APPROXIMATE ITERATIONS FOR TENSOR-STRUCTURED MATRICES

Eugene Tyrtyshnikov

tee@inm.ras.ru

Institute of Numerical Mathematics
Russian Academy of Sciences

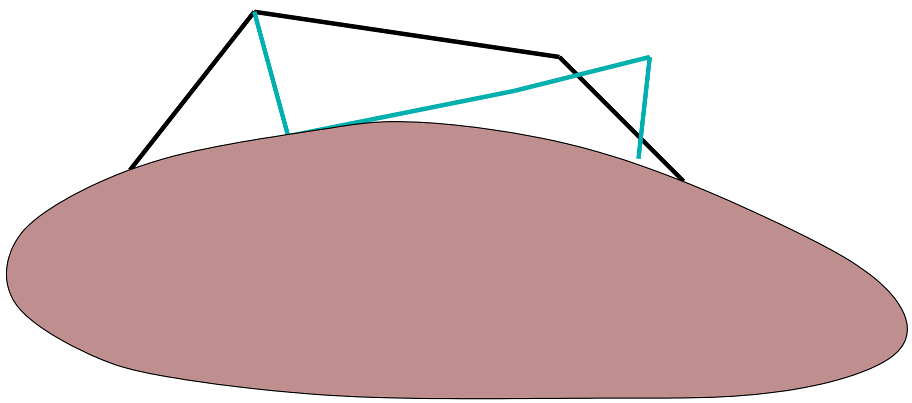
Newton–Hotelling–Schultz iterations:

$$\boldsymbol{X_k = 2X_{k-1} - X_{k-1}AX_{k-1}} \;\rightarrow\; \boldsymbol{A^{-1}}$$

$$\boldsymbol{I - AX_k = (I - AX_{k-1})^2} \quad \Rightarrow \quad \text{quadratic convergence}$$

Assume that $\boldsymbol{A}$ and $\boldsymbol{A^{-1}}$ are structured matrices
with a data-sparse (low-parametric) representation.

Truncated iterations: $\qquad \boldsymbol{X_k := R(X_k)}$



Can we keep quadratic convergence in the truncated iterations?

$\boldsymbol{V}$ a normed space

$\boldsymbol{B} \in \boldsymbol{V}$ the target of computaion

ITERATIVE PROCESS: $\quad \boldsymbol{X_k = \Phi_k(X_{k-1})}$

LEMMA.
Assume $\exists\ \boldsymbol{\alpha > 1,\ \varepsilon_\Phi,\ c_\Phi}$ s.t.

$$\boldsymbol{||X - B|| \le \varepsilon_\Phi \quad \Rightarrow}$$

$$\boldsymbol{||\Phi_k(X) - B|| \ \le\ c_\Phi ||X - B||^\alpha.}$$

Then

$$\boldsymbol{||X_0 - B|| < \varepsilon \quad \Rightarrow}$$

$$\boldsymbol{||X_k - B|| \ \le\ c^{-1}\ (c\,||X_0 - B||\,)^{\alpha^k}\,, \quad k = 1,\, 2\ldots}$$

$$\boldsymbol{\varepsilon = \min\left(\varepsilon_\Phi,\, c^{-1}\right), \quad c = c_\Phi^{\frac{1}{\alpha-1}}}$$

$\boldsymbol{S} \subset \boldsymbol{V}$ a subset of "structured" elements (e.g. structured matrices)

$\boldsymbol{R} : \boldsymbol{V} \longrightarrow \boldsymbol{S}$ a truncation operator.

$\boldsymbol{X} \in \boldsymbol{S} \quad \Rightarrow \quad \boldsymbol{R}(\boldsymbol{X}) = \boldsymbol{X}$.

TRUNCATED ITERATIVE PROCESS:

$$\boldsymbol{Y}_0 = \boldsymbol{R}(\boldsymbol{X}_0)$$

$$\boldsymbol{Y}_k = \boldsymbol{R}(\boldsymbol{\Phi}_k(\boldsymbol{Y}_{k-1}))$$

THEOREM 1.

Assume that

(1) Premises of Lemma are fulfilled.

(2) $R(B) = B$.

(3) $||X - B|| \le \varepsilon_{\Phi} \quad \Rightarrow \quad ||X - R(X)|| \le c_R \, ||X - B||$.

Then $\exists \;\; \delta > 0$ s.t.

$$Y_0 = R(Y_0), \quad ||Y_0 - B|| < \delta \quad \Rightarrow$$

$$||Y_k - B|| \;\le\; c_{R\Phi} \, ||Y_{k-1} - B||^{\alpha}, \quad k = 1, 2, \ldots$$

$$c_{R\Phi} = (c_R + 1) c_{\Phi}$$

W. Hackbusch, B.N. Khoromskij, E. Tyrtyshnikov. *Approximate iteration for structured matrices.* Preprint no. 112, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig 2005. Numer. Math., DOI 10.1007/s00211-008-0143-0, 2008.

