Eugene Tyrtyshnikov

Institute of Numerical Mathematics
Russian Academy of Sciences

E-mail: tee@inm.ras.ru

Tensor approximations and computation of inverse matrices

Practical problems (esp. using integral equations) may lead to matrices which are

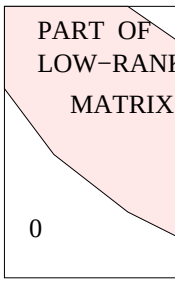
- **huge in size (up to $10^5 \sim 10^6$)**
- **with no noticable structure**

The only way is to treat such matrices implicitly – via approximation by some well-structured matrices.

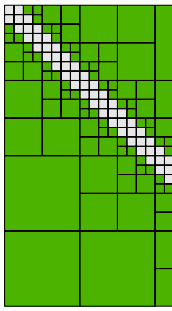
The key idea of structures is **data-sparsity**.

CONTENTS

- Tensor approximations $A \approx \sum_{i=1}^{r} U_i \otimes V_i$ with structured factors:
  - Wavelet sparsification
  - Profile-low-rank structure
  - Block-low-rank structure
- Application example (integral equation for potential flow problem)



- Computation of inverse matrices:
  - Theorem on the Newton method with approximations (applicable to multidimensional tensor approximations, wavelet sparsifications, block-low-rank truncations, etc.)
  - Self-accelerated modification
  - Numerical examples
- Construction of non-standard wavelets (related to irregular grids)



TWO-LEVEL MATRIX

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

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

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

DEFINITION OF TENSOR (KRONECKER) PRODUCT

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

REARRANGEMENT OF ENTRIES: $\boldsymbol{\mathcal{V}\left([a_{ij}^{km}]\right) = [a_{ik}^{jm}]}$

(This is not permutation.)

IMPORTANT OBSERVATION: (Van Loan - Pitsianis)

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

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-WAVELET SOLVER FOR $Au = f$

(1) Approximate A by B in the tensor format

$$B = \sum_{k=1}^{r} U_k \otimes V_k, \qquad ||B - A||_F \leq \varepsilon ||A||_F$$

$U_k,\ V_k$ are of order p (A is of order $n = p^2$). r **is called tensor ε-rank.**

(2) Apply a Discrete Wavelet Transform (DFT) to each tensor factor:

$$P_k = W U_k W^T,\ Q_k = W V_k W^T, \quad 1 \leq k \leq r. \qquad (W \text{ is the DFT matrix.})$$

Sparsify P_k and Q_k using a threshold $\tau = \tau(\varepsilon, P_k, Q_k)$:

$$C = (W^{-T} \otimes W^{-T}) D (W \otimes W) \approx B, \quad D = \sum_{k=1}^{r} P_k^{\tau} \otimes Q_k^{\tau}, \qquad ||B - C||_F \leq \varepsilon ||B$$

(3) Construct a preconditioner H^{-1} for A. Options:

- **Incomplete scaled block-circulant preconditioner.**
- **Appoximations to the inverse matrix (by the Newton method).**

(4) Apply GMRES for $H^{-1}Cy = H^{-1}f$ and set $u \approx y$.

TENSOR APPROXIMATIONS

- Use tensor format when ε-rank $r \ll n$. Then B 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 funct *Sbornik: Mathematics*, Vol. 194, No. 6 (2003), pp. 147–160.
  - E. E. Tyrtyshnikov, Kronecker-product approximations for some function-related matrices, *Linear Al Appl.*, 379 (2004), 423–437.
  - W. Hackbusch W., B. N. Khoromskij, E. E. Tyrtyshnikov, *Hierarchical Kronecker tensor-product app mations*, Max-Panck-Institut für Mathematik in den Naturwissenschaften, Leipzig, Preprint No. 35, 2
- Efficient construction of low-tensor-rank approximations is provided by *incomp 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, *L Algebra Appl.* 261: 1–21 (1997).
  - E. E. Tyrtyshnikov, Incomplete cross approximation in the mosaic-skeleton method. – *Computing*, 64 4 (2000), 367–380.
  - S. A. Goreinov, E. E. Tyrtyshnikov, The maximal-volume concept in approximation by low-rank mat *Contemporary Mathematics*, Vol. 208, 2001, 47-51.
  - J. M. Ford, E. E. Tyrtyshnikov, Combining Kronecker product approximation with discrete wavelet t forms 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, *Computin* 1–24, 2003.
- Matrix-by-vector complexity is reduced by wavelet sparsification of the tensor tors. (Alternatives can use profile-low-rank or block-low-rank structures.)

