Accurate linear algebra (ALA) in computational methods for system and control theory

Z. Drmač

University of Zagreb

Research supported by Croatian Science Fundation under HRZZ-9345

drmac@math.hr

SIAM Conference on Applied Linear Algebra, Atlanta

October 29. 2015.

Overview

1. Introduction - motivating examples and goals
2. Preliminaries: accurate computation with ill-conditioned matrices
   - Accurate SVD and scaling invariant condition number
   - Rank revealing (pivoted) QR factorization
   - An interesting connection: RRQR and DIME
   - PSVD, RRD and Cauchy and Vandermonde SVD
3. Hankel matrices
   - Mission impossible
   - Exploiting Vandermonde product representation
4. Case study: Matrix valued rational LS approximation
   - Sanathanan–Koerner iterations and Vector Fitting. $\mathcal{H}_2$ MOR
   - Details of the VF algorithm
   - Accurate LS for more robust VF
5. Concluding remarks

$\mathcal{H}_2$, L_2 rational matrix valued approximations

Suppose the dynamics of an n–dimensional LTI stable system

$$\begin{aligned} \dot{x}(t) &= \mathbf{A}x(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \iff \mathbf{H}(s) = C(sI - \mathbf{A})^{-1}B$$

is inaccessible to direct modeling, but the input-output relationships may be observed in the frequency domain yielding $\mathbf{H}(\xi_i) \in \mathbb{C}^{p\times m}$ for selected $\xi_j \in \mathbf{i}\mathbb{R}$, $j = 1, \ldots, \ell$. The task is to construct an empirical dynamical system model represented as a stable matrix-valued rational function $\mathbf{H}_r(s)$ that fits the measured frequency response data $\mathbf{H}(\xi_i)$.

Problem: Minimize over $\mathcal{R}_r$

$$\min_{\mathbf{H}_r \in \mathcal{R}_r} \sum_{i=1}^{\ell} \rho_i \left\| \mathbf{H}_r(\xi_i) - \mathbf{H}(\xi_i) \right\|_F^2 \ , \quad \mathbf{H}_r \text{ stable.}$$

$\mathcal{R}_r$ consists of $p \times m$ matrix valued functions with entries that are strictly proper rational functions of McMillan degree r.

Basic Iterations (Sanathanan–Koerner, Kalman)

Sanathanan–Koerner (SK) Iterations

Compute a sequence of $\mathbf{H}_r^{(k)}(s) = \mathbf{N}^{(k)}(s)/d^{(k)}(s)$, where $\mathbf{N}^{(k)}(s)$ is a $p \times m$ matrix of polynomials of degree $r-1$ or less and $d^{(k)}(s)$ is a (scalar-valued) polynomial of degree r. Iterate for $k = 0, 1, 2, \ldots$

$$\epsilon^{(k)} = \sum_{i=1}^{\ell} \frac{\rho_i}{|d^{(k)}(\xi_i)|^2} \left\| \mathbf{N}^{(k+1)}(\xi_i) - d^{(k+1)}(\xi_i)\mathbf{H}(\xi_i) \right\|_F^2 \longrightarrow \min .$$

Barycentric type representation

$$\mathbf{H}_r^{(k)}(s) = \frac{\mathbf{N}^{(k)}(s)}{d^{(k)}(s)} \equiv \frac{\sum_{j=1}^{r} \mathbf{\Phi}_j^{(k)}/(s-\lambda_j^{(k)})}{1+\sum_{j=1}^{r} \varphi_j^{(k)}/(s-\lambda_j^{(k)})}, \quad \begin{array}{l} \mathbf{\Phi}_j^{(k)} \in \mathbb{C}^{p\times m} \\ \varphi_j^{(k)}, \lambda_j^{(k)} \in \mathbb{C} \end{array}$$

Vector Fitting (Gustavsen–Semlyen)

SK iterations in barycentric form

$$\epsilon^{(k)} = \sum_{i=1}^{\ell} \frac{\rho_i}{|d^{(k)}(\xi_i)|^2} \left\| \sum_{j=1}^{r} \frac{\mathbf{\Phi}_j^{(k+1)}}{\xi_i - \lambda_j^{(k)}} - \mathbf{H}(\xi_i) \left(1 + \sum_{j=1}^{r} \frac{\varphi_j^{(k+1)}}{\xi_i - \lambda_j^{(k)}} \right) \right\|_F^2 \rightarrow \min$$

The $\lambda_j^{(k+1)}$'s are computed as the zeros of $1 + \sum_{j=1}^{r} \varphi_j^{(k+1)} / (s - \lambda_j^{(k)})$. Once the poles are fixed at λ_j's, the residues follow by solving

$$\epsilon = \sum_{i=1}^{\ell} w_i \left\| \sum_{j=1}^{r} \frac{\mathbf{\Phi}_j}{\xi_i - \lambda_j} - \mathbf{H}(\xi_i) \right\|_F^2 \longrightarrow \min$$

Consider first the scalar SISO case ($m = p = 1$), $\mathbf{\Phi}_j \equiv \phi_j \in \mathbb{C}$, Then the LS problems for $\epsilon^{(k)}$ and ϵ have particular Cauchy-type structures.

Motivation

Hence, we have to solve $\|W(Ax-h)\|_2 \longrightarrow \min$, where

$$A = \begin{pmatrix} \frac{1}{\xi_1-\lambda_1} & \frac{1}{\xi_1-\lambda_2} & \cdots & \frac{1}{\xi_1-\lambda_r} & \frac{-H(\xi_1)}{\xi_1-\lambda_1} & \frac{-H(\xi_1)}{\xi_1-\lambda_2} & \cdots & \frac{-H(\xi_1)}{\xi_1-\lambda_r} \\ \frac{1}{\xi_2-\lambda_1} & \frac{1}{\xi_2-\lambda_2} & \cdots & \frac{1}{\xi_2-\lambda_r} & \frac{-H(\xi_2)}{\xi_2-\lambda_1} & \frac{-H(\xi_2)}{\xi_2-\lambda_2} & \cdots & \frac{-H(\xi_2)}{\xi_2-\lambda_r} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\xi_\ell-\lambda_1} & \frac{1}{\xi_\ell-\lambda_2} & \cdots & \frac{1}{\xi_\ell-\lambda_r} & \frac{-H(\xi_\ell)}{\xi_\ell-\lambda_1} & \frac{-H(\xi_\ell)}{\xi_\ell-\lambda_2} & \cdots & \frac{-H(\xi_\ell)}{\xi_\ell-\lambda_r} \end{pmatrix},$$

$$h = \begin{pmatrix} H(\xi_1) & H(\xi_2) & \cdots & H(\xi_\ell) \end{pmatrix}^T, \quad x = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_r & \varphi_1 & \varphi_2 & \cdots & \varphi_r \end{pmatrix}^T .$$

$$A = \begin{pmatrix} \mathscr{C} & D_\sigma \mathscr{C} \end{pmatrix}, \quad \mathscr{C} = \left(\frac{1}{\xi_i-\lambda_j} \right)_{i,j=1}^{\ell,r}, \quad D_\sigma = -\mathrm{diag}(H(\xi_i))_{i=1}^{\ell}$$

Ill-conditioned weighted Cauchy-type least squares problem

Solve $\|W(Ax-h)\|_2^2 \longrightarrow \min$ where

$$A = W \begin{pmatrix} \mathscr{C} & D_\sigma \mathscr{C} \end{pmatrix}, \quad \text{or} \quad A = W\mathscr{C}$$

Cauchy-Vandermonde type ill-conditioning

- The coefficient matrices from previous examples are of Cauchy type, which, together with Vandermonde matrices, are among the most notoriously ill-conditioned matrices.
- The condition number of an arbitrary 100×100 real Vandermomde matrix is larger than $3 \cdot 10^{28}$, and the condition number of the 100×100 Hilbert (Cauchy) matrix is more than 10^{150}.

condition number(condition number) = condition number, D. J. Higham

```
>> cond(hilb(100))
ans = 4.6226e+19
```

Goal: high acuracy numerical linear algebra

Accurate SVD of any Cauchy or Vandermonde matrix is feasible without resorting to higher precision arithmetic. Need only a few basic building blocks for a uniform approach to many classes of ill-conditioned matrices.

Plan

1. Introduction - motivating examples and goals
2. Preliminaries: accurate computation with ill-conditioned matrices
   - Accurate SVD and scaling invariant condition number
   - Rank revealing (pivoted) QR factorization
   - An interesting connection: RRQR and DIME
   - PSVD, RRD and Cauchy and Vandermonde SVD
3. Hankel matrices
   - Mission impossible
   - Exploiting Vandermonde product representation
4. Case study: Matrix valued rational LS approximation
   - Sanathanan–Koerner iterations and Vector Fitting. $\mathcal{H}_2$ MOR
   - Details of the VF algorithm
   - Accurate LS for more robust VF
5. Concluding remarks

backward error $\longleftrightarrow$ perturbation theory

Let $\mathrm{rank}(A) = n \le m$, $D = \mathrm{diag}(\|A(:,i)\|_2)$, and

$$A \mapsto A + \delta A = (I + \delta A A^\dagger)A \implies \sigma_j \mapsto \sigma_j + \delta\sigma_j .$$

$$\max_j \frac{|\tilde{\sigma}_j - \sigma_j|}{\sigma_j} \le \|\delta A A^\dagger\|_2, \;\; \|\delta A A^\dagger\|_2 \le \begin{cases} \dfrac{\|\delta A\|_2}{\|A\|_2}(\|A^\dagger\|_2\|A\|_2) = \epsilon \cdot \kappa(A), \\ \|\delta A D^{-1}\|_2 \|(AD^{-1})^\dagger\|_2 . \end{cases}$$

- $\|\delta A D^{-1}\|_2 \le \sqrt{n}\max_j \frac{\|\delta A(:,j)\|_2}{\|A(:,j)\|_2}$; column-wise small backward error
- $\|(AD^{-1})^\dagger\|_2 \equiv \|B^\dagger\|_2 \le \kappa_2(B) \le \sqrt{n}\min_{\Delta=diag}\kappa(A\Delta)$
- Possible: $\|B^\dagger\|_2 \ll \kappa(A)$; always $\|B^\dagger\|_2 \le \sqrt{n}\kappa(A)$.
- Jacobi SVD: $\max_j \frac{\|\delta A(:,j)\|_2}{\|A(:,j)\|_2} \le \varepsilon \longrightarrow \|B^\dagger\|_2 \longrightarrow$ more accurate .
- bidiagonal SVD: $\frac{\|\delta A\|_2}{\|A\|_2} \le \varepsilon \longrightarrow \kappa(A) \longrightarrow$ less accurate , bidiagonalization provokes $\kappa(A)$.

Demmel and Veselić: Jacobi's method is more accurate than QR. (1992.)

