Eugene Tyrtyshnikov

Institute of Numerical Mathematics
Russian Academy of Sciences

E-mail: tee@inm.ras.ru

Tensor approximation and inversion of structured matrices

Part of this work is joint with V. Olshevsky and I. Oseledets.

TWO PERVASIVE STRUCTURES

- **Shift-invariance:** Toeplitz etc.

- **Separation of variables:** semiseparability, low-rank chops (mosaic etc.)

CHALLENGES FOR NUMERICALLY VIABLE STRUCTURES

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

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

  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: $O(n^{3/2})$.

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

  $$A = \begin{bmatrix} a_1^0 & a_1^1 & & & \\ a_2^{-1} & a_2^0 & a_2^1 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{p-1}^{-1} & a_{p-1}^0 & a_{p-1}^1 \\ & & & a_p^{-1} & a_p^0 \end{bmatrix}, \quad a_i^j = \begin{bmatrix} a_{i,1}^{j,0} & a_{i,1}^{j,1} & & & \\ a_{i,2}^{j,-1} & a_{i,2}^{j,0} & a_{i,2}^{j,1} & & \\ & \ddots & \ddots & \ddots & \\ & & a_{i,p-1}^{j,-1} & a_{i,p-1}^{j,0} & a_{i,p-1}^{j,1} \\ & & & a_{i,p}^{j,-1} & a_{i,p}^{j,0} \end{bmatrix}.$$

  A can be viewed as a block tridiagonal matrix $\Rightarrow$ Block semiseparable. Easy to adapt the formula

  $$A^{-1} = L \circ E + U \circ E^T - \frac{1}{2}\mathrm{diag}(L+U),$$

  $$E = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ \dots & \dots & \dots & \\ 1 & \dots & 1 & 1 \end{bmatrix}, \qquad \mathrm{rank}\, L = \mathrm{rank}\, U = 1, \quad \mathrm{diag}(L) = \mathrm{diag}(U).$$

  Again, too many parmaters: $O(n^{3/2})$.

TWO NEW MESSAGES: EFFECT OF CORTONA MEETINGS !

- **General structured classes are inherited by the factors of optimal Kronecker approximations.**
  (Inspired by an approach of J.Kamm-J.Nagy)

- **Many types of two-level structured matrices can be inverted by fast Newton iterations.**
  (Inspired by approaches of V.Pan and D.Bini-B.Meini)

TWO-LEVEL MATRIX

$$Au = f, \qquad A = [a_{ij}^{km}], \qquad p^2 \times p^2$$

Entry a_{ij}^{km} is on Row $(i-1)p+j$ and Column $(k-1)p+m$.

Components u_{ij}, f_{km} occupy positions $(i-1)p+j$, $(k-1)p+m$ in the vector-columns $u = [u_{ij}]$, $f = [f_{km}]$.

DEFINITION OF TENSOR (KRONECKER) PRODUCT

$$[c_{ij}] \times [d_{km}] = [b_{ik}^{jm}], \quad b_{ik}^{jm} = c_{ij} d_{km} \qquad (B = C \times D)$$

RESHAPING OPERATOR: $\mathcal{V}\left([a_{ij}^{km}]\right) = [a_{ik}^{jm}]$

(This is not permutation.)

IMPORTANT OBSERVATION: (Van Loan - Pitsianis)

$$B = C \times D \quad \Longleftrightarrow \quad \mathcal{V}(B) = \mathrm{vec}(C)(\mathrm{vec}(D))^\top$$

EXAMPLE:

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

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

STATEMENT: $$\mathcal{V}\left(\sum_{k=1}^{r} U_k \times V_k\right) = \sum_{k=1}^{r} (\mathrm{vec}(U_k))(\mathrm{vec}(V_k))^T$$

$$\left\| A - \sum_{k=1}^{r} U_k \times V_k \right\|_F = \left\| \mathcal{V}(A) - \sum_{k=1}^{r} (\mathrm{vec}(U_k))(\mathrm{vec}(V_k))^T \right\|_F$$

The best Kronecker approximation (in Frobenius norm) can be computed by the standard SVD method applied to $\mathcal{V}(A)$.

