Matrix and vector extrapolation methods for linear discrete ill-posed problems

Hassane SADOK

L.M.P.A
Université du Littoral - Calais

joint work with
A. Bouhamidi, K. Jbilou and L. Reichel

SIAM-LAA-2015

Table of contents

definition

Consider the linear system of equations

$$Ax = b$$

where A is a complex nonsingular $m \times m$ matrix and b is a given complex vector.

- For the GMRES method, the iterates $\{x_k\}$ are defined by the following conditions

GMRES

$$x_k - x_0 \in \mathcal{K}_k(A, r_0),$$

$$(A^i r_0, r_k) = 0 \qquad \text{for} \quad i = 1, \ldots, k,$$

Krylov matrix and GMRES

Let us first define the Krylov matrix

- $K_k = [r_0, \ldots, A^{k-1}\, r_0]$

So we have :

The residual norm of GMRES

1. $r_k^G = r_0 - W_k\,(W_k^H\, W_k)^{-1}\, W_k^H\, r_0 = r_0 - W_k\,(W_k)^{\dagger}\, r_0 = P_k^G\, r_0$

Properties :

1. $(P_k^G)^2 = P_k^G$
2. $(P_k^G)^H = P_k^G$
3. $||r_k^G|| = \min_{z \in K_k(A, r_0)} ||b - A\,(x_0 + z)||$

Krylov subspace Methods

Let Y_k be the matrix $Y_k = [y_1, \dots, y_k]$, we define
$(W_k)^L = (Y_k^H W_k)^{-1} Y_k^H$,

- $r_k^K = r_0 - W_k (Y_k^H W_k)^{-1} Y_k^H r_0 = r_0 - W_k (W_k)^L r_0 = P_k^K r_0$ and $x_k^K = x_0 + K_k (Y_k^H A K_k)^{-1} Y_k^H r_0$

Property

- $(P_k^K)^2 = P_k^K$

1. If $y_i = A^{i-1} r_0$, we obtain the Orthogonal Residula method (FOM, Arnoldi, Conjugate Gradient).
2. If $y_i = A^i r_0$, we obtain the Minimal Residual method (GMRES, Orthodir, Orthomin, GCR) .
3. If we set $y_i = {A^{i-1}}^H y$, we obtain the Lanczos method (BCG,).

Krylov matrix and Arnoldi process

We consider now the QR factorization of the Krylov matrix K_k.
Let V_k^G be an orthogonal matrix i.e. $(V^G)_k^H V_k^G = I_k$ and R_k be an upper triangular matrix of order k such that $K_k = V_k^G R_k$.
we have

- $A\, V_k^G = V_k^G\, H_k^G + h_{k+1,k}^G\, v_{k+1}^G\, (e_k^{(k)})^T = V_{k+1}^G\, H_{k+1,k}^G$.
- $r_k^G = r_0 - W_k\, (W_k^H\, W_k)^{-1}\, W_k^H\, r_0$
- $r_k^G = \|r_0\|\, V_{k+1}^G (I - H_{k+1,k}^G\, {H_{k+1,k}^G}^{\dagger})\, e_1$.
- $\|r_k^G\| = \|r_0\|\, \|(I - H_{k+1,k}^G\, {H_{k+1,k}^R}^{\dagger})\, e_1\|$.

Krylov matrix : Implicit QR factorization

We consider now the QR factorization of the Krylov matrix K_k.
Let V_k be an orthogonal matrix i.e. $V_k^H V_k = I_k$ and $\widetilde{R}_k$ be an upper triangular matrix of order k such that $K_k = V_k \widetilde{R}_k$.
Using the QR factorizations of the matrices K_k and K_{k+1} together, we get K_{k+1} and K_k, and the fact that

$$K_{k+1} \begin{bmatrix} 0 \\ I_k \end{bmatrix} = V_{k+1} \widetilde{R}_{k+1} \begin{bmatrix} 0 \\ I_k \end{bmatrix} = A K_k = A V_k \widetilde{R}_k,$$

The Arnoldi Algorithm

- Goal: to compute an orthogonal basis of $K_k(A, r_0)$.
- Input: Initial vector r_0, set $v_1 = \frac{1}{\|r_0\|} r_0$ and k.
- Arnoldi's procedure

For $j = 1, ..., k$ do

Compute $w := Av_j$

For $i = 1, \ldots, j$, do $\begin{cases} h_{i,j} := (w, v_i) \\ w := w - h_{i,j} v_i \end{cases}$

$h_{j+1,j} = \|w\|_2; \qquad v_{j+1} = w / h_{j+1,j}$

End.

Hessenberg process with pivoting strategy

1. $p = (1, 2, \cdots, n)^T$;
   Determine i_0 such that $|(l_1)_{i_0}| = \|r_0\|_\infty$;
   $\alpha = (l_1)_{i_0}$; $l_1 = r_0/\alpha$; $p_1 \longleftrightarrow p_{i_0}$;
2. for $k = 1, \cdots, m$

   $u = Al_k$;
   for $j = 1, \cdots, k$
   $h_{j,k} = (u)_{p_j}$; $(u)_{p_j} = 0$;
   $(u)_{p_j:p_n} = (u)_{p_j:p_n} - h_{j,k}(l_j)_{p_j:p_n}$;
   end
   Determine i_0 such that $|(u)_{p_{i_0}}| = \|(u)_{p_{k+1}:p_n}\|_\infty$;
   $h_{k+1,k} = (u)_{i_0}$; $l_{k+1} = u/h_{k+1,k}$; $p_{k+1} \longleftrightarrow p_{i_0}$;

   end

Solving linear systems

Consider the system of linear equations

$$Cx = f \tag{1}$$

where C is a real nonsingular $N \times N$ matrix, f is a vector of $\mathbf{R}^N$ and x^* denotes the unique solution.
Instead of applying the extrapolation methods for solving (1), we will use them for the preconditioned linear system

$$M^{-1} Cx = M^{-1} f$$

where M is a nonsingular matrix.

linearly generated Sequences

Starting from an initial vector s_0, we construct the sequence (s_j) by

$$s_{j+1} = Bs_j + b; \;\; j = 0, 1, \ldots \tag{2}$$

with $B = I - A$; $A = M^{-1}\, C$ and $b = M^{-1}\, f$.

