On the numerical solution of certain linear multiterm matrix equations

Valeria Simoncini

Dipartimento di Matematica
Alma Mater Studiorum - Università di Bologna
valeria.simoncini@unibo.it

The matrix equation problem

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

$A_i \in \mathbb{R}^{n\times n}$, $B_i \in \mathbb{R}^{m\times m}$, $\boldsymbol{X}$ unknown matrix

Possibly large dimensions, structured coefficient matrices

The problem in its full generality is far from tractable, although the transformation to a matrix-vector equation [...] allows us to use the considerable arsenal of numerical weapons currently available for the solution of such problems.

Peter Lancaster, SIAM Rev. 1970

The matrix equation problem

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

$A_i \in \mathbb{R}^{n\times n}$, $B_i \in \mathbb{R}^{m\times m}$, $\boldsymbol{X}$ unknown matrix

Possibly large dimensions, structured coefficient matrices

> *The problem in its full generality is far from tractable, although the transformation to a matrix-vector equation [...] allows us to use the considerable arsenal of numerical weapons currently available for the solution of such problems.*
>
> *Peter Lancaster, SIAM Rev. 1970*

Multiterm linear matrix equation. Classical device

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

Kronecker formulation

$$\left(B_1^\top \otimes A_1 + \ldots + B_\ell^\top \otimes A_\ell\right) \boldsymbol{x} = c \qquad \Leftrightarrow \qquad \mathcal{A}\boldsymbol{x} = c$$

with $c = \mathrm{vec}(C)$, $\boldsymbol{x} = \mathrm{vec}(\boldsymbol{X})$.

Iterative methods: matrix-matrix multiplications and rank truncation

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

$$\text{Kronecker product}: \quad M \otimes P = \begin{bmatrix} m_{11}P & \ldots & m_{1n}P \\ \vdots & \ddots & \vdots \\ m_{n1}P & \ldots & m_{nn}P \end{bmatrix} \text{ and } \mathrm{vec}(AXB) = (B^\top \otimes A)\mathrm{vec}(X)$$

Applications:

- Control
- Deterministic and stochastic, and time dependent PDEs
- Inverse problems and optimization

Multiterm linear matrix equation. Classical device

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

Kronecker formulation

$$\left(B_1^\top \otimes A_1 + \ldots + B_\ell^\top \otimes A_\ell\right) \boldsymbol{x} = c \qquad \Leftrightarrow \qquad \mathcal{A}\boldsymbol{x} = c$$

with $c = \mathrm{vec}(C)$, $\boldsymbol{x} = \mathrm{vec}(\boldsymbol{X})$.

Iterative methods: matrix-matrix multiplications and rank truncation
(Benner, Breiten, Bouhamidi, Chehab, Damm, Grasedyck, Jbilou, Kressner, Matthies, Nagy, Onwunta, Raydan, Stoll, Tobler, Wedderburn, Zander, ...)

$$\text{Kronecker product}: \quad M \otimes P = \begin{bmatrix} m_{11}P & \ldots & m_{1n}P \\ \vdots & \ddots & \vdots \\ m_{n1}P & \ldots & m_{nn}P \end{bmatrix} \text{ and } \mathrm{vec}(AXB) = (B^\top \otimes A)\mathrm{vec}(X)$$

Applications:

- Control
- Deterministic and stochastic, and time dependent PDEs
- Inverse problems and optimization

Multiterm linear matrix equation

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

Alternative approaches to the Kronecker form:

- Fixed point iterations (an "evergreen"...)
- Projection-type methods $\Rightarrow$ low rank approximation
- Ad-hoc problem-dependent procedures
- etc.

A sample of these methodologies on different problems:

- Stochastic PDE
- PDEs on polygonal domains
- All-at-once PDE-constrained optimization problem
- Bilinear control problems
- ....

Multiterm linear matrix equation

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

Alternative approaches to the Kronecker form:

- Fixed point iterations (an "evergreen"...)
- Projection-type methods ⇒ low rank approximation
- Ad-hoc problem-dependent procedures
- etc.

A sample of these methodologies on different problems:

- Stochastic PDE
- PDEs on polygonal domains
- All-at-once PDE-constrained optimization problem
- Bilinear control problems
- ....

Some elementary cases

We consider

$$AX + XB + \sigma_j X = C, \quad j = 1, \ldots, s$$

$A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times m}$

Solution strategies

Consider, e.g., $(A + \sigma_j I)X + XB = C$

- **Small scale:** (Bartels-Stewart method)
  - Schur decompositions: $A = QRQ^*$, $B^* = USU^*$
  - For each j, solve $(R + \sigma_j I)(Q^* X U) + Q^* X U)S^* = Q^* CU$
- **Large scale: with** $m \approx n$
  - Construct approximation spaces with A and B
  - Solve a distinct projected problem for each σ_j

Some elementary cases

We consider

$$A\boldsymbol{X} + \boldsymbol{X}B + \sigma_j \boldsymbol{X} = C, \quad j = 1, \ldots, s$$