PROOF IS ELEMENTARY!

$$||\boldsymbol{Y}_{k-1} - \boldsymbol{B}|| \leq \varepsilon \quad \Rightarrow \quad ||\boldsymbol{\Phi}_k(\boldsymbol{Y}_{k-1}) - \boldsymbol{B}|| \leq \varepsilon \quad \Rightarrow$$

$$\begin{aligned} ||\boldsymbol{Y}_k - \boldsymbol{B}|| \; &= \; ||(\boldsymbol{R}(\boldsymbol{\Phi}_k(\boldsymbol{Y}_{k-1})) - \boldsymbol{\Phi}_k(\boldsymbol{Y}_{k-1})) + (\boldsymbol{\Phi}_k(\boldsymbol{Y}_{k-1}) - \boldsymbol{B})|| \\ &\leq \; (\boldsymbol{c}_R + 1)\, ||\boldsymbol{\Phi}_k(\boldsymbol{Y}_{k-1}) - \boldsymbol{B}|| \end{aligned}$$

$$||\boldsymbol{\Phi}_k(\boldsymbol{Y}_{k-1}) - \boldsymbol{B}|| \leq \boldsymbol{c}_{\Phi}\, ||\boldsymbol{Y}_{k-1} - \boldsymbol{B}||^{\alpha}$$

Choose

$$\boldsymbol{\delta} = \min\left(\varepsilon,\, \boldsymbol{C}^{-1}\right), \quad \boldsymbol{C} = \boldsymbol{c}_{R\Phi}^{\frac{1}{\alpha-1}}, \quad \boldsymbol{c}_{R\Phi} = (\boldsymbol{c}_R + 1)\boldsymbol{c}_{\Phi}$$

$$\Rightarrow \quad ||\boldsymbol{Y}_k - \boldsymbol{B}|| \leq \varepsilon \quad \forall \;\; \boldsymbol{k}$$

COROLLARY.

$$||\boldsymbol{Y}_0 - \boldsymbol{B}|| < \boldsymbol{\delta} \quad \Rightarrow \quad ||\boldsymbol{Y}_k - \boldsymbol{B}|| \; \leq \; \boldsymbol{C}^{-1}\,(\boldsymbol{C}\,||\boldsymbol{Y}_0 - \boldsymbol{B}||)^{\alpha^k}, \quad \boldsymbol{k} = 1, 2, \ldots .$$

THEOREM 2.

Assume $\exists$ ε_Φ, c_R, ε_{RB} s.t

$$||X - B|| \le \varepsilon_\Phi \quad \Rightarrow$$
$$||X - R(X)|| \le c_R\, ||X - B|| + \varepsilon_{RB}.$$

Let m be the minimal k s.t.

$$e^\alpha_{k-1} \le \frac{\varepsilon_{RB}}{c_{R\Phi}}, \qquad c_{R\Phi} = (c_R + 1)c_\Phi.$$

Then errors $e_k = ||Y_k - B||$ of trun. iter. decrease superlinearly until $k \le m$:

$$k \le m - 1 \quad \Rightarrow \quad e_k \le 2c_{R\Phi}\, e^\alpha_{k-1},$$

$$k \ge m \quad \Rightarrow \quad e_m \le 2\varepsilon_{RB}.$$

PROOF. $Z_k := \Phi_k(Y_{k-1})$

$$||Y_k - B|| \le ||Y_k - Z_k|| + ||Z_k - B|| \le (c_R + 1)||Z_k - B|| + \varepsilon_{RB} \quad \Rightarrow$$

$$e_k \le c_{R\Phi}\, e^\alpha_{k-1} + \varepsilon_{RB} \le 2c_{R\Phi}\, e^\alpha_{k-1}.$$

TRUNCATION OPERATOR PROPERTY

$$||\boldsymbol{X} - \boldsymbol{R}(\boldsymbol{X})|| \leq \boldsymbol{c_R}\, ||\boldsymbol{X} - \boldsymbol{B}||$$

LEMMA

Let $\boldsymbol{B} = \boldsymbol{R}(\boldsymbol{B})$, and assume that $\boldsymbol{R}$ is a Lipschitz operator.
Then the truncation operator property holds.

COROLLARY

The truncation operator property is fulfilled as soon as
$\boldsymbol{B} = \boldsymbol{R}(\boldsymbol{B})$ and $\boldsymbol{R}$ is a bounded linear operator.

REMARK

It is sufficient to weaken the Lipschitz property so that it is required only for pairs of elements of which one is fixed and equal to $\boldsymbol{B}$.

EXAMPLE

$\boldsymbol{R}(\boldsymbol{X})$ keeps the entries of $\boldsymbol{X}$ inside a sparsity pattern
and sets zeroes elsewhere.

