New Frontiers in Numerical Analysis and Scientific Computing
A conference on the occasion of Lothar Reichel's 60th birthday and on the 20th anniversary of ETNA
Kent State University
April 19–20, 2013

ADI for Sylvester Equations Lothars Contributions and New Results

Peter Benner
with (a lot of help from) Tobias Breiten and Patrick Kürschner
in parts joint work with Heike Faßbender and Jens Saak

Computational Methods in Systems and Control Theory
Max Planck Institute for Dynamics of Complex Technical Systems
Magdeburg, Germany

20 Years of ETNA

(20 years of Peter Benner in science . . .)

Electronic Transactions on Numerical Analysis (ETNA)

Volume 1 (1993)

60 Years of Lothar

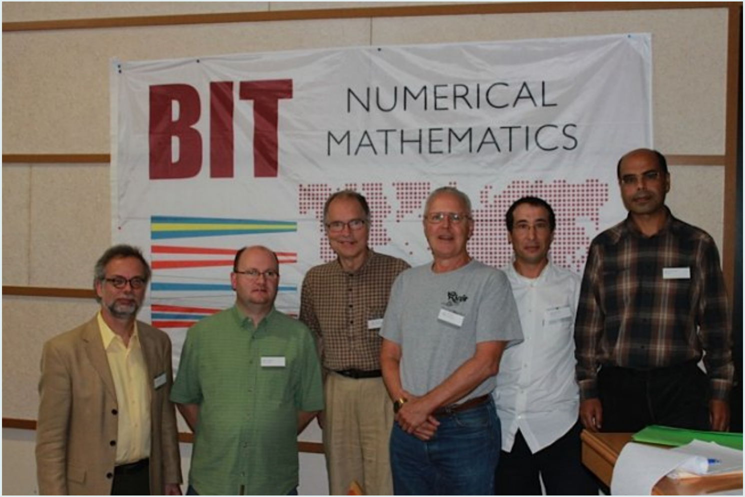
Some time ago. . .

60 Years of Lothar

Not so long ago . . . with friends at Prof. Varga's 80th

60 Years of Lothar

Just a little bit ago . . . at Luminy 2012



60 Years of Lothar

. . . and always vibrant!

Lothar and Sylvester Equations

- Hu, D. Y.; **Reichel, L.**, *Krylov-subspace methods for the Sylvester equation*. Linear Algebra Appl. 172 (1992), 283-313.

  Investigates Galerkin and Minimal Residual methods for Sylvester equations.

  Most (83 in Scopus) or 2nd most (41 in MathSciNet, 151 in Google Scholar) frequently cited of Lothar's papers!

Lothar and Sylvester Equations

- Hu, D. Y.; **Reichel, L.**, *Krylov-subspace methods for the Sylvester equation*. LINEAR ALGEBRA APPL. 172 (1992), 283-313.

  Investigates Galerkin and Minimal Residual methods for Sylvester equations.

  Most (83 in Scopus) or 2nd most (41 in MathSciNet, 151 in Google Scholar) frequently cited of Lothar's papers!

- Levenberg, N.; **Reichel, L.**, *A generalized ADI iterative method.* NUMER. MATH. 66 (1993), 215-233.

  Proposes a variable alternating directions implicit (ADI) scheme, allows bias towards one direction.

Lothar and Sylvester Equations

- Hu, D. Y.; **Reichel, L.**, *Krylov-subspace methods for the Sylvester equation*. LINEAR ALGEBRA APPL. 172 (1992), 283-313.

  Investigates Galerkin and Minimal Residual methods for Sylvester equations.

  Most (83 in Scopus) or 2nd most (41 in MathSciNet, 151 in Google Scholar) frequently cited of Lothar's papers!

- Levenberg, N.; **Reichel, L.**, *A generalized ADI iterative method.* NUMER. MATH. 66 (1993), 215-233.

  Proposes a variable alternating directions implicit (ADI) scheme, allows bias towards one direction.

- Calvetti, D.; **Reichel, L.**, *Application of ADI iterative methods to the restoration of noisy images*. SIAM J. MATRIX ANAL. APPL. 17 (1996), 165-186.

  Applies the variable ADI scheme to the image restoration problem.

Lothar and Sylvester Equations

- Hu, D. Y.; **Reichel, L.**, *Krylov-subspace methods for the Sylvester equation*. Linear Algebra Appl. 172 (1992), 283-313.

  Investigates Galerkin and Minimal Residual methods for Sylvester equations.

  Most (83 in Scopus) or 2nd most (41 in MathSciNet, 151 in Google Scholar) frequently cited of Lothar's papers!

- Levenberg, N.; **Reichel, L.**, *A generalized ADI iterative method.* Numer. Math. 66 (1993), 215-233.

  Proposes a variable alternating directions implicit (ADI) scheme, allows bias towards one direction.

- Calvetti, D.; **Reichel, L.**, *Application of ADI iterative methods to the restoration of noisy images*. SIAM J. Matrix Anal. Appl. 17 (1996), 165-186.

  Applies the variable ADI scheme to the image restoration problem.

- Calvetti, D.; Levenberg, N.; **Reichel, L.**, *Iterative methods for $X - AXB = C$.* J. Comput. Appl. Math. 86 (1997), 73-101.

  Analyzes the variable ADI scheme for the special Sylvester equation arising in image restoration.

Lothar and Sylvester Equations

- Hu, D. Y.; **Reichel, L.**, *Krylov-subspace methods for the Sylvester equation*. LINEAR ALGEBRA APPL. 172 (1992), 283-313.