$A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times m}$

Solution strategies

Consider, e.g., $(A + \sigma_j I)\boldsymbol{X} + \boldsymbol{X}B = C$

- **Small scale:** (Bartels-Stewart method)
  - Schur decompositions: $A = QRQ^*$, $B^* = USU^*$
  - For each j, solve $(R + \sigma_j I)(Q^*\boldsymbol{X}U) + Q^*\boldsymbol{X}U)S^* = Q^*CU$
- **Large scale: with** $m \approx n$
  - Construct approximation spaces with A and B
  - Solve a distinct projected problem for each σ_j

Some elementary cases

We consider

$$A\boldsymbol{X} + \boldsymbol{X}B + \sigma_j \boldsymbol{X} = C, \quad j = 1, \ldots, s$$

$A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times m}$

Solution strategies

Consider, e.g., $(A + \sigma_j I)\boldsymbol{X} + \boldsymbol{X}B = C$

- **Small scale:** (Bartels-Stewart method)
  - Schur decompositions: $A = QRQ^*$, $B^* = USU^*$
  - For each j, solve $(R + \sigma_j I)(Q^*\boldsymbol{X}U) + Q^*\boldsymbol{X}U)S^* = Q^*CU$
- **Large scale: with** $m \approx n$
  - Construct approximation spaces with A and B
  - Solve a distinct projected problem for each σ_j

A special multiterm problem. 1

Let

$$A\boldsymbol{X} + \boldsymbol{X}B + f(\boldsymbol{X})M = C \qquad (\bullet)$$

with $f : \mathbb{R}^{n\times m} \to \mathbb{R}$ a linear function

Closed form solution:

Let Z_1, Z_2 be the solutions to the Sylvester equations $AZ + ZB = C$ and $AZ + ZB = M$, respectively. Assume that $1 + f(Z_2) \neq 0$. Then the solution to $(\bullet)$ is given by

$$\boldsymbol{X} = Z_1 - \alpha Z_2, \quad \alpha = \frac{f(Z_1)}{1 + f(Z_2)}$$

A special multiterm problem. 1

Let

$$A\boldsymbol{X} + \boldsymbol{X}B + f(\boldsymbol{X})M = C \qquad (\bullet)$$

with $f : \mathbb{R}^{n\times m} \to \mathbb{R}$ a linear function

Closed form solution:

Let Z_1, Z_2 be the solutions to the Sylvester equations $AZ + ZB = C$ and $AZ + ZB = M$, respectively. Assume that $1 + f(Z_2) \neq 0$. Then the solution to $(\bullet)$ is given by

$$\boldsymbol{X} = Z_1 - \alpha Z_2, \quad \alpha = \frac{f(Z_1)}{1 + f(Z_2)}$$

A special multiterm problem. 2

Some samples:

1. $A\boldsymbol{X} + \boldsymbol{X}B + \mathrm{trace}(\boldsymbol{X})M = C$. Then

$$\boldsymbol{X} = Z_1 - \alpha Z_2, \quad \alpha = \frac{\mathrm{trace}(Z_1)}{1 + \mathrm{trace}(Z_2)}$$

2. $A\boldsymbol{X} + \boldsymbol{X}B + (v^T\boldsymbol{X}u)M = C$. Then

$$\boldsymbol{X} = Z_1 - \alpha Z_2, \quad \alpha = \frac{v^T Z_1 u}{1 + v^T Z_2 u}$$

A special multiterm problem. 2

Some samples:

1. $A\boldsymbol{X} + \boldsymbol{X}B + \mathrm{trace}(\boldsymbol{X})M = C$. Then

$$\boldsymbol{X} = Z_1 - \alpha Z_2, \quad \alpha = \frac{\mathrm{trace}(Z_1)}{1 + \mathrm{trace}(Z_2)}$$

2. $A\boldsymbol{X} + \boldsymbol{X}B + (v^T\boldsymbol{X}u)M = C$. Then

$$\boldsymbol{X} = Z_1 - \alpha Z_2, \quad \alpha = \frac{v^T Z_1 u}{1 + v^T Z_2 u}$$

A special multiterm problem. 3

⇒ The approach solves a seemingly unrelated problem

Let

$$A\boldsymbol{X} + \boldsymbol{X}B + M_1\boldsymbol{X}M_2 = C, \qquad M_1, M_2 \quad \text{rank-one} \quad \text{matrices}$$

Letting $M_i = u_i v_i^T$, $i = 1, 2$, then

$$M_1\boldsymbol{X}M_2 = u_1 v_1^T \boldsymbol{X} u_2 v_2^T = (v_1^T \boldsymbol{X} u_2) u_1 v_2^T \equiv f(\boldsymbol{X}) M$$

♣ The closed form is just the (vector) Sherman-Morrison formula in disguise

Current generalizations

- Multiterm case
$$A\boldsymbol{X} + \boldsymbol{X}B + f_1(\boldsymbol{X})M_1 + \ldots + f_\ell(\boldsymbol{X})M_\ell = C$$
with $f_j : \mathbb{R}^{n\times m} \to \mathbb{R}$, $j = 1, \ldots, \ell$ linear functions