We have

$$\Delta s_j = s_{j+1} - s_j = Bs_j + b - s_j = b - A\, s_j = r(s_j).$$

Note also that, since $\Delta^2 s_n = -A\, \Delta s_n$, we have $\Delta^2 S_k = -A\, \Delta S_k$, and

$$\Delta s_k = (I - A)^k r(s_0) \qquad \text{and} \qquad \Delta^k s_n = (-1)^{k-1} A^{k-1}\, \Delta s_n.$$

We deduce that
$span\{\Delta s_0, \Delta s_1, \ldots, \Delta s_{k-1}\} = span\{\Delta s_0, \Delta^2 s_0, \ldots, \Delta^k s_0\}$ and

$$span\{\Delta s_0, \Delta s_1, \ldots, \Delta s_{k-1}\} = span\{\Delta s_0, A\Delta s_0, \ldots, A^{k-1}\Delta s_0\}.$$

Consequently since $x_0 = s_0$, then

$$x_k = s_0 - \Delta S_k (\Delta^2 S_k)^{\dagger} \Delta s_0,$$

$$r(x_k) = b - Ax_k = \Delta s_0 - \Delta^2 S_k (\Delta^2 S_k)^{\dagger} \Delta s_0,$$

we deduce that for the GMRES method, the iterates $\{x_k\}$ are defined by the following conditions

GMRES

$$x_k - s_0 \in span\{\Delta s_0, \Delta s_1, \ldots, \Delta s_{k-1}\},$$

$$\Delta^2 S_k^T r(x_k) = 0.$$

Polynomials methods

Let $\{s_n\}_{n\geq 0}$ be a sequence of vectors in $\mathbb{R}^N$, and define the first and the second forward differences

$$\Delta s_n := s_{n+1} - s_n \quad \text{and} \quad \Delta^2 s_n := \Delta s_{n+1} - \Delta s_n.$$

When applied to the sequence $\{s_n\}_{n\geq 0}$, the polynomials vector extrapolation methods MPE, RRE, and MMPE produce approximations $t_n^{(q)}$ of the limit or antilimit of the s_n as $n \to \infty$ of the form

$$t_n^{(q)} := \sum_{j=0}^{q} \gamma_n^{(j)} s_{n+j},$$

where

$$\sum_{j=0}^{q} \gamma_n^{(j)} = 1, \quad \text{and} \quad \sum_{j=0}^{q} \eta_{ij}^{(n)} \gamma_n^{(j)} = 0, \quad 0 \leq i < q, \qquad (3)$$

Convergence of RRE : $y_{i+1}^{(n)} := \Delta^2 s_{n+i}$

We have

RRE method

$$t_n^{(q)} = s_n - \Delta S_{q,n}(\Delta^2 S_{q,n})^\dagger \Delta s_n,$$

If we consider a vector sequence such that
$s_n = s + \lambda_1^n v_1 + \lambda_2^n v_2 + \ldots + \lambda_k^n v_k + \ldots + \lambda_m^n v_m$, where
$0 \leq |\lambda_m| \leq \ldots < |\lambda_1|$ and $|\lambda_{k+1}| < |\lambda_k|$ then

$$t_n^{(k)} = s + O((\lambda_{k+1})^n).$$

Implementation of RRE : $y_{i+1}^{(n)} := \Delta^2 s_{n+i}$

We set $n = 0$ and we denote the matrices $\Delta^i S_{q,0}$ by $\Delta^i S_q$, $1 \leq i \leq 2$, and the vectors $y_q^{(0)}$ and $t_0^{(q)}$ by y_q and t_q, respectively. Then

$$t_q = s_0 - \Delta S_q (\Delta^2 S_q^T \Delta^2 S_q)^{-1} \Delta^2 S_q^T \Delta s_0,$$

The system of equations (3) can be written as

$$\begin{cases} \gamma_0^{(0)} + \ldots + \gamma_q^{(0)} = 1 \\ \gamma_0^{(0)}(\Delta^2 s_0, \Delta s_0) + \ldots + \gamma_q^{(0)}(\Delta^2 s_0, \Delta s_q) = 0 \\ \gamma_0^{(0)}(\Delta^2 s_1, \Delta s_0) + \gamma_q^{(0)}(\Delta^2 s_1, \Delta s_q) = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \gamma_0^{(0)}(\Delta^2 s_{q-1}, \Delta s_0) + \ldots + \gamma_q^{(0)}(\Delta^2 s_{q-1}, \Delta s_q) = 0 \end{cases}$$

Assume now that $\gamma_0^{(0)}, \gamma_1^{(0)}, \ldots, \gamma_q^{(0)}$ have been calculated, and introduce the new variables

$$\alpha_0^{(0)} = 1 - \gamma_0^{(0)}, \quad \alpha_j^{(0)} = \alpha_{j-1}^{(0)} - \gamma_j^{(0)}, \quad 1 \leq j < q, \text{ and } \alpha_{q-1}^{(0)} = \gamma_q^{(0)},$$

so that the vector t_q can be expressed as

$$t_q = s_0 + \sum_{j=0}^{q-1} \alpha_j^{(0)} \Delta s_j = s_0 + \Delta S_{q-1}\, \alpha^{(q)},$$

where $\alpha^{(q)} = [\alpha_0^{(0)}, \ldots, \alpha_{q-1}^{(0)}]^T$.