GENERAL FORM OF TRUNCATION OPERATOR

$$\boldsymbol{R} = \boldsymbol{L}^{-1} \circ \boldsymbol{\Pi} \circ \boldsymbol{L}$$

TRUNCATION LEMMA

Let $\boldsymbol{V}$ and $\boldsymbol{W}$ be normed spaces,
$\boldsymbol{L} : \boldsymbol{V} \longrightarrow \boldsymbol{W}$ a bounded linear operator with a bounded inverse.
Given $\boldsymbol{B} \in \boldsymbol{V}$, assume that $\boldsymbol{\Pi} : \boldsymbol{W} \longrightarrow \boldsymbol{W}$ possesses the following property:

$$||\boldsymbol{Z} - \boldsymbol{\Pi}(\boldsymbol{Z})|| \leq \boldsymbol{c}_{\boldsymbol{\Pi}} ||\boldsymbol{Z} - \boldsymbol{L}(\boldsymbol{B})||$$

for all $\boldsymbol{Z} \in \boldsymbol{W}$ s.t. $||\boldsymbol{L}^{-1}(\boldsymbol{Z}) - \boldsymbol{B}|| \leq \boldsymbol{\varepsilon}_{\boldsymbol{\Phi}}$.

Then the truncation operator property holds with

$$\boldsymbol{c}_{\boldsymbol{R}} = \boldsymbol{c}_{\boldsymbol{\Pi}} ||\boldsymbol{L}|| \, ||\boldsymbol{L}^{-1}||.$$

EXAMPLE 1. DISPLACEMENT-STRUCTURE TRANSFORM

$$\boldsymbol{L}(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{U} - \boldsymbol{V}\boldsymbol{X}$$

EXAMPLE 2. DISCRETE WAVELET TRANSFORM

$$\boldsymbol{L}(\boldsymbol{X}) = \boldsymbol{Q}^{\top}\boldsymbol{X}\boldsymbol{Q}$$

EXAMPLE 3. RESHAPING FOR A 2-LEVEL MATRIX

Vectorization of a matrix: $\boldsymbol{A} \longrightarrow \boldsymbol{\mathcal{V}}(\boldsymbol{A})$ (pick up entries column by column)

$$\boldsymbol{A} \mapsto \boldsymbol{L}(\boldsymbol{A}) = [\boldsymbol{\mathcal{V}}(\boldsymbol{A}_{11}), \boldsymbol{\mathcal{V}}(\boldsymbol{A}_{21}), \ \ldots \ , \boldsymbol{\mathcal{V}}(\boldsymbol{A}_{pp})]^{\boldsymbol{T}}.$$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \ldots & \boldsymbol{A}_{1p} \\ \ldots & \ldots & \ldots \\ \boldsymbol{A}_{p1} & \ldots & \boldsymbol{A}_{pp} \end{bmatrix} \in \mathbb{C}^{(pq)\times(pq)} \quad \longrightarrow \quad \boldsymbol{L}(\boldsymbol{A}) \in \mathbb{C}^{p^2 \times q^2}$$

(e.g., $\boldsymbol{A} = \boldsymbol{U} \otimes \boldsymbol{V}$ with $\boldsymbol{U} \in \mathbb{C}^{p\times p}$ and $\boldsymbol{V} \in \mathbb{C}^{q\times q}$)

Important that $||\boldsymbol{A}||_F = ||\boldsymbol{L}(\boldsymbol{A})||_F$.

KRONECKER AND LOW-RANK APPROXIMATIONS

The best Kronecker approximation (in Frobenius norm) can be computed by the SVD applied to $\boldsymbol{L}(\boldsymbol{A})$.

EXAMPLE

$$\boldsymbol{U} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}, \quad \boldsymbol{V} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}$$

$$\boldsymbol{L}(\boldsymbol{U} \times \boldsymbol{V}) = \begin{bmatrix} u_{11} \\ u_{21} \\ u_{12} \\ u_{22} \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & v_{31} & v_{12} & v_{22} & v_{32} & v_{13} & v_{23} & v_{33} \end{bmatrix}$$

APPLICATIONS OF THE TRUNCATION LEMMA

Define a suitable system of normed spaces $\boldsymbol{W_1, \ldots, W_N}$ and set

$$\boldsymbol{W = W_1 \times \ldots \times W_N = \{H = (H_1, \ldots, H_N) : H_i \in W_i\},}$$
$$\boldsymbol{||H|| = \left(\sum_{i=1}^{N} ||H_i||^2\right)^{1/2}.}$$

Let $\boldsymbol{W_i}$ be endowed a truncation operator $\boldsymbol{\Pi_i : W_i \longrightarrow W_i}$ s.t.

$$\boldsymbol{||H_i - \Pi_i(H_i)|| \leq c_i||H_i - Z_i||, \quad 1 \leq i \leq N,}$$

where $\boldsymbol{H_i \in W_i}$ are arbitrary elements and $\boldsymbol{Z_i \in W_i}$ are some fixed elements.

LEMMA.

Assume that

(a) $\boldsymbol{L : V \longrightarrow W}$ is a bounded linear operator with bounded inverse

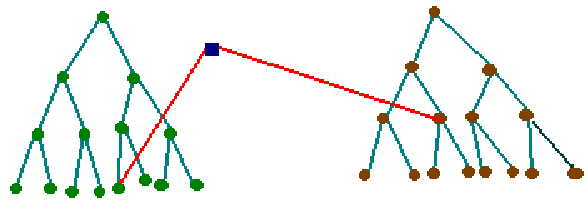
(b) $\boldsymbol{Z_i \in W_i}$ are s.t. $\boldsymbol{L(B) = (Z_1, \ldots, Z_N)}$, $\boldsymbol{B}$ is the target element.