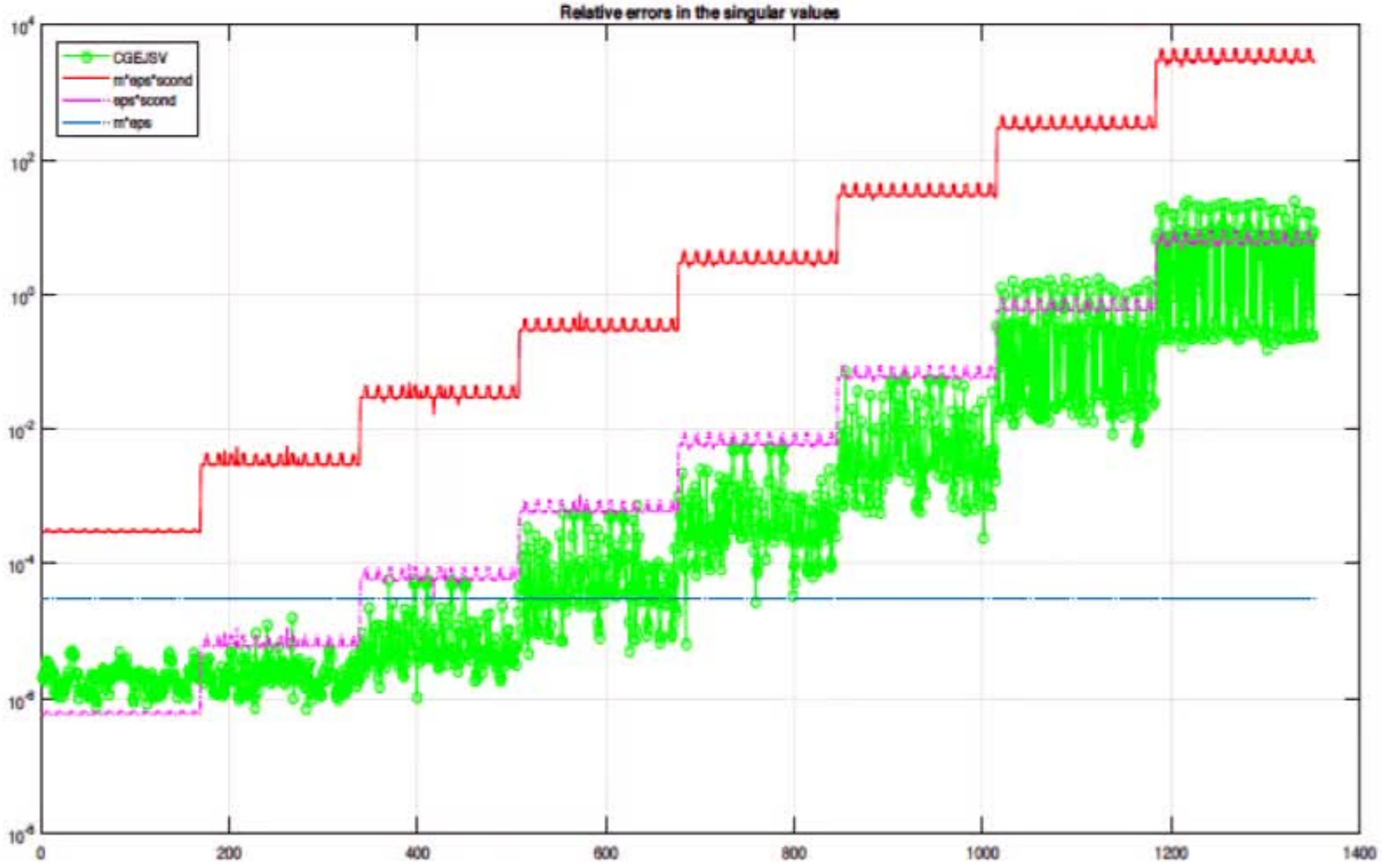
Jacobi SVD (Z. D., K. Veselić 2008., 2015.)

- $(\Pi A)P = Q\begin{pmatrix} R \\ 0 \end{pmatrix}$; Rank Revealing Decomposition (RRD)
  - $R(1:\rho, 1:n)^T = Q_1\begin{pmatrix} R_1 \\ 0 \end{pmatrix}$; $\rho = \mathrm{rank}(R)$,
- $X = R_1^T = \begin{pmatrix} \blacksquare & 0 \\ \blacksquare & \blacksquare \end{pmatrix}$; $X^T X - \xi I$ quasi–definite; entropy based decisions
  - $X_\infty \equiv U_x \Sigma = X \underbrace{\langle J_1 J_2 \cdots J_\infty \rangle}_{V_x}$ Jacobi rotations
  - $V_x = R_1^{-T}(X_\infty)$
- $U = \Pi^T Q\begin{pmatrix} U_x & 0 \\ 0 & I_{m-\rho} \end{pmatrix}$; $V = PQ_1\begin{pmatrix} V_x & 0 \\ 0 & I_{n-\rho} \end{pmatrix}$
- if $\rho = n$, $Q_1 V_x = R^{-1} X_\infty$

Delivers provably accurate SVD if A can be written as $A = BD$ with some diagonal D and well conditioned B. If $A = D_1 C D_2$ with D_1, D_2 diagonal and C well conditioned, the results are also accurate but theoretical bounds are lacking. $\kappa_2(D)$, $\kappa_2(D_1)$, $\kappa_2(D_2)$ irrelevant.

Numerical test: accuracy and scaled condition number

Complex $A \in \mathbb{C}^{m \times n}$; $m = 500$, $n = 456$, single precision



RRD: QRCP with Businger–Golub pivoting

$$\underbrace{A}_{m\times n}\overbrace{P}^{\text{permutation}} = Q\begin{pmatrix} R \\ 0 \end{pmatrix},\; R = \left(\begin{array}{cccccc} \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ 0 & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \hline 0 & 0 & \blacksquare & \bullet & \blacksquare & \blacklozenge \\ 0 & 0 & 0 & \bullet & \blacksquare & \blacklozenge \\ 0 & 0 & 0 & 0 & \blacksquare & \blacklozenge \\ 0 & 0 & 0 & 0 & 0 & \blacklozenge \end{array}\right)$$

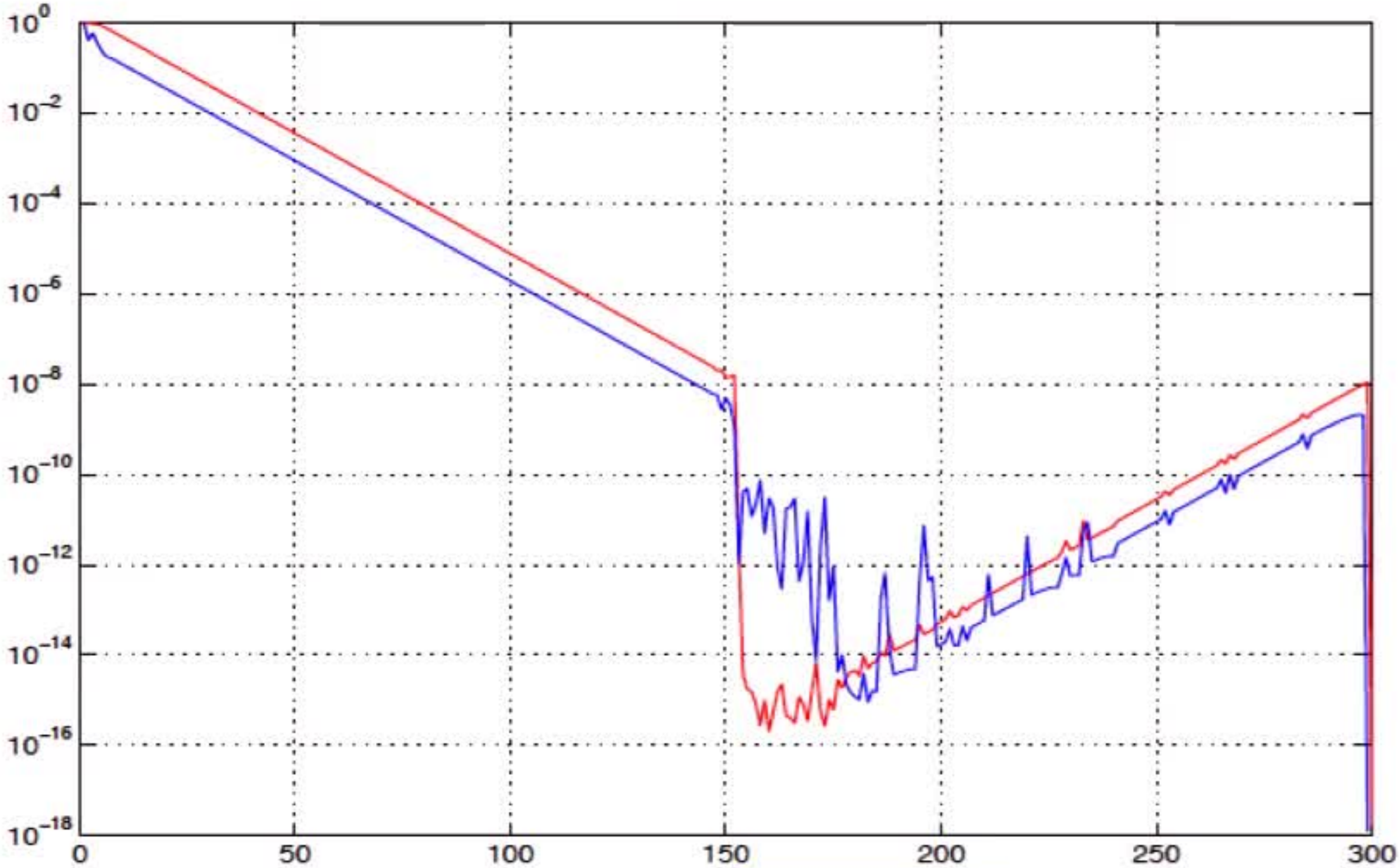
$$Q^* Q = I_m.$$

$$|R_{ii}| \geq \sqrt{\sum_{k=i}^{j} |R_{kj}|^2}, \text{ for all } 1 \leq i \leq j \leq n. \qquad (2.1)$$

$$|R_{11}| \geq |R_{22}| \geq \cdots \geq |R_{\rho\rho}| \gg |R_{\rho+1,\rho+1}| \geq \cdots \geq |R_{nn}| \qquad (2.2)$$

The structure (2.1), (2.2) may not be rank revealing but it must be guaranteed by the software (e.g. LAPACK, Matlab). Implemented in LINPACK in 1971., adopted by (Sca)LAPACK and used in many packages.

Examples of failure (since LINPACK 1971., until LAPACK 3.1, 2006.) of: minpack, odrpack, SLICOT, Matlab, ...

