Computational methods for large-scale matrix equations: recent advances

V. Simoncini

Dipartimento di Matematica, Università di Bologna (Italy)

valeria.simoncini@unibo.it

Some matrix equations

- Sylvester matrix equation

$$A\mathbf{X} + \mathbf{X}B + D = 0$$

Eigenvalue pbs and tracking, Control, MOR, Assignment pbs, Riccati eqn

Some matrix equations

- Sylvester matrix equation

$$A\mathbf{X} + \mathbf{X}B + D = 0$$

Eigenvalue pbs and tracking, Control, MOR, Assignment pbs, Riccati eqn

- Lyapunov matrix equation

$$A\mathbf{X} + \mathbf{X}A^\top + D = 0, \qquad D = D^\top$$

Stability analysis in Control and Dynamical systems, Signal processing, eigenvalue computations

Some matrix equations

- Sylvester matrix equation

$$A\mathbf{X} + \mathbf{X}B + D = 0$$

Eigenvalue pbs and tracking, Control, MOR, Assignment pbs, Riccati eqn

- Lyapunov matrix equation

$$A\mathbf{X} + \mathbf{X}A^\top + D = 0, \qquad D = D^\top$$

Stability analysis in Control and Dynamical systems, Signal processing, eigenvalue computations

- Algebraic Riccati equation

$$A\mathbf{X} + \mathbf{X}A^\top - \mathbf{X}BB^\top\mathbf{X} + D = 0, \qquad D = D^\top$$

Lancaster-Rodman '95, Konstantinov-Gu-Mehrmann-Petkov, '02, Bini-Iannazzo-Meini '12

Some matrix equations

- Sylvester matrix equation

$$A\mathbf{X} + \mathbf{X}B + D = 0$$

Eigenvalue pbs and tracking, Control, MOR, Assignment pbs, Riccati eqn

- Lyapunov matrix equation

$$A\mathbf{X} + \mathbf{X}A^\top + D = 0, \qquad D = D^\top$$

Stability analysis in Control and Dynamical systems, Signal processing, eigenvalue computations

- Algebraic Riccati equation

$$A\mathbf{X} + \mathbf{X}A^\top - \mathbf{X}BB^\top\mathbf{X} + D = 0, \qquad D = D^\top$$

Lancaster-Rodman '95, Konstantinov-Gu-Mehrmann-Petkov, '02, Bini-Iannazzo-Meini '12

Focus: All or some of the matrices are large (and possibly sparse)

Solving the Lyapunov equation. The problem

Approximate X in:

$$AX + XA^\top + BB^\top = 0$$

$A \in \mathbb{R}^{n \times n}$ neg.real $\qquad B \in \mathbb{R}^{n \times p}, \qquad 1 \le p \ll n$

Solving the Lyapunov equation. The problem

Approximate X in:

$$AX + XA^\top + BB^\top = 0$$

$A \in \mathbb{R}^{n\times n}$ neg.real $\qquad B \in \mathbb{R}^{n\times p}, \qquad 1 \le p \ll n$

Time-invariant linear system:

$$\mathbf{x}'(t) = A\mathbf{x}(t) + B\mathbf{u}(t), \qquad \mathbf{x}(0) = x_0$$

Closed form solution:

$$X = \int_0^\infty e^{-tA} BB^\top e^{-tA^\top} dt$$

$\Rightarrow \qquad X$ symmetric semidef.

see, e.g., Antoulas '05, Benner '06

Linear systems vs linear matrix equations

Large linear systems:

$$Ax = b, \quad A \in \mathbb{R}^{n \times n}$$

- Krylov subspace methods (CG, MINRES, GMRES, BiCGSTAB, etc.)
- Preconditioners: find P such that

$$AP^{-1}\widetilde{x} = b \qquad x = P^{-1}\widetilde{x}$$

is easier and fast to solve

Linear systems vs linear matrix equations

Large linear systems:

$$Ax = b, \quad A \in \mathbb{R}^{n \times n}$$

- Krylov subspace methods (CG, MINRES, GMRES, BiCGSTAB, etc.)
- Preconditioners: find P such that

$$AP^{-1}\widetilde{x} = b \qquad x = P^{-1}\widetilde{x}$$

  is easier and fast to solve

Large linear matrix equations:

$$AX + XA^\top + BB^\top = 0$$

- No preconditioning - to preserve symmetry
- X is a large, dense matrix $\Rightarrow$ low rank approximation