MOTIVATION: 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 $\|A - A_r\|_F \le \varepsilon \|A\|_F$.

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

$f = 1/\rho$ **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

SEPARATION OF VARIABLES

LOW-RANK APPROXIMATIONS: $f(x_1, x_2;\ y_1, y_2) \approx \sum_{i=1}^{r} u_i(x_1, x_2) v_i(y_1, y_2)$

TENSOR APPROXIMATIONS: $f(x_1, x_2;\ y_1, y_2) \approx \sum_{i=1}^{r} \phi_i(x_1, y_1) \psi_i(x_2, y_2)$

FUNCTION APPROXIMATION THEOREM.

- $f = (u, v)$ $(u = x_1 - y_1,\ v = x_2 - y_2)$ is an *asymptotically smooth* function:
  $\exists\ g$ and $c, t > 0$: $\left|\frac{\partial^{p_1}}{(\partial\zeta)^{p_1}} \frac{\partial^{p_2}}{(\partial\eta)^{p_2}} F\right| \leq c\ t^{p_1+p_2}\ (p_1 + p_2)!\ (\zeta^2 + \eta^2)^{(g-p_1-p_2)/2}, \quad \zeta^2 + \eta^2 > 0$
  any integer $p_1, p_2 \geq 0$.
- $\Pi_h = [-1, 1]^2 \backslash (-h, h)^2, \quad 0 < h < 1.$
- $0 < \gamma < 1.$

$\Longrightarrow \quad \forall\ p = 1, 2, \ \ldots \quad \exists F_r(u, v) \approx F(u, v)$ such that

$$r \ \leq\ (c_0 + c_1\ \log h^{-1})p,$$

$$|F(u, v) - F_r(u, v)| \leq c_2\ \gamma^p\ (u^2 + v^2)^{g/2}, \quad (u, v) \in \Pi_h,$$

where c_0, c_1 and c_2 are positive constants depending on γ.

- E. E. Tyrtyshnikov, Kronecker-product approximations for some function-related matr
  Linear Algebra Appl., 379 (2004), 423–437.
- E. E. Tyrtyshnikov, Tensor approximations of matrices associated with asymptotically sm
  functions, *Sbor. Math.*, Vol. 194, no. 6, pp. 147–160 (2003).

POTENTIAL FLOW AROUND THIN AIRFOIL

$\triangle\Phi = 0, \quad x \in \Pi = [0,\ 1]\times[0,\ 1]; \qquad \dfrac{\partial\Phi}{\partial n} = -V_z,\ \ x \in \partial\Pi; \qquad \Phi,\ \nabla\Phi \to 0,\ \ x \to$

Prandtl equation:

$$-\int_0^1\int_0^1 \frac{U(x,y)}{[(x-x_0)^2+(y-y_0)^2]^{3/2}}dxdy = F(x_0,y_0) \quad (\ = V_z(x_0,y_0,0)\ ).$$

Method of discrete vortices:

$$U_p(x,y) = \sum_{i=1}^{p}\sum_{j=1}^{p} u_{ij}\,\varphi_{ij}(x,y), \qquad \varphi_{ij}(x,y) = \begin{cases} 1, & (x,y) \in (x_{i-1},\ x_i)\times(y_{j-1}, \\ 0, & \text{otherwise.} \end{cases}$$