(c) $\boldsymbol{\Pi : W \longrightarrow W}$ is defined by

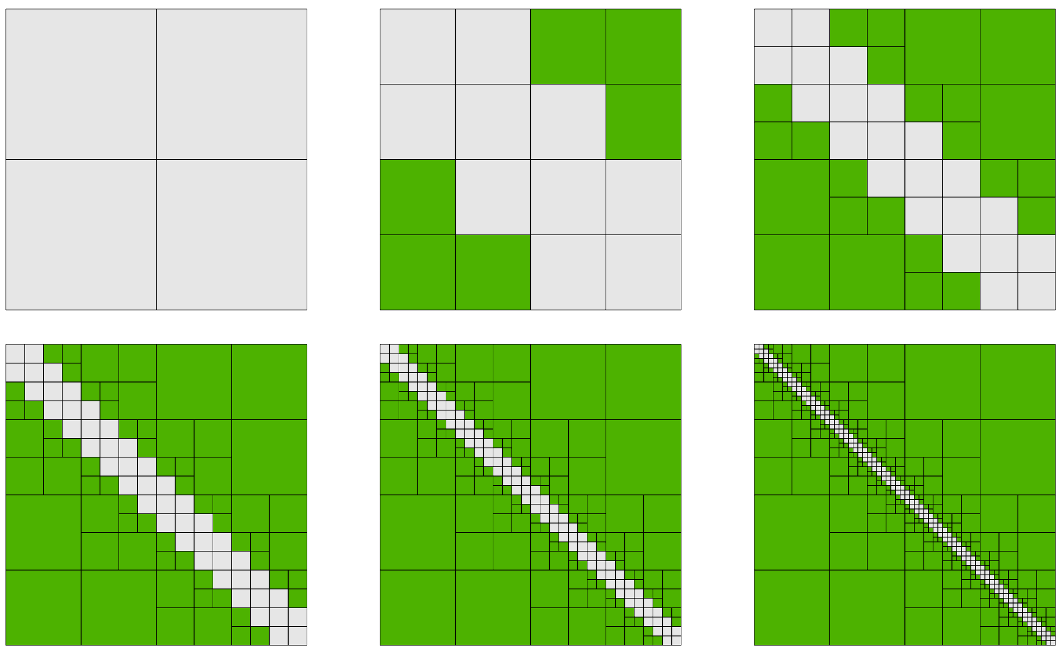
$$\boldsymbol{\Pi(H) = (\Pi_1(H_1), \ldots, \Pi_N(H_N)).}$$

Then $\boldsymbol{R = L^{-1} \circ R \circ L}$ possesses the truncation operator property.

TREE OF SUBDOMAINS



MULTILEVEL MATRICES



LEMMA. Let $\boldsymbol{W}$ be a normed space of all matrices of a fixed size and $\boldsymbol{S} \subset \boldsymbol{W}$ consist of all matrices with rank $\leq \boldsymbol{r}$. Then for any $\boldsymbol{H} \in \boldsymbol{W}$ there exists a matrix $\boldsymbol{T} \in \boldsymbol{S}$ s.t.

$$||\boldsymbol{H} - \boldsymbol{T}|| = \min_{\operatorname{rank} \boldsymbol{Z} \leq \boldsymbol{r}} ||\boldsymbol{H} - \boldsymbol{Z}||.$$

PROOF. Consider a minimizing sequence $\boldsymbol{Z_k} \in \boldsymbol{S}$ s.t.

$$\lim_{k \to \infty} ||\boldsymbol{H} - \boldsymbol{Z_k}|| = \boldsymbol{\rho} \equiv \inf_{\operatorname{rank} \boldsymbol{Z} \leq \boldsymbol{r}} ||\boldsymbol{H} - \boldsymbol{Z}||.$$

$\boldsymbol{Z_k}$ is bounded in the norm $\quad \Rightarrow \quad \exists \; \boldsymbol{Z_{k_l}} \to \boldsymbol{T}$.
By transition to the limit,

$$||\boldsymbol{H} - \boldsymbol{T}|| = \boldsymbol{\rho}.$$

Why $\boldsymbol{T} \in \boldsymbol{S}$?
This is because a matrix of rank $\boldsymbol{p}$ possesses
a vicinity wherein any matrix is of rank $\geq \boldsymbol{p}$.

COROLLARY.
For any norm, let $\boldsymbol{\Pi}$ be the best approximant.
Then $\boldsymbol{\Pi}$ satisfies the truncation operator property.