  Investigates Galerkin and Minimal Residual methods for Sylvester equations.

  Most (83 in Scopus) or 2nd most (41 in MathSciNet, 151 in Google Scholar) frequently cited of Lothar's papers!

- Levenberg, N.; **Reichel, L.**, *A generalized ADI iterative method.* NUMER. MATH. 66 (1993), 215-233.

  Proposes a variable alternating directions implicit (ADI) scheme, allows bias towards one direction.

- Calvetti, D.; **Reichel, L.**, *Application of ADI iterative methods to the restoration of noisy images*. SIAM J. MATRIX ANAL. APPL. 17 (1996), 165-186.

  Applies the variable ADI scheme to the image restoration problem.

- Calvetti, D.; Levenberg, N.; **Reichel, L.**, *Iterative methods for $X - AXB = C$*. J. COMPUT. APPL. MATH. 86 (1997), 73-101.

  Analyzes the variable ADI scheme for the special Sylvester equation arising in image restoration.

- Calvetti, D.; Lewis, B.; **Reichel, L.**, *On the solution of large Sylvester-observer equations*. NUMER. LINEAR ALGEBRA APPL. 8 (2001), 435-451.

  Analyzes a method to solve the Sylvester-observer equation suggested by Y. Saad and B. Datta, leading to suggestions for parameter choices.

Large-Scale Sylvester Equations

Problem Setting

Sylvester Equations

Find $X \in \mathbb{R}^{n \times m}$ solving

$$AX - XB = FG^T,$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{m \times m}$, $F \in \mathbb{R}^{n \times r}$, $G \in \mathbb{R}^{m \times r}$.

Applications (just for now):

- control theory (e.g., Luenberger observer, model reduction, ...),
- image restoration,
- fluid queue models, solving transport equations (Newton step for solving nonsymmetric Riccati equations)
- ...

Large-Scale Sylvester Equations

Properties & Algorithms

Sylvester Equations

$$AX - XB = FG^T$$

- Unique solvability ensured if

$$\Lambda(A) \cap \Lambda(B) = \emptyset.$$

- Reduces to Lyapunov equation if $B = A^T$, $G = F$.
- Algorithms for small to moderately sized problems based on
  - Schur, spectral, or Hessenberg decompositions of A and B (Bartels-Stewart,...),
  - sign function iteration (Roberts,...),
  - **alternating directions implicit (ADI) iteration** (Wachspress,...).

Large-Scale Sylvester Equations

Low-rank Phenomena

Sylvester Equations

$$AX - XB = FG^T$$

In this talk:

- both $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times m}$ large and sparse,
- $F \in \mathbb{R}^{n\times r}$, $G \in \mathbb{R}^{m\times r}$ with $r \ll n, m$.

Large-Scale Sylvester Equations

Low-rank Phenomena

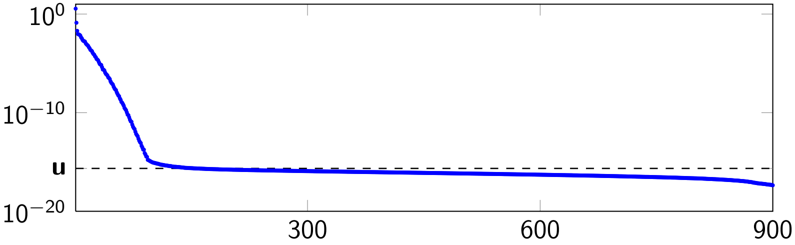
Sylvester Equations

$$AX - XB = FG^T$$

In this talk:

- both $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times m}$ large and sparse,
- $F \in \mathbb{R}^{n\times r}$, $G \in \mathbb{R}^{m\times r}$ with $r \ll n, m$.

Plot of singular values of solution X for artificial example with $n = 1600$, $m = 900$ and $r = 4$.



Large-Scale Sylvester Equations

Low-rank Phenomena

Sylvester Equations

$$AX - XB = FG^T$$

In this talk:

- both $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$ large and sparse,
- $F \in \mathbb{R}^{n \times r}$, $G \in \mathbb{R}^{m \times r}$ with $r \ll n, m$.

Observation: X has small numerical rank

[Penzl '99, Ant./Sor./Zhou '02, Grasedyck '04]

$$\operatorname{rank}(X, \tau) = f \ll \min(n, m)$$

$\rightsquigarrow$ Compute low-rank solution factors $Z \in \mathbb{R}^{n \times f}$, $Y \in \mathbb{R}^{m \times f}$, $D \in \mathbb{R}^{f \times f}$, $f \ll \min(n, m)$ such that $X \approx ZDY^T$.

Large-Scale Sylvester Equations

Low-rank Phenomena & methods

As for Lyapunov equations, there are mainly three classes of methods for computing Z, Y, D

1. (Petrov-)Galerkin-projection methods based on (rational) Krylov subspaces, e.g.,

$$\operatorname{span}\{Z\} \subseteq \mathcal{K}(A, F, k), \quad \operatorname{span}\{Y\} \subseteq \mathcal{K}(B, G, k)$$
$$\text{and } D \text{ solves } (Z^T AZ)D - D(Y^T BY) = (Z^T F)(G^T Y).$$

[Jbilou et al '02,...]