$$X \approx \widetilde{X} = ZZ^\top, \quad Z \text{ tall}$$

Linear systems vs linear matrix equations

Large linear systems:

$$Ax = b, \quad A \in \mathbb{R}^{n\times n}$$

- Krylov subspace methods (CG, MINRES, GMRES, BiCGSTAB, etc.)
- Preconditioners: find P such that

$$AP^{-1}\widetilde{x} = b \qquad x = P^{-1}\widetilde{x}$$

is easier and fast to solve

Large linear matrix equations:

$$AX + XA^\top + BB^\top = 0$$

Kronecker formulation:

$$(A \otimes I + I \otimes A)x = b \qquad x = \mathrm{vec}(X)$$

Projection-type methods

Given an approximation space $\mathcal{K}$,

$$X \approx X_m \qquad \mathrm{col}(X_m) \in \mathcal{K}$$

Galerkin condition: $\quad R := AX_m + X_mA^\top + BB^\top \quad \perp \quad \mathcal{K}$

$$V_m^\top R V_m = 0 \qquad\qquad \mathcal{K} = \mathrm{Range}(V_m)$$

Projection-type methods

Given an approximation space $\mathcal{K}$,

$$X \approx X_m \qquad \mathrm{col}(X_m) \in \mathcal{K}$$

Galerkin condition: $\quad R := AX_m + X_m A^\top + BB^\top \quad \perp \quad \mathcal{K}$

$$V_m^\top R V_m = 0 \qquad\qquad \mathcal{K} = \mathrm{Range}(V_m)$$

Assume $V_m^\top V_m = I_m$ and let $X_m := V_m Y_m V_m^\top$.

Projected Lyapunov equation:

$$V_m^\top (A V_m Y_m V_m^\top + V_m Y_m V_m^\top A^\top + BB^\top) V_m \quad = \quad 0$$

Projection-type methods

Given an approximation space $\mathcal{K}$,

$$X \approx X_m \qquad \mathrm{col}(X_m) \in \mathcal{K}$$

Galerkin condition: $\quad R := AX_m + X_m A^\top + BB^\top \quad \perp \quad \mathcal{K}$

$$V_m^\top R V_m = 0 \qquad\qquad \mathcal{K} = \mathrm{Range}(V_m)$$

Assume $V_m^\top V_m = I_m$ and let $X_m := V_m Y_m V_m^\top$.

Projected Lyapunov equation:

$$\begin{aligned} V_m^\top (A V_m Y_m V_m^\top + V_m Y_m V_m^\top A^\top + BB^\top) V_m &= 0 \\ (V_m^\top A V_m) Y_m + Y_m (V_m^\top A^\top V_m) + V_m^\top B B^\top V_m &= 0 \end{aligned}$$

Early contributions: Saad '90, Jaimoukha & Kasenally '94, for $\mathcal{K} = \mathcal{K}_m(A, B) = \mathrm{Range}([B, AB, \ldots, A^{m-1}B])$

More recent options as approximation space

Enrich space to decrease space dimension

- Extended Krylov subspace

$$\mathcal{K} = \mathcal{K}_m(A, B) + \mathcal{K}_m(A^{-1}, A^{-1}B),$$

that is, $\mathcal{K} = \mathrm{Range}([B, A^{-1}B, AB, A^{-2}B, A^2, A^{-3}B, \ldots,])$

(Druskin & Knizhnerman ’98, Simoncini ’07)

More recent options as approximation space

Enrich space to decrease space dimension

- Extended Krylov subspace

$$\mathcal{K} = \mathcal{K}_m(A, B) + \mathcal{K}_m(A^{-1}, A^{-1}B),$$

that is, $\mathcal{K} = \mathrm{Range}([B, A^{-1}B, AB, A^{-2}B, A^2, A^{-3}B, \ldots,])$

(Druskin & Knizhnerman '98, Simoncini '07)

- Rational Krylov subspace

$$\mathcal{K} = \mathrm{Range}([B, (A - s_1 I)^{-1}B, \ldots, (A - s_m I)^{-1}B])$$

usually, $\{s_1, \ldots, s_m\} \subset \mathbb{C}^+$ chosen a-priori

More recent options as approximation space

Enrich space to decrease space dimension

- Extended Krylov subspace

$$\mathcal{K} = \mathcal{K}_m(A, B) + \mathcal{K}_m(A^{-1}, A^{-1}B),$$

that is, $\mathcal{K} = \mathrm{Range}([B, A^{-1}B, AB, A^{-2}B, A^2, A^{-3}B, \ldots,])$