In order to determine the $\gamma_i^{(0)}$, we first have to compute the $\beta_i^{(0)}$ by solving the nonsingular linear system of equations (4).

Solving non linear systems

Consider the system of nonlinear equations

$$G(x) = x \tag{5}$$

where $G: \mathbf{R}^N \Longrightarrow \mathbf{R}^N$ and let x^* be a solution of (5).
For any arbitrary vector x, the residual is defined by

$$r(x) = G(x) - x.$$

Let $(s_j)_j$ be the sequence of vectors generated from an initial guess s_0 as follows

$$s_{j+1} = G(s_j),\ j = 0, 1, \ldots. \tag{6}$$

Note that

$$r(s_j) = \Delta s_j,\ j =, 1, \ldots$$

In practice, it is recommended to restart the algorithms after a fixed number of iterations. Another important remark is the fact that the extrapolation methods are more efficient if they are applied to a preconditioned nonlinear system

$$\tilde{G}(x) = x$$

where the function $\tilde{G}$ is obtained from G by some preconditioning nonlinear technique.

Vector extrapolation for non linear system

An extrapolation algorithm for solving the nonlinear problem is summarized as follows

1- $k = 0$, choose x_0 and the integers p and m.

2- Basic iteration

set $t_0 = x_0$

$w_0 = t_0$

$w_{j+1} = \tilde{G}(w_j),\ j = 0, \ldots, p-1.$

3- Extrapolation phase

$s_0 = w_p;$

if $\| s_1 - s_0 \| < \epsilon$ stop;

otherwise generate $s_{j+1} = \tilde{G}(s_j),\ j = 0, \ldots, m,$

compute the approximation t_m by RRE, MPE or MMPE;

4- set $s_0 = t_m$, $k = k + 1$ and go to 2.

Numerical example

We consider now the following nonlinear partial differential equation

$$-u_{xx} - u_{yy} + 2p_1 u_x + 2p_2 u_y - p_3 u + 5e^{u(x,y)} = \phi(x,y) \quad \text{on } \Omega$$

$$u(x,y) = 1 + xy \quad \text{on } \partial\Omega,$$

over the unit square of $\mathbb{R}^2$ with Dirichlet boundary condition. This problem is discretized by a standard five-point central difference formula with uniform grid of size $h = 1/(n+1)$. We get the following nonlinear system of dimension $N \times N$, where $N = n^2$,

$$AX + 5e^X - b = 0. \tag{5.4}$$

The right hand-side function $\phi(x,y)$ was chosen so that the true solution is $u(x,y) = 1 + xy$ in Ω

The sequence (s_j) is generated by using the nonlinear SSOR method. Hence we have $s_{j+1} = G(s_j)$, where

$$G(X) = B_\omega X + \omega(2-\omega)(D-\omega U)^{-1} D (D-\omega L)^{-1}(b - 5e^X),$$

the matrix

$B_\omega = (D-\omega U)^{-1}(\omega L + (1-\omega)D)(D-\omega L)^{-1}(\omega U + (1-\omega)D)$

and $A = D - L - U$, the classical splitting decomposition.

The stopping criterion was $||x_k - G(x_k)|| < 10^{-8}$.

In our tests, we choose $n = 72$, $p_1 = 1$, $p_2 = 1$, $p_3 = 10$, $N = 4900$.

With $m = 20$ and $\omega = 0.5$, we obtain the results of Table 3.

Table 3

Method	MMPE	MPE	RRE
Number of restarts	20	18	19
residual norms	2.9d-09	9.2d-08	2.8d-08

Stopping Criterion

We need to evaluate $\|t_{k+1} - t_k\|$ for our stopping criterion. From the last formula we deduce that

$$t_{k+1} - t_k = \sum_{j=1}^{k} \frac{(\alpha_{j-1}^{(k+1)} - \alpha_{j-1}^{(k)})}{\sqrt{\delta_j}} v_j + \frac{\alpha_k^{(k+1)}}{\sqrt{\delta_{k+1}}} v_{k+1}.$$

Since the vectors v_j, $1 \leq j \leq k+1$, are orthonormal, it follows that

$$\|t_{k+1} - t_k\| = \sqrt{\sum_{j=1}^{k} \frac{|\alpha_{j-1}^{(k+1)} - \alpha_{j-1}^{(k)}|^2}{\delta_j} + \frac{|\alpha_k^{(k+1)}|^2}{\delta_{k+1}}}.$$

RRE-TSVD algorithm

The RRE-TSVD algorithm is summarized as follows:

The RRE-TSVD algorithm

- Compute the SVD of the matrix A:
  $[U, \Sigma, V] = svd(A)$.
  Set $s_0 = 0$, $s_1 = \frac{u_1^T b}{\sigma_1} v_1$, and $t_1 = s_1$, with
  $u_i = U(:, i)$ and $v_i = V(:, i)$ for $i = 1, \ldots, n$.
- For $k = 2, \ldots, n$
  1. Compute s_k .
  2. Compute the $\gamma_i^{(k)}$ and $\alpha_i^{(k)}$ for $i = 0, \ldots, k-1$.
  3. Form the approximation t_k.
  4. If $\|t_k - t_{k-1}\| / \|t_{k-1}\| < tol$, stop.
- End

We consider linear discrete ill-posed problems of the form

$$A_1 X A_2^T = B, \tag{7}$$

where at least one of the matrices $A_1, A_2 \in \mathbf{R}^{n \times n}$ is of ill-determined rank.
The right-hand side $B \in \mathbf{R}^{n \times n}$ represents observations that are contaminated by measurement errors, i.e.,

$$B = \widetilde{B} + E, \tag{8}$$

where $\widetilde{B}$ denotes the unavailable error-free right-hand side.
The norm of the error E is not assumed to be known.