2. iterative Krylov subspace methods for equivalent linear system

$$\left(I_n \otimes A - B^T \otimes I_m\right) \operatorname{vec}(X) = \operatorname{vec}(FG^T).$$

3. Smith & alternating directions implicit (ADI) type methods.

[B./Li/Truhar '09]

Low-rank Sylvester ADI

Derivation

Continuous and discrete time Sylvester equations

Let $\alpha \neq \beta$ with $\alpha \notin \Lambda(B)$, $\beta \notin \Lambda(A)$, then

$$\underbrace{AX - XB = FG^T}_{\text{cont.-time Sylv. Eq.}} \quad \Leftrightarrow \quad \underbrace{X = \mathcal{A}X\mathcal{B} + (\beta - \alpha)\mathcal{F}\mathcal{G}^H}_{\text{disc.-time Sylv. Eq.}},$$

where

$$\begin{aligned}
\mathcal{A} &:= (A - \beta I_n)^{-1}(A - \alpha I_n), \\
\mathcal{B} &:= (B - \alpha I_m)^{-1}(B - \beta I_m), \\
\mathcal{F} &:= (A - \beta I_n)^{-1}F, \\
\mathcal{G} &:= (B - \alpha I_m)^{-H}G.
\end{aligned}$$

Low-rank Sylvester ADI

Derivation

The equivalent discrete-time Sylvester equation

$$X = \mathcal{A}X\mathcal{B} + (\beta - \alpha)\mathcal{F}\mathcal{G}^H$$

motivates the iteration for $k \geq 1$

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k)\mathcal{F}_k \mathcal{G}_k^H,$$

where

$$\begin{aligned}
\mathcal{A}_k &:= (A - \beta_k I_n)^{-1}(A - \alpha_k I_n), \\
\mathcal{B}_k &:= (B - \alpha_k I_m)^{-1}(B - \beta_k I_m), \\
\mathcal{F}_k &:= (A - \beta_k I_n)^{-1} F, \\
\mathcal{G}_k &:= (B - \alpha_k I_m)^{-H} G.
\end{aligned}$$

for $\alpha_k \neq \beta_k$ with $\alpha_k \notin \Lambda(B)$, $\beta_k \notin \Lambda(A)$, $X_0 \in \mathbb{R}^{n \times m}$.

Low-rank Sylvester ADI

Derivation

The equivalent discrete-time Sylvester equation

$$X = \mathcal{A}X\mathcal{B} + (\beta - \alpha)\mathcal{F}\mathcal{G}^H$$

motivates the **ADI** iteration for $k \geq 1$ [Wachspress '88]

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k)\mathcal{F}_k \mathcal{G}_k^H,$$

where

$$\begin{aligned}
\mathcal{A}_k &:= (A - \beta_k I_n)^{-1}(A - \alpha_k I_n), \\
\mathcal{B}_k &:= (B - \alpha_k I_m)^{-1}(B - \beta_k I_m), \\
\mathcal{F}_k &:= (A - \beta_k I_n)^{-1} F, \\
\mathcal{G}_k &:= (B - \alpha_k I_m)^{-H} G.
\end{aligned}$$

for $\alpha_k \neq \beta_k$ with $\alpha_k \notin \Lambda(B)$, $\beta_k \notin \Lambda(A)$, $X_0 \in \mathbb{R}^{n \times m}$.

Low-rank Sylvester ADI

Derivation

Sylvester ADI iteration

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k) \mathcal{F}_k \mathcal{G}_k^H$$

Now set $X_0 = 0$, $X_k = Z_k D_k Y_k^H$, and observe

$$X_1 = \mathcal{A}_1 X_0 \mathcal{B}_1 + (\beta_1 - \alpha_1) \mathcal{F}_1 \mathcal{G}_1^H$$

Low-rank Sylvester ADI

Derivation

Sylvester ADI iteration

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k)\mathcal{F}_k \mathcal{G}_k^H$$

Now set $X_0 = 0$, $X_k = Z_k D_k Y_k^H$, and observe

$$X_1 = (\beta_1 - \alpha_1)(A - \beta_1 I_n)^{-1} F G^T (B - \alpha_1 I_m)^{-1}$$

$$\Rightarrow Z_1 = (A - \beta_1 I_n)^{-1} F, \quad Y_1 = (B - \alpha_1 I_m)^{-H} G, \quad D_1 = (\beta_1 - \alpha_1) I_r.$$

Low-rank Sylvester ADI

Derivation

Sylvester ADI iteration

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k) \mathcal{F}_k \mathcal{G}_k^H$$

Now set $X_0 = 0$, $X_k = Z_k D_k Y_k^H$, and observe

$$X_2 = \mathcal{A}_2 X_1 \mathcal{B}_2 + (\beta_2 - \alpha_2) \mathcal{F}_2 \mathcal{G}_2^H$$

Low-rank Sylvester ADI

Derivation

Sylvester ADI iteration

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k)\mathcal{F}_k \mathcal{G}_k^H$$

Now set $X_0 = 0$, $X_k = Z_k D_k Y_k^H$, and observe

$$\begin{aligned} X_2 = {} & (\beta_1 - \alpha_1)(A - \beta_2 I_n)^{-1}(A - \alpha_2 I_n)(A - \beta_1 I_n)^{-1} F \times \\ & \times G^T (B - \alpha_1 I_m)^{-1}(B - \beta_2 I_m)(B - \alpha_2 I_m)^{-1} \\ & + (\beta_2 - \alpha_2)(A - \beta_2 I_n)^{-1} F G^T (B - \alpha_2 I_m)^{-1} \end{aligned}$$

Low-rank Sylvester ADI

Derivation

Sylvester ADI iteration

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k)\mathcal{F}_k \mathcal{G}_k^H$$