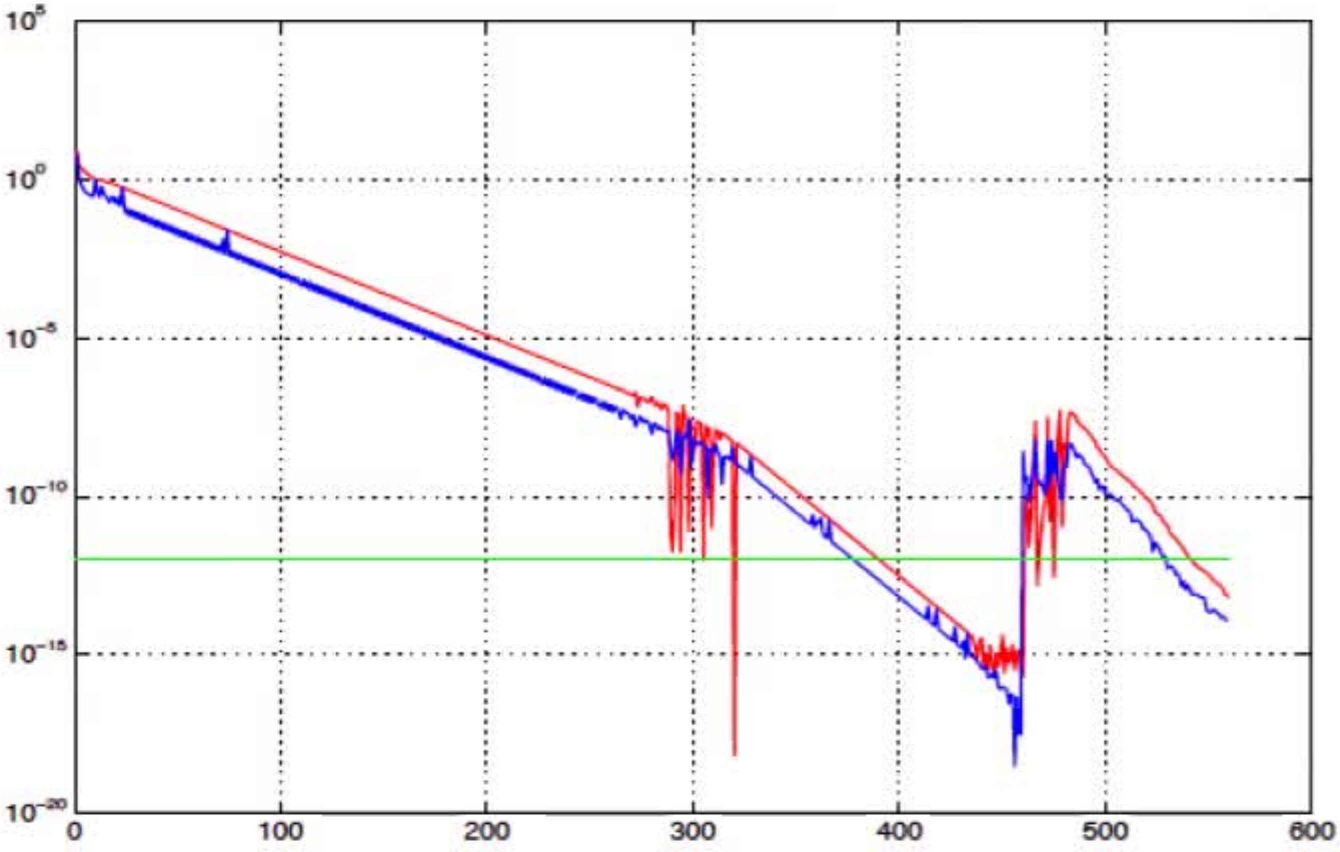


$|R_{ii}|$, $\max_{j\geq i}\sqrt{\sum_{k=i}^{j}|R_{kj}|^2}$

Consequences (since LINPACK 1971., until LAPACK 3.1, 2006.) ... sparse QRCP, windowed QRCP at risk as well

$\|Ax - d\|_2 \rightarrow \min;\ x = A\backslash d$

```
Warning:  Rank deficient, rank = 304 tol = 1.0994e-012.
```



`rank(`A`,1.0994e-12)` returns 466

L(IN+A)PACK update, xGEQPF, xGEQP3

```
      DO 30 J = I+1, N
         IF ( WORK( J ).NE.ZERO ) THEN
            TEMP = ONE - ( ABS( A( I, J ) ) / WORK( J ) )**2
            TEMP = MAX( TEMP, ZERO )
            TEMP2 = ONE + 0.05*TEMP*( WORK( J ) / WORK( N+J ) )**2
            WRITE(*,*) TEMP2
            IF( TEMP2.EQ.ONE ) THEN
               IF( M-I.GT.0 ) THEN
                  WORK( J ) = SNRM2( M-I, A( I+1, J ), 1 )
                  WORK( N+J ) = WORK( J )
               ELSE
                  WORK( J ) = ZERO
                  WORK( N+J ) = ZERO
               END IF
            ELSE
               WORK( J ) = WORK( J )*SQRT( TEMP )
            END IF
         END IF
30    CONTINUE ...
```

A strategically placed WRITE(*,*) statement may change the computed numerical rank substantially (!!) and thus completely change LS solution, computed properties of a dynamical system (e.g. staircase form). Numerical catastrophes in mission critical applications! Detailed analysis and solution by Z.D. and Z. Bujanović, ACM TOMS 2006., LAPACK 3.1.

An interesting new application of QRCP

- Discrete Empirical Interpolation Method: Chaturantabut/Sorensen, 2010
- Discrete variation of the EIM algorithm (Barrault, Maday, Nguyen, Patera; 2004)
- Given are: $\mathbf{f} : \mathcal{T} \subset \mathbb{R}^d \longrightarrow \mathbb{R}^n$ and a basis matrix $\mathrm{U} \in \mathbb{R}^{n\times m}$
- U is the POD basis for $F = [\mathbf{f}(t_1)\ \mathbf{f}(t_2), \ldots, \mathbf{f}(t_N)]$; $\mathrm{U}^T\mathrm{U} = \mathbb{I}_m$.
- The goal is: $\mathbf{f}(t) \approx \mathrm{U}\,\mathbf{c}(t)$ where $\mathbf{c}(t) \in \mathbb{R}^m$

  DEIM approximation is $\widehat{\mathbf{f}}(t) = \mathrm{U}(\mathbb{P}^T\mathrm{U})^{-1}\mathbb{P}^T\mathbf{f}(t)$,

  where $\mathbb{P}$ is $n \times m$ matrix obtained by selecting columns of the identity $\mathbb{I}_n$.
- Note that $\mathbb{P}^T\mathbf{f}(t) = \mathbb{P}^T\,\widehat{\mathbf{f}}(t)$, i.e., interpolation at the selected rows.
- How to pick $\mathbb{P}$?

The key: best conditioned submatrix

Lemma (Chaturantabut/Sorensen, 2010)

Let $\mathrm{U} \in \mathbb{R}^{n\times m}$ *be orthonormal* ($\mathrm{U}^*\mathrm{U} = \mathbb{I}_m$, $m < n$) *and let*

$$\widehat{f} = \mathrm{U}(\mathbb{P}^T\mathrm{U})^{-1}\mathbb{P}^T f \tag{2.3}$$

be the DEIM projection $f \in \mathbb{R}^n$, *with* $\mathbb{P}$ *computed by DEIM. Then*

$$\|f - \widehat{f}\|_2 \leq C\|(\mathbb{I} - \mathrm{U}\mathrm{U}^*)f\|_2, \;\; C = \|(\mathbb{P}^T\mathrm{U})^{-1}\|_2, \tag{2.4}$$

where

$$C \leq \frac{(1+\sqrt{2n})^{m-1}}{\|u_1\|_\infty} \leq \sqrt{n}(1+\sqrt{2n})^{m-1}.$$

- If $\mathcal{R}(\mathrm{U})$ captures the behavior of **f** well, and if $\mathbb{P}$ results in a moderate C, the DEIM approximation will succeed.
- Optimality related to picking the submatrix of maximal volume (Goreinov, Tyrtyshnikov and Zamarshkin; Mikhalev and Oseledets)

Discrete Empirical Interpolation Method (DEIM)

DEIM

INPUT: $u_1, \ldots, u_m \in \mathbb{C}^n$ (linearly independent)

OUTPUT: $\wp_1, \ldots, \wp_m$

```
• $[\rho \;\; \wp_1] = max|u_1|$
  $\mathrm{U} = [u_1]$, $\vec{\wp} = [\wp_1]$, $\mathbb{P} = [e_{\wp_1}]$

• for $j = 2$ to $m$
    1 $u \leftarrow u_j$
    2 Solve $(\mathbb{P}^T \mathrm{U})z = \mathbb{P}^T u$ for $z$
    3 $r = u - \mathrm{U}z$
    4 $[\rho \;\; \wp_j] = max\{|r|\}$
    5 $\mathrm{U} \leftarrow [\mathrm{U} \;\; u]$, $\vec{\wp} \leftarrow \begin{bmatrix} \vec{\wp} \\ \wp_j \end{bmatrix}$,
      $\mathbb{P} \leftarrow [\mathbb{P} \;\; e_{\wp_j}]$
```

Q_DEIM, Z.D., S. Gugercin 2015.

```
function S = q_deim(U) ;
[~,~,P] = qr( U', 'vector' ) ;
S = P(1:size(U,2)) ;
end
```

Q_DEIM properties:

- simple, efficient, blocked, parallelizable, numerically robust code already available
- basis independent
- better error bounds
- close to optimal volume property
- randomized sampling QRCP enhanced version possible

QRCP as preconditioner

Let $AP = Q\left(\begin{smallmatrix} R \\ 0 \end{smallmatrix}\right)$; $A_c = A \cdot \mathrm{diag}(1/\|A(:,1)\|_2, \ldots, 1/\|A(:,n)\|_2)$;
$R_c = R \cdot \mathrm{diag}(1/\|R(:,1)\|_2, \ldots, 1/\|R(:,n)\|_2) = \begin{pmatrix} \downarrow & \downarrow & \downarrow \\ 0 & \downarrow & \downarrow \\ 0 & 0 & \downarrow \end{pmatrix}$;
$R_r = \mathrm{diag}(1/\|R(1,:)\|_2, \ldots, 1/\|R(n,:)\|_2) \cdot R = \begin{pmatrix} \rightarrow & \rightarrow & \rightarrow \\ 0 & \rightarrow & \rightarrow \\ 0 & 0 & \rightarrow \end{pmatrix}$.

Theorem

Let $AP = Q\left(\begin{smallmatrix} R \\ 0 \end{smallmatrix}\right)$, where $|R_{ii}| \geq \sqrt{\sum_{k=i}^{j} |R_{kj}|^2}$, $1 \leq i \leq j \leq n$. Then $\|\,|R_r^{-1}|\,\|_2 \leq \sqrt{n}\|\,|R_c^{-1}|\,\|_2$, $\kappa_2(R_r) \leq n^{3/2}\kappa_2(A_c)$. Moreover, $\|R_r^{-1}\|_2$ is bounded by $O(2^n)$, independent of A. With exception of rare pathological cases, $\|R_r^{-1}\|_2$ is below $O(n)$ for any A. RR^* is more diagonal than R^*R.