$\boldsymbol{Au = f}$: $\quad \sum_{i=1}^{p}\sum_{j=1}^{p} u_{ij}a_{ij}^{km} = f_{km} \equiv F(x_{0k},y_{0m}), \quad 1 \le k,m \le p,$

$$\boldsymbol{a_{ij}^{km} = -\int_{x_{i-1}}^{x_i}\int_{y_{j-1}}^{y_j} \frac{dxdy}{[(x-x_{0k})^2+(y-y_{0m})^2]^{3/2}}}$$

$$= -\frac{\sqrt{(x_{0k}-x_{i-1})^2+(y_{0m}-y_j)^2}}{(x_{0k}-x_{i-1})(y_{0m}-y_j)} + \frac{\sqrt{(x_{0k}-x_i)^2+(y_{0m}-y_j)^2}}{(x_{0k}-x_i)(y_{0m}-y_j)}$$

$$-\frac{\sqrt{(x_{0k}-x_i)^2+(y_{0m}-y_{j-1})^2}}{(x_{0k}-x_i)(y_{0m}-y_{j-1})} + \frac{\sqrt{(x_{0k}-x_{i-1})^2+(y_{0m}-y_{j-1})^2}}{(x_{0k}-x_{i-1})(y_{0m}-y_{j-1})}.$$

STABILITY

Diagonal entries of A are positive, and off-diagonal are negative.
A is diagonally dominant.

$$\sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}^{km} = -\int_0^1\int_0^1 \frac{dxdy}{[(x-x_{0k})^2+(y-y_{0m})^2]^{3/2}} =$$

$$= \frac{\sqrt{x_{0k}^2+(y_{0m}-1)^2}}{x_{0k}(1-y_{0m})} + \frac{\sqrt{(x_{0k}-1)^2+(y_{0m}-1)^2}}{(1-x_{0k})(1-y_{0m})} +$$

$$+\frac{\sqrt{(x_{0k}-1)^2+y_{0m}^2}}{(1-x_{0k})y_{0m}} + \frac{\sqrt{x_{0k}^2+y_{0m}^2}}{x_{0k}y_{0m}} = h(x_{0k},y_{0m}) > 0.$$

Function $h(x,y)$ attains minimal value $8\sqrt{2}$ at $(\frac{1}{2},\frac{1}{2})$.
The eigenvalues of A lie in the union of the Gershgorin disks

$$\{z : |a_{km}^{km} - z| \le a_{km}^{km} - h(x_{0k},y_{0m}) \le a_{km}^{km} - 8\sqrt{2}\}.$$

Hence,

$$|\lambda| \ge 8\sqrt{2}.$$

If A is symmetric (uniform grids) then

$$||A^{-1}||_2 \le \frac{1}{8\sqrt{2}}.$$

CONVERGENCE FOR UNIFORM GRIDS (I.Lifanov, L.Poltavsky)

$$||U - U_p||_{L_1(\Pi)} \to 0, \quad p \to \infty$$

Uniform convergence away from $\partial\Pi$ in the distance δ:

$$|U(x_{0k}, y_{0m}) - U_p(x_{0k}, y_{0m})| < A_\delta |\ln h|^{9/4} h^{1/4-\varepsilon}$$

$$0 < \varepsilon < 1/4, \qquad h = 1/p \text{ is the step-size.}$$

TYPES OF GRIDS

♠ **Uniform grid:**

$$x_i = y_i = i/p, \quad i = 0, \ldots, p,$$
$$x_{0i} = y_{0i} = (i - 0.5)/p, \quad i = 1, \ldots, p.$$

♠ **Non-uniform Chebyshev grid:**

$$x_i = y_i = (1 - \cos\frac{\pi i}{p})/2, \quad i = 0, \ldots, p,$$
$$x_{0i} = y_{0i} = (1 - \cos\frac{\pi(i - 0.5)}{p})/2, \quad i = 1, \ldots, p.$$

WHY INVERSE MATRICES?

- Nonstationary flow problems require to solve $Au = f$ for many different f.
  Thus, we need a convenient approximation to A^{-1}
  (with small storage and matrix-by-vector complexity).

- A rougher approximation to A^{-1} may serve as a preconditioner.