Now set $X_0 = 0$, $X_k = Z_k D_k Y_k^H$, and observe

$$\begin{aligned} X_2 = (\beta_2 - \alpha_2)&(A - \beta_1 I_n)^{-1}(A - \alpha_1 I_n)(A - \beta_2 I_n)^{-1} F \times \\ &\times G^T (B - \alpha_2 I_m)^{-1}(B - \beta_1 I_m)(B - \alpha_1 I_m)^{-1} \\ &+ (\beta_1 - \alpha_1)(A - \beta_1 I_n)^{-1} F G^T (B - \alpha_1 I_m)^{-1} \end{aligned}$$

Low-rank Sylvester ADI

Derivation

Sylvester ADI iteration

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k)\mathcal{F}_k \mathcal{G}_k^H$$

Now set $X_0 = 0$, $X_k = Z_k D_k Y_k^H$, and observe

$$\begin{aligned} X_2 = {} & (\beta_2 - \alpha_2)(A - \beta_2 I_n)^{-1}(A - \alpha_1 I_n)(A - \beta_1 I_n)^{-1} F \times \\ & \times G^T (B - \alpha_1 I_m)^{-1}(B - \beta_1 I_m)(B - \alpha_2 I_m)^{-1} \\ & + (\beta_1 - \alpha_1)(A - \beta_1 I_n)^{-1} F G^T (B - \alpha_1 I_m)^{-1} \end{aligned}$$

Low-rank Sylvester ADI

Derivation

Sylvester ADI iteration

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k) \mathcal{F}_k \mathcal{G}_k^H$$

Now set $X_0 = 0$, $X_k = Z_k D_k Y_k^H$, and observe

$$\begin{aligned} X_2 &= (\beta_2 - \alpha_2)(A - \beta_2 I_n)^{-1}(A - \alpha_1 I_n)(A - \beta_1 I_n)^{-1} F \times \\ &\quad \times G^T (B - \alpha_1 I_m)^{-1}(B - \beta_1 I_m)(B - \alpha_2 I_m)^{-1} \\ &\quad + (\beta_1 - \alpha_1)(A - \beta_1 I_n)^{-1} F G^T (B - \alpha_1 I_m)^{-1} \\ &= (\beta_2 - \alpha_2) \underbrace{(A - \alpha_1 I_n)(A - \beta_2 I_n)^{-1} Z_1}_{=: V_2} \times \\ &\quad \times \underbrace{Y_1^H (B - \alpha_2 I_m)^{-1}(B - \beta_1 I_m)}_{=: W_2^H} \quad + (\beta_1 - \alpha_1) Z_1 Y_1^H \end{aligned}$$

Low-rank Sylvester ADI

Derivation

Sylvester ADI iteration

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k) \mathcal{F}_k \mathcal{G}_k^H$$

Now set $X_0 = 0$, $X_k = Z_k D_k Y_k^H$, and observe

$$\begin{aligned}
X_2 &= (\beta_2 - \alpha_2)(A - \beta_2 I_n)^{-1}(A - \alpha_1 I_n)(A - \beta_1 I_n)^{-1} F \times \\
&\quad \times G^T (B - \alpha_1 I_m)^{-1}(B - \beta_1 I_m)(B - \alpha_2 I_m)^{-1} \\
&\quad + (\beta_1 - \alpha_1)(A - \beta_1 I_n)^{-1} F G^T (B - \alpha_1 I_m)^{-1} \\
&= (\beta_2 - \alpha_2) \underbrace{(A - \alpha_1 I_n)(A - \beta_2 I_n)^{-1} Z_1}_{=: V_2} \times \\
&\quad \times \underbrace{Y_1^H (B - \alpha_2 I_m)^{-1}(B - \beta_1 I_m)}_{=: W_2^H} \quad + (\beta_1 - \alpha_1) Z_1 Y_1^H
\end{aligned}$$

$$\Rightarrow \; Z_2 = [Z_1, V_2], \quad Y_2 = [Y_1, W_2], \quad D = \operatorname{diag}\left(D_1, (\beta_2 - \alpha_2) I_r\right)$$

Low-rank Sylvester ADI

Algorithm [B. '05, Li/Truhar '08, B./Li/Truhar '09]

Algorithm 1 Low-rank Sylvester ADI (factored ADI)

Input: Matrices defining $AX - XB = FG^T$ and shift parameters $\{\alpha_1, \ldots, \alpha_{k_{\max}}\}$, $\{\beta_1, \ldots, \beta_{k_{\max}}\}$.

Output: Z, Y, D such that $ZDY^H \approx X$.

1: $Z_1 = V_1 = (A - \beta_1 I_n)^{-1} F$
2: $Y_1 = W_1 = (B - \alpha_1 I_m)^{-H} G$
3: $D_1 = (\beta_1 - \alpha_1) I_r$
4: **for** $k = 2, \ldots, k_{\max}$ **do**
5: $\quad V_k = V_{k-1} + (\beta_k - \alpha_{k-1})(A - \beta_k I_n)^{-1} V_{k-1}$.
6: $\quad W_k = W_{k-1} + \overline{(\alpha_k - \beta_{k-1})}(B - \alpha_k I_n)^{-H} W_{k-1}$.
7: **end for**
8: Update solution factors

$$Z_k = [Z_{k-1}, V_k], \quad Y_k = [Y_{k-1}, W_k], \quad D_k = \operatorname{diag}\left(D_{k-1}, (\beta_k - \alpha_k) I_r\right).$$