Discrete ill-posed problems of the form (7) arise from the discretization of Fredholm integral equations of the first kind in two space-dimensions,

$$\iint_{\Omega} K(x, y, s, t) f(s, t) ds dt = g(x, y), \qquad (x, y) \in \Omega', \qquad (9)$$

where Ω and Ω' are rectangles in $\mathbb{R}^2$ and the kernel is separable,

$$K(x, y, s, t) = k_1(x, s)\, k_2(y, t), \qquad (x, y) \in \Omega', \qquad (s, t) \in \Omega.$$

Discretization of (9) gives a matrix equation of the form (7).

Definition of Matrix Extrapolation Methods

Let (S_p) be a sequence of matrices in $\mathbf{R}^{N\times s}$ and consider the transformation T_q, $q \geq 1$ defined by

$$T_q : \mathbf{R}^{N\times s} \longrightarrow \mathbf{R}^{N\times s}$$

$$S_p \to T_q^{(p)}$$

with

$$T_q^{(p)} = S_p + \sum_{i=1}^{q} \mathbf{a_i^{(p)}}\, G_i(p),\ p \geq 0$$

where the auxiliary sequences $(G_i(p))_p$; $i = 1, \ldots, q$ are given.
Let $\tilde{T}_q$ denotes the new transformation obtained from T_q as follows

$$\tilde{T}_q^{(p)} = S_{p+1} + \sum_{i=1}^{q} \mathbf{a_i^{(p)}}\, G_i(p+1),\ p \geq 0.$$

We define the generalized residual of $T_q^{(p)}$ by

$$\tilde{R}(T_q^{(p)}) = \tilde{T}_q^{(p)} - T_q^{(p)} = \Delta S_p + \sum_{i=1}^{q} a_i^{(p)} \Delta G_i(p).$$

The coefficients $a_i^{(p)}$ are obtained from the orthogonality relation

$$\tilde{R}(T_q^{(p)}) \perp_F \; span\{Y_1^{(p)}, \ldots, Y_q^{(p)}\}$$

where $\perp_F$ means the orhtogonality with respect to the Frobenius inner product.

- $G_i(p) = \Delta S_{p+i-1} = S_{p+i} - S_{p+i-1}$
- $Y_i^{(p)} = \Delta S_{p+i-1}$: Matrix Minimal Poly.Extrapolation (MPE)
- $Y_i^{(p)} = \Delta^2 S_{p+i-1}$: Matrix Reduced Rank Extrapolation (RRE)
- $Y_i^{(p)} = Y_i$: Matrix Modified MPE (MMPE) (Pugachev

If we set

- $\widetilde{\mathbb{V}}_{q,p} = span\{\Delta S_p, \ldots, \Delta S_{p+q-1}\}$
- $\widetilde{\mathbb{W}}_{q,p} = span\{\Delta^2 S_p, \ldots, \Delta^2 S_{p+q-1}\}$ and
- $\widetilde{\mathbb{Y}}_{q,p} = span\{Y_1^{(p)}, \ldots, Y_q^{(p)}\}$

then we have the following relations

$$\begin{cases} \tilde{R}(T_q^{(p)}) - \Delta S_p \in \widetilde{\mathbb{W}}_{q,p} \\ \tilde{R}(T_q^{(p)}) \perp_F \widetilde{\mathbb{Y}}_{q,p}. \end{cases}$$

$\perp_F$ means the orthogonality with respect to the Frobenius inner product.
$\tilde{R}(T_q^{(p)})$ is obtained from an oblique projection.

This gives the following expression

$$\widetilde{R}(T_p^{(q)}) = \Delta S_p + \mathbb{W}_{q,p}(\alpha^{(p)} \otimes I_s), \tag{10}$$

and

$$\alpha^{(p)} = -(\mathbb{Y}_{q,p}^T \diamond \mathbb{W}_{q,p})^{-1} \, (\mathbb{Y}_{q,p}^T \diamond \Delta S_p).$$

The approximation $T_p^{(q)}$ is given by

$$T_p^{(q)} = S_p + \mathbb{V}_{q,p}(\alpha^{(p)} \otimes I_s).$$

Let $A = [A_1, A_2, \ldots, A_p]$ and $B = [B_1, B_2, \ldots, B_l]$ be matrices of dimension $n \times ps$ and $n \times ls$ respectively where A_i and B_j $(i = 1, \ldots, p; j = 1, \ldots, l)$ are $N \times s$ matrices. Then the $p \times l$ matrix $A^T \diamond B$ is defined by:

$$A^T \diamond B = \begin{pmatrix} \langle A_1, B_1 \rangle_F & \langle A_1, B_2 \rangle_F & \ldots & \langle A_1, B_l \rangle_F \\ \langle A_2, B_1 \rangle_F & \langle A_2, B_2 \rangle_F & \ldots & \langle A_2, B_l \rangle_F \\ \vdots & \vdots & \vdots & \vdots \\ \langle A_p, B_1 \rangle_F & \langle A_p, B_2 \rangle_F & \ldots & \langle A_p, B_l \rangle_F \end{pmatrix},$$

where

$$\langle A_i, B_j \rangle_F = trace(A_i^T B_j).$$

The matrx A is F-orthonormal if

$$A^T \diamond A = I.$$

Some properties of the $\diamond$ product:

Let $A, B, C \in \mathbf{R}^{N\times ps}$, $D \in \mathbf{R}^{N\times N}$, $L \in \mathbf{R}^{p\times p}$. Then we have

1. $(A^T \diamond B)^T = B^T \diamond A$.
2. $(DA)^T \diamond B = A^T \diamond (D^T B)$.
3. $A^T \diamond (B(L \otimes I_s)) = (A^T \diamond B)L$.
4. If $s = 1$ then $A^T \diamond B = A^T B$.
5. If $X \in \mathbf{R}^{N\times s}$, then $X^T \diamond X = \|X\|_F^2$.

Let $\mathcal{T}_q^{(p)}$ be the matrix given by

$$\mathcal{T}_q^{(p)} = \begin{pmatrix} S_p & \mathbb{V}_{q,p} \\ (\mathbb{Y}_{q,p}^T \diamond \Delta S_p) \otimes I_s & (\mathbb{Y}_{q,p}^T \diamond \mathbb{W}_{q,p}) \otimes I_s. \end{pmatrix}. \qquad (11)$$

The approximation $T_q^{(p)}$ is then expressed as the Schur complement

$$T_q^{(p)} = \left(\mathcal{T}_q^{(p)} / \ (\mathbb{Y}_{q,p}^T \diamond \mathbb{W}_{q,p}) \otimes I_s) \right).$$

With

- $\mathbb{V}_{q,p} = [\Delta S_p, \ldots, \Delta S_{p+q-1}]$
- $\mathbb{V}_{q,p} = [\Delta^2 S_p, \ldots, \Delta^2 S_{p+q-1}]$
- $\mathbb{Y}_{q,p} = [Y_1^{(p)}, \ldots, Y_q^{(p)}]$
- $Y_i^{(p)} = \Delta S_{p+i-1}$: Matrix Minimal Poly. Extrapolation (M-MPE)
- $Y_i^{(p)} = \Delta^2 S_{p+i-1}$: Matrix Reduced Rank Extrapolation (M-RRE)
- $Y_i^{(p)} = Y_i$: Matrix Modified MPE (MMPE)

Extrapolating the TSVD sequence by the matrix Reduced Rank Extrapolation method

The TSVD of a Kronecker product

Let the matrices A_1 and A_2 in (7) have the singular value decompositions

$$A_1 = U_1 \Sigma_1 V_1^T, \qquad A_2 = U_2 \Sigma_2 V_2^T,$$

respectively.

$$U_k = [u_{1,k}, u_{2,k}, \ldots, u_{n,k}] \in \mathbb{R}^{n \times n}, \qquad V_k = [v_{1,k}, v_{2,k}, \ldots, v_{n,k}] \in \mathbb{R}^{n \times n}$$

and

$$\Sigma_k = \mathrm{diag}[\sigma_{1,k}, \sigma_{2,k}, \ldots, \sigma_{n,k}] \in \mathbb{R}^{n \times n}$$

with

$$\sigma_{1,k} \geq \sigma_{2,k} \geq \ldots \geq \sigma_{n,k} \geq 0, \qquad k = 1, 2;$$

The singular value decomposition of the matrix A, defined by

$$A = A_2 \otimes A_1$$

is given by

$$A = U\Sigma V^T.$$

$$U = U_2 \otimes U_1,\ V = V_2 \otimes V_1, \text{ and } \Sigma = \Sigma_2 \otimes \Sigma_1, \text{ where}$$

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\ell_0} > \sigma_{\ell_0+1} = \ldots = \sigma_{n^2} = 0. \tag{12}$$

with

$$\sigma_\ell = \sigma_{j(\ell),2}\sigma_{i(\ell),1}, \qquad 1 \leq \ell \leq n^2, \tag{13}$$

$i(\ell)$ and $j(\ell)$ are nondecreasing functions of ℓ with range $\{1, 2, \ldots, n\}$. The columns of the orthogonal matrices U and V are given by

$$u_\ell = u_{j(\ell),2} \otimes u_{i(\ell),1}, \qquad v_\ell = v_{j(\ell),2} \otimes v_{i(\ell),1}, \qquad 1 \leq \ell \leq n^2; \tag{14}$$

The rank-k approximation $\widetilde{A}_k$ of $A = A_2 \otimes A_1$ is defined by

$$\widetilde{A}_k = \sum_{\ell=1}^{k} \sigma_\ell u_\ell v_\ell^T;$$

and its Moore-Penrose pseudoinverse can be expressed as

$$\widetilde{A}_k^\dagger = \sum_{\ell=1}^{k} \sigma_\ell^{-1} v_\ell u_\ell^T.$$

The Matrix RRE for TSVD sequences

Consider the least-squares problem

$$\min_X \|A_1 X A_2^T - B\|_F. \tag{15}$$

The minimal-norm solution of (15) is given by

$$S_p = \sum_{\ell=1}^{p} \frac{\left(u_{i(\ell),1}^T B u_{j(\ell),2}\right)}{\sigma_{i(\ell),2}\sigma_{j(\ell),1}} (v_{i(\ell),1} v_{j(\ell),2}^T) = \sum_{l=1}^{p} \delta_l V_l.$$

The vector $s_p = vec(S_p)$ is given by

$$s_p = \sum_{\ell=1}^{p} \frac{\left(u_{j(\ell),2}^T \otimes u_{i(\ell),1}^T\right) b}{\sigma_{j(\ell),2}\sigma_{i(\ell),1}} v_{j(\ell),2} \otimes v_{i(\ell),1}. \tag{16}$$

where $b = vec(B)$.