Example ($A = Hilbert(100)$. $\kappa_2(A) > 10^{150} \gg \mathrm{cond}(A) \approx 4.6\mathrm{e}19$)

$\kappa_2(A_c) = \kappa_2(R_c) \gg 10^{19}$, $\kappa_2(R_r) \approx 48.31$. Repeat with $A \leftarrow R^T$, $P = I$, to get new $\kappa_2(R_r) \approx 3.22$.

PSVD($X \cdot D \cdot Y^*$), D diagonal; X, Y well conditioned

1. $XDY^* = \begin{pmatrix} \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \end{pmatrix} \Delta_X D \begin{pmatrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{pmatrix} = \widetilde{X}\widetilde{Y}^*$; $\widetilde{X}$ has unit columns; $\Delta_X = \mathrm{diag}(\|X(:,i)\|)$, $\widetilde{Y}^* = \Delta_X D Y^*$.
2. QRCP: $\widetilde{Y}P = Q\begin{pmatrix} R \\ 0 \end{pmatrix}$; it holds that $XDY^* = (\widetilde{X}P)\begin{pmatrix} R^* & 0 \end{pmatrix} Q^*$;
3. Need SVD of $\widetilde{A} = (\widetilde{X}P)R^*$, where $R^* = \begin{pmatrix} \blacksquare & & \\ \blacksquare & \blacksquare & \\ \blacksquare & \blacksquare & \blacksquare \end{pmatrix}$ has dominant diagonal and $\min_{\Delta=\mathrm{diag}} \kappa_2(\widetilde{A}\Delta) \leq \kappa_2(\widetilde{X}) \min_{\Delta=\mathrm{diag}} \kappa_2(R^*\Delta)$
4. $[U, \Sigma, V_1]$=SVD($\widetilde{A}$); Jacobi SVD of **explicitly computed** $\widetilde{A}$
5. With $V = Q\begin{pmatrix} V_1 & 0 \\ 0 & I_{n-p} \end{pmatrix}$, the SVD is $XDY^* = U\Sigma V^*$

The SVD will be accurate if $\min_{\Delta=\mathrm{diag}} \kappa_2(\mathbf{X\Delta})$ and $\min_{\Delta=\mathrm{diag}} \kappa_2(\mathbf{Y\Delta})$ are moderate. Detailed analysis in Z.D. 1998.

PSVD application: Hankel SVD, $\sigma_i = \sqrt{\lambda_i(HM)}$

$\dot{x}(t) = Ax(t) + Bu(t)$, $y(t) = Cx(t)$
Grammians $H = \int_0^\infty e^{tA} BB^T e^{tA^T} dt$, $M = \int_0^\infty e^{tA^T} C^T C e^{tA} dt$ via
Lyapunov equations $AH + HA^T = -BB^T$, $A^T M + MA = -C^T C$.
Let $H = L_H L_H^T$, $M = L_M L_M^T$, where L_H, L_M are the Cholesky factors computed by the Hammarling algorithm. Solve $HMx = \lambda x$ via the SVD of $L_M^T L_H$, using the PSVD($L_M^T L_H^T$) algorithm. The algorithm solves

$$(H + \delta H)(M + \delta M)x = \tilde{\lambda} x \ \text{ exactly, with symmetric } \delta H,\ \delta M,$$

$$\frac{|\delta H_{ij}|}{\sqrt{H_{ii} H_{jj}}} \le f(n) \cdot \varepsilon, \ \ \frac{|\delta M_{ij}|}{\sqrt{M_{ii} M_{jj}}} \le g(n) \cdot \varepsilon, \ \ 1 \le i, j \le n$$

$$\frac{|\delta \lambda|}{\lambda} \le h(n)(\sqrt{\|H_s^{-1}\|_2} + \sqrt{\|M_s^{-1}\|_2}) \cdot \varepsilon, \ \ \varepsilon = \mathrm{eps}.$$

$H_s = \mathrm{diag}(H)^{-1/2} H \mathrm{diag}(H)^{-1/2}$, $\kappa_2(H_s) \le n \min_{D=diag} \kappa_2(DHD)$.
Accuracy invariant under changes of physical units in state variables.

Example: $\boxtimes\!\rightsquigarrow\!\blacksquare\!\leftrightsquigarrow\!\blacksquare\!\leftrightsquigarrow\!\blacksquare$ early loss of definiteness

The stiffness matrix of a mass spring system with 3 masses $\boxtimes\!\rightsquigarrow\!\blacksquare\!\leftrightsquigarrow\!\blacksquare\!\leftrightsquigarrow\!\blacksquare$ with spring constants $k_1 = k_3 = 1$, $k_2 = \varepsilon/2$

$$K = \begin{pmatrix} k_1 + k_2 & -k_2 & 0 \\ -k_2 & k_2 + k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{pmatrix},$$

Let $\varepsilon \leq$ eps $=$ **round-off**. Then the true and the computed matrix are

$$K = \begin{pmatrix} 1+\frac{\varepsilon}{2} & -\frac{\varepsilon}{2} & 0 \\ -\frac{\varepsilon}{2} & 1+\frac{\varepsilon}{2} & -1 \\ 0 & -1 & 1 \end{pmatrix}, \widetilde{K} = \begin{pmatrix} 1 & -\frac{\varepsilon}{2} & 0 \\ -\frac{\varepsilon}{2} & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}, \max_{i,j} \frac{|\widetilde{K}_{ij} - K_{ij}|}{|K_{ij}|} < \varepsilon/2.$$

$\widetilde{K}$ is the best machine representation of K. However:

K is **positive definite** with $\lambda_{\min}(K) \approx \varepsilon/4$,
$\tilde{K}$ is **indefinite** with $\lambda_{\min}(\widetilde{K}) \approx -\varepsilon^2/8$.
Too late for $\lambda_{\min}(K)$, even in exact computation with $\widetilde{K}$. :(

Example ... implicit formulation

On the other hand, $K = A^T A$ with

$$A = \begin{pmatrix} \sqrt{k_1} & 0 & 0 \\ -\sqrt{k_2} & \sqrt{k_2} & 0 \\ 0 & -\sqrt{k_3} & \sqrt{k_3} \end{pmatrix} = \begin{pmatrix} \sqrt{k_1} & 0 & 0 \\ 0 & \sqrt{k_2} & 0 \\ 0 & 0 & \sqrt{k_3} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

clearly separates physical parameters and the geometry of the connections. The problem reduces to the SVD of A.
Since A is bidiagonal, for any choice of k_1, k_2, k_3, the singular values of A can be computed (zero-shift bidiagonal QR SVD of Demmel and Kahan) to nearly the same number of accurate digits to which the spring constants are given. Hence, in this formulation, the initial eigenvalue problem $Kx = \lambda x$ is perfectly well conditioned.
This is an example of preserving the important qualitative property (definiteness) exactly, using an implicit formulation of the problem and an accurate algorithm.

SVD($D_1 \times$ Cauchy $\times D_2$)

Given Cauchy matrix $C = C(x, y)$ and any two diagonal matrices D_1, D_2, the SVD of $G = D_1 C D_2$ can be computed to nearly fully precision as follows (Demmel):

1. Compute the LDU, $P_1 G P_2 = LDU$ using explicit determinant based formulas to update the Schur complement. This is entry wise forward stable computation of L, D, U. Moreover, $\kappa(L)$, $\kappa(U)$ are moderate. (Small $\|\delta L\|/\|L\|$, $\|\delta U\|/\|U\|$, $|\delta D_{ii}|/|D_{ii}|$ is also OK) ($G = hilb(100)$, $\kappa_2(G) > \mathbf{10}^{150}$, $\kappa_2(L) = \kappa_2(U) \approx 72.24$, $\kappa_2(D) \approx 2.32 \cdot \mathbf{10}^{149}$)
2. Compute the SVD of the product LDU using the Jacobi type PSVD algorithm (Z.D). The forward error is determined by $\max(\kappa(L), \kappa(U))$. The backward errors $\|\Delta L\|/\|L\|$, $\|\Delta U\|/\|U\|$, $\Delta D_{ii}/D_{ii}$ are small.

Key: ~~$C = (C_{ij})_{i,j=1}^{m,n}$~~. Instead, $C = C(x, y)$ and compute LDU as function of x, y and the diagonal entries of D_1, D_2.

Example in applications: con–eigenvalues

In the con–eigenvalue problem $Cu = \lambda\overline{u}$, or equivalently,

$$\overline{C}\,Cu = |\lambda|^2 u, \;\; C = (\frac{\sqrt{\alpha_i}\sqrt{\overline{\alpha_j}}}{1-\gamma_i\overline{\gamma_j}})$$

C is factored as $C = XD^2X^*$. The problem reduces to computing the SVD of the product $G = DX^TXD$. Accurate SVD via the PSVD based on the Jacobi SVD. Haut and Beylkin tested the accuracy with $\kappa_2(C) > 10^{200}$ and using Mathematica with 300 hundred digits for reference values. Over 500 test examples of size 120, the maximal error in IEEE 16 digit arithmetic ($\varepsilon \approx 2.2\cdot 10^{-16}$) was

$$\frac{|\tilde{\lambda}_i - \lambda_i|}{|\lambda_i|} < 5.2\cdot 10^{-12}, \;\; \frac{\|\tilde{u}_i - u_i\|_2}{\|u_i\|_2} < 5.4\cdot 10^{-12}.$$

Successfully used in reducing the order of the approximation to the viscous Burgers' equation.

Plan

1. Introduction - motivating examples and goals
2. Preliminaries: accurate computation with ill-conditioned matrices
   - Accurate SVD and scaling invariant condition number
   - Rank revealing (pivoted) QR factorization
   - An interesting connection: RRQR and DIME
   - PSVD, RRD and Cauchy and Vandermonde SVD
3. Hankel matrices
   - Mission impossible
   - Exploiting Vandermonde product representation
4. Case study: Matrix valued rational LS approximation
   - Sanathanan–Koerner iterations and Vector Fitting, $\mathcal{H}_2$ MOR
   - Details of the VF algorithm
   - Accurate LS for more robust VF
5. Concluding remarks

Hankel matrices

$$\mathcal{H} \equiv \mathcal{H}(h) = \begin{pmatrix} h_1 & h_2 & h_3 & \cdot & h_n \\ h_2 & h_3 & \cdot & h_n & h_{n+1} \\ h_3 & \cdot & \cdot & h_{n+1} & \cdot \\ \cdot & h_n & h_{n+1} & \cdot & h_{2n-2} \\ h_n & h_{n+1} & \cdot & h_{2n-2} & h_{2n-1} \end{pmatrix} = \mathcal{H}^T \in \mathbb{C}^{n \times n}.$$

Ubiquitous in matrix theory, rational approximation theory, signal processing, control theory, computational geometry, algebraic coding theory. Related with Toeplitz, Vandermonde, Bernstein, Cauchy, Pascal, Krylov, companion and other structured matrices. Triangular Hankel matrices key in the Carathéodory-Feyér rational approximation theory.

Autonne-Takagi-Schur factorization

Let $A \in \mathbb{C}^{n \times n}$. Then its SVD can be written as $A = \mathbb{W}\Sigma\mathbb{W}^T$, where $\mathbb{W}$ is unitary and Σ is a diagonal matrix carrying the singular values of A.