Low-rank Sylvester ADI

The Residual

Sylvester ADI iteration

$$X_k = \mathcal{A}_k X_{k-1} \mathcal{B}_k + (\beta_k - \alpha_k)\mathcal{F}_k \mathcal{G}_k^H$$

yields

$$X_k - X = \mathcal{A}_k (X_{k-1} - X)\mathcal{B}_k = \Big[\prod_{j=1}^{k} \mathcal{A}_j\Big](X_0 - X)\Big[\prod_{j=1}^{k} \mathcal{B}_j\Big].$$

Hence, the Sylvester residual is given by

$$\begin{aligned}\mathcal{S}(X_k) &= AX_k - X_k B - FG^T = A(X_k - X) - (X_k - X)B \\ &= \Big[\prod_{j=1}^{k} \mathcal{A}_j\Big]\mathcal{S}(X_0)\Big[\prod_{j=1}^{k} \mathcal{B}_j\Big].\end{aligned}$$

Low-rank Sylvester ADI

The Residual

For $X_0 = 0$ (as in the low-rank Sylvester ADI) we have

$$\mathcal{S}(X_k) = AX_k - X_k B - FG^T = -\hat{V}_k \hat{W}_k,$$

$$\hat{V}_k := \Big[\prod_{j=1}^{k} \mathcal{A}_j\Big] F, \quad \hat{W}_k := \Big[\prod_{j=1}^{k} \mathcal{B}_j\Big]^H G.$$

Lemma

$\mathrm{rank}(\mathcal{S}(X_k)) \leq r$ (and semidefinite in the Lyapunov case).

Low-rank Sylvester ADI

The Residual

For $X_0 = 0$ (as in the low-rank Sylvester ADI) we have

$$\mathcal{S}(X_k) = AX_k - X_k B - FG^T = -\hat{V}_k \hat{W}_k,$$

$$\hat{V}_k := \Big[\prod_{j=1}^{k} \mathcal{A}_j\Big] F, \quad \hat{W}_k := \Big[\prod_{j=1}^{k} \mathcal{B}_j\Big]^H G.$$

Lemma

$\mathrm{rank}(\mathcal{S}(X_k)) \leq r$ (and semidefinite in the Lyapunov case).

Moreover,

$$\begin{aligned} V_k &= (A - \alpha_{k-1} I_n)(A - \beta_k I_n)^{-1} V_{k-1} \\ &= (A - \beta_k I_n)^{-1}(A - \alpha_{k-1} I_n)(A - \alpha_{k-2} I_n)(A - \beta_{k-1} I_n)^{-1} V_{k-2} \\ &= (A - \beta_k I_n)^{-1} \mathcal{A}_{k-1}(A - \alpha_{k-2} I_n) V_{k-2} \\ &= \ldots = (A - \beta_k I_n)^{-1} \Big[\prod_{j=1}^{k-1} \mathcal{A}_j\Big] F \quad \Rightarrow \quad \hat{V}_k = (A - \alpha_k I_n) V_k. \end{aligned}$$

Low-rank Sylvester ADI

The Residual

Furthermore,

$$V_k = (A - \beta_k I_n)^{-1} \Big[\prod_{j=1}^{k-1} \mathcal{A}_j \Big] F$$

$$\hat{V}_k = (A - \alpha_k I_n) V_k = \Big[\prod_{j=1}^{k} \mathcal{A}_j \Big] F$$

Low-rank Sylvester ADI

The Residual

Furthermore,

$$V_k = (A - \beta_k I_n)^{-1} \Big[\prod_{j=1}^{k-1} \mathcal{A}_j \Big] F = (A - \beta_k I_n)^{-1} \hat{V}_{k-1},$$

$$\hat{V}_k = (A - \alpha_k I_n) V_k = \Big[\prod_{j=1}^{k} \mathcal{A}_j \Big] F$$

Low-rank Sylvester ADI

The Residual

Furthermore,

$$V_k = (A - \beta_k I_n)^{-1} \Big[\prod_{j=1}^{k-1} \mathcal{A}_j \Big] F = (A - \beta_k I_n)^{-1} \hat{V}_{k-1},$$

$$\begin{aligned} \hat{V}_k &= (A - \alpha_k I_n) V_k = \Big[\prod_{j=1}^{k} \mathcal{A}_j \Big] F \\ &= (A - \alpha_k I_n)(A - \beta_k I_n)^{-1} \hat{V}_{k-1} \end{aligned}$$

Low-rank Sylvester ADI

The Residual

Furthermore,

$$V_k = (A - \beta_k I_n)^{-1} \Big[\prod_{j=1}^{k-1} \mathcal{A}_j \Big] F = (A - \beta_k I_n)^{-1} \hat{V}_{k-1},$$

$$\begin{aligned} \hat{V}_k &= (A - \alpha_k I_n) V_k = \Big[\prod_{j=1}^{k} \mathcal{A}_j \Big] F \\ &= (A - \alpha_k I_n)(A - \beta_k I_n)^{-1} \hat{V}_{k-1} = \hat{V}_{k-1} + (\beta_k - \alpha_k) V_k, \end{aligned}$$

Low-rank Sylvester ADI

The Residual

Furthermore,

$$
\begin{aligned}
V_k &= (A - \beta_k I_n)^{-1}\Big[\prod_{j=1}^{k-1} \mathcal{A}_j\Big] F = (A - \beta_k I_n)^{-1} \hat{V}_{k-1}, \\
\hat{V}_k &= (A - \alpha_k I_n) V_k = \Big[\prod_{j=1}^{k} \mathcal{A}_j\Big] F \\
&= (A - \alpha_k I_n)(A - \beta_k I_n)^{-1} \hat{V}_{k-1} = \hat{V}_{k-1} + (\beta_k - \alpha_k) V_k, \\
W_k &= (B - \alpha_k I_m)^{-H} \hat{W}_{k-1}, \\
\hat{W}_k &= \hat{W}_{k-1} - \overline{(\beta_k - \alpha_k)} W_k,
\end{aligned}
$$

where $\hat{V}_0 = F$, $\hat{W}_0 = G$.