- Thus, S_p is the p-TSVD approximate solution of the original matrix problem .
- It is important to choose a suitable value of the truncation index p. This task is much simplified by **extrapolating** the sequence $\{S_p\}_{p\geq 0}$ before selecting an index.
- Here, we applied the matrix RRE (M-RRE) method.

Starting with S_0, $(p = 0)$ the new matrix sequence (T_k)

$$T_k = T_k^{(0)} = S_0 - \Delta\mathcal{S}_{k-1}(\alpha^{(k)} \otimes I_s),$$

with
$\Delta\mathcal{S}_{k-1} = [\Delta S_0, \ldots, \Delta S_{k-1}]$; $\Delta^2\mathcal{S}_{k-1} = [\Delta^2 S_0, \ldots, \Delta^2 S_{k-1}]$.

The vector $\alpha^{(k)} = (\alpha_1^{(k)}, \ldots, \alpha_k^{(k)})$ solves the linear system of equations

$$(\Delta^2\mathcal{S}_{k-1}{}^T \diamond \Delta^2\mathcal{S}_{k-1})\,\alpha^{(k)} = -\Delta^2\mathcal{S}_{k-1}{}^T \diamond \Delta S_0$$

Using the fact that $\Delta S_{j-1} = \delta_j V_j$, the matrix $\Delta \mathcal{S}_{k-1}$ can be factored according to

$$\Delta \mathcal{S}_{k-1} = [\delta_1 V_1, \ldots, \delta_k V_k] = \mathcal{V}_k \, (\mathrm{diag}[\delta_1, \ldots, \delta_k] \otimes I_s),$$

and

$$\Delta^2 \mathcal{S}_{k-1} = \mathcal{V}_{k+1} \, (\begin{bmatrix} -\delta_1 & & & \\ \delta_2 & -\delta_2 & & \\ & \ddots & \ddots & \\ & & \delta_k & -\delta_k \\ & & & \delta_{k+1} \end{bmatrix} \otimes I_s). \qquad (17)$$

Since $\mathcal{V}_{k+1}$ is F-orthogonal ($\mathcal{V}_{k+1}^T \diamond \mathcal{V}_{k+1} = I$), it follows that

$$\Delta^2 \mathcal{S}_{k-1}^T \diamond \Delta^2 \mathcal{S}_{k-1} = tridiag(-\delta_i^2, \delta_i^2 + \delta_{i+1}^2, -\delta_{i+1}^2)$$

the extrapolated matrix T_k is expressed as

$$T_k = \sum_{\ell=1}^{k} \alpha_\ell^{(k)} \frac{u_{i(\ell),1}^T \, B \, u_{j(\ell),2}}{\sigma_{i(\ell),1}\sigma_{j(\ell),2}} \, v_{i(\ell),1} v_{j(\ell),2}^T. \qquad (18)$$

where $\alpha^{(k)}$ is the solution of the linear system of equations

$$(\Delta^2 \mathcal{S}_{k-1}^T \diamond \Delta^2 \mathcal{S}_{k-1})\alpha^{(k)} = -\Delta^2 \mathcal{S}_{k-1}^T \diamond \Delta S_0 = \delta_1^2 [1, 0, \ldots, 0]^T$$

The expression (18) shows that the matrix RRE method acts as a **filter** on the TSVD sequence.

The expression $\| T_{k+1} - T_k \|$ can be helpful to determine when to terminate the computations. We have the following relation

$$\| T_{k+1} - T_k \|_F = \sqrt{\sum_{j=1}^{k} \frac{|\alpha_{j-1}^{(k+1)} - \alpha_{j-1}^{(k)}|^2}{\delta_j} + \frac{|\alpha_k^{(k+1)}|^2}{\delta_{k+1}}},$$

In the computed examples, we also used the norm of the generalized residual $\tilde{R}_k$, to determine a suitable truncation index.