Variational characterization using the bilinear form $x^T A x$, $x \in \mathbb{C}^n$. (Danciger 2006.)

Mission impossible

Severely ill-conditioned: $\kappa_2(H \succ 0) \geq 3 \cdot 2^{n-6}$ (Tyrtyshnikov 1994)

Ill–conditioning caused by a connection with Vandermonde matrices. Accurate SVD of $\mathcal{H}$ as a function of the h_i's is impossible.

Theorem

It is not possible to compute the SVD of $\mathcal{H}$ to guaranteed high relative accuracy as a function of any input $h \in \mathbb{C}^{2n-1}$, for n big enough.

This follows from a similar claim for Toeplitz matrices:

Theorem (Demmel, Dimitriu, Holtz)

The determinant of a Toeplitz matrix is irreducible over any field. Hence, the determinant of a complex Toeplitz matrix T cannot be evaluated accurately. Thus, accurate SVD of T is impossible.

Vandermonde product representation of Hankel matrices

If $(h_k)_k$ is a signal generated by $h_k = \sum_{\ell=1}^{r} d_\ell x_\ell^k$, $k = 0, 1, \ldots$, or, if

$$\mathcal{H}_{ij} = \int_a^b x^{i+j-2} d\mu(x) \approx \sum_{\ell=1}^{r} d_\ell x_\ell^{i+j-2} = h_{i+j-2},$$

we can write $\mathcal{H}$ implicitly as $\mathcal{H} = \mathcal{V}^T D \mathcal{V}$,

$$\mathcal{H} = \begin{pmatrix} 1 & 1 & \cdot & 1 & 1 \\ x_1 & x_2 & \cdot & x_{n-1} & x_n \\ x_1^2 & x_2^2 & \cdot & x_{n-1}^2 & x_n^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_1^{n-1} & x_2^{n-1} & \cdot & x_{n-1}^{n-1} & x_n^{n-1} \end{pmatrix} \begin{pmatrix} d_1 & & & \\ & d_2 & & & \\ & & \cdot & & \\ & & & d_{n-1} & \\ & & & & d_n \end{pmatrix} \begin{pmatrix} 1 & x_1 & x_1^2 & \cdot & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdot & x_2^{n-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n-1} & x_{n-1}^2 & \cdot & x_{n-1}^{n-1} \\ 1 & x_n & x_n^2 & \cdot & x_n^{n-1} \end{pmatrix}.$$

This decomposition naturally arises if the underlying finite rank Hankel operator $\mathfrak{H} : \ell^2 \longrightarrow \ell^2$ is defined by its rational symbol that is represented in the pole-residue form $\chi(z) = \sum_{j=1}^{n} d_j/(z - x_j)$. In that case, $\mathcal{V}$ is replaced by the infinite matrix $\mathcal{V}_\infty = (\mathcal{V}, D^n\mathcal{V}, D^{2n}\mathcal{V}, \ldots)$, and $\mathcal{V}^T D \mathcal{V}$ is the leading $n \times n$ submatrix of $\mathfrak{H} = \mathcal{V}_\infty^T D \mathcal{V}_\infty$. (Gragg, Reichel, 1989.)

Accurate SVD of $\mathcal{V}^T D \mathcal{V}$

Any Vandermonde matrix $\mathcal{V}_n(x) = \mathcal{V} = (x_i^{j-1})_{i,j=1}^n$ can be written as

$$\mathcal{V} = D_1 \mathscr{C} D_2 \mathbb{F}^*, \quad \mathbb{F}_{ij} = \omega^{(i-1)(j-1)}/\sqrt{n}, \omega = \mathbf{e}^{2\pi \mathbf{i}/n},$$

where D_1 and $\mathcal{D}_2$ are diagonal, and $\mathscr{C}$ is a Cauchy matrix. More precisely,

$$(\mathcal{V}\mathbb{F})_{ij} = \left[\frac{1-x_i^n}{\sqrt{n}}\right]\left[\frac{1}{\omega^{1-j}-x_i}\right]\left[\frac{1}{\omega^{j-1}}\right] \equiv (D_1)_{ii}\,\mathscr{C}_{ij}\,(\mathcal{D}_2)_{jj},\ 1 \le i,j \le n.$$

In our setting, this gives

$$\mathbb{F}^T \mathcal{H} \mathbb{F} = \mathcal{D}_2 \mathscr{C}^T D_1 D D_1 \mathscr{C} \mathcal{D}_2 = \mathcal{D}_2 \mathscr{C}^T D_3 \mathscr{C} \mathcal{D}_2.$$

Since $\mathbb{F}$ and $\mathcal{D}_2$ are unitary, it remains to compute the SVD of the complex symmetric $M = \mathscr{C}^T D_3 \mathscr{C}$, where $D_3 = D_1^2 D$. Note that M is given implicitly and its factors $\mathscr{C}$ and D_3 are given to full accuracy.

Algorithm: $(\Sigma, \mathbb{U}, \mathbb{V}) = \text{SVD of}$ $\mathcal{H}(x,d) \equiv \mathcal{V}(x)^T \mathrm{diag}(d)\mathcal{V}(x);\ x, d \in \mathbb{C}^n$

Global structure of the algorithm:

1. $\omega = \mathbf{e}^{2\pi\mathbf{i}/n}$; $\vec{\omega} = (\omega^{1-i})_{i=1}^{n}$; $d_1 = (1 - x.^n)/\sqrt{n}$; $d_2 = (\omega^{j-1})_{j=1}^{n}$;
2. $[L, d_4, U, \pi_1, \pi_2] = \mathrm{CauchyLDU}(-x, \vec{\omega}, d_1. * \sqrt{d}, ones(n,1))$;
3. $A = (L\mathrm{diag}(d_4))^T L(\mathrm{diag}(d_4))$; Diagonal scaling and cross-product.
4. $[X_A, d_A, Y_A, \pi_3, \pi_4] = \mathrm{XDY}^*(A)$; Here $A = X_A \mathrm{diag}(d_A) Y_A^*$.
5. $[Q, R, \pi_5] = \mathrm{qr}(U^T \overline{Y_A} \mathrm{diag}(d_A))$; Pivoted QR factorization.
6. $S = U^T X_A(:, \pi_5) R^T$;
7. SVD of S: $S = U_s \Sigma W_s^*$; Jacobi SVD algorithm.
8. Assemble the singular vectors:
   $\mathbb{U} = \overline{\mathbb{F}}\mathrm{diag}(d_2)\Pi_2 U_s$; $\mathbb{V} = \mathbb{F}\mathrm{diag}(\overline{d_2})\Pi_2 \overline{Q} W_s$.

Many nontrivial details. Complicated implementation and analysis (Z.D., ETNA, to appear). Output includes reliable error estimates.

Example: $\varepsilon \equiv \mathrm{eps} \approx 2.22e-016$, $\kappa_2(\mathcal{H}) \approx 1.40e+260$

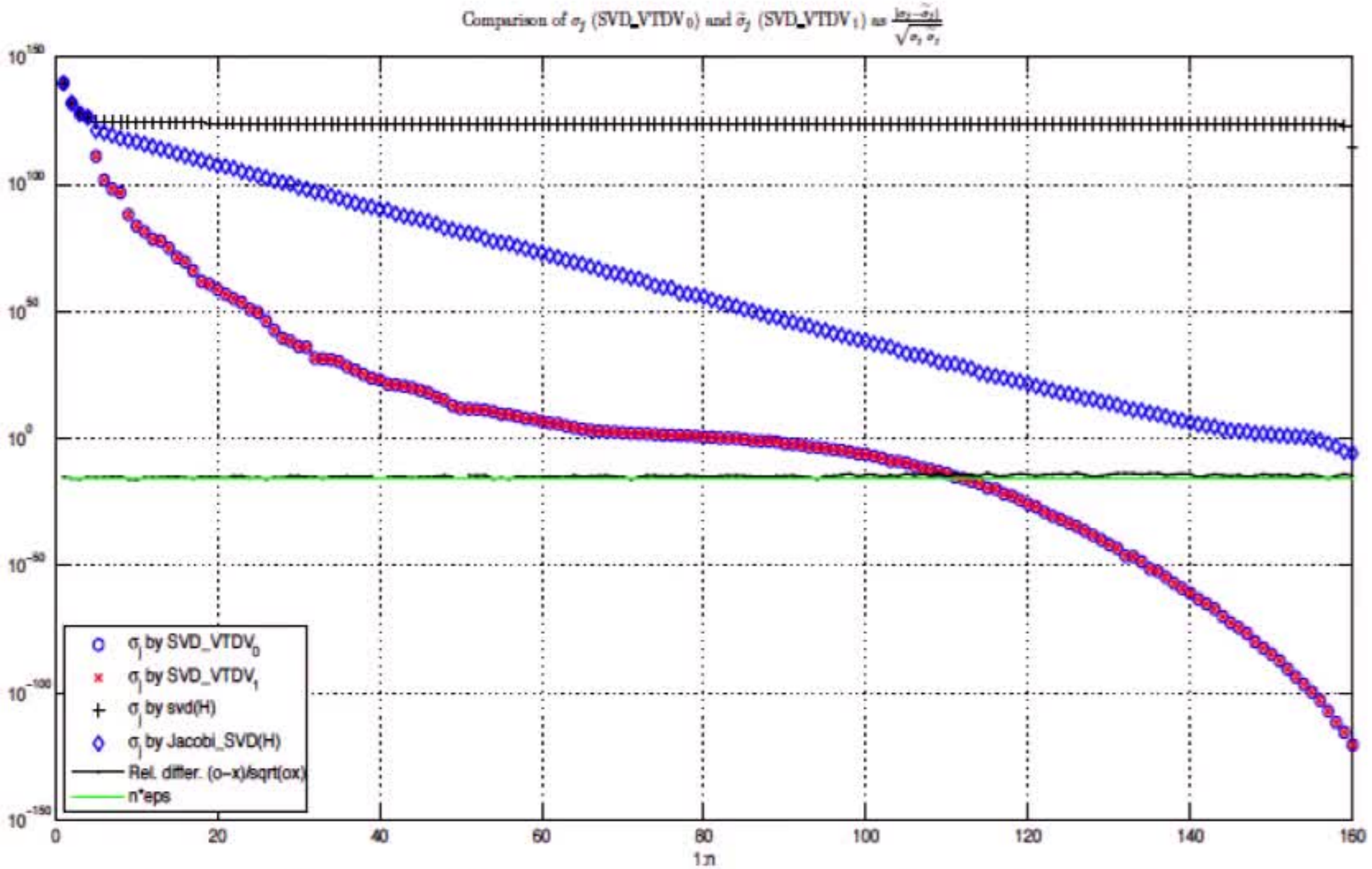


Figure: The relative differences $|\sigma_j - \widetilde{\sigma}_j| / \sqrt{\sigma_j \widetilde{\sigma}_j}$ are all at the level of $n \cdot \mathrm{eps}$. Reference values using Advanpix multiprecision toolbox (300 digits).