- For uniform grids, A is doubly Toeplitz.
  In this case A^{-1} can be dealt with via the Gohberg-Heinig formulas.

  BUT: they still contain $O(n^{3/2})$ parameters!

  For non-uniform grids, A^{-1} possesses no explicit structure.

NEWTON METHOD FOR INVERSE MATRICES (cf Hotelling and Sch

Consider nonlinear equation

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

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

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

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

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

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

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*

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

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

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

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)** $||A^{-1} - Z_k|| \;\leq\; ||A||\,||A^{-1} - X_{k-1}||$

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
r-rank approximation:

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

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
rank of $L(A) = AU - VA$ for special matrices U, V. Our theorem generalizes a result obta
for this particular case of $L(A)$:

- V. Y. Pan, Y. Rami, Newton's iteration for the inversion of structured matrices, *Structured Matrices: R Developments in Theory and Computation* (Eds. Bini D.A., Tyrtyshnikov E.E., Yalamov P.), Nova Sc 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 $2c\delta_0 <$ then the Newton with approximations provides quadratic decrease of errors on 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 invo
~ 200 products of sparse matrices of order p. For reduction of complexity consider the follo
modification:

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

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 $X_0 = \alpha A$, $\alpha < 2/||A||_2$. $\Rightarrow$ $||I - X_0||_2 < 1$ $\Rightarrow$ $\lim_{k\to\infty} Y_k = \frac{A}{\alpha}$

Our modification of the truncated Newton:

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

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 NEWTO

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 $\varepsilon = 10^{-3}$

EXPERIMENTAL CONVERGENCE

As "exact" solution, use that for $p = 511$. To trace how the convergence rate depends o
consider the L_1-norm error in the double logarithmic scale.

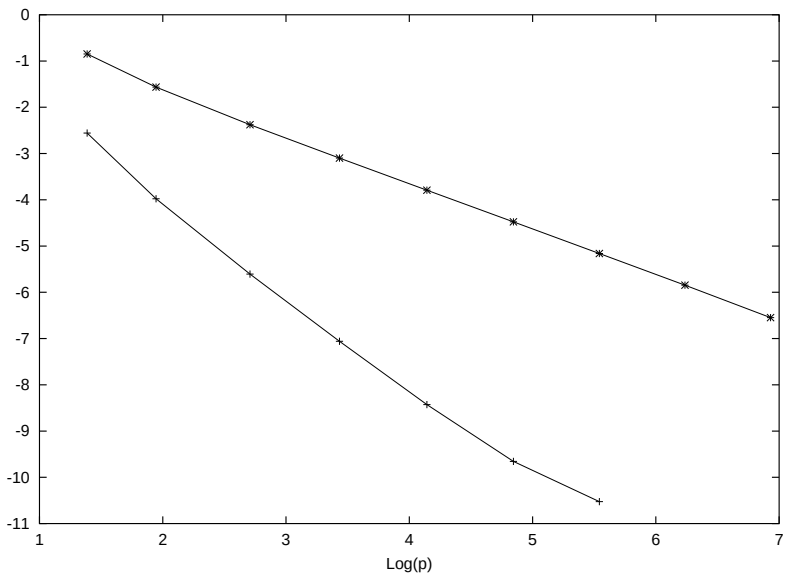


Figure 1: L_1-norm error vs p. Uniform grids: * . Chebyshev grids: $\times$. On x: $\ln p$, on y: $\ln((||u - u_p||_{L_1}/||u_p$
.

Numerical results support the following hypotheses:

(Uniform grids) $||u - u_p||_{L_1} \leq C_1 h ||u_p||_{L_1}$ (Chebyshev grids) $||u - u_p||_{L_1} \leq C_2 h^2 ||u_p||$