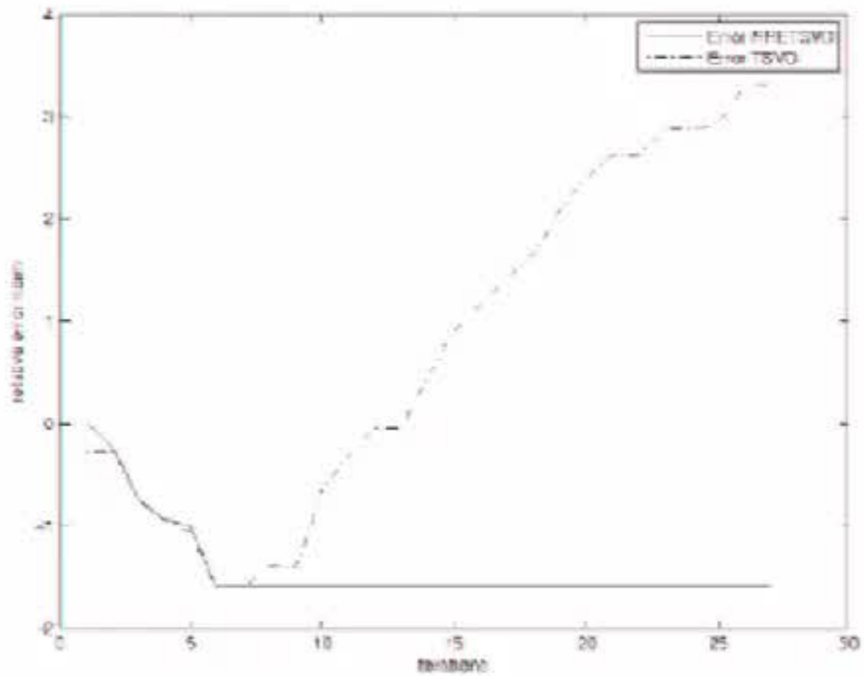
This norm easily can be evaluated using

$$\| \tilde{R}_k \|_F = \frac{1}{\sqrt{\sum_{j=0}^{k} \frac{1}{\delta_{j+1}^2}}}.$$

Example 1. The nonsymmetric matrices $A_1, A_2 \in \mathbb{R}^{1500\times1500}$ are:

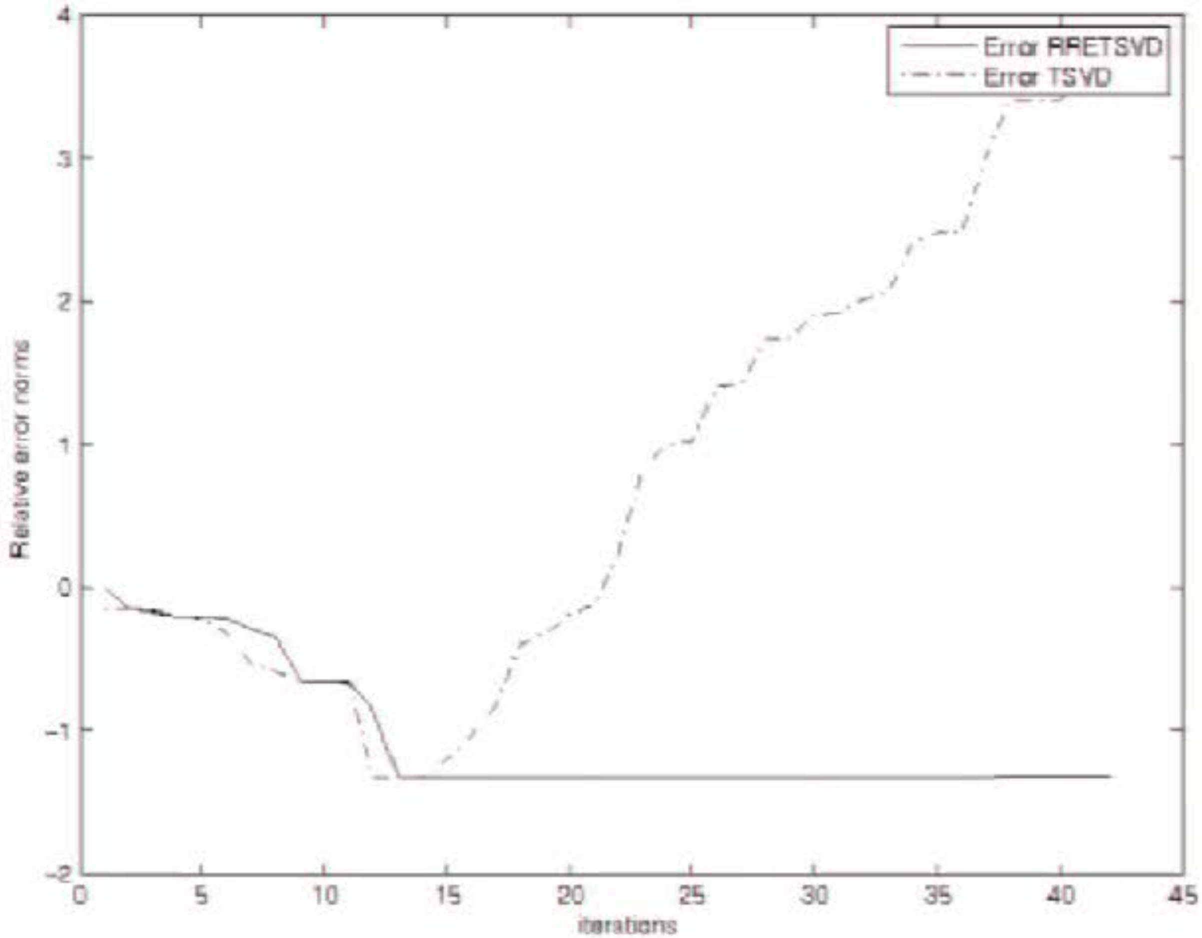
$$A_1 = \texttt{baart(1500)} \text{ and } A_2 = \texttt{foxgood(1500)}.$$

$\kappa(A_1) = 2\cdot10^{18}$ and $\kappa(A_2) = \cdot10^{13}$.
The noise-level in the right-hand side is $\nu = 1.2\cdot10^{-2}$.



We remark that is much easier to determine an accurate approximation of $\hat{X}$ from the extrapolated sequence $\{T_k\}_{k>0}$ than from the sequence $\{S_k\}_{k>0}$; it suffices to choose $k \geq 6$.

Example 2. $A_1 = baart(n)$ and $A_2 = usrsell(n)$; $n = 2000$.