(Druskin & Knizhnerman '98, Simoncini '07)

- Rational Krylov subspace

$$\mathcal{K} = \mathrm{Range}([B, (A - s_1 I)^{-1}B, \ldots, (A - s_m I)^{-1}B])$$

usually, $\{s_1, \ldots, s_m\} \subset \mathbb{C}^+$ chosen a-priori

In both cases, for Range$(V_m) = \mathcal{K}$, projected Lyapunov equation:

$$(V_m^\top A V_m)Y_m + Y_m(V_m^\top A^\top V_m) + V_m^\top BB^\top V_m = 0$$

$X_m = V_m Y_m V_m^\top$

Rational Krylov Subspaces. A long tradition...

In general,

$$K_m(A, B, \mathbf{s}) = \text{Range}([(A - s_1 I)^{-1} B, (A - s_2 I)^{-1} B, \ldots, (A - s_m I)^{-1} B])$$

- Eigenvalue problems (Ruhe, 1984)
- Model Order Reduction (transfer function evaluation)
- In Alternating Direction Implicit iteration (ADI) for linear matrix equations

Rational Krylov Subspaces in MOR. Choice of poles.

$$K_m(A, B, \mathbf{s}) = \mathrm{Range}([(A - s_1 I)^{-1}B, (A - s_2 I)^{-1}B, \dots, (A - s_m I)^{-1}B])$$

cf. General discussion in Antoulas, 2005.

Many contributions:

- Gallivan, Grimme, Van Dooren (1996–, ad-hoc poles)
- Penzl (1999-2000, ADI shifts - preprocessing, Ritz values)
- ....
- Sabino (2006 - tuning within preprocessing)
- IRKA – Gugercin, Antoulas, Beattie (2008)
- Druskin, Lieberman, Simoncini, Zaslavski (adaptive greedy procedure)
- Güttel, Knizhnerman (black-box for matrix functions)
- ....

Alternating Direction Implicit iteration (ADI) - Wachspress

(see, e.g., Li 2000, Penzl 2000)

$$\begin{aligned} X_0 = 0,\ X_j &= -2p_j(A+p_jI)^{-1}BB^\top(A+p_jI)^{-\top} \quad j=1,\ldots,\ell \\ &+(A+p_jI)^{-1}(A-p_jI)X_{j-1}(A-p_jI)^\top(A+p_jI)^{-\top} \end{aligned}$$

with

$$\phi_\ell(t) = \prod_{j=1}^{\ell}(t-p_j), \quad \{p_1,\ldots,p_\ell\} = \operatorname{argmin} \max_{t\in\Lambda(A)} \left|\frac{\phi_\ell(t)}{\phi_\ell(-t)}\right|$$

Implementation aspects: Benner, Saak, Quintana-Ortì[2], ...

Convergence depends on choice of poles $\{p_j\}$

More advanced approach: Galerkin-Projection Accelerated ADI (Benner, Saak, tr 2010)

ADI and Rational Krylov subspaces

Let $B = b$ (vector). Main consideration (see, e.g., Li, Wright 2000)

$$\mathrm{col}(X_m^{(ADI)}) \in K_m(A, b, \mathbf{s})$$

and also, for $U_m = [(A - s_1 I)^{-1} b, \ldots, (A - s_m I)^{-1} b]$,

$$X_m^{(ADI)} = U_m \boldsymbol{\alpha}^{-1} U_m^*$$

with $\boldsymbol{\alpha}$ Cauchy matrix

ADI and Rational Krylov subspaces

Let $B = b$ (vector). Main consideration (see, e.g., Li, Wright 2000)

$$\mathrm{col}(X_m^{(ADI)}) \in K_m(A, b, \mathbf{s})$$

and also, for $U_m = [(A - s_1 I)^{-1} b, \ldots, (A - s_m I)^{-1} b]$,

$$X_m^{(ADI)} = U_m \boldsymbol{\alpha}^{-1} U_m^*$$

with $\boldsymbol{\alpha}$ Cauchy matrix

Equivalence between ADI and RKSM:

ADI coincides with the Galerkin solution X_m in Rational Krylov space if and only if

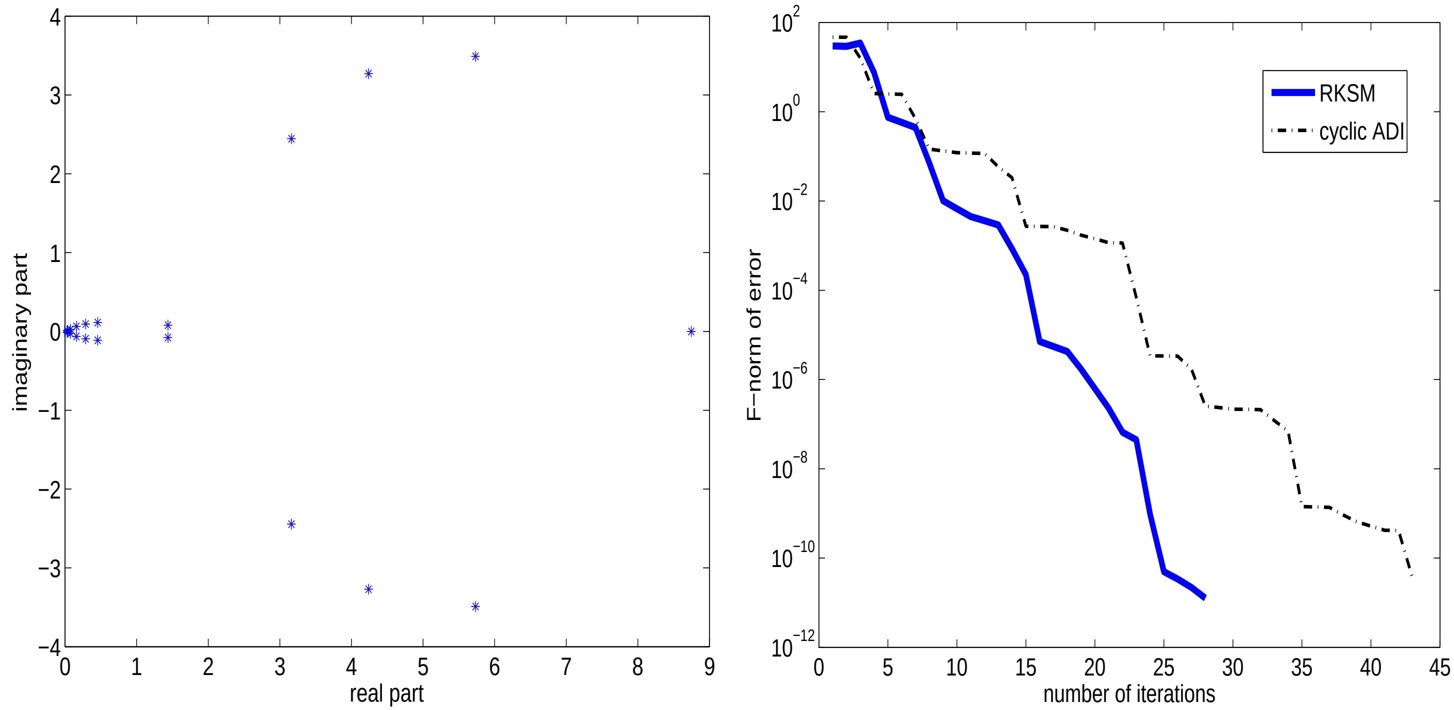
$$s_j = -\bar{\lambda}_j$$

where $\lambda_j = \mathrm{eigs}(V_m^* A V_m)$ Ritz values (suitably ordered)