An extreme case: $\kappa_2(\mathcal{H}) \approx 0.28 \cdot 10^{615}$

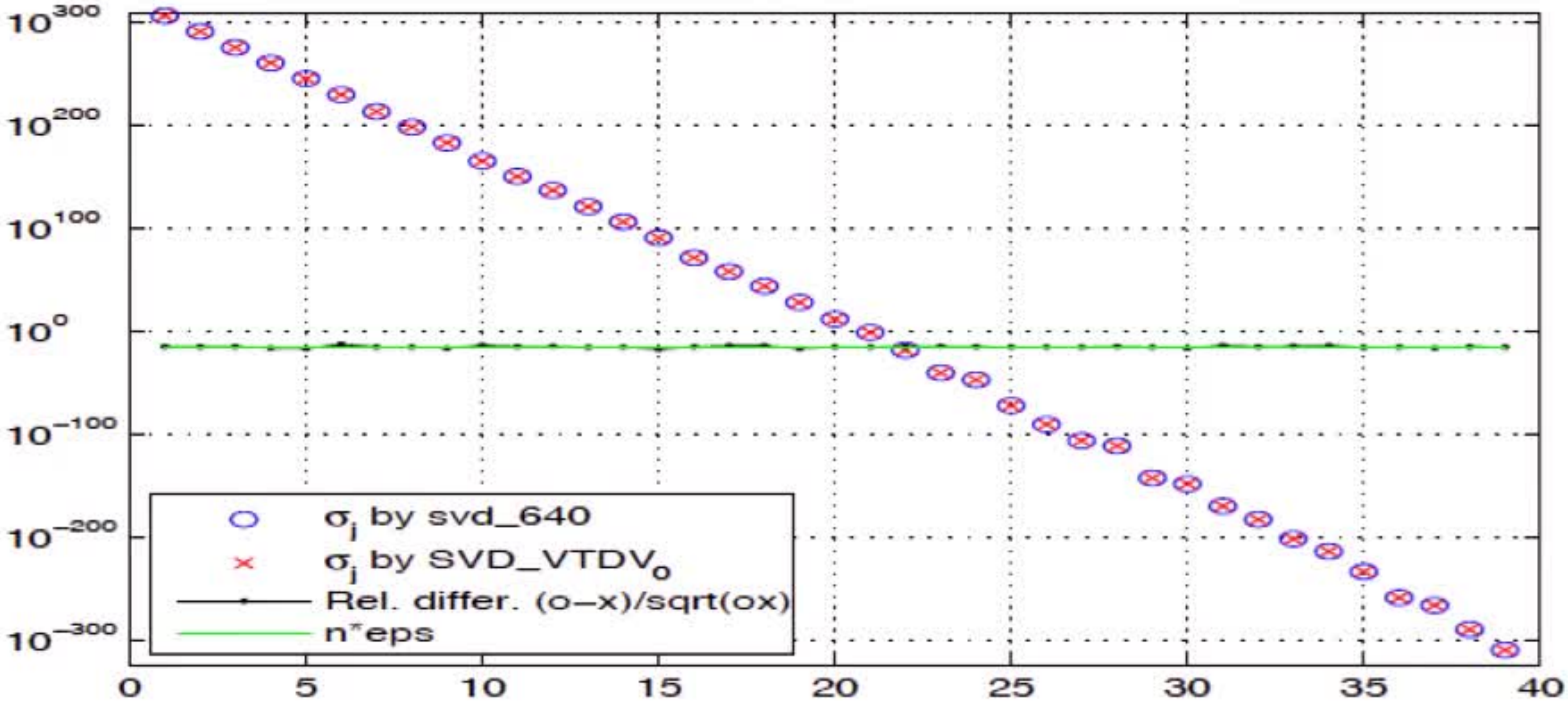


Figure: The singular values of the 39×39 product $\mathcal{H} = \mathcal{V}^T D \mathcal{V}$, computed in 16 digit arithmetic and the reference values computed in 640 digit arithmetic. The extreme singular values were $\sigma_1 \approxeq 1.659563214356268e{+}306$, $\sigma_{39} \approxeq 5.752792768736278e - 309$. The maximal measured relative error over all singular values was $8.632997535220512e - 013$.

Plan

1. Introduction - motivating examples and goals
2. Preliminaries: accurate computation with ill-conditioned matrices
   - Accurate SVD and scaling invariant condition number
   - Rank revealing (pivoted) QR factorization
   - An interesting connection: RRQR and DIME
   - PSVD, RRD and Cauchy and Vandermonde SVD
3. Hankel matrices
   - Mission impossible
   - Exploiting Vandermonde product representation
4. Case study: Matrix valued rational LS approximation
   - Sanathanan–Koerner iterations and Vector Fitting. $\mathcal{H}_2$ MOR
   - Details of the VF algorithm
   - Accurate LS for more robust VF
5. Concluding remarks

Basic Iterations (Sanathanan–Koerner, Kalman)

Sanathanan–Koerner (SK) Iterations

Compute a sequence of $\mathbf{H}_r^{(k)}(s) = \mathbf{N}^{(k)}(s)/d^{(k)}(s)$, where $\mathbf{N}^{(k)}(s)$ is a $p \times m$ matrix of polynomials of degree $r-1$ or less and $d^{(k)}(s)$ is a (scalar-valued) polynomial of degree r. Iterate for $k = 0, 1, 2, \ldots$

$$\epsilon^{(k)} = \sum_{i=1}^{\ell} \frac{\rho_i}{|d^{(k)}(\xi_i)|^2} \left\| \mathbf{N}^{(k+1)}(\xi_i) - d^{(k+1)}(\xi_i)\mathbf{H}(\xi_i) \right\|_F^2 \longrightarrow \min .$$

SK iterations in barycentric form

$$\epsilon^{(k)} = \sum_{i=1}^{\ell} \frac{\rho_i}{|d^{(k)}(\xi_i)|^2} \left\| \sum_{j=1}^{r} \frac{\mathbf{\Phi}_j^{(k+1)}}{\xi_i - \lambda_j^{(k)}} - \mathbf{H}(\xi_i) \left(1 + \sum_{j=1}^{r} \frac{\varphi_j^{(k+1)}}{\xi_i - \lambda_j^{(k)}} \right) \right\|_F^2 .$$

Implementation details D. Deschrijver, B. Gustavsen, T. Dhaene.

Vector Fitting (Gustavsen–Semlyen)

Vector Fitting: replace scaling by pole relocation

Let $d^{(k)}(s) = \prod_{j=1}^{r}(s - \lambda_j^{(k+1)}) / \prod_{=1}^{r}(s - \lambda_j^{(k)})$. Then, using the $\lambda_j^{(k+1)}$'s,

$$\epsilon^{(k)} = \sum_{i=1}^{\ell} \rho_i \left\| \sum_{j=1}^{r} \frac{\mathbf{\Phi}_j^{(k+1)}}{\xi_i - \lambda_j^{(k+1)}} - \mathbf{H}(\xi_i)\left(1 + \sum_{j=1}^{r} \frac{\varphi_j^{(k+1)}}{\xi_i - \lambda_j^{(k+1)}}\right) \right\|_F^2 .$$

Compare with

$$\epsilon^{(k)} = \sum_{i=1}^{\ell} \frac{\rho_i}{|d^{(k)}(\xi_i)|^2} \left\| \sum_{j=1}^{r} \frac{\mathbf{\Phi}_j^{(k+1)}}{\xi_i - \lambda_j^{(k)}} - \mathbf{H}(\xi_i)\left(1 + \sum_{j=1}^{r} \frac{\varphi_j^{(k+1)}}{\xi_i - \lambda_j^{(k)}}\right) \right\|_F^2 .$$

New development jointly with Chris Beattie and Serkan Gugercin (Virginia Tech): new VF and connection to MOR (IRKA).

A bigger picture: The Hardy space $\mathcal{H}_{2,+}^{p\times m}$

The algebraic least squares error is closely related to the $\mathcal{H}_2$ system norm. More precisely, consider the space $\mathcal{H}_{2,+}^{p\times m}$ of $p \times m$ matrix functions $\mathbf{M}(s)$, analytic in the open right half-plane $\mathbb{C}_+ = \{s \in \mathbb{C} \,:\, \Im(s) > 0\}$, such that $\sup_{x>0} \int_{\infty}^{\infty} \|\mathbf{M}(x + \imath y)\|_F^2 dy < \infty$.

$\mathcal{H}_{2,+}^{p\times m}$, $\langle\cdot,\cdot\rangle$

The space $\mathcal{H}_{2,+}^{p\times m}$is a Hilbert space with the associated inner product and norm defined by

$$
\begin{aligned}
\langle \mathbf{M}_1, \mathbf{M}_2 \rangle_{\mathcal{H}_2} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{Trace}\left(\overline{\mathbf{M}_1(\imath\omega)} \mathbf{M}_2(\imath\omega)^T \right) \, d\omega, \\
\|\mathbf{M}\|_{\mathcal{H}_2} &= \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \|\mathbf{M}(\imath\omega)\|_F^2 \, d\omega \right)^{1/2}.
\end{aligned}
$$

By a Fatou theorem, $M(s)$ can be identified with its boundary function $M(\imath\omega)$, $\omega \in \mathbb{R}$.

Distretized $\mathcal{H}_2$: Quadrature driven LS on $L_2(\imath\mathbb{R})$

Discretized $\mathcal{H}_2$ error as algebraic least squares error

$$\begin{aligned}\int_{-\infty}^{+\infty}\|\mathbf{H}(\imath\omega)-\mathbf{H}_r(\imath\omega)\|_F^2 d\omega \;&\approx\; \sum_{j=1}^{\ell}\rho_j^2\|\mathbf{H}(\xi_j)-\mathbf{H}_r(\xi_j)\|_F^2 \\ &+\;\; \rho_+^2\,|M_+[\mathbf{H}-\mathbf{H}_r]|^2+\rho_-^2\,|M_-[\mathbf{H}-\mathbf{H}_r|]|^2\end{aligned}$$

$M_\pm[H-Hr]=\lim_{\omega\to\pm\infty}\imath\omega[H-H_r](\imath\omega)$

An adapted Clenshaw-Curtis scheme (Boyd) :

$$\int_{-\infty}^{\infty}f(\omega)d\omega=\int_0^{\pi}f(L\cot t)\frac{dt}{\sin^2 t}\approx\sum_{j=0}^{\ell+1}w_j f(L\cot t_j),\;\; t_j=\frac{j\pi}{\ell+1};$$

$$w_j=\begin{cases}\rho_j^2=\dfrac{L\pi}{(\ell+1)\sin^2 t_j},\; j=1,\ldots,\ell\\[2ex] \rho_\pm^2=\dfrac{L\pi}{(2\ell+2)\sin^2 t_j},\; j=0,\; j=\ell+1\end{cases}$$

VF details - computing the new residues $\varphi_j^{(k)}$ and $\mathbf{\Phi}_j^{(k)}$

Data structure

Samples $\mathbb{S}(:,:,i) = S^{(i)} = \mathbf{H}(\xi_i) + \mathcal{E}_i$, $i = 1, \ldots, \ell$, $\mathbb{S} \in \mathbb{C}^{p \times m \times \ell}$.
Residues $\mathcal{F}^{(k)}(:,:,j) = \mathbf{\Phi}_j^{(k)}$, $j = 1, \ldots, r$, $\mathcal{F} \in \mathbb{C}^{p \times m \times r}$.