(Generalization of Lyapunov case discussed in [B./Saak/Kürschner '13].)

Low-rank Sylvester ADI

Low-rank Sylvester ADI Reloaded [B./Kürschner '13]

Algorithm 2 Reformulated Factored ADI iteration (fADI 2.0)

Input: A, B, F, G defining the Sylvester equation and shift parameters $\{\alpha_1, \dots, \alpha_{k_{\max}}\}$, $\{\beta_1, \dots, \beta_{k_{\max}}\}$, tolerance $\tau > 0$.

Output: $Z_{k_{\max}} \in \mathbb{C}^{n \times rk_{\max}}$, $Y_{k_{\max}} \in \mathbb{C}^{m \times rk_{\max}}$, $D_{k_{\max}} \in \mathbb{C}^{rk_{\max} \times rk_{\max}}$ such that $Z_{k_{\max}} D_{k_{\max}} Y_{k_{\max}}^H \approx X$.

1: $\hat{V}_0 = F,\ \hat{W}_0 = G,\ k = 0$

2: **while** $\|\hat{W}_k^T \hat{V}_k\|_F \geq \tau \|G^T F\|_F$ **do**

3: $\gamma_k = \beta_k - \alpha_k$

4: $V_k = (A - \beta_k I_n)^{-1} \hat{V}_{k-1},\ W_k = (B - \alpha_k I_m)^{-H} \hat{W}_{k-1}$

5: $\hat{V}_k = \hat{V}_{k-1} + \gamma_k V_k,\ \hat{W}_k = \hat{W}_{k-1} - \overline{\gamma_k} W_k$

6: Update solution factors

$$Z_k = [Z_{k-1}, V_k], \quad Y_k = [Y_{k-1}, W_k], \quad D_k = \operatorname{diag}(D_{k-1}, \gamma_k I_r)$$

7: $k{+}{+}$

8: **end while**

Low-rank Sylvester ADI

Some further improvements [B./Kürschner '13]

- Non-symm. A, B usually lead to complex shifts $\rightsquigarrow$ complex iterates. Can be avoided, only one linear system per complex conjugate pair of shifts necessary $\rightsquigarrow$ acceleration of factor 2.

Low-rank Sylvester ADI

Some further improvements [B./Kürschner '13]

- Non-symm. A, B usually lead to complex shifts $\rightsquigarrow$ complex iterates. Can be avoided, only one linear system per complex conjugate pair of shifts necessary $\rightsquigarrow$ acceleration of factor 2.
- Self-tuning of shifts by selecting Ritz values corresponding to current rational Krylov bases V_k, W_k $\rightsquigarrow$ acceleration factor 1–4.

Low-rank Sylvester ADI

Some further improvements [B./Kürschner '13]

- Non-symm. A, B usually lead to complex shifts $\rightsquigarrow$ complex iterates. Can be avoided, only one linear system per complex conjugate pair of shifts necessary $\rightsquigarrow$ acceleration of factor 2.
- Self-tuning of shifts by selecting Ritz values corresponding to current rational Krylov bases V_k, W_k $\rightsquigarrow$ acceleration factor 1–4.
- In summary, accelerated ADI iteration for Sylvester equation by factor between 2 and 10.

Low-rank Sylvester ADI

Some further improvements [B./Kürschner '13]

- Non-symm. A, B usually lead to complex shifts $\rightsquigarrow$ complex iterates. Can be avoided, only one linear system per complex conjugate pair of shifts necessary $\rightsquigarrow$ acceleration of factor 2.
- Self-tuning of shifts by selecting Ritz values corresponding to current rational Krylov bases V_k, W_k $\rightsquigarrow$ acceleration factor 1–4.
- In summary, accelerated ADI iteration for Sylvester equation by factor between 2 and 10.
- Special variants for special types of Sylvester equations:
  - Generalized Sylvester equations:
  $$AXC - EXB = FG^T.$$
  - Cross-Gramian Sylvester equation:
  $$AXE + EXA = -FG^T.$$
  - Discrete-time Lyapunov equation (Stein equation):
  $$AXA^T - EXE^T = -FF^T.$$
  - Discrete time Sylvester equation:
  $$AXB - EXC = -FG^T.$$

Optimal Low-Rank Solutions

Rank-1 right hand side

[B./Breiten '12/'13]

Let us consider the generalized Sylvester equation

$$AXD + CXB - uv^T = 0,$$

with $A, C \in \mathbb{R}^{n\times n}$, $B, D \in \mathbb{R}^{m\times m}$, $u \in \mathbb{R}^n, v \in \mathbb{R}^m$.

Assume that $A = A^T, B = B^T, C = C^T, D = D^T \succ 0$.