TENSOR APPROXIMATIONS: $A \approx B = \sum_{k=1}^{r} U_k \otimes V_k$

- Use tensor format when $r \ll n$. Then $B \ \ (\approx A)$ is stored in $2rn \ll n^2$ units.
- In many cases $r = O(\log^{\alpha} n \log^{\beta} \varepsilon^{-1})$:
  - E. E. Tyrtyshnikov, Tensor approximations of matrices generated by asymptotically smooth functions, *Sbornik: Mathematics*, Vol. 194, No. 6 (2003), pp. 147–160.
  - E. E. Tyrtyshnikov, Kronecker-product approximations for some function-related matrices, *Linear Algebra Appl.*, 379 (2004), 423–437.
  - W. Hackbusch W., B. N. Khoromskij, E. E. Tyrtyshnikov, *Hierarchical Kronecker tensor-product approximations*, Max-Panck-Institut für Mathematik in den Naturwissenschaften, Leipzig, Preprint No. 35, 2003.
- Efficient construction of low-tensor-rank approximations is provided by *incomplete cross approximation* using only $O(r)$ rows and $O(r)$ columns of A.
  - S.A.Goreinov, E.E.Tyrtyshnikov, and N.L.Zamarashkin, A theory of pseudo-skeleton approximations, *Linear Algebra Appl.* 261: 1–21 (1997).
  - E. E. Tyrtyshnikov, Incomplete cross approximation in the mosaic-skeleton method. – *Computing*, 64, no. 4 (2000), 367–380.
  - S. A. Goreinov, E. E. Tyrtyshnikov, The maximal-volume concept in approximation by low-rank matrices, *Contemporary Mathematics*, Vol. 208, 2001, 47-51.
  - 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.
  - M. Bebendorf and S. Rjasanow, Adaptive low-rank approximation of collocation matrices, *Computing*, 70: 1–24, 2003.
- Matrix-by-vector complexity is further reduced by wavelet sparsification or block-low-rank structure of the tensor factors.

SOME NUMERICAL EXAMPLES

$a_{ij} = f(z_i, z_j), \qquad z_1, \ \ldots \ , \ z_N \in [0, \ 1]^2, \qquad N = n^2$

EXAMPLE 1. $f(z', z) = 1/\rho, \quad \rho = \sqrt{(x'-x)^2 + (y'-y)^2}.$

Definition of r: the smallest such that $\|A - A_r\|_F \leq \varepsilon \|A\|_F$.

$n = 20 \quad (N = 400)$

$f = 1/\rho$ **uniform grids**

ε	10^{-2}	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-12}
r	4	7	9	11	13	15
ε_r	$6 \cdot 10^{-3}$	$3 \cdot 10^{-5}$	$5 \cdot 10^{-7}$	$5 \cdot 10^{-9}$	$3 \cdot 10^{-11}$	$7 \cdot 10^{-14}$

$f = 1/\rho$ **nonuniform grids**

ε	10^{-2}	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-12}
r	5	9	13	16	20	24
ε_r	$7 \cdot 10^{-3}$	$7 \cdot 10^{-5}$	$4 \cdot 10^{-7}$	$7 \cdot 10^{-9}$	$5 \cdot 10^{-11}$	$3 \cdot 10^{-14}$

EXAMPLE 2. Other functions f: uniform and nonuniform grids

f	$1/\rho^2$	$1/\rho^3$	$1/\sqrt{\rho}$	$\log \rho$	f	$1/\rho^2$	$1/\rho^3$	$1/\sqrt{\rho}$	$\log \rho$
$\varepsilon = 10^{-4}$	7	6	7	7	$\varepsilon = 10^{-4}$	8	6	8	8
$\varepsilon = 10^{-12}$	14	14	14	15	$\varepsilon = 10^{-12}$	23	22	23	22

STRUCTURED CLASSES:

$\mathbf{S}$: a sequence of linear subspaces $\mathbf{S}^1, \mathbf{S}^2, \ldots$ where $\mathbf{S}^n$ is a subspace of $n \times n$ matrices.

EXAMPLES: Toeplitz, circulants, diagonal, three-diagonal, banded, etc.

