RANK DECOMPOSITIONS PROVIDE RANK STRUCTURES

$$A = \sum_{\alpha=1}^{r} u_\alpha v_\alpha^\top = \sum_{\alpha=1}^{r} \begin{bmatrix} u_{1\alpha} \\ \vdots \\ u_{m\alpha} \end{bmatrix} \begin{bmatrix} v_{1\alpha} & \dots & v_{n\alpha} \end{bmatrix}$$

$$a(i,j) = \sum u(i,\alpha) v(j,\alpha)$$

$$(m+n)r \ll mn$$

GAUSSIAN ELIMINATION FOR LOW-RANK MATRICES

If the first $r = \operatorname{rank} A$ columns are linearly independent, then the elimination with column pivoting finishes after r steps. Moreover, it uses only the elements of the first r columns and some r rows.

What about the practical case when only a perturbed matrix is of rank r?

COLUMN-AND-ROW INTERPOLATION OF MATRICES

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad A_{11} \text{ is } r \times r$$

A can be interpolated on the first r columns and rows by

$$A_r = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} A_{11}^{-1} \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$$

INTERPOLATION ERROR

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} - \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} A_{11}^{-1} \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & A_{22} - A_{21} A_{11}^{-1} A_{12} \end{bmatrix}$$

MAXIMAL VOLUME PRINCIPLE

THEOREM (Goreinov, Tyr.) *Let*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A_{11} \text{ is } r \times r$$

with maximal volume (determinant in modulus) among all $r \times r$ blocks in A, and set

$$A_r = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} A_{11}^{-1} \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}.$$

Then

$$||A - A_r||_C \leqslant (r+1)^2 \min_{\operatorname{rank} B \leqslant r} ||A - B||_C.$$

A WAY TO RECURSION
to be explained later

$$A \approx A_r = Q \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$$

$$Q = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} A_{11}^{-1}$$

THEOREM:

$$|Q_{ij}| \leqslant 1$$

PROOF

$$Q = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ q_{r+1,1} & \cdots & q_{r+1,r} \\ \cdots & \cdots & \cdots \\ q_{n1} & \cdots & q_{nr} \end{bmatrix}$$

Necessary for the maximal volume:
$|q_{ij}| \leqslant 1, \quad r+1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant r.$

Otherwise, swapping the rows increases the volume!

CROSS INTERPOLATION HISTORY

1985 Knuth: Semi-optimal bases for linear dependencies

1995 Tyr., Goreinov, Zamarashkin: $A = CGR$ pseudoskeleton

2000 Tyr.: incomplete cross approximation with ALS maxvol

2000 Bebendorf: ACA = Gaussian elimination

2001 Tyr., Goreinov: maximum volume principle,
quasioptimality $\| \text{cross} \|_C \leq (r+1) \| \text{best} \|_2$

2006 Mahoney et al: randomized CUR algorithm

2008 Oseledets, Savostyanov, Tyr.: Cross3D

2009 Oseledets, Tyr.: TT-Cross

2010 J.Schneider: function-related quasioptimality
$\| \text{cross} \|_C \leq (r+1)^2 \| \text{best} \|_C$

2011 Tyr., Goreinov: quasioptimality
$\| \text{cross} \|_C \leq (r+1)^2 \| \text{best} \|_C$

2013 Ballani, Grasedyck, Kluge: HT-Cross

2013 Townsend, Trefethen -- Chebfun2

CANONICAL POLYADIC DECOMPOSITION

$$a(i_1 \dots i_d) = \sum u_1(i_1 \alpha) \dots u_d(i_d \alpha)$$

TUCKER DECOMPOSITION

$$a(i_1 \dots i_d) = \sum g(\alpha_1 \dots \alpha_d) u_1(i_1 \alpha_1) \dots u_d(i_d \alpha_d)$$

NEW TENSOR DECOMPOSITIONS REDUCE TENSORS TO MATRICES

Canonical polyadic and Tucker decompositions are of limited use for our purposes (by different reasons).

New decompositions in numerical anaysis:

- TT (Tensor Train) – Moscow, INM (2009)
- HT (Hierarchical Tucker) – Leipzig, MPI (2009)

Both use *low-rank matrices*.

Both use the same *dimensionality reduction tree*.