Hence, the generalized Sylvester operator

$$\mathcal{L}_S = D \otimes A + B \otimes C$$

naturally defines an inner product $\langle \cdot, \cdot \rangle_{\mathcal{L}_S}$ via

$$\mathbb{R}^{n\times m} \times \mathbb{R}^{n\times m} \mapsto \mathbb{R}, \quad (Y_1, Y_2) \to \langle Y_1, Y_2 \rangle_{\mathcal{L}_S} := \langle \mathcal{L}_S \operatorname{vec}(Y_1), \operatorname{vec}(Y_2) \rangle.$$

Goal: Find a rank-n approximation $X_{\hat{n}}$ that is optimal w.r.t.

$$||X - X_{\hat{n}}||_{\mathcal{L}_S} := \sqrt{\langle X - X_{\hat{n}}, X - X_{\hat{n}} \rangle_{\mathcal{L}_S}}.$$

Optimal Low-Rank Solutions

An objective function and the error set [B./Breiten '12/'13]

Let

$$\Sigma = (A, B, C, D, u, v)$$

denote a generalized Sylvester equation with solution X.

Define an objective function $f(\Sigma) = u^T X v$ and the error set Σ_{err} as

$$\mathcal{A} = \begin{bmatrix} -A & 0 \\ 0 & \hat{A} \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} -B & 0 \\ 0 & \hat{B} \end{bmatrix}, \quad \mathcal{C} = \begin{bmatrix} -C & 0 \\ 0 & \hat{C} \end{bmatrix},$$

$$\mathcal{E} = \begin{bmatrix} -E & 0 \\ 0 & \hat{E} \end{bmatrix}, \quad \mathcal{U} = \begin{bmatrix} u \\ \hat{u} \end{bmatrix}, \quad \mathcal{V} = \begin{bmatrix} v \\ \hat{v} \end{bmatrix},$$

where the reduced set is given as

$$\hat{A} = V^T A V, \quad \hat{C} = V^T C V, \quad \hat{u} = V^T u,$$
$$\hat{B} = W^T B W, \quad \hat{D} = W^T D W, \quad \hat{v} = W^T v,$$

with orthonormal matrices $V \in \mathbb{R}^{n \times \hat{n}}$ and $W \in \mathbb{R}^{m \times \hat{n}}$.

Optimal Low-Rank Solutions

An auxiliary result and optimality conditions [B./Breiten '12/'13]

Deriving first-order optimality conditions for f (SISO) now leads to

$$H(-\lambda_i) = \hat{H}(-\lambda_i),$$
$$H'(-\lambda_i) = \hat{H}'(-\lambda_i),$$

with $H(s) = v^T (sD - B)^{-1} v$ and λ_i denoting the eigenvalues of $(\hat{A}, \hat{C})$.

Optimal Low-Rank Solutions

An auxiliary result and optimality conditions [B./Breiten '12/'13]

Deriving first-order optimality conditions for f (SISO) now leads to

$$G(-\mu_i) = \hat{G}(-\mu_i),$$
$$G'(-\mu_i) = \hat{G}'(-\mu_i),$$

with $G(s) = u^T(sC - A)^{-1}u$ and μ_i denoting the eigenvalues of $(\hat{B}, \hat{D})$.

Optimal Low-Rank Solutions

An auxiliary result and optimality conditions [B./Breiten '12/'13]

Deriving first-order optimality conditions for f (SISO) now leads to

$$G(-\mu_i) = \hat{G}(-\mu_i),$$
$$G'(-\mu_i) = \hat{G}'(-\mu_i),$$

with $G(s) = u^T(sC - A)^{-1}u$ and μ_i denoting the eigenvalues of $(\hat{B}, \hat{D})$.

Lemma

- $\Sigma = (A, B, C, D, u, v)$ symmetric set of matrices.
- $\hat{\Sigma} = (\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{u}, \hat{v})$ reduced set of matrices obtained by V, W.
- $X, \hat{X}$ solutions of associated Sylvester equations.

$\Rightarrow f(\Sigma_{err}) \leq f(\Sigma) - f(\hat{\Sigma})$.

$\Rightarrow f(\Sigma_{err}) = f(\Sigma) - f(\hat{\Sigma}) \Leftrightarrow \hat{\Sigma}$ fulfills optimality conditions.

Optimal Low-Rank Solutions

The main result

Theorem

- $\Sigma = (A, B, C, D, u, v)$ symmetric set of matrices.
- $\hat{\Sigma} = (\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{u}, \hat{v})$ reduced set of matrices obtained by V, W.
- $X, \hat{X}$ solutions of associated Sylvester equations.

$\Rightarrow \left\| X - V\hat{X}W^T \right\|_{\mathcal{L}_S} \geq f(\Sigma_{err})$.

$\Rightarrow \left\| X - V\hat{X}W^T \right\|_{\mathcal{L}_S} = f(\Sigma_{err}) \Leftrightarrow \hat{\Sigma}$ fulfills optimality conditions.

Optimal Low-Rank Solutions

The main result

Theorem

- $\Sigma = (A, B, C, D, u, v)$ symmetric set of matrices.
- $\hat{\Sigma} = (\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{u}, \hat{v})$ reduced set of matrices obtained by V, W.
- $X, \hat{X}$ solutions of associated Sylvester equations.

$\Rightarrow \left\| X - V\hat{X}W^T \right\|_{\mathcal{L}_S} \geq f(\Sigma_{err})$.

$\Rightarrow \left\| X - V\hat{X}W^T \right\|_{\mathcal{L}_S} = f(\Sigma_{err}) \Leftrightarrow \hat{\Sigma}$ fulfills optimality conditions.

Consequence: f-optimality of $\hat{\Sigma}$ implies $\mathcal{L}_S$-optimality of $V\hat{X}W^T$.

Goal: Find a local minimizer of $f(\Sigma_{err})$!