Druskin, Knizhnerman, S. '11, Beckermann '11, Flagg '09, Gugercin, Flagg '12

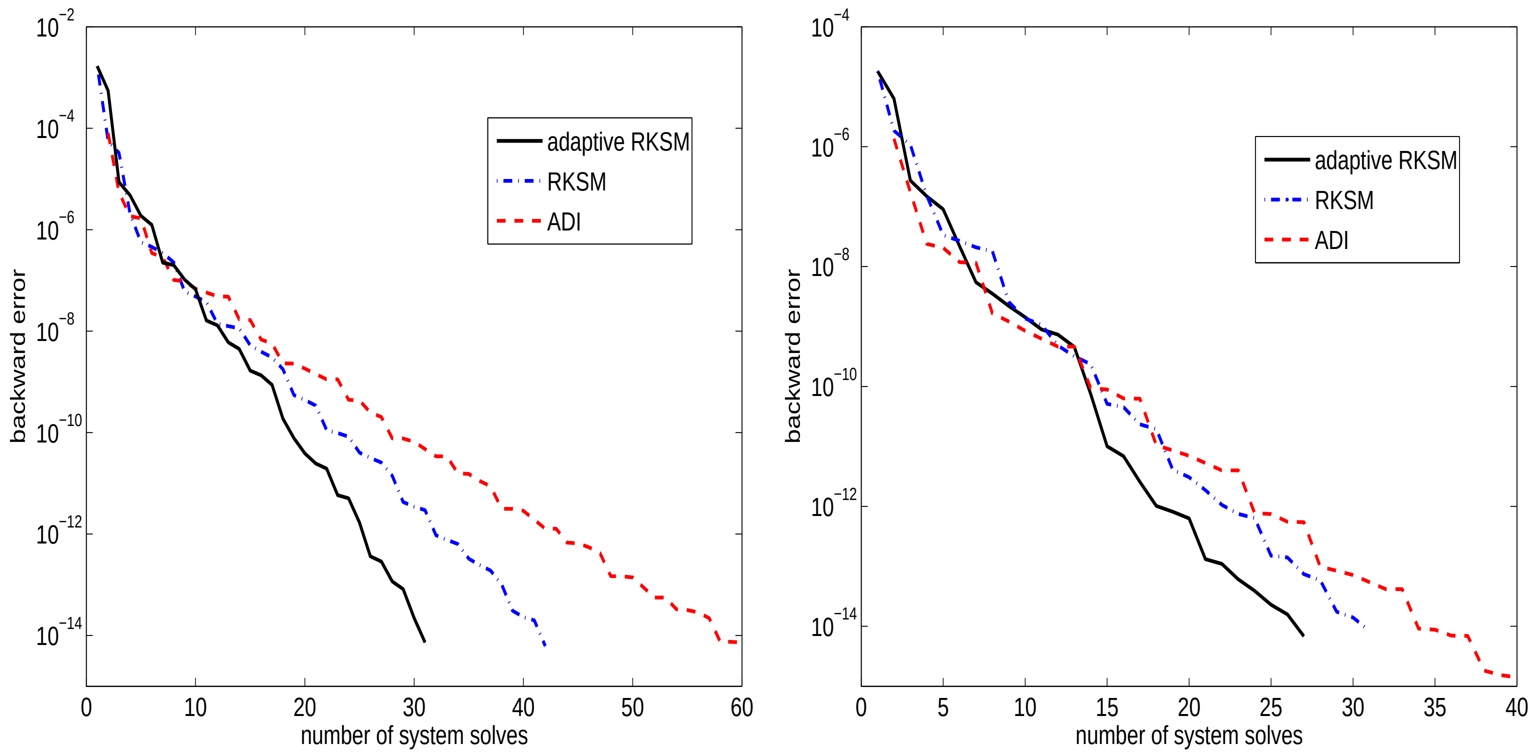
Typical behavior of ADI and generic RKSM for the same poles

Operator: $L(u) = -\Delta u + (50xu_x)_x + (50yu_y)_y$ on $[0,1]^2$



Same non-optimal 20 poles, repeated cyclically.

Expected performance (from Oberwolfach Collection)



Left: `rail` problem, A symmetric.

Right: `flow_meter_model_v0.5` problem, A nonsymmetric.

ADI and RKSM use 10 non-optimal poles cyclically (computed a-priori with `lyapack`, Penzl 2000)

A minimal residual approach for $AX + XA^\top + BB^\top = 0, \qquad B = b$

$X \approx X_m^{MR} = V_m Y_m^{MR} V_m^\top$ where

$$Y_m^{MR} = \arg \min_{Y_m \in \mathbb{R}^{m\times m}} \|AV_mY_mV_m^\top + V_mY_mV_m^\top A^\top + BB^\top\|_F.$$