The residual for the input–output pair $(u, v) \in \{1, \ldots, p\} \times \{1, \ldots, m\}$ is

$$\left\| \mathcal{D}_\rho \left(\mathscr{C}^{(k+1)}, \quad -D^{(uv)} \mathscr{C}^{(k+1)} \right) \begin{pmatrix} \mathcal{F}^{(k+1)}(u, v, :) \\ \varphi^{(k+1)} \end{pmatrix} - \mathcal{D}_\rho \mathbb{S}(u, v, :) \right\|_2^2 ,$$

$\mathcal{D}_\rho = \mathrm{diag}(\sqrt{\rho_i})$, $D^{(uv)} = \mathrm{diag}(S_{uv}^{(i)})_{i=1}^{\ell}$, $\varphi^{(k+1)} = (\varphi_1^{(k+1)}, \ldots, \varphi_r^{(k+1)})^T$.

$$\mathscr{C}^{(k+1)} = \begin{pmatrix} \frac{1}{\xi_1 - \lambda_1^{(k+1)}} & \frac{1}{\xi_1 - \lambda_2^{(k+1)}} & \cdots & \frac{1}{\xi_1 - \lambda_r^{(k+1)}} \\ \frac{1}{\xi_2 - \lambda_1^{(k+1)}} & \frac{1}{\xi_2 - \lambda_2^{(k+1)}} & \cdots & \frac{1}{\xi_2 - \lambda_r^{(k+1)}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\xi_\ell - \lambda_1^{(k+1)}} & \frac{1}{\xi_\ell - \lambda_2^{(k+1)}} & \cdots & \frac{1}{\xi_\ell - \lambda_r^{(k+1)}} \end{pmatrix}, \ \mathcal{F}^{(k+1)}(u, v, :) = \begin{pmatrix} (\mathbf{\Phi}_1^{(k+1)})_{uv} \\ (\mathbf{\Phi}_2^{(k+1)})_{uv} \\ \vdots \\ (\mathbf{\Phi}_{r-1}^{(k+1)})_{uv} \\ (\mathbf{\Phi}_r^{(k+1)})_{uv} \end{pmatrix}. \tag{4.1}$$

VF details - assembling the global errror $A^{(k+1)}x^{(k+1)} - b$

$$\underbrace{\begin{pmatrix} \div&\div&\div&\ast&\ast&\ast\\ \div&\div&\div&\ast&\ast&\ast\\ \div&\div&\div&\ast&\ast&\ast\\ \div&\div&\div&\ast&\ast&\ast\\ \div&\div&\div&\ast&\ast&\ast\\ \div&\div&\div&\ast&\ast&\ast \end{pmatrix}},$$

$\left(\mathscr{C}^{(k+1)},\quad -D^{(uv)}\mathscr{C}^{(k+1)}\right)$

paired Cauchy structure can be exploited for more accurate computation

$$\underbrace{\left(\begin{array}{ccccccccccccccc}
\div&\div&\div&&&&&&&&&&\ast&\ast&\ast\\
\div&\div&\div&&&&&&&&&&\ast&\ast&\ast\\
\div&\div&\div&&&&&&&&&&\ast&\ast&\ast\\
\div&\div&\div&&&&&&&&&&\ast&\ast&\ast\\
\div&\div&\div&&&&&&&&&&\ast&\ast&\ast\\
\div&\div&\div&&&&&&&&&&\ast&\ast&\ast\\
\div&\div&\div&&&&&&&&&&\ast&\ast&\ast\\ \hline
&&&\div&\div&\div&&&&&&&\ast&\ast&\ast\\
&&&\div&\div&\div&&&&&&&\ast&\ast&\ast\\
&&&\div&\div&\div&&&&&&&\ast&\ast&\ast\\
&&&\div&\div&\div&&&&&&&\ast&\ast&\ast\\
&&&\div&\div&\div&&&&&&&\ast&\ast&\ast\\
&&&\div&\div&\div&&&&&&&\ast&\ast&\ast\\
&&&\div&\div&\div&&&&&&&\ast&\ast&\ast\\ \hline
&&&&&&\div&\div&\div&&&&\ast&\ast&\ast\\
&&&&&&\div&\div&\div&&&&\ast&\ast&\ast\\
&&&&&&\div&\div&\div&&&&\ast&\ast&\ast\\
&&&&&&\div&\div&\div&&&&\ast&\ast&\ast\\
&&&&&&\div&\div&\div&&&&\ast&\ast&\ast\\
&&&&&&\div&\div&\div&&&&\ast&\ast&\ast\\
&&&&&&\div&\div&\div&&&&\ast&\ast&\ast\\ \hline
&&&&&&&&&\div&\div&\div&\ast&\ast&\ast\\
&&&&&&&&&\div&\div&\div&\ast&\ast&\ast\\
&&&&&&&&&\div&\div&\div&\ast&\ast&\ast\\
&&&&&&&&&\div&\div&\div&\ast&\ast&\ast\\
&&&&&&&&&\div&\div&\div&\ast&\ast&\ast\\
&&&&&&&&&\div&\div&\div&\ast&\ast&\ast\\
&&&&&&&&&\div&\div&\div&\ast&\ast&\ast
\end{array}\right)}_{\mathbf{A}^{(k+1)},\ell=7,r=3,p=m=2},$$

$p \cdot m$ QR factorization(s) of size $\ell \times 2 \cdot r$

$$\begin{pmatrix} (R^{(k+1)})_{11} & (R_{uv}^{(k+1)})_{12} \\ 0 & (R_{uv}^{(k+1)})_{22} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} * & * & * & \times & \times & \times \\ 0 & * & * & \times & \times & \times \\ 0 & 0 & * & \times & \times & \times \\ 0 & 0 & 0 & \star & \star & \star \\ 0 & 0 & 0 & 0 & \star & \star \\ 0 & 0 & 0 & 0 & 0 & \star \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & & & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\underbrace{\left(\begin{array}{ccccccccccccccc}
* & * & * & & & & & & & & & & \times & \times & \times \\
0 & * & * & & & & & & & & & & \times & \times & \times \\
0 & 0 & * & & & & & & & & & & \times & \times & \times \\
0 & 0 & 0 & & & & & & & & & & \star & \star & \star \\
0 & 0 & 0 & & & & & & & & & & 0 & \star & \star \\
0 & 0 & 0 & & & & & & & & & & 0 & 0 & \star \\
0 & 0 & 0 & & & & & & & & & & 0 & 0 & 0 \\ \hline
 & & & * & * & * & & & & & & & \times & \times & \times \\
 & & & 0 & * & * & & & & & & & \times & \times & \times \\
 & & & 0 & 0 & * & & & & & & & \times & \times & \times \\
 & & & 0 & 0 & 0 & & & & & & & \star & \star & \star \\
 & & & 0 & 0 & 0 & & & & & & & 0 & \star & \star \\
 & & & 0 & 0 & 0 & & & & & & & 0 & 0 & \star \\
 & & & 0 & 0 & 0 & & & & & & & 0 & 0 & 0 \\ \hline
 & & & & & & * & * & * & & & & \times & \times & \times \\
 & & & & & & 0 & * & * & & & & \times & \times & \times \\
 & & & & & & 0 & 0 & * & & & & \times & \times & \times \\
 & & & & & & 0 & 0 & 0 & & & & \star & \star & \star \\
 & & & & & & 0 & 0 & 0 & & & & 0 & \star & \star \\
 & & & & & & 0 & 0 & 0 & & & & 0 & 0 & \star \\
 & & & & & & 0 & 0 & 0 & & & & 0 & 0 & 0 \\ \hline
 & & & & & & & & & * & * & * & \times & \times & \times \\
 & & & & & & & & & 0 & * & * & \times & \times & \times \\
 & & & & & & & & & 0 & 0 & * & \times & \times & \times \\
 & & & & & & & & & 0 & 0 & 0 & \star & \star & \star \\
 & & & & & & & & & 0 & 0 & 0 & 0 & \star & \star \\
 & & & & & & & & & 0 & 0 & 0 & 0 & 0 & \star \\
 & & & & & & & & & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right)}_{(\mathcal{Q}^{(k+1)})^* \mathbf{A}^{(k+1)}}$$

single pivoted QR factorization $*$

QR factorizations with column pivoting $\star$

not computed $\times$

---

faster implementation, LAPACK, BLAS

numerically more robust; details follow ...

LS solution for $\varphi^{(k+1)}$

$$
\underbrace{\left(\begin{array}{ccccccccccccccc}
\ast & \ast & \ast & & & & & & & & & & \times & \times & \times \\
0 & \ast & \ast & & & & & & & & & & \times & \times & \times \\
0 & 0 & \ast & & & & & & & & & & \times & \times & \times \\
 & & & \ast & \ast & \ast & & & & & & & \times & \times & \times \\
 & & & 0 & \ast & \ast & & & & & & & \times & \times & \times \\
 & & & 0 & 0 & \ast & & & & & & & \times & \times & \times \\
 & & & & & & \ast & \ast & \ast & & & & \times & \times & \times \\
 & & & & & & 0 & \ast & \ast & & & & \times & \times & \times \\
 & & & & & & 0 & 0 & \ast & & & & \times & \times & \times \\
 & & & & & & & & & \ast & \ast & \ast & \times & \times & \times \\
 & & & & & & & & & 0 & \ast & \ast & \times & \times & \times \\
 & & & & & & & & & 0 & 0 & \ast & \times & \times & \times \\
\hline\hline
 & & & & & & & & & & & & \star & \star & \star \\
 & & & & & & & & & & & & 0 & \star & \star \\
 & & & & & & & & & & & & 0 & 0 & \star \\
 & & & & & & & & & & & & \star & \star & \star \\
 & & & & & & & & & & & & 0 & \star & \star \\
 & & & & & & & & & & & & 0 & 0 & \star \\
 & & & & & & & & & & & & \star & \star & \star \\
 & & & & & & & & & & & & 0 & \star & \star \\
 & & & & & & & & & & & & 0 & 0 & \star \\
 & & & & & & & & & & & & \star & \star & \star \\
 & & & & & & & & & & & & 0 & \star & \star \\
 & & & & & & & & & & & & 0 & 0 & \star \\
 & & & & & & & & & & & & 0 & 0 & 0 \\
 & & & & & & & & & & & & 0 & 0 & 0 \\
 & & & & & & & & & & & & 0 & 0 & 0 \\
 & & & & & & & & & & & & 0 & 0 & 0
\end{array}\right)}_{(\mathcal{Q}^{(k+1)})^{*}\mathbf{A}^{(k+1)},\ \text{row permuted}}
= \begin{pmatrix} \mathbf{B}_{[11]}^{(k+1)} & \mathbf{B}_{[12]}^{(k+1)} \\ 0 & \mathbf{B}_{[22]}^{(k+1)} \\ 0 & 0 \end{pmatrix}
$$

$$
\begin{pmatrix} \mathbf{B}_{[11]}^{(k+1)} & \mathbf{B}_{[12]}^{(k+1)} \\ 0 & \mathbf{B}_{[22]}^{(k+1)} \\ 0 & 0 \end{pmatrix}
\begin{pmatrix} \mathbf{\Phi}^{(k+1)} \\ \varphi^{(k+1)} \end{pmatrix}
\approxeq
\begin{pmatrix} \mathbf{s}_1^{(k+1)} \\ \mathbf{s}_2^{(k+1)} \\ \mathbf{s}_3^{(k+1)} \end{pmatrix}
$$

$$
\|\mathbf{B}_{[22]}^{(k+1)}\varphi^{(k+1)} - \mathbf{s}_2^{(k+1)}\|_2 \longrightarrow \min
$$