TENSOR PRODUCT OF STRUCTURED CLASSES:

$\mathbf{S}_\alpha^n \otimes \mathbf{S}_\beta^m$: a subspace of two-level matrices (with level sizes n,m) such that

$$A = [a_{(i_1,i_2)(j_1,j_2)}] \in \mathbf{S}_\alpha^n \otimes \mathbf{S}_\beta^m$$

iff

$$A^2_{i_2 j_2} \equiv [a_{(i_1,i_2)(j_1,j_2)}]_{i_1 j_1 = 0}^{n-1} \in \mathbf{S}_\alpha^n \quad \forall\, 0 \le i_2, j_2 \le m-1,$$
$$A^1_{i_1 j_1} \equiv [a_{(i_1,i_2)(j_1,j_2)}]_{i_2 j_2 = 0}^{m-1} \in \mathbf{S}_\beta^m \quad \forall\, 0 \le i_1, j_1 \le n-1.$$

$\mathbf{S}_\alpha \otimes \mathbf{S}_\beta$: a sequence $\mathbf{S}_\alpha^n \otimes \mathbf{S}_\beta^m$ with two indices $n, m = 1, 2, \ldots$.

Call $\mathbf{S}_\alpha \otimes \mathbf{S}_\beta$ the *tensor product of structured classes* $\mathbf{S}_\alpha$ and $\mathbf{S}_\beta$.

EXAMPLES: two-level Toeplitz, two-level circulants, block Toeplitz with circulant blocks, etc.

TENSOR APPROXIMATION WITH STRUCTURES: two-level Toeplitz

Reshaping operator result: $\mathcal{V}(A) = [a_{(i_1-j_1)(i_2-j_2)}]$ remains two-level Toeplitz. $\Rightarrow$

Consider (J.Jamm-J.Nagy)

$$W(A) = [a_{\mu\nu}], \quad 1-n_1 \le \mu \le n_1 - 1, \ 1 - n_2 \le \nu \le n_2 - 1.$$

Find a low-rank approximation

$$W(A) \approx W(A_r) \equiv \sum_{k=1}^{r} u^k (v^k)^\top \quad \Rightarrow \quad A \ \approx \ A_r = \sum_{k=1}^{r} U^k \otimes V^k$$

$$u^k = [u^k_\mu], \quad 1-n_1 \le \mu \le n_1 - 1, \qquad v^k = [v^k_\nu], \quad 1-n_2 \le \nu \le n_2 - 1.$$

$$U^k = [u^k_{i_1-j_1}], \quad 0 \le i_1, j_1 \le n_1 - 1, \qquad V^k = [v^k_{i_2-j_2}], \quad 0 \le i_2, j_2 \le n_2 - 1,$$

ERROR ESTIMATE:

$$||A - A_r||_{\mathbf{T}\otimes\mathbf{T}} = ||W(A) - W(A_r)||_F, \qquad ||A||_{\mathbf{T}\otimes\mathbf{T}} \equiv ||W(A)||_F.$$

GENERAL CASE OF TENSOR APPROXIMATIONS

Any matrix $P \in \mathbf{S}^n$ is defined by some *free parameters* w_k, $1 \le k \le \dim \mathbf{S}^n$:

$$P_{i_1 j_1} = \sum_{k=1}^{p} \alpha_{i_1 j_1}^{k} w_k.$$

$P \leftrightarrow W(P) \equiv [w_k]$ is a bijection: $\quad \mathrm{vec}(P) = A_S W(P)$.
Call A_S a *structure-frame matrix*.

Consider two classes with the structure-frame matrices

$$A_{\mathbf{S}_1} = [a_1^{\mathbf{S}_1}, ..., a_p^{\mathbf{S}_1}], \quad A_{\mathbf{S}_2} = [a_1^{\mathbf{S}_2}, ..., a_q^{\mathbf{S}_2}], \qquad p = \dim S_1,\ q = \dim S_2.$$