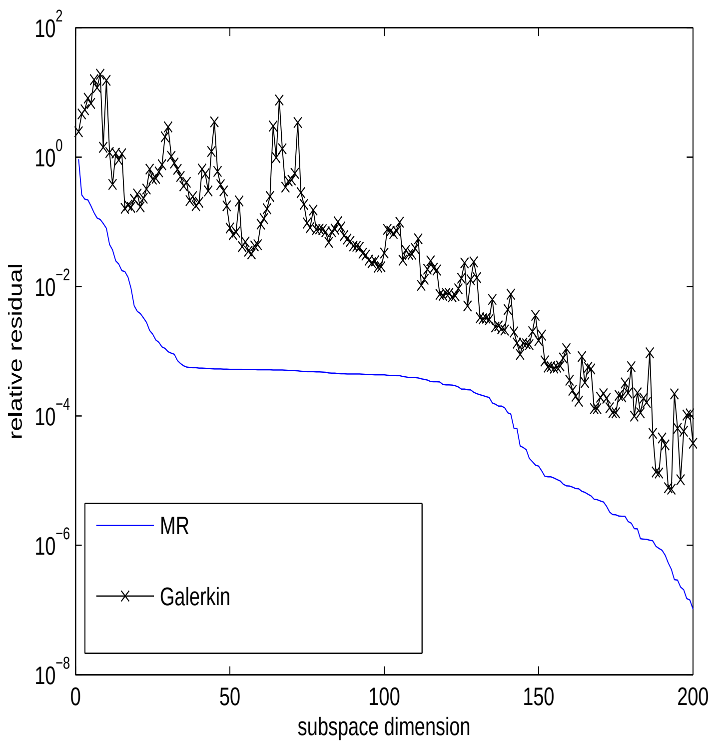
Equivalent to a "Petrov-Galerkin" condition on residual matrix

Hu-Reichel (1992), Jaimoukha-Kasenally (1994), Lin-Simoncini (2013)

A minimal residual approach for $AX + XA^\top + BB^\top = 0, \qquad B = b$

$X \approx X_m^{MR} = V_m Y_m^{MR} V_m^\top$ where

$$Y_m^{MR} = \arg \min_{Y_m \in \mathbb{R}^{m\times m}} \|AV_mY_mV_m^\top + V_mY_mV_m^\top A^\top + BB^\top\|_F.$$



An extreme example:

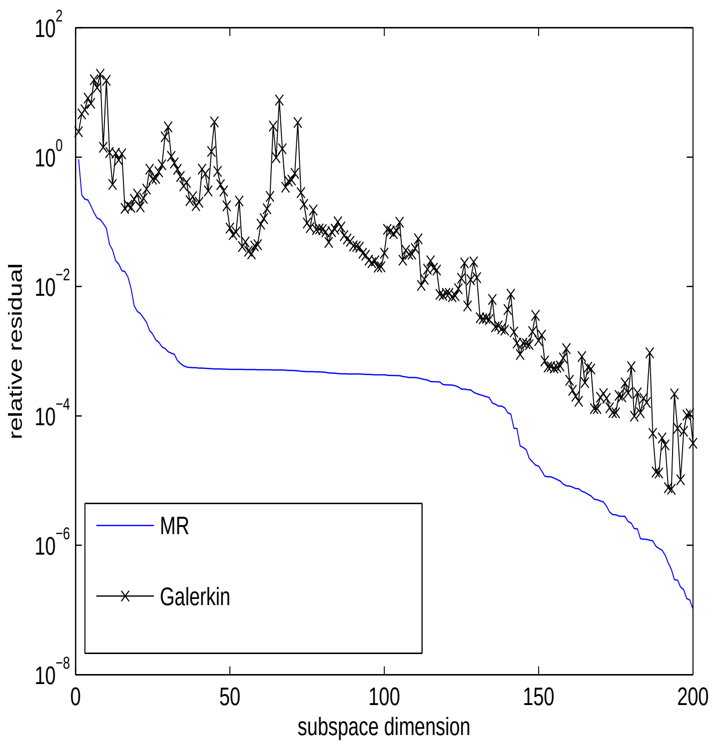
the ISS matrix

K_m = Rational Krylov space

A minimal residual approach for $AX + XA^\top + BB^\top = 0, \qquad B = b$

$X \approx X_m^{MR} = V_m Y_m^{MR} V_m^\top$ where

$$Y_m^{MR} = \arg \min_{Y_m \in \mathbb{R}^{m\times m}} \|AV_mY_mV_m^\top + V_mY_mV_m^\top A^\top + BB^\top\|_F.$$