ASSUME SEPARATION OF VARIABLES

Tensor converts into a matrix (many ways!):

$$I = \{1, \dots, d\} = I_1 \sqcup I_2, \quad b(I_1, I_2) := a(i_1, \dots, i_d)$$

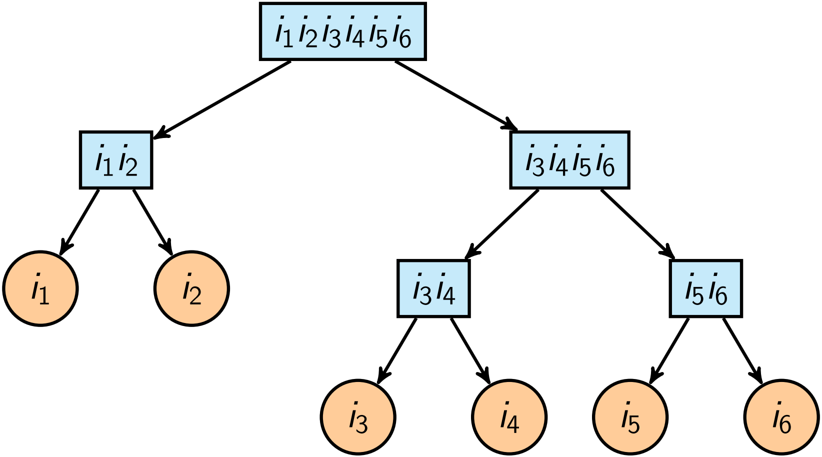
This matrix is assumed to be of low rank:

$$b(I_1, I_2) = \sum u(I_1, \alpha) v(\alpha, I_2)$$

Next idea is to repeat same for $u(I_1, \alpha)$ and $v(\alpha, I_2)$.

If straightforwardly, then too many α's arise.

REDUCTION OF DIMENSIONALITY



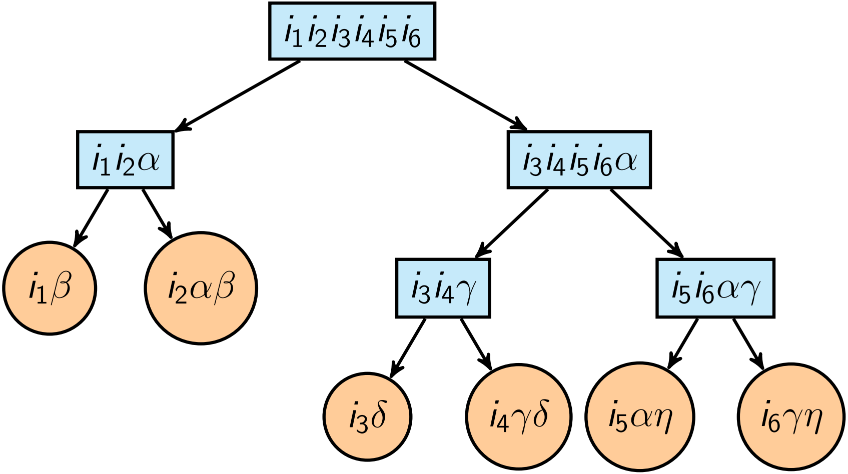
THE FIRST STEP IS ESSENTIALLY SAME

$$a(i_1 i_2\,;\; i_3 i_4 i_5 i_6) = \sum u(i_1 i_2\,;\; \alpha) v(\alpha\,;\; i_3 i_4 i_5 i_6)$$

Tensor reduces to smaller dimensionality tensors.

The α index is no longer viewed as a parameter!

SCHEME FOR TT



WHERE TT AND HT START TO DIFFER

In TT, we relegate α and γ to different descendants.

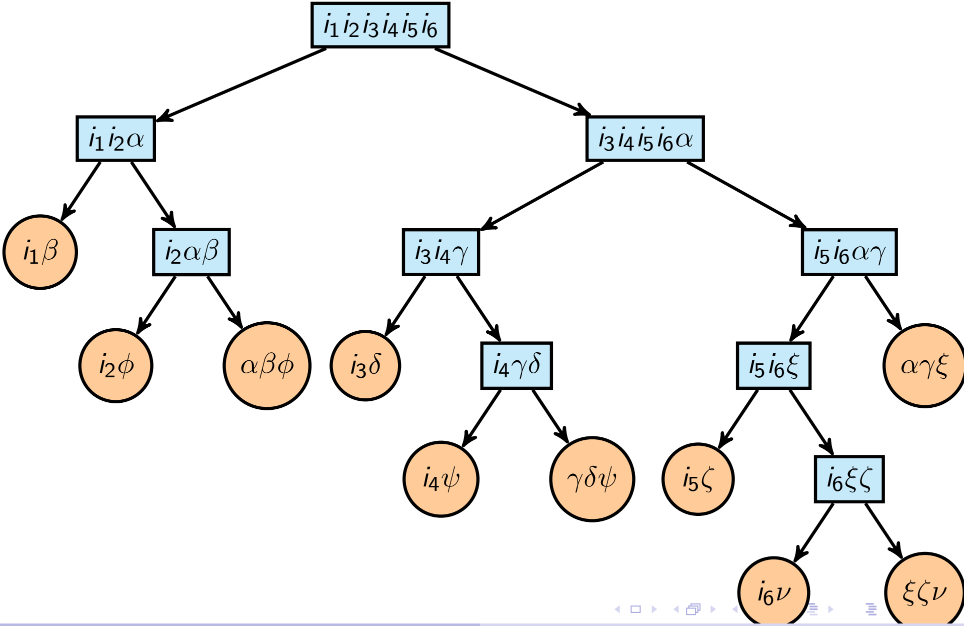
$$a(i_5 i_6 \alpha\gamma) = \sum u(i_5\alpha\,;\,\eta)v(\eta\,;\,i_6\gamma)$$