Then

$$\mathbf{S}_1^{n_1} = \mathrm{span}\{W_1^{-1}(a_1^{\mathbf{S}_1}),\ \dots,\ W_1^{-1}(a_p^{\mathbf{S}_1})\},$$
$$\mathbf{S}_2^{n_2} = \mathrm{span}\{W_2^{-1}(a_1^{\mathbf{S}_2}),\ \dots,\ W_2^{-1}(a_q^{\mathbf{S}_2})\},$$
$$\mathbf{S}_1^{n_1} \otimes \mathbf{S}_2^{n_2} = \mathrm{span}\{W_1^{-1}(a_k^{\mathbf{S}_1}) \otimes W_2^{-1}(a_l^{\mathbf{S}_2}), \quad k = 1,\ \dots,\ p,\ l = 1,\ \dots,\ q\}.$$

The free-parameters defining $A \in \mathbf{S}_1^{n_1} \otimes \mathbf{S}_2^{n_2}$ can be considered as the entries of a rectangular matrix of size $p \times q$. Denote this matrix by $W(A)$ and write $W = W_1 \otimes W_2$ to refer to the corresponding bijection $A \leftrightarrow W(A)$.

THEOREM.
Let $A \in \mathbf{S}_1^{n_1} \otimes \mathbf{S}_2^{n_2}$, W_1 and W_2 be the free-parameter bijections for $\mathbf{S}_1^{n_1}$ and $\mathbf{S}_2^{n_2}$, and $W = W_1 \otimes W_2$. Then $\mathcal{V}(A)$ is $n_1^2 \times n_2^2$ and can be written as

$$\mathcal{V}(A) = A_{\mathbf{S}_1} W(A) A_{\mathbf{S}_2}^T.$$

Proof.

$$W_1^{-1}(a_k^{\mathbf{S}_1}) = [\alpha_{i_1 j_1}^k], \quad W_2^{-1}(a_l^{\mathbf{S}_2}) = [\beta_{i_2 j_2}^l].$$

Assume that

$$A = [a_{(i_1 i_2)(j_1 j_2)}] \in \mathbf{S}_1^{n_1} \otimes \mathbf{S}_2^{n_2}$$

$$W(A) = [w_{kl}], \quad 1 \le k \le p,\ 1 \le l \le q.$$

Then

$$a_{(i_1 i_2)(j_1 j_2)} = \sum_{k=1}^{p} \alpha_{i_1 j_1}^k \sum_{l=1}^{q} \beta_{i_2 j_2}^l w_{kl} = \sum_{k=1}^{p} \sum_{l=1}^{q} \alpha_{i_1 j_1}^k \beta_{i_2 j_2}^l w_{kl}.$$

OBSERVATION.

A low-rank approximation for $\mathcal{V}(A)$ can be obtained via the SVD (singular value decomposition) of a matrix of the form $A_{S_1} W(A) A_{S_2}^\top$. It can be done in the following way:

1. Compute QR-factorization for matrices A_{S_1} and A_{S_2}:
$$A_{\mathbf{S}_1} = Q_1 R_1, \quad A_{\mathbf{S}_2} = Q_2 R_2.$$
2. Compute SVD for matrix $R_1 W(A) R_2^\top$:
$$R_1 W(A) R_2^\top = U \Sigma V^\top.$$
3. Then, the SVD of $\mathcal{V}(A)$ reads
$$\mathcal{V}(A) = (Q_1 U) \Sigma (Q_2 V)^\top.$$

The cost of the QR factorization is $O(pn^2 + qn^2)$ operations; it can be computed only once for a given structured class, on the *preprocessing stage*. Given a particular matrix A in this structured class, we have to compute the SVD of a smaller rectangular matrix of size $p \times q$.

THEOREM. (NEW MESSAGE !)

Let $A \in \mathbf{S}_1^{n_1} \otimes \mathbf{S}_2^{n_2}$, *and assume that*

$$A_r = \sum_{k=1}^{r} A_k^1 \otimes A_k^2,$$

is the optimal approximation to A *such that*

$$||A - A_r||_F = \min_{B_r} ||A - B_r||_F$$

over all matrices B_r *of tensor rank* r. *Then* $A_k^1 \in S_1^{n_1}$ *and* $A_k^2 \in S_2^{n_2}$.

Corollary.

Assume that $A \in \mathbf{S}_1^{n_1} \otimes \mathbf{S}_2^{n_2}$. *Then, any left singular vector* u *of* $\mathcal{V}(A)$ *is such that* $W_1^{-1}(u) \in \mathbf{S}_1^{n_1}$, *and any right singular vector* v *of the same matrix is such that* $W_2^{-1}(v) \in \mathbf{S}_2^{n_2}$.