An extreme example:

the ISS matrix

K_m = Rational Krylov space

Cause: Origin close or inside field of values of A

The numerical solution of the minimum residual problem

If V_m is such that $AV_m = \check{V}_{m+1}\underline{H}_m$ and $B = V_1 R_B$

($V_m, \check{V}_{m+1}$ with orth.columns, $\underline{H}_m = [H_m; h]$)

Then

$$\min_{Y_m} \|AV_m Y_m V_m^\top + V_m Y_m V_m^\top A^\top + BB^\top\|_F$$

is equivalent to

$$\min_{Y_m} \|\underline{H}_m Y \left[I, \, 0\right] + \begin{bmatrix} I \\ 0 \end{bmatrix} Y \underline{H}_m^\top + \begin{bmatrix} R_B R_B^\top & 0 \\ 0 & 0 \end{bmatrix} \|_F$$

$\Rightarrow$ Reduced *matrix* least squares problem

The numerical solution of the matrix least squares problem

$$Y_m^{MR} = \arg\min_{Y_m} \| \underline{H}_m Y \begin{bmatrix} I, & 0 \end{bmatrix} + \begin{bmatrix} I \\ 0 \end{bmatrix} Y \underline{H}_m^\top + \begin{bmatrix} R_B R_B^\top & 0 \\ 0 & 0 \end{bmatrix} \|_F$$

Numerical solution strategies: (Lin-Simoncini, '13)

- Kronecker formulation $O(m^4)$: $\min_y \|e_1 \beta_0^2 + \mathcal{H} y\|$,
  $\mathcal{H} = \underline{H} \otimes \underline{I} + \underline{I} \otimes \underline{H}$
- Revise Kronecker formulation to exploit structure, $O(m^3)$
- Iterative method for normal matrix equation:

$$\underline{H}^\top \underline{H} \mathbf{Y} + \mathbf{Y} \underline{H}^\top \underline{H} + H \mathbf{Y} H + H^\top \mathbf{Y} H^\top = C_{NE}$$

Computational Cost comparable to that of Galerkin method

Multiterm linear matrix equation

$$A_1 X B_1 + A_2 X B_2 + \ldots + A_\ell X B_\ell = C$$

Applications:

- Matrix least squares
- Control
- Stochastic PDEs
- ...

Multiterm linear matrix equation

$$A_1 X B_1 + A_2 X B_2 + \ldots + A_\ell X B_\ell = C$$

Applications:

- Matrix least squares
- Control
- Stochastic PDEs
- ...

Main device: Kronecker formulation

$$\left(B_1^\top \otimes A_1 + \ldots + B_\ell^\top \otimes A_\ell\right) x = c$$

Iterative methods: matrix-matrix multiplications and rank truncation

(Benner, Breiten, Bouhamidi, Chehab, Damm, Grasedyck, Jbilou, Kressner, Matthies, Onwunta, Raydan, Stoll, Tobler, Zander, ...)

Bilinear systems of linear matrix equation

$$A_1 X + Y B_1 = C_1$$
$$A_2 X + Y B_2 = C_2$$

...very few numerical procedures available.

A "special" case: Constrained Sylvester equation

$$\begin{aligned} A_1 X + X A_2 - Y C &= 0 \\ X B &= 0 \end{aligned} \tag{1}$$

(X, Y) unknown matrices

(1) constraint

Constrained Sylvester equation

$$A_1 X + X A_2 - Y C = 0$$
$$X B = 0$$

Typically: B low column rank, C low row rank

New formulation: **unconstrained** Sylvester eqn (Shank-Simoncini '13)

For *any* $Y_2 \neq 0$, matrix X solves the linear matrix equation

$$A_1 X + X A_2 (I - P)\Pi = Y_2 Q_2^T C \Pi$$

while $Y = [Y_1, Y_2]$, with $Y_1 = X A_2 U_1 R^{-1}$, where

P: projector onto Range(B) and orthogonal to Range$(C^T C B)$

Π: orth. projector onto null$(B^\top)$

U_1: spans B

R: such that $C U_1 = [Q_1, Q_2][R; 0]$

Numerics for the unconstrained Sylvester eqn (Shank-Simoncini '13)

For *any* $Y_2 \neq 0$, matrix X solves the linear equation

$$A_1 X + X A_2 (I - P)\Pi = Y_2 Q_2^T C \Pi$$

Computational considerations:

- Choose Y_2 so that $Y_2 Q_2^T C \Pi$ is low rank !

$$A_1 X + X \underbrace{A_2 (I - P)\Pi}_{\widehat{A}_2} = d_1 d_2^\top \qquad \text{with } d_1 d_2^\top = Y_2 Q_2^T C \Pi$$

Numerics for the unconstrained Sylvester eqn (Shank-Simoncini '13)

For *any* $Y_2 \neq 0$, matrix X solves the linear equation

$$A_1 X + X A_2 (I - P)\Pi = Y_2 Q_2^T C \Pi$$

Computational considerations:

- Choose Y_2 so that $Y_2 Q_2^T C\Pi$ is low rank !

$$A_1 X + X \underbrace{A_2(I-P)\Pi}_{\widehat{A}_2} = d_1 d_2^\top \qquad \text{with } d_1 d_2^\top = Y_2 Q_2^T C \Pi$$

- $\widehat{A}_2$ singular and not explicitly available!

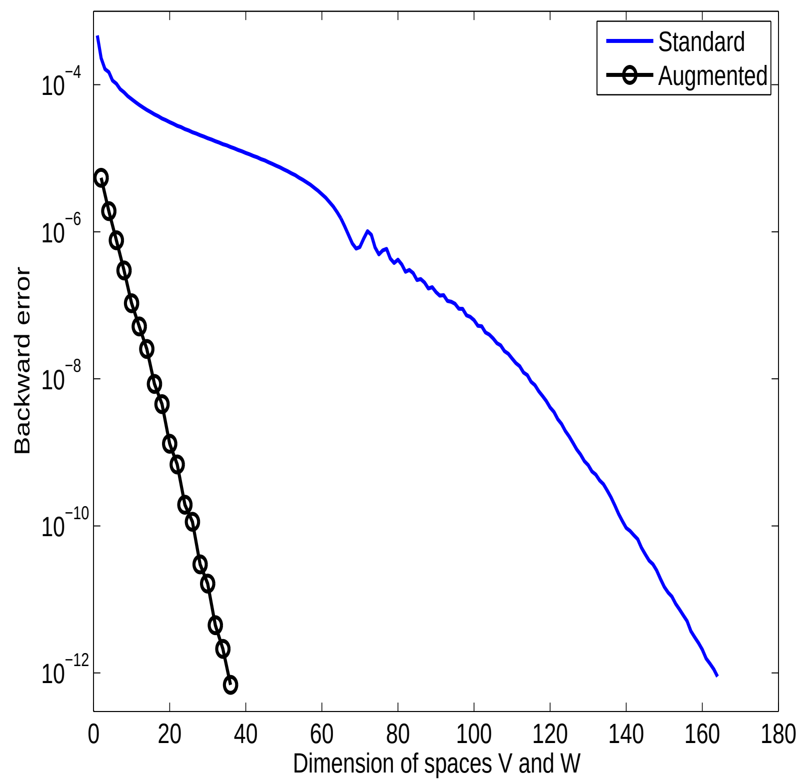
  For building a rational Krylov subspace, one would have to apply

$$(\widehat{A}_2 + \sigma I)^{-1} = (A_2(I-P)\Pi + \sigma I)^{-1}$$

  which is not available...

  $\Rightarrow$ effective alternatives lead to "augmented" methods

A sample plot



Convergence history of standard and augmented Krylov solvers

A_2: FLOW dataset, $n = 9669$, $\quad A_1$: Laplace operator, $n = 9604$

Other related matrix equations

More "exotic" linear matrix equations

- Sylvester-like

$$BX + f(X)A = C$$

typically (but not only!)

$$f(X) = \bar{X}, \quad f(X) = X^\top, \quad \text{or} \quad f(X) = X^*$$

(Bevis, Braden, Byers, Chiang, De Terán, Dopico, Duan, Feng, Guillery, Hall, Hartwig, Ikramov, Kressner, Montealegre, Reyes, Schröder, Vorntsov, Watkins, Wu, ...)

Conclusions

- Large advances in solving really large linear matrix equations
- Second order (matrix) challenges rely on strength and maturity of linear system solvers

Reference for a survey:

⋆ V. S., *Computational methods for linear matrix equations*,

March 2013, Submitted (currently under revision)

available at `www.dm.unibo.it/~simoncin`