UNITARILY INVARIANT NORMS

$$\boldsymbol{\Sigma}(\boldsymbol{H}) = \begin{bmatrix} \boldsymbol{\sigma}_1(\boldsymbol{H}) & & \\ & \boldsymbol{\sigma}_2(\boldsymbol{H}) & \\ & & \ddots \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\boldsymbol{r}}(\boldsymbol{H}) = \begin{bmatrix} \boldsymbol{\sigma}_1(\boldsymbol{H}) & & & & \\ & \ddots & & & \\ & & \boldsymbol{\sigma}_{\boldsymbol{r}}(\boldsymbol{H}) & & \\ & & & \mathbf{0} & \\ & & & & \ddots \end{bmatrix}$$

$\boldsymbol{H} = \boldsymbol{Q}_1\boldsymbol{\Sigma}(\boldsymbol{H})\boldsymbol{Q}_2 \quad \Rightarrow \quad \boldsymbol{\Pi}(\boldsymbol{H}) = \boldsymbol{Q}_1\boldsymbol{\Sigma}_{\boldsymbol{r}}(\boldsymbol{H})\boldsymbol{Q}_2$ is the best possible approximant to $\boldsymbol{H}$ over matrices of rank $\leq \boldsymbol{r}$. The norm is arbitrary but unitarily invariant.

It can be readily deduced from the *Mirsky theorem*:

$$||\boldsymbol{\Sigma}(\boldsymbol{H}) - \boldsymbol{\Sigma}(\boldsymbol{Z})|| \leq ||\boldsymbol{H} - \boldsymbol{Z}||.$$

If $\boldsymbol{Z} \in \boldsymbol{S}$ then $\boldsymbol{\sigma}_{\boldsymbol{i}}(\boldsymbol{Z}) = \mathbf{0}$ for $\boldsymbol{i} \geq \boldsymbol{r} + \mathbf{1}$. Using this together with the monotonicity we obtain

$$||\boldsymbol{H} - \boldsymbol{\Pi}(\boldsymbol{H})|| = ||\boldsymbol{\Sigma}(\boldsymbol{H}) - \boldsymbol{\Sigma}_{\boldsymbol{r}}(\boldsymbol{H})|| \leq ||\boldsymbol{\Sigma}(\boldsymbol{H}) - \boldsymbol{\Sigma}(\boldsymbol{Z})||.$$

By Mirski the latter norm is upper bounded by $||\boldsymbol{H} - \boldsymbol{Z}||$.

SPECTRAL AND FROBENIUS NORMS

$$\min_{\mathrm{rank}\, Z \le r} ||H - Z||_2 = \sigma_{r+1}(H), \quad \min_{\mathrm{rank}\, Z \le r} ||H - Z||_F = \left(\sum_{i \ge r+1} \sigma_i^2(H) \right)^{1/2}$$

Truncation operator property is granted by existence of the best approximation element. Sometimes (e.g. canonical approximations of tensors) it is not the case.

LEMMA.

Let $\rho(H) = \inf_{T \in S} ||H - T||$. Given a fixed $\varepsilon > 0$, denote by $\Pi(H)$ an ε-optimal approximation to H:

$$\rho(H) \le ||H - \Pi(H)|| \le \rho(H) + \varepsilon.$$

Then

$$||H - \Pi(H)|| \le ||H - Z|| + \varepsilon \quad \forall \; Z \in S.$$

Take into account that $\rho(H) \le ||H - Z||$. Then

$$||H - \Pi(H)|| - ||H - Z|| \le (\rho(H) + \varepsilon) - \rho(H) = \varepsilon.$$

APPROXIMATE INVERSION OF MATRICES

$$\int_0^1 \int_0^1 \frac{u(x,y)}{(x_0-x)^2+(y_0-y)^2)^{3/2}} dxdy \;=\; f(x_0,y_0)$$

Uniform grids

$n = p^2$	4096	16384	65536	262144
ε	10^{-5}	10^{-5}	10^{-5}	10^{-5}
Tensor rank of A	8	8	9	10
Inversion time	2.68 sec	15.39 sec	1.47 min	7.29 min
Number of iterations	7	8	9	10
Tensor rank of A^{-1}	14	15	15	15
Residual	$4 \cdot 10^{-6}$	$3 \cdot 10^{-6}$	$1.5 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$

$n = p^2$	4096	16384
ε	10^{-5}	10^{-5}
Tensor rank of A	15	16
Inversion time	18.1 sec	1.4 min
Number of iterations	12	15
Tensor rank of A^{-1}	21	21
Residual	$8 \cdot 10^{-5}$	$4 \cdot 10^{-5}$

Chebyshev grids

$$x_i = (1 - \cos \frac{\pi i}{p})/2$$

$$x_{0i} = (1 - \cos \frac{\pi(i-0.5)}{p})/2$$

I. V. Oseledets, E. E. Tyrtyshnikov: Approximate inversion of matrices in the process of solving a hypersingular integral equation, *Comp. Math. and Math. Phys.*, Vol. 45, No. 2 (2005), 302–313 (translated from *JVM i MF*, Vol. 45, No. 2 (2005), 315–326).