In HT, we construct a descendant with 3 auxilliary indices α, γ and ξ:

$$a(i_5 i_6 \alpha\gamma) = \sum u(i_5 i_6\,;\,\xi)v(\xi\,;\,\alpha\gamma)$$

The only difference: auxilliary summation indices are treated in different ways!

SCHEME FOR HT



WHAT IS OUR CLASS OF TENSORS?

$$A_k = [a(i_1 \dots i_k \,;\, i_{k+1} \dots i_d)] =$$

$$\left[\sum u_k(i_1 \dots i_k \,;\, \alpha_k)\; v_k(\alpha_k \,;\, i_{k+1} \dots i_d)\right] = U_k V_k^\top$$

$$u_k(i_1 \dots i_k \alpha_k) = \sum g_1(i_1 \alpha_1) \dots g_k(\alpha_{k-1} i_k \alpha_k)$$

$$v_k(\alpha_k i_{k+1} \dots i_d) = \sum g_{k+1}(\alpha_k i_{k+1} \alpha_{k+1}) \dots g_d(\alpha_{k-1} i_d)$$

THE MAIN PROPERTY OF THE CLASS:
all matrices A_k must be (close to) low-rank matrices.

WHAT IS OUR CLASS OF TENSORS?

THEOREM (Oseledets-Tyr.'2009)

Given a tensor A, assume that $\mathrm{rank}\,(A_k + E_k) = r_k$. Then a tensor train T exists with ranks $r_1, \ldots, r_{d-1}$ s.t.

$$||A - T||_F \leqslant \sqrt{\sum_{i=1}^{d-1} ||E_k||_2^2}$$

L.Graesedyck: a similar result for HT.

HOW USEFUL ARE TENSOR TRAINS?

CHEMICAL MASTER EQUATION: Gillespie'1976 vs TT

$$\frac{d\psi(\mathbf{i},t)}{dt} = \sum_{m=1}^{M} w^m(\mathbf{i}-\mathbf{z}^m)\psi(\mathbf{i}-\mathbf{z}^m,t) - w^m(\mathbf{i})\psi(\mathbf{i},t)$$

$$\frac{d\psi(\mathbf{i},t)}{dt} = A\psi(\mathbf{i},t)$$

Ammar, Cueto, Chinesta' 2011. Hegland' 2011. Dolgov, Khoromskij' 2014.

V.Kazeev, M.Khammash, M.Nip, Ch.Schwab'2014. Dolgov'2014.

FROM ONE DAY ON A SUPERCOMPUTER TO MINUTES ON A WORKSTATION

Dolgov, Tyr.' 2015 (under support of RSCF):

SSA	DMRG, hp-DG	Full	DMRG, Euler	AMEn, Euler
$2 \cdot 10^8$	$1.52 \cdot 10^4$	$7.05 \cdot 10^3$	$4.97 \cdot 10^3$	$9.23 \cdot 10^2$

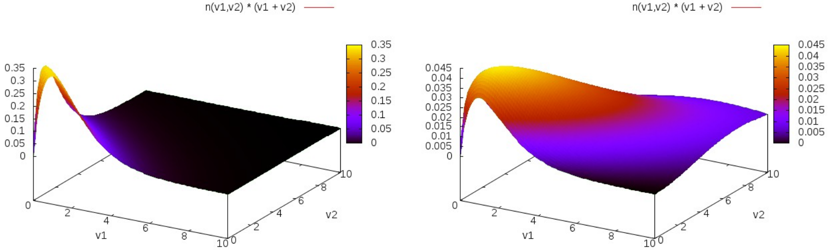
SMOLUCHOWSKI EQUATION

- $\overline{v} = (v_1, \ldots, v_d)$ – объёмы различных веществ в составе частицы
- t – время
- $n(\overline{v}, t)$ – функция концентрации объёмных компонент частицы