Numerical evaluation of the constants yields

$$C_1 = 1.64, C_2 = 1.06.$$

EXPERIMENTAL POINTWISE CONVERGENCE

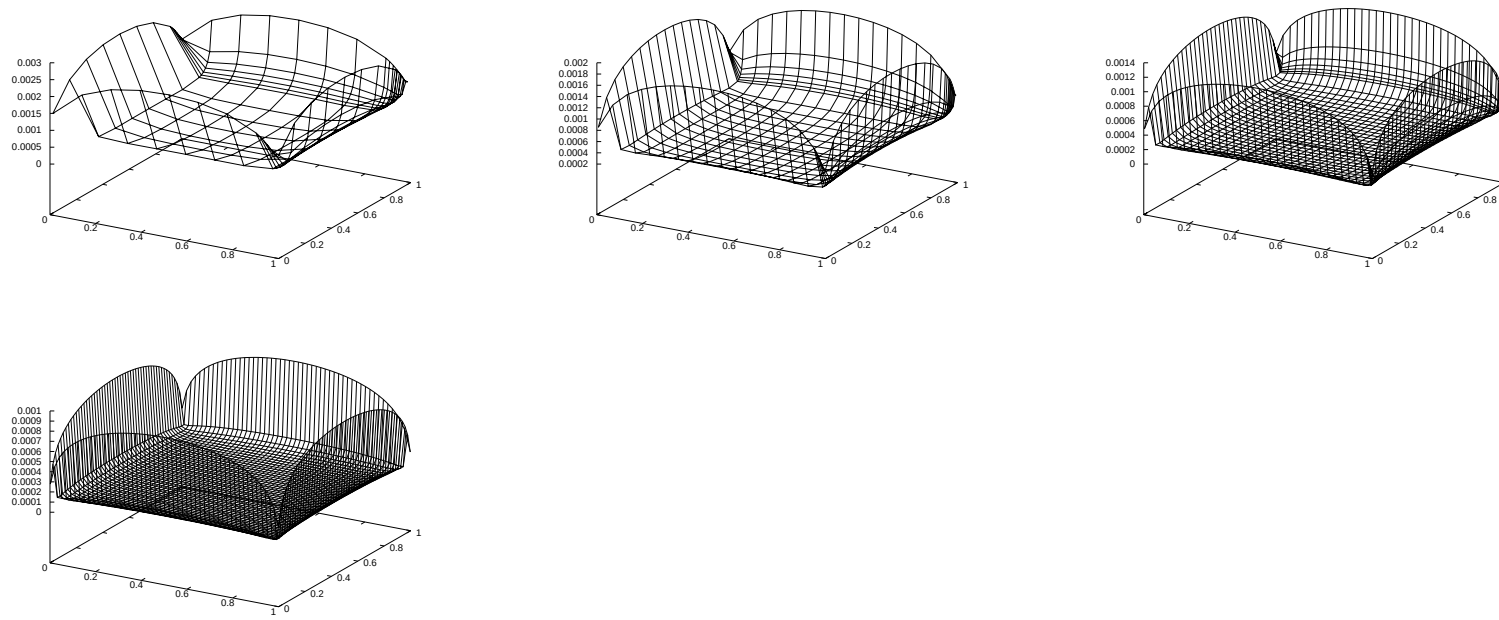


Figure 2: Pointwise error for uniform grids: p = 63,127,255,511

EXPERIMENTAL POINTWISE CONVERGENCE

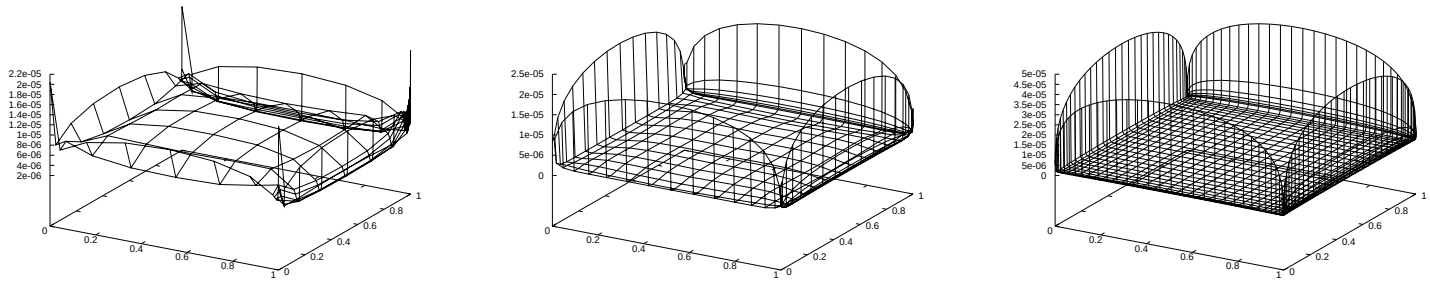


Figure 3: Pointwise error for Chebyshev grids: p = 63,127,255

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

NON-STANDARD WAVELET TRANSFORMS (nonuniform grids)

Given a grid x_i $(i = 1, \ldots, n + k + 1)$ and a subgrid $\widetilde{x}_i$ $(i = 1, \ldots, N + k + 1)$ w
$N < n$, introduce B-splines of order k:

$$B_i(x) = [x_i; \ldots; x_{i+k+1}](y - x)_+^k, \quad i = 1, \ldots, n,$$
$$\widetilde{B}_i(x) = [\widetilde{x}_i; \ldots; \widetilde{x}_{i+k+1}](y - x)_+^k, \quad i = 1, \ldots, N.$$

$$(z)_+ = \begin{cases} z, & \text{if } z \geq 0, \\ 0, & \text{otherwise} \end{cases}$$

Define V and $\widetilde{V}$ as the spans of B_i and $\widetilde{B}_i$ $\Rightarrow$ $\widetilde{V} \subset V$ $\Rightarrow$ $\widetilde{B}_i = \sum_s r_{is} B_s$.

(r_{is} are called *refinement* coefficie

Take *any space* W (wavelets) such that $\widetilde{V} + W = V$ and "improve" it using *the lifting scheme* of Sweldens:

$$\widetilde{\psi}_i = \sum_s \beta_{is} B_s - \sum_{j=j_{min}}^{j=j_{max}} \alpha_{ij} \widetilde{B}_j, \quad i = 1, \ldots, n - N.$$

The lifting coefficients α_{ij} must provide m vanishing moments:

$$\int \widetilde{\psi}_i x^p dx = 0, \quad p = 0, \ldots, m.$$

For consistency set $j_{max} = j_{min} + m$.

THEOREM. (I. Oseledets)

The lifting coefficients satisfying (M) read

$$\alpha_{ij} = \sum_s \beta_{is}[x_i; \ldots; x_{i+k+1}] P_j(x),$$

where $P_j(x)$ is a polynomial of degree $m+k+1$ such that

$$P_j(\widetilde{x}_r) = \begin{cases} q_j(\widetilde{x}_r), & j_{\min} \le r \le j, \\ 0, & j < r \le j_{\max} + k + 1, \end{cases}$$

$$q_j(x) = \begin{cases} (\widetilde{x}_j - \widetilde{x}_{j+k+1}) \prod_{l=j+1}^{j+k} (x - \widetilde{x}_l), & k \ge 1, \\ (\widetilde{x}_j - \widetilde{x}_{j+1}), & k = 0. \end{cases}$$

- J. M. Ford, I. V. Oseledets, E. E. Tyrtyshnikov, Matrix approximations and solvers using sor products and non-standard wavelet transforms related to irregular grids, *Rus. J. Nu Anal. and Math. Modelling*, Vol. 19, No. 2 (2004), 185-204.
- I. V. Oseledets, Application of divided differences and B-splines to construction of fast disc wavelet-like transforms on irregular grids, *Math. Notes*, in press (2004).

ALGORITHM.