2D TENSOR-WAVELET SOLVER FOR $Au = f$

(1) Finde a canonical approximation for $\boldsymbol{A}$:

$$\boldsymbol{B} = \sum_{k=1}^{r} \boldsymbol{U}_k \otimes \boldsymbol{V}_k, \qquad ||\boldsymbol{B} - \boldsymbol{A}||_F \leq \varepsilon ||\boldsymbol{A}||_F$$

(2) Apply a discrete wavelet transform (DWT) with matrix $\boldsymbol{W}$ to each tensor factor to make them pseudo=sparse:

$$\boldsymbol{P}_k = \boldsymbol{W} \boldsymbol{U}_k \boldsymbol{W}^T, \ \boldsymbol{Q}_k = \boldsymbol{W} \boldsymbol{V}_k \boldsymbol{W}^T, \quad 1 \leq k \leq r.$$

Choose a sparsification threshold $\boldsymbol{\tau} = \boldsymbol{\tau}(\varepsilon, \boldsymbol{P}_k, \boldsymbol{Q}_k)$ and neglect small entries:

$$\boldsymbol{C} = (\boldsymbol{W}^{-T} \otimes \boldsymbol{W}^{-T}) \boldsymbol{D} (\boldsymbol{W} \otimes \boldsymbol{W}) \approx \boldsymbol{B}, \quad \boldsymbol{D} = \sum_{k=1}^{r} \boldsymbol{P}_k^{\tau} \otimes \boldsymbol{Q}_k^{\tau}$$

$$||\boldsymbol{B} - \boldsymbol{C}||_F \leq \varepsilon ||\boldsymbol{B}||_F.$$

Different DWT could be used.
Could be adapted to a given non-uniform grid.

2D TENSOR-WAVELET SOLVER FOR $\boldsymbol{Au} = \boldsymbol{f}$

(3) Construct a preconditioner $\boldsymbol{H}^{-1}$ for $\boldsymbol{A}$, e.g. a block circulant preconditioner with scaling or a Newton-based approximation for the inverse.

(4) Apply iterations (GMRES) to solve

$$\boldsymbol{H}^{-1}\boldsymbol{C}\boldsymbol{y} = \boldsymbol{H}^{-1}\boldsymbol{f}. \tag{1}$$

(5) Finish with $\boldsymbol{u} \approx \boldsymbol{y}$.

J. M. Ford, E. E. Tyrtyshnikov, Combining Kronecker product approximation with discrete wavelet transforms to solve dense, function-related systems, *SIAM J. Sci. Comp.*, Vol. 25, No. 3 (2003), 961–981.

J.M.Ford, I.V.Oseledets, E.E.Tyrtyshnikov, Matrix approximations and solvers using tensor products and non-standard wavelet transforms related to irregular grids, *Rus. J. Numer. Anal. and Math. Modelling*, Vol. 19, No. 2 (2004), 185–204.

MODIFIED NEWTON (I. Oseledets, E. T.)

$\boldsymbol{Y_0 = I}$

$\boldsymbol{X_0}$ is nonsingular s.t. $\texttt{spectral radius}(\boldsymbol{I - X_0}) < 1$

$$\boldsymbol{X_k \; = \; X_{k-1}(2I - X_{k-1}), \quad Y_k \; = \; Y_{k-1}(2I - X_{k-1})}$$

PROOF. $\boldsymbol{\lim_{k\to\infty} X_k = I \;\; \Rightarrow \;\; X_{k+1} = X_0 Y_k \;\; \Rightarrow \;\; \lim_{k\to\infty} Y_k = X_0^{-1}.}$

If $\boldsymbol{A}$ is spd, then $\boldsymbol{X_0 := \alpha A}$ with $\boldsymbol{\alpha < 2/||A||_2 \;\; \Rightarrow}$

$$\boldsymbol{|I - X_0||_2 < 1, \qquad \lim_{k\to\infty} Y_k \; = \; A^{-1}/\alpha.}$$

TRUNCATED MODIFIED NEWTON

$$\boldsymbol{X_k \; = \; R(X_{k-1}(2I - X_{k-1})), \quad Y_k \; = \; R(Y_{k-1}(2I - X_{k-1}))}$$

In practice we find it 3 times faster. But, proof is lacking!

EXAMPLE OF LINEAR STRUCTURE

$$\boldsymbol{A} = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{1-n} \\ a_1 & a_0 & a_{-1} & \cdots & a_{2-n} \\ a_2 & a_1 & a_0 & \cdots & a_{3-n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_0 \end{bmatrix} \quad \text{(Toeplitz matrix)}$$

RELATION WITH A BILINEAR STRUCTURE

$$Z_s = \begin{bmatrix} 0 & \dots & 0 & s \\ 1 & & & 0 \\ & \ddots & & \\ & & 1 & 0 \end{bmatrix} \quad \Rightarrow \quad \operatorname{rank}(A - Z_s A Z_t^\top) \leq 2$$

Operator $A \mapsto A - Z_s A Z_t^\top = GH^\top$ is invertible for $st \neq 1$:

$$(1 - st)A = \sum_{k=1}^{r} C_s(g_k) C_t^\top(h_k)$$

$$G = [g_1, \dots, g_r], \quad H = [h_1, \dots, h_r]$$

$$[C_s(v)]_{ij} = \begin{cases} v_{i-j}, & i - j \geq 0, \\ s\, v_{n+i-j}, & i - j < 0. \end{cases} \qquad (s\text{-circulant})$$

FORMULAS FOR THE INVERSE MATRICES

$$\begin{aligned} \mathbf{rank}(A - Z_s A Z_t^\top) &= \mathbf{rank}(A^{-1} - Z_t^\top A^{-1} Z_s) \\ &= \mathbf{rank}(A^{-\top} - Z_t A^{-\top} Z_s^\top) \end{aligned}$$

Trench, Gohberg–Sementsul, Krupnik, Sahnovich, Heinig, Kailath, ...