Weighted augmented Cauchy least squares problem

Extracting residues from weighted Cauchy LS problems

$$\|\mathcal{D}_\rho\,(\mathscr{C}\Phi(i,j,1:\ell) - \mathbb{S}(i,j,1:\ell))\,\|_2 \longrightarrow \min, \;\; i=1:p, \;\; j=1:m.$$

To simplify the notation, write $\|\mathcal{D}_\rho\mathscr{C}x - h\|_2 \longrightarrow \min$, where $\mathscr{C} = \mathscr{C}_{\xi,\boldsymbol{\lambda}}$ is a Cauchy matrix, h is the corresponding scaled right-hand side, $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_r)$ is closed under conjugation and and the solution vector should also be closed under conjugation. Consider equivalent augmented unconstrained LS

$$\left\| \begin{pmatrix} \mathcal{D}_\rho\mathscr{C}_{\xi,\boldsymbol{\lambda}} \\ \mathcal{D}_\rho\mathscr{C}_{\overline{\xi},\boldsymbol{\lambda}} \end{pmatrix} x - \begin{pmatrix} h \\ \overline{h} \end{pmatrix} \right\|_2 \equiv \|\widehat{\mathscr{C}}x - \widehat{h}\|_2 \longrightarrow \min$$

with the coefficient matrix again of the diagonally scaled Cauchy structure, $\widehat{\mathscr{C}} = (\mathcal{D}_\rho \oplus \mathcal{D}_\rho)\mathscr{C}_{(\xi,\overline{\xi}),\boldsymbol{\lambda}}$.

Accurate regularized LS solution

Let $\widehat{\mathscr{C}} = W\Sigma V^*$ be the SVD and let the unique[1] LS solution be $x = V\Sigma^\dagger W^* = \sum_{i=1}^r v_i(w_i^*\widehat{h})/\sigma_i$. Unfortunately, an accurate SVD is not enough to have the LS solution computed to high relative accuracy, and additional regularization techniques must be deployed. This is in particular important if the right-hand side is contaminated by noise. In the Tichonov regularization, we choose $\mu \geq 0$ and use the solution of $\|\widehat{\mathscr{C}}x - \widehat{h}\|_2^2 + \mu^2\|x\|_2^2 \rightarrow \min$, explicitly computable as

$$x_\mu = \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 + \mu^2}(w_i^*\widehat{h})v_i. \qquad (4.2)$$

The parameter μ can be further adjusted using the Morozov discrepancy principle, i.e., to achieve $\|\widehat{\mathscr{C}}x_\mu - \widehat{h}\|_2 \approx \nu$, where ν is the estimated level of noise $\delta\widehat{h}$ in the right-hand side, $\nu \approx \|\delta\widehat{h}\|_2$.

[1] Since all nodes are distinct and the poles are assumed simple, the matrix is of full column rank.

Test drive of the new implementation (mimoVF)

For the resulting rational approximation $\mathbf{H}_r$, define $\mathbb{S}_r(:,:,i) = \mathbf{H}_r(\xi_i)$, $i = 1, \ldots, \ell$, and the relative LS error as

$$\gamma = \|\mathbb{S} - \mathbb{S}_r\|_F / \|\mathbb{S}\|_F .$$

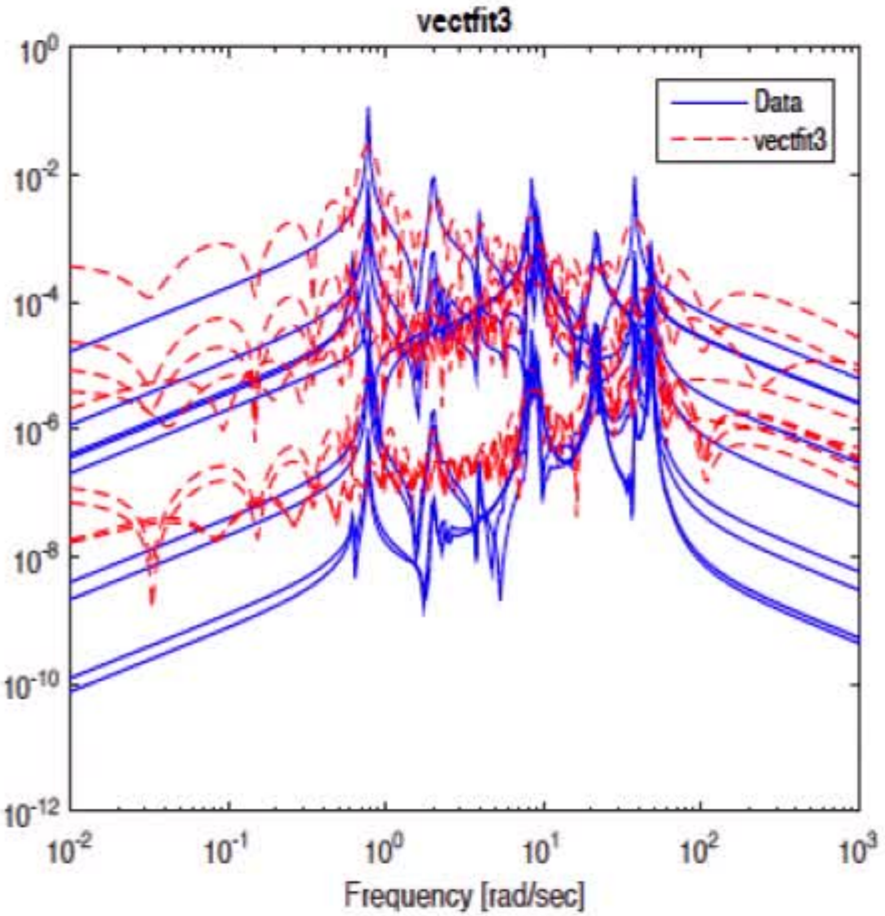
Recall that $\mathbb{S}(:,:,i) = \mathbf{H}(\xi_i) \in \mathbb{C}^{p\times m}$, $i = 1, \ldots, \ell$, contains the original samples that are either measurements, or computed from a state space realization of the underlying LTI dynamical system.

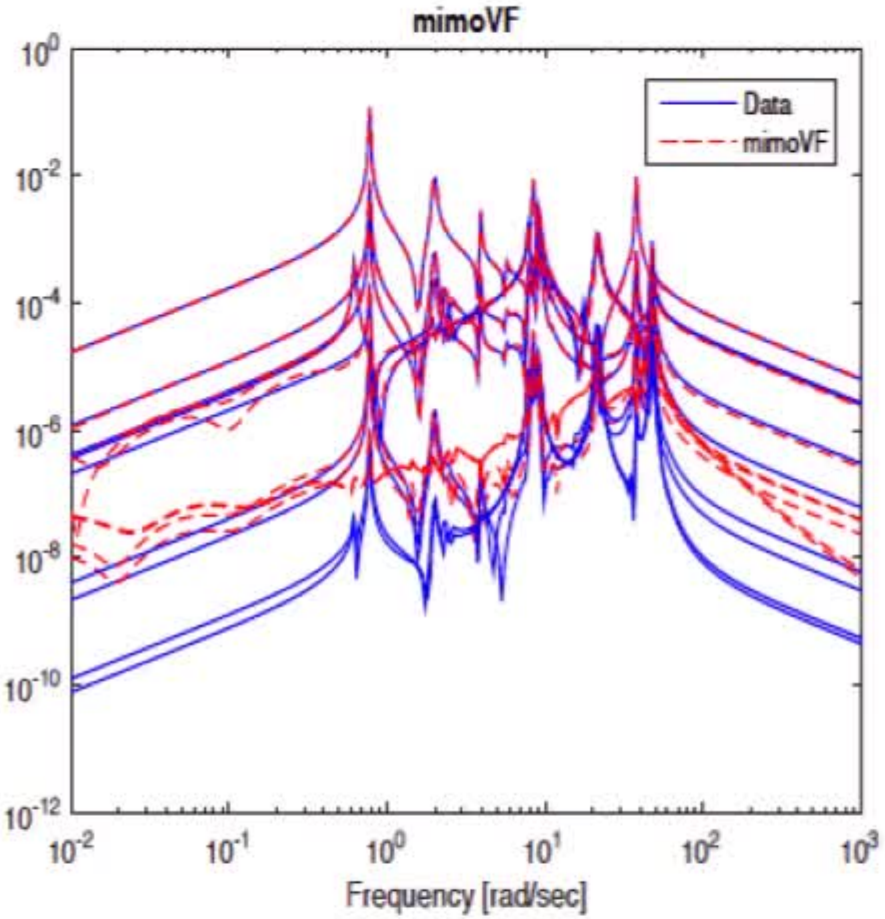
Example (1R module, International Space Station)

Dynamical system of order $n = 270$, with $m = 3$ inputs and $p = 3$ outputs. It is very hard to approximate and is challenging for model order reduction. We take $r = 50$ and use $\ell = 300$ samples.

Compare with vectfit3 (http://www.sintef.no/Projectweb/VECTFIT/)

Stress test: ISS R1, difficult poles, good subspace in $\mathcal{R}_r$





$\gamma(\text{vectfit3}) \approx 14.1,$ $\qquad$ $\gamma(\text{mimoVF}) \approx 6.45 \cdot 10^{-3}.$

Plan

1. Introduction - motivating examples and goals
2. Preliminaries: accurate computation with ill-conditioned matrices
   - Accurate SVD and scaling invariant condition number
   - Rank revealing (pivoted) QR factorization
   - An interesting connection: RRQR and DIME
   - PSVD, RRD and Cauchy and Vandermonde SVD
3. Hankel matrices
   - Mission impossible
   - Exploiting Vandermonde product representation
4. Case study: Matrix valued rational LS approximation
   - Sanathanan–Koerner iterations and Vector Fitting. $\mathcal{H}_2$ MOR
   - Details of the VF algorithm
   - Accurate LS for more robust VF
5. Concluding remarks