NEWTON METHOD FOR INVERSE MATRICES (cf Hotelling and Schults)

Consider nonlinear equation

$$\Phi(X) \equiv A - X^{-1} = 0$$

and solve it by the Newton method with the kth iteration of the form

$$(\Phi'(X_{k-1}))(X_k - X_{k-1}) = -\Phi(X_{k-1}).$$

$$\Phi(X + \delta X) - \Phi(X) = X^{-1} - (X + \delta X)^{-1} = X^{-1}\, \delta X\, (X + \delta X)^{-1} \quad \Rightarrow$$

$$(\Phi'(X_{k-1}))(\delta X) = X_{k-1}^{-1}\, \delta X\, X_{k-1}^{-1} \quad \Rightarrow \quad X_k - X_{k-1} = -X_{k-1}(A - X_{k-1}^{-1})X_{k-1} \quad \Rightarrow$$

$$\boldsymbol{X_k = 2X_{k-1} - X_{k-1}AX_{k-1}}$$

Set $R_k = I - AX_k$, then $R_{k+1} = R_k^2$. $\Rightarrow$

Quadratic convergence if $\rho(R_0) < 1$ (ρ is spectral radius).

Initial guess can be $X_0 = \alpha A^*$ for an appropriate $\alpha > 0$.

Then $||A^{-1} - X_k||_2/||A^{-1}||_2 \le \varepsilon$ with the number of iterations as much as

$$\log_2(c^2 + 1) + \log_2 \ln \frac{1}{\varepsilon} \qquad (c \text{ is the spectral condition number for } A).$$

NEWTON METHOD WITH APPROXIMATIONS

Let $R(X)$ be a nonlinear Lipschitz operator (on $n \times n$ matrices):

$$||(X - R(X)) - (Y - R(Y))|| \leq M||X - Y||$$

Say that $R(X)$ is an approximation of X.

Transform $X_{k-1} = R(Z_{k-1})$ to $X_k = R(Z_k)$ (for some Z_{k-1} and Z_k):

$$\boldsymbol{Z_k = 2X_{k-1} - X_{k-1}AX_{k-1}, \quad X_k = R(Z_k) \quad (*)}$$

THEOREM. *Assume that*

$$R(A^{-1}) = A^{-1}. \tag{1}$$

Then for any initial guess $X_0 = R(X_0)$ sufficiently close to A^{-1}, the matrices X_k in $(*)$ converge to A^{-1} quadratically:

$$||A^{-1} - X_k|| \leq (1 + M)\,||A||\,||A^{-1} - X_{k-1}||^2, \qquad k = 1, 2, \ldots . \tag{2}$$

Proof. **(A)** $||X_k - Z_k|| = ||R(Z_k) - Z_k|| = ||R(Z_k) - R(A^{-1}) + A^{-1} - Z_k|| \;\leq\; M\,||Z_k - A^{-1}||$.

By the triangle inequality, $||A^{-1} - X_k|| \;\leq\; ||A^{-1} - Z_k|| + ||Z_k - X_k|| \;\leq\; (1 + M)||A^{-1} - Z_k||$.

Then, $(*)$ implies $A^{-1} - Z_k = (A^{-1} - X_{k-1})\,A\,(A^{-1} - X_{k-1})$. Hence, **(B)** $||A^{-1} - Z_k|| \;\leq\; ||A||\,||A^{-1} - X_{k-1}||^2$.

Put (A) and (B) together.