Gohberg–Sementsul:

$$A^{-1} = x_0^{-1} \begin{bmatrix} x_0 & & & \\ x_1 & x_0 & & \\ \dots & \dots & \dots & \\ x_n & \dots & \dots & x_0 \end{bmatrix} \begin{bmatrix} u_0 & u_1 & \dots & u_n \\ & u_0 & \dots & u_{n-1} \\ & & \dots & \dots \\ & & & u_0 \end{bmatrix} -$$

$$- x_0^{-1} \begin{bmatrix} 0 & & & & \\ y_0 & 0 & & & \\ y_1 & y_0 & 0 & & \\ \dots & \dots & \dots & \dots & \\ y_{n-1} & \dots & \dots & y_0 & 0 \end{bmatrix} \begin{bmatrix} 0 & v_0 & v_1 & \dots & v_{n-1} \\ & 0 & v_0 & \dots & v_{n-2} \\ & & \dots & \dots & \dots \\ & & & 0 & v_0 \\ & & & & 0 \end{bmatrix}$$

$$A \begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \dots \\ 0 \end{bmatrix}, \quad A \begin{bmatrix} y_0 \\ \dots \\ y_{n-1} \\ y_n \end{bmatrix} = \begin{bmatrix} 0 \\ \dots \\ 0 \\ 1 \end{bmatrix}, \quad u_i = y_{n-i}, \quad w_i = x_{n-1}, \qquad 0 \le i \le n.$$

A CHALLENGE OF 2D TOEPLITZ MATRICES

Two-level Toeplitz matrix of order $n = p^2$:

$$A = [a_{k_1-l_1}], \quad 0 \le k_1, l_1 \le p-1,$$

$$a_{k_1-l_1} = [a_{k_1-l_1,k_2-l_2}], \qquad 0 \le k_2, l_2 \le p-1.$$

A can be viewed as a block Toeplitz matrix with $p \times p$ blocks $\Rightarrow$
Gohberg-Heinig formulas for A^{-1}.
But too many parameters: $O(n^{3/2})$.

SUBLINEAR COMPLEXITY FOR 2D TOEPLITZ INVERSION

Consider a 5-point Laplacian of order $\boldsymbol{N = n^2}$ (2-level Toeplitz matrix). The inverse matrix is approximated by the Newton–Schultz method

$$\boldsymbol{X_{k+1} = \mathtt{APPROXIMATION}(2X_k - X_k A X_k)}$$

with a rank-structured approximation of all computed matrices: by matrices of limited tensor rank and limited displacement rank of each block.

n	$\mathbf{64^2}$	$\mathbf{128^2}$	$\mathbf{256^2}$	$\mathbf{512^2}$
Tensor rank $\boldsymbol{A^{-1}}$	9	10	11	12
Averaged displacement rank of $\boldsymbol{A^{-1}}$	13.5	13.5	16.8	18.6

Inversion of the 5-point Laplacian

Time behaves as $\boldsymbol{\mathcal{O}(\sqrt{N} r_{\mathrm{mean}}^2)}$,
$\boldsymbol{r_{\mathrm{mean}}}$ = averaged displacement rank.

V.Olshevsky, I.Oseledets, E.Tyrtyshnikov,
Superfast inversion of two-level Toeplitz matrices using Newton iteration and tensor-displacement structure, *Operator Theory Advances and Applications*, vol. 179, pp. 229–240 (2007).

Low-parametric representations of inverse matrices contain $o(N)$ parameters.

Hence, all difficulties are relegated to the representation of vectors, not matrices!

TENSOR STRUCTURE IN VECTORS

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{bmatrix} \qquad \Leftrightarrow \qquad \boldsymbol{X} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

$$\boldsymbol{X} = \mathtt{MATRIX}(\boldsymbol{x}) \qquad \Leftrightarrow \qquad \boldsymbol{x} = \mathtt{VECTOR}(\boldsymbol{X})$$

$$\boldsymbol{X} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \end{bmatrix} = \boldsymbol{u}_1 \boldsymbol{v}_1^\top + \boldsymbol{u}_2 \boldsymbol{v}_2^\top$$

$$\boldsymbol{x} = \boldsymbol{v}_1 \otimes \boldsymbol{u}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{u}_2$$

EIGENVECTOR STRUCTURE

$$\boldsymbol{M} = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{C} \otimes \boldsymbol{D}$$

If $\boldsymbol{A}$, $\boldsymbol{C}$ and $\boldsymbol{B}$, $\boldsymbol{D}$ are two pairs of commuting matrices, then any eigenvector has a tensor rank-1 structure.