Optimal Low-Rank Solutions

An iterative algorithm [B./Breiten '12/'13]

Algorithm 3 IRKA for symmetric Sylvester equations ((Sy)^{2}IRKA)

Input: Interpolation points σ_i and μ_i for $i = 1, \ldots, \hat{n}$

Output: $X_{\hat{n}} = V\hat{X}W^T$ locally minimizing the $\mathcal{L}_S$-norm

1: **while** relative change in $\{\sigma_i, \mu_i\} > tol$ **do**
2: $V = \mathrm{span}\{(\sigma_1 C - A)^{-1}u, \ldots, (\sigma_{\hat{n}} C - A)^{-1}u\}, \quad V^T V = I.$
3: $W = \mathrm{span}\{(\mu_1 D - B)^{-1}v, \ldots, (\mu_{\hat{n}} D - B)^{-1}v\}, \quad W^T W = I.$
4: $\hat{A} = V^T AV, \quad \hat{C} = V^T CV, \quad \hat{B} = W^T BW, \quad \hat{D} = W^T DW$
5: Assign $\sigma_i \leftarrow -\lambda_i(\hat{B}, \hat{D})$ and $\mu_i \leftarrow -\lambda_i(\hat{A}, \hat{C})$ for $i = 1, \ldots, \hat{n}$.
6: **end while**
7: Solve $\hat{A}\hat{X}\hat{D} + \hat{C}\hat{X}\hat{B} - \hat{u}\hat{v}^T$, with $\hat{u} = V^T u, \quad v = W^T v$.
8: Set $X_{\hat{n}} = V\hat{X}W^T$.

Remark: Steps 2 and 3 can be replaced by solving

$$AX\hat{D} + CX\hat{B} - u\hat{v}^T = 0, \quad DY\hat{A} + BY\hat{C} - v\hat{u}^T = 0$$

$\rightarrow$ straigthforward extension to general r.h.s.

Application: Image reconstruction

A tribute to Lothar's "Application of ADI ... to restoration of noisy images"

Tikhonov regularization

$$\hat{x} = \min \left\| \begin{bmatrix} H \\ \lambda L \end{bmatrix} x - \begin{bmatrix} g \\ 0 \end{bmatrix} \right\|_2^2 .$$

Minimizer is given as solution of

$$(H_\lambda = H^T H + \lambda^2 L^T L)\hat{x} = H^T g.$$

Assume $H = H_2 \otimes H_1$ and $L = L_2 \otimes L_1$, with $H_1, L_1 \in \mathbb{R}^{n\times n},\ H_2, L_2 \in \mathbb{R}^{m\times m}$.

Can be written as

$$A\hat{X}D + \lambda^2 C\hat{X}B = E,$$

where $A = H_1^T H_1, B = L_2^T L_2, C = L_1^T L_1, D = H_2^T H_2, E = H_1^T G H_2$.

In the following, $H_1 = [h_{ij}]$ is the Toeplitz matrix with
$h_{ij} = \frac{1}{2r-1}, |i-j| \le r,\ r = 20.$

Moreover, L_1 is *tridiag*$(1,2,1)$ and $G = \hat{G} + N$, with Gaussian Noise
$||N||_F/||\hat{G}||_F = 10^{-3},\ \hat{G} = H_1 X_{or} H_2$.

Image reconstruction

The problem

An evening with Lothar

At 8 pm

Image reconstruction

The problem

An evening with Lothar

At 9 pm

Image reconstruction

The problem

An evening with Lothar

At 10 pm

Image reconstruction

The problem

An evening with Lothar

At 11 pm

Image reconstruction

The problem

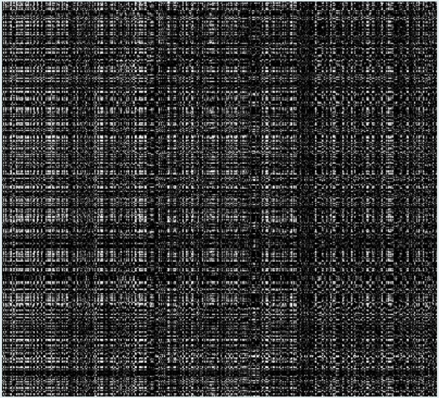
An evening with Lothar

At midnight

Image reconstruction

The solution

Early morning — who, again, was the guy I was drinking with?



(Sy)^{2}IRKA step 1, $r = 40$

Image reconstruction

The solution

Early morning — who, again, was the guy I was drinking with?

(Sy)^{2}IRKA step 2, $r = 40$

Image reconstruction

The solution

Early morning — who, again, was the guy I was drinking with?

$(\mathrm{Sy})^2$IRKA step 3, $r = 40$

Image reconstruction

The solution

Early morning — who, again, was the guy I was drinking with?

$(\mathrm{Sy})^2$IRKA step 4, $r = 40$

Image reconstruction

The solution

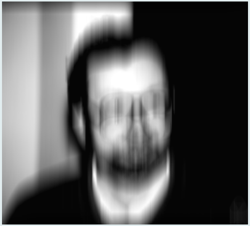
Early morning — who, again, was the guy I was drinking with?

$(\mathrm{Sy})^2$IRKA step 5, $r = 40$

Image reconstruction

The solution

Early morning — who, again, was the guy I was drinking with?



. . . well. . . recovered

(Not) The End.

HAPPY BIRTHDAY, Lothar . . .

(Not) The End.

HAPPY BIRTHDAY, Lothar . . .

. . . and keep flying!

(Not) The End.

HAPPY BIRTHDAY, Lothar ...

... and keep flying!

Last but not least, HAPPY BIRTHDAY, ETNA!