IMPORTANT EXAMPLE OF $R(A)$

Let $||\cdot||$ be any unitary invariant norm (spectral, Frobenius, etc.) and $\Pi_r(A)$ the best r-rank approximation:

$$\rho_{r+1}(A) \equiv ||A - \Pi_r(A)|| = \min_{\mathrm{rank} B \le r} ||A - B||.$$

Let L be a linear invertible operator on $n \times n$ matrices. Fix some r and define

$$R(A) = L^{-1}(\Pi_r(L(A))).$$

Lipschitz property of R:

$$||R(X) - R(Y)|| \ \le \ ||L^{-1}||\, ||\Pi_r(LX) - \Pi_r(LY)|| \ \le \ ||L^{-1}||\, ||L||\, ||X - Y||.$$

Low-tensor-rank approximations fit this framework:

$$A = [a_{ij}^{km}], \quad 1 \le i, j, k, m \le p. \quad \Rightarrow \quad L(A) = [b_{ij}^{km}], \quad b_{ij}^{km} = a_{ik}^{jm}.$$

Tensor rank of A coincides with rank of $L(A)$. (Van Loan - Pitsianis)

Remarkable generalizations of Toeplitz (Hankel, Cauchy etc.) matrices are all related with low rank of $L(A) = AU - VA$ for special matrices U, V. Our theorem generalizes a result obtained for this particular case of $L(A)$:

- V. Y. Pan, Y. Rami, Newton's iteration for the inversion of structured matrices, *Structured Matrices: Recent Developments in Theory and Computation* (Eds. Bini D.A., Tyrtyshnikov E.E., Yalamov P.), Nova Science Publishers, Huntington, New York, 2001, 79-90.

TRUNCATED NEWTON IN THE CASE $A^{-1} \approx R(A^{-1})$

Now $\varepsilon = ||A^{-1} - R(A^{-1})|| \neq 0$.

$$||Z_k - X_k|| \leq M\,||Z_k - A^{-1}|| + ||A^{-1} - R(A^{-1})||$$

Set $c = (1+M)\,||A||$. Then the values $\delta_k \equiv ||A^{-1} - X_k||$ satisfy

$$\delta_k \leq c\,\delta_{k-1}^2 + \varepsilon.$$

THEOREM.
Let m be the maximal index such that $\sqrt{\varepsilon/c} \leq \delta_{k-1}$ for all $1 \leq k \leq m$. If $2c\delta_0 < 1$, then the Newton with approximations provides quadratic decrease of errors on the iterations $1 \leq k \leq m$:

$$2c\delta_k \leq (2c\delta_{k-1})^2,$$

and finally $\delta_m \leq 2\varepsilon$.

Proof. $\sqrt{\varepsilon/c} \leq \delta_{k-1}$ implies $c\varepsilon \leq (c\delta_{k-1})^2$. Hence, $c\delta_k \leq (c\delta_{k-1})^2 + c\varepsilon \leq 2(c\delta_{k-1})^2$.

IMPROVING EFFICIENCY OF NEWTON

$$M^1 = \sum_{i=1}^{r_1} A_i^1 \otimes B_i^1, \qquad M^2 = \sum_{i=1}^{r_2} A_i^2 \otimes B_i^2. \quad \Rightarrow \quad M^1 M^2 = \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} (A_i^1 A_j^2) \otimes (B_i^1 B_j^2).$$

Recompression (a lower-rank approximation to a low-rank matrix):

$$X_k^\varepsilon = R_\varepsilon(X_k), \qquad ||X_k - X_k^\varepsilon||_F \le \varepsilon ||X_k||_F$$

If tensor ε-ranks of A and A^{-1} are $\sim 15-20$, then (as $X_k \approx A^{-1}$) each Newton iteration involves ~ 200 products of sparse matrices of order p. For reduction of complexity consider the following modification:

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

where $Y_0 = I$ and X_0 is a nonsingular matrix such that $\rho(I - X_0) < 1$. Then

$$\lim_{k\to\infty} X_k = I, \quad \lim_{k\to\infty} Y_k = X_0^{-1}. \qquad (\text{Since } X_{k+1} = X_0 Y_k.)$$

Let A be s.p.d. $\Rightarrow$ Take $X_0 = \alpha A$, $\alpha < 2/||A||_2$. $\Rightarrow$ $||I - X_0||_2 < 1$ $\Rightarrow$ $\lim_{k\to\infty} Y_k \;=\; \frac{A^{-1}}{\alpha}$.

Our modification of the truncated Newton:

$$X_k \;=\; R_\varepsilon(X_{k-1}(2I - X_{k-1})), \quad Y_k \;=\; R_\varepsilon(Y_{k-1}(2I - X_{k-1})), \quad k = 1, 2, \ldots.$$

Since $X_k \to I$, their tensor ranks become small. So the method is self-accelerated.