DISCRETE LAPLACIAN CASE

$$\boldsymbol{M}\boldsymbol{u} = \boldsymbol{\lambda}\boldsymbol{u}, \quad \boldsymbol{M} = \boldsymbol{A} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{A}$$

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \cdots & \cdots & \cdots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}$$

$$\boldsymbol{x}_{kl} = \boldsymbol{u}^k \otimes \boldsymbol{v}^l \qquad u_s^k = \sin \frac{\pi k s}{n+1}, \qquad v_t^l = \sin \frac{\pi l t}{n+1}$$

$$\lambda_{kl} = 4 \sin^2 \frac{\pi k}{2(n+1)} + 4 \sin^2 \frac{\pi l}{2(n+1)}, \quad 1 \le k, l \le n$$

USE TENSOR VECTORS IN EIGENSOLVERS

LANCZOS:

- Choose an initial vector $\boldsymbol{p_1}$ with $||\boldsymbol{p_1}|| = \mathbf{1}$ and set $\boldsymbol{p_0} = \mathbf{0}$, $\boldsymbol{b_0} = \mathbf{0}$.
- For $\boldsymbol{k = 1, 2, \ldots}$ compute

$$\begin{aligned}
\boldsymbol{z_k} &= \boldsymbol{M p_k} \\
\boldsymbol{a_k} &= (\boldsymbol{z_k}, \boldsymbol{p_k}) \\
\boldsymbol{q_k} &= \boldsymbol{z_k} - \boldsymbol{a_k p_k} - \boldsymbol{b_{k-1} p_{k-1}} \\
\boldsymbol{b_k} &= ||\boldsymbol{q_k}|| \\
\boldsymbol{p_{k+1}} &= \boldsymbol{q_k / b_k}
\end{aligned}$$

- Compute the Rietz values as the eigenvalues of a projection $\boldsymbol{k} \times \boldsymbol{k}$ matrix (a symmetric tridiagonal matrix consisting of the values $\boldsymbol{a_k}, \boldsymbol{b_k}$)

$$\boldsymbol{M_k} = \boldsymbol{P_k^\top M P_k}, \quad \boldsymbol{P_k} = [\boldsymbol{p_1}, \ldots, \boldsymbol{p_k}].$$

TENSOR LANCZOS

- Choose an initial vector $\boldsymbol{p_1}$ with $||\boldsymbol{p_1}|| = \mathbf{1}$ and set $\boldsymbol{p_0} = \mathbf{0}$, $\boldsymbol{b_0} = \mathbf{0}$.
- For $\boldsymbol{k} = \mathbf{1, 2, \ldots}$ compute

$$
\begin{aligned}
\boldsymbol{z_k} &= \boldsymbol{M p_k} \\
\boldsymbol{a_k} &= (\boldsymbol{z_k}, \boldsymbol{p_k}) \\
\boldsymbol{q_k} &= \boldsymbol{T_\varepsilon}(\boldsymbol{z_k} - \boldsymbol{a_k p_k} - \boldsymbol{b_{k-1} p_{k-1}}) \\
\boldsymbol{b_k} &= ||\boldsymbol{q_k}|| \\
\boldsymbol{p_{k+1}} &= \boldsymbol{q_k} / \boldsymbol{b_k}
\end{aligned}
$$

- Compute the Rietz values using $\boldsymbol{k} \times \boldsymbol{k}$ projection matrices.

STANDARD VERSUS TENSOR (50 iterations)

n	1000	2000	4000	6000
Lanczos time (sec)	2.8	12.1	76.7	224.9
Tensor Lanczos time (sec)	0.4	0.7	1.5	2.2

For $\boldsymbol{n} = \mathbf{6000}$ we observe a $\mathbf{100}$ times acceleration.

MORE EXAMPLES

$$\boldsymbol{M_r = M - \sum_{t=1}^{\rho} D_t \otimes D_t}$$

$\boldsymbol{M} = -$Laplacian. $\boldsymbol{D_t}$ are diagonal matrices with positive entries.

Compare maximal eigenvalues on the 50th iteration.
$\boldsymbol{n = 300^2}$, truncation rank $= \mathbf{10}$, $\boldsymbol{\varepsilon = 10^{-2}}$.

$\boldsymbol{\rho}$	1	3	5	7	9
Standard Lanczos	7.989	7.957	7.925	7.900	7.893
Tensor Lanczos	7.977	7.940	7.917	7.893	7.906

Entries of $\boldsymbol{D_t}$ are uniform grid values of $\boldsymbol{(1 + T_t(x))/10}$,
$\boldsymbol{T_t}$ is the Chebyshev polynomial of degree $\boldsymbol{t}$.

$\boldsymbol{\rho}$	1	3	5	7	9
Standard Lanczos	7.862	7.615	7.302	6.800	6.460
Tensor Lanczos	7.852	7.608	7.292	6.789	6.452

Entries of $\boldsymbol{D_t}$ come from random vectors with uniform distribution on $\boldsymbol{[0, 1]}$.

SUBLINEAR COMPLEXITY

For 2-dimensional problems time grows as SQUARE ROOT of total number of nodes!

For d-dimensional problems time should grow as ROOT OF DEGREE d of total number of nodes!