$$\frac{\partial n(\overline{v}, t)}{\partial t} = \frac{1}{2} \int_0^{v_1} du_1 \ldots \int_0^{v_d} K(\overline{v} - \overline{u}; \overline{u}) n(\overline{u}, t) n(\overline{v} - \overline{u}, t) du_d -$$

$$-n(\overline{v}, t) \int_0^{\infty} du_1 \ldots \int_0^{\infty} K(\overline{v}; \overline{u}) n(\overline{u}, t) du_d,$$

$$n(\overline{v}, 0) = n_0(\overline{v}).$$



Двумерное уравнение с константным ядром и начальным условием $n_0(v) = e^{(-v_1 - v_2)}$ в виде массовой концентрации $(v_1 + v_2)n(v_1, v_2, t)$. Моменты времени $T = 0.1,\ 5.0$. Относительная погрешность в L_2-норме 0.034.

MULTI-DIMENSIONAL SMOLUCHOWSKI EQUATION

Двумерное уравнение с баллистическим ядром,
$t \in [0, 1], h = 10^{-1}, \tau = 10^{-1}$.

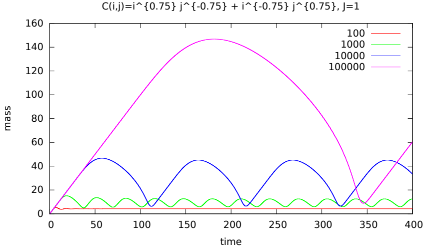
Число узлов	Предиктор-корректор, сек	Новый метод, сек	R
50	108.1	175.8	7
100	1 684	1 200	8
200	22 511	8 463	10
400	425 182	45 141	11

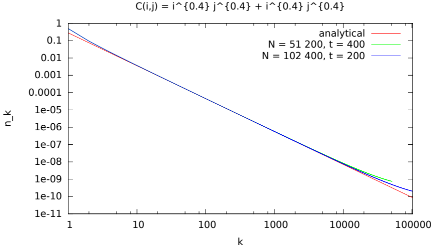
Многомерное уравнение с константным ядром
$t \in [0, 1], h = 10^{-1}, \tau = 10^{-1}, N = 100, n_0(\vec{v}) = \exp\left(-\sum_{i=0}^{d} v_i\right)$.

Размерность	Новый метод, сек	Максимальный ранг R
3	817	3
4	2 973	3
5	6 301	3

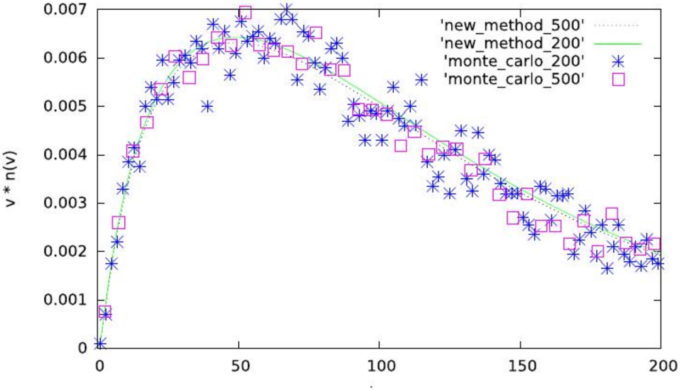
1D EQUATION WITH BALLISTIC KERNEL, $t \in [0, 10]$

V_{max}	Предиктор-корректор, сек	Адаптация к одномерной задаче, сек	Монте частиц)
100	115.08	0.55	526.68
200	459.45	1.29	526.68
500	2 877.71	5.35	526.68





Динамика изменения полной массы решения для модели необратимой коагуляции с источником мономеров и стоком крупных частиц. Справа сравнение численных и аналитический решений для модели с источником мономеров и стоком крупных частиц при $N = 51200, 102400$.



Сходимость решений одномерного уравнения Смолуховского с баллистическим ядром, полученных новым методом (непрерывные линии), и решений методом Монте-Карло(точки) при увеличении максимально допустимого размера частиц.

WELCOME THE BLESSING OF DINENSIONALITY

- Fokker-Planck, Smoluchovski equations
- Differential equations with parameters
- Green functions in integral equations
- Spin dynamics
- Global optimization algorithms
- Many others

Publications of the INM group:
http://pub.inm.ras.ru

CONJECTURE FOR RANK-ONE TENSORS

For any tensor A whose rank is less than the generic rank there is a rank-one tensor B s.t.

$$\mathrm{rank}\,(A + B) = \mathrm{rank}\,(A) + 1.$$

Still open. However, we can prove that this is true for *almost any* tensor.