Given

- a grid $x_i, \quad i = 1, \ldots, n+k+1,$
- a subgrid $\widetilde{x}_i, \quad i = 1, \ldots, N+k+1,$
- arrays $j_{min}(i), \quad i = 1, \ldots, n-N$, and $\beta_{is}, \; i = 1, \ldots, n-N, \; s = s_0, \ldots, s_1,$

calculate the lifting coefficients by the following FORTRAN-like code:

```
do i = 1, n - N
  do j = j_min(i), j_min(i) + m
    do r = j_min(i), j_min(i) + m + k + 1
      Calculate q_j(x̃_r).
    end do
    Calculate P_j(x) using Newton interpolation formulas.
    Calculate α_ij = Σ_{s=s_0}^{s_1} β_is [x_i; ...; x_{i+k+1}] P_j(x).
  end do
end do
```

The same algorithm may give the refinement coefficients because they satisfy similar equatio

$$\int \widetilde{B}_i x^p dx = \sum_{s=s_0}^{s_1} \int r_{is} B_s x^p dx, \quad p = 0, \ldots, s_1 - s_0 - 1.$$

CONSTRUCTION OF DISCRETE TRANSFORMS

$$\Phi_i = \sum_s r_{is}\phi_s, \quad \Psi_l = \sum_s \beta_{ls}\phi_s - \sum_j \alpha_{lj}\Phi_j \qquad (\phi \leftrightarrow B,\ \Phi \leftrightarrow \tilde{B},\ \Psi \leftrightarrow \tilde{\psi})$$

$$f = \sum_s v_s\phi'_s = \sum_i V_i\Phi'_i + \sum_l W_l\Psi'_l \quad \overset{\Longrightarrow}{\text{by duality}} \quad v_s = (f,\ \phi_s),\ V_i = (f,\ \Phi_i),\ W_l =$$

$$\begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} I & 0 \\ -A & I \end{bmatrix} \begin{bmatrix} R_1 & R_2 \\ 0 & I \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$[R_1\ R_2] = [r_{is}], \qquad R_1 = \text{diagonal}, \qquad A = [\alpha_{lj}]$$

EXAMPLES OF FUNCTION-RELATED MATRICES

$$a_{ij} = \begin{cases} 0.0 & \text{if } i = j \\ 1/|x_i - x_j|^p & \text{otherwise} \end{cases} \qquad x_i = 1 - \cos(i\pi/2n).$$

Consider the non-standard transforms with $k = 1$ and $m = 4$. Sparsification threshold is set to 10^{-6}. The computational comple transforming vectors and matrices using the non-standard wavelets is approx. the same (actually a little less) as for a Daubechies order 3 tran The storage for non-zeroes with non-standard wavelets is less than 1/3 of that using the Daubechies order 4 transform and only about 1/2 using Daubechies 3 wavelets. Figure 4 illustrates this by showing the Frobenius norm of the error (obtained by applying an inverse transform sparsified matrix and comparing with the original dense matrix), plotted against the compression rate (i.e. the fraction of non-zero entries sparse matrix).

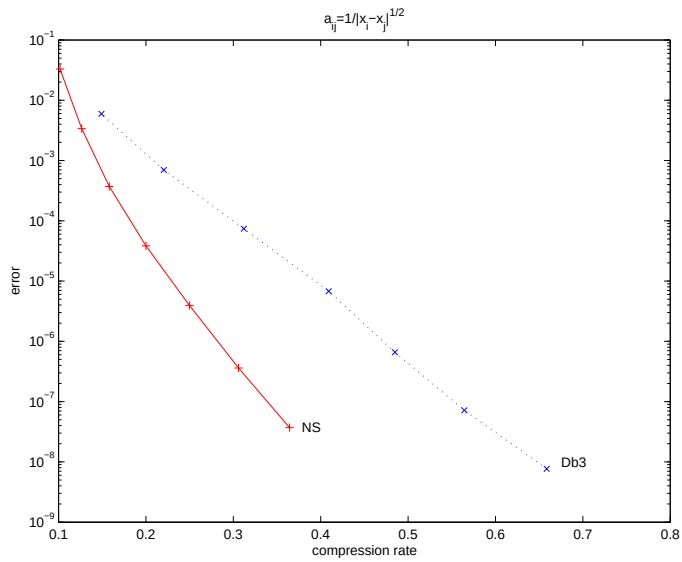


Figure 4: Frobenius norm a posteriori error plotted against compression ratio for function-related matrix from Example 1 with p using non-standard (NS) and Daubechies order 3 (Db3) wavelets.