COMPUTATION OF INVERSE MATRICES: numerical results

$n=p^2$	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	$4 \cdot 10^{-6}$	$3 \cdot 10^{-6}$	$1.5 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$

Case of uniform grids

$n=p^2$	4096	16384	65536
ε	10^{-5}	10^{-5}	10^{-3}
Tensor rank of A	15	16	18
Inversion time	18.1 sec	1.4 min	10 min
Number of iterations	12	15	16
Tensor rank of A^{-1}	21	21	21
Residual	$8 \cdot 10^{-5}$	$4 \cdot 10^{-5}$	$8 \cdot 10^{-2}$

Case of non-uniform Chebyshev grids

Compare the orthodox and modified Newton:

for uniform grids with $p = 256$ ($n = 65536$), the orthodox Newton has taken 12 iterations to achieve

$$||AX_k - I||_F = 4.88 \cdot 10^{-3}.$$

It has taken ~ 4.5 min, which is about 3 times longer than the modified Newton's time.

TIMINGS FOR THE TENSOR-WAVELET SOLVER

Choose r.h.s. to be the sum of the matrix columns 1, 5, 10.

n	16129	65025	261121	1046529
r	11	12	12	13
ε	$8.4 \cdot 10^{-8}$	$6.4 \cdot 10^{-8}$	$9.4 \cdot 10^{-7}$	$1.9 \cdot 10^{-7}$
Matrix-by-vector time	0.05 sec	0.28 sec	1.43 sec	7.6 sec
Number of iterations	17	21	22	22
Construction of preconditioner	1.8 sec	7.2 sec	29.1 sec	1.9 min
Solution time	2.3 sec	7.9 sec	41.6 sec	3.3 min
Relative error	$3.1 \cdot 10^{-6}$	$5.8 \cdot 10^{-6}$	$1.93 \cdot 10^{-5}$	$1.1 \cdot 10^{-5}$

Using block-circulant-scaled preconditioner. Uniform grids.

n	16129	65025	261121	1046529
r	20	22	22	23
ε	$8.3 \cdot 10^{-7}$	$6.0 \cdot 10^{-8}$	$9.8 \cdot 10^{-8}$	$9.4 \cdot 10^{-8}$
Matrix-by-vector time	0.09 sec	0.43 sec	2.08 sec	13.02 sec
Number of iterations	17	17	19	20
Construction of preconditioner	1.8 sec	7.2 sec	29.1 sec	1.9 min
Solution time	2.9 sec	8.6 sec	47.9 sec	4.8 min
Relative error	$1.66 \cdot 10^{-5}$	$2.2 \cdot 10^{-5}$	$1.34 \cdot 10^{-5}$	$1.1 \cdot 10^{-5}$

Using block-circulant-scaled preconditioner. Chebyshev grids.

APPROXIMATE INVERSE PRECONDITIONERS VIA NEWTON

Matrices A are approximated in the tensor-wavelet format with accuracy of 10^{-7}.

The Newton accuracy: 10^{-5} (uniform grids) and 10^{-3} (Chebyshev grids).

n	16129	65025	261121
Construction of preconditioner	18 sec	99 sec	342.3 sec
Number of iterations	3	3	4
Solution time	0.4 sec	2 sec	13.5 sec

Uniform grids

n	3969	16129	65025
Construction of preconditioner	18 sec	84 sec	737 sec
Number of iterations	3	3	5
Solution time	0.2 sec	0.9 sec	6.6 sec

Chebyshev grids

Incomplete Kronecker Product precondtioner - IKP(r):

construct approximate inverses to rougher approximations to A
(of a smaller tensor rank r).

r	1	2	3	5
Construction of preconditioner	17 sec	30 sec	41 sec	58 sec
Number of iterations	33	15	10	8
Solution time	82.1 sec	39.8 sec	27.7 sec	22.2 sec

Results with IKP(r), $p = 511$, $n = 261121$. The Newton accuracy is $\varepsilon = 10^{-3}$.

- I. V.Oseledets, E. E. Tyrtyshnikov, Approximate inversion of matrices with solution of the hypersingular integral equation, *J. Comput. Math. and Math. Physics*, to appear.