Example 3. In this example, we consider the Fredholm integral equation

$$\iint_{\Omega} K(x,y,s,t) f(s,t)\, ds dt = g(x,y), \qquad (x,y) \in \Omega', \quad (19)$$

where $\Omega = [0, \pi/2] \times [0, \pi/2]$ and $\Omega' = [0, \pi] \times [0, \pi]$. Let the kernel be given by

$$K(x,y,s,t) = k_1(x,s)\, k_2(y,t), \quad (x,y) \in \Omega', \quad (s,t) \in \Omega,$$

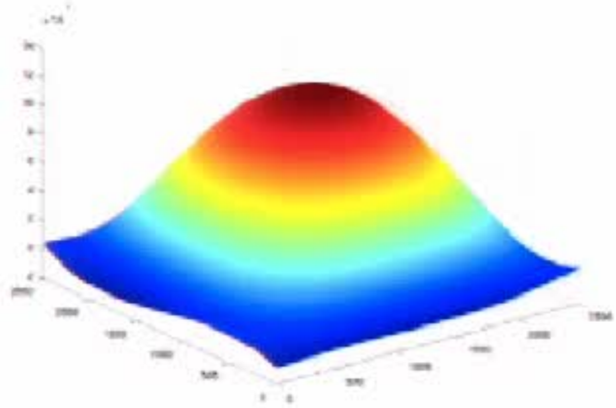
and define

$$g(x,y) = g_1(x)\, g_2(y),$$

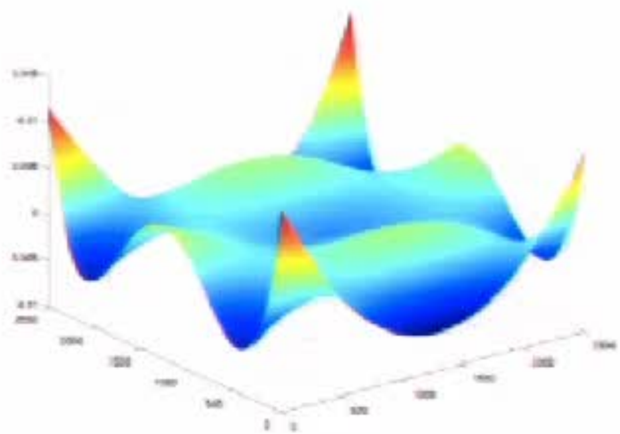
where

$$k_i(s,x) = \exp(s\cos(x)), \qquad g_i(s) = 2\sinh(s)/s, \qquad i = 1,2.$$

We obtain two matrices $A_1, A_2 \in \mathbb{R}^{2500\times 2500}$ and a scaled approximation $\hat{X}$ of the exact solution $f(t,s) = \sin(t)\sin(s)$. The error-free right-hand side of (7) is determined by $\widetilde{B} = A_1\, \hat{X}\, A_2^T$. Adding an error with noise-level $\nu = 1\cdot 10^{-2}$, we obtain the right-hand B.



Approximation T_{23} by the matrix RRE-TSVD method.



Approximation S_{23} determined by the TSVD method

Matrix extrapolations and Tikhonov regularization

Here, we consider the Tikhonov regulrization problem

$$\min_{X}(\|A_1 X A_2^T - B\|_F^2 + \lambda^2 \|X\|_F{}^2). \quad (20)$$

where λ is a parameter to be chosen. The problem (20) is equivalent to solving the nonsymmetric Stein matrix equation

$$\mathcal{A}X\mathcal{C} - X + \mathcal{F} = 0.$$

where $\mathcal{A} = A_1^T A_1$, $\mathcal{C} = (1/\lambda^2) A_2^T A_2$, $\mathcal{F} = -(1/\lambda^2) A_1^T B A_2$.

If the eigenvalues of $\mathcal{A}$ and $\mathcal{C}$ are inside the unit disc, the solution X could be expressed as

$$X = \sum_{i=0}^{\infty} \mathcal{A}^i \mathcal{F} \mathcal{C}^i$$

Then, we generate the following matrix Smith iteration

$$S_0 = 0; \;\; S_j = \mathcal{F} + \mathcal{A} S_{j-1} \mathcal{C}.$$

or the Squared Smith iteration defined as

$$S_0 = 0; \;\; S_j = S_{j-1} + \mathcal{A}_{j-1} S_{j-1} \mathcal{C}_{j-1}; \; \mathcal{A}_j = \mathcal{A}_{j-1}^2; \mathcal{C}_j = \mathcal{C}_{j-1}^2.$$

As the convergence of the Smith iteration is very slow, we can apply the Matrix RRE extrapolation method to the sequence (S_j).

Example 4.

- The original image is denoted by $\widehat{X}$;
- The vector $\widehat{B} = A_1 \widehat{X} A_2^T$ represents the associated blurred and noise-free image.
- We generated a blurred and noisy image: $B = \widehat{B} + \mathbf{N}$, where $\mathbf{N}$ is a noise chosen such that $||\mathbf{N}||/||\widehat{B}|| = 10^{-2}$.

The blurring matrix A is given by $A = A_2 \otimes A_1 \in \mathbb{R}^{256^2 \times 256^2}$, where $A_1 = A_2 = [a_{ij}]$: Toeplitz matrix given by

$$a_{ij} = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(i-j)^2}{2\sigma^2}\right), & |i-j| \leq r, \\ 0, & \text{otherwise.} \end{cases}$$

We used: $r = 4$ and $\sigma = 5$;
$\lambda_{opt} = 0.0014586$ (computed by the GCV method). The restored image corresponds to the approximation T_2 (with matrix RRE).