- Nonlinear case
$$A\boldsymbol{X} + \boldsymbol{X}B + f(\boldsymbol{X})M = C$$
with $f(\boldsymbol{X}) = g(\phi(\boldsymbol{X}))$, where ϕ is a real valued *linear* function, and g real function.
(work in progress)

(joint work with Margherita Porcelli, Università di Bologna)

Current generalizations

- Multiterm case

$$A\boldsymbol{X} + \boldsymbol{X}B + f_1(\boldsymbol{X})M_1 + \ldots + f_\ell(\boldsymbol{X})M_\ell = C$$

  with $f_j : \mathbb{R}^{n\times m} \to \mathbb{R}$, $j = 1, \ldots, \ell$ linear functions

- Nonlinear case

$$A\boldsymbol{X} + \boldsymbol{X}B + f(\boldsymbol{X})M = C$$

  with $f(\boldsymbol{X}) = g(\phi(\boldsymbol{X}))$, where ϕ is a real valued *linear* function, and g real function.
  (work in progress)

(joint work with Margherita Porcelli, Università di Bologna)

Multiterm low-rank linear matrix equation

Consider

$$A\mathbf{X} + \mathbf{X}A^T + M\mathbf{X}M^T = C, \qquad M = UV^T \qquad (\bullet)$$

$U, V \in \mathbb{R}^{n\times s}$ (without loss of generality, we consider this simplified form)

Let $\mathcal{U} = U \otimes U$, $\mathcal{V} = V \otimes V$. Then

$$M\mathbf{X}M^T = UV^T\mathbf{X}VU^T \quad \to \quad \mathcal{U}\mathcal{V}^T\mathbf{x}$$

Setting $\mathcal{A} = A \otimes I + I \otimes A$, then $(\bullet)$ becomes

$$(\mathcal{A} + \mathcal{U}\mathcal{V}^T)\mathbf{x} = c, \qquad c = \mathrm{vec}(C)$$

Sherman-Morrison-Woodbury (vector) formula:

$$\mathbf{x} = (\mathcal{A} + \mathcal{U}\mathcal{V}^T)^{-1}c = \mathcal{A}^{-1}c - \mathcal{A}^{-1}\mathcal{U}(I + \mathcal{V}^T\mathcal{A}^{-1}\mathcal{U})^{-1}\mathcal{V}^T\mathcal{A}^{-1}c$$

(joint work with Yue Hao, Lanzhou University, China)

Multiterm low-rank linear matrix equation

Consider

$$A\boldsymbol{X} + \boldsymbol{X}A^T + M\boldsymbol{X}M^T = C, \qquad M = UV^T \qquad (\bullet)$$

$U, V \in \mathbb{R}^{n\times s}$ (without loss of generality, we consider this simplified form)

Let $\mathcal{U} = U \otimes U$, $\mathcal{V} = V \otimes V$. Then

$$M\boldsymbol{X}M^T = UV^T\boldsymbol{X}VU^T \quad \rightarrow \quad \mathcal{U}\mathcal{V}^T\boldsymbol{x}$$

Setting $\mathcal{A} = A \otimes I + I \otimes A$, then $(\bullet)$ becomes

$$(\mathcal{A} + \mathcal{U}\mathcal{V}^T)\boldsymbol{x} = c, \qquad c = \mathrm{vec}(C)$$

Sherman-Morrison-Woodbury (vector) formula:

$$\boldsymbol{x} = (\mathcal{A} + \mathcal{U}\mathcal{V}^T)^{-1}c = \mathcal{A}^{-1}c - \mathcal{A}^{-1}\mathcal{U}(I + \mathcal{V}^T\mathcal{A}^{-1}\mathcal{U})^{-1}\mathcal{V}^T\mathcal{A}^{-1}c$$

(joint work with Yue Hao, Lanzhou University, China)

Multiterm low-rank linear matrix equation

Consider

$$A\boldsymbol{X} + \boldsymbol{X}A^T + M\boldsymbol{X}M^T = C, \qquad M = UV^T \qquad (\bullet)$$

$U, V \in \mathbb{R}^{n\times s}$ (without loss of generality, we consider this simplified form)

Let $\mathcal{U} = U \otimes U$, $\mathcal{V} = V \otimes V$. Then

$$M\boldsymbol{X}M^T = UV^T\boldsymbol{X}VU^T \quad \to \quad \mathcal{U}\mathcal{V}^T\boldsymbol{x}$$

Setting $\mathcal{A} = A \otimes I + I \otimes A$, then $(\bullet)$ becomes

$$(\mathcal{A} + \mathcal{U}\mathcal{V}^T)\boldsymbol{x} = c, \qquad c = \mathrm{vec}(C)$$

Sherman-Morrison-Woodbury (vector) formula:

$$\boldsymbol{x} = (\mathcal{A} + \mathcal{U}\mathcal{V}^T)^{-1}c = \mathcal{A}^{-1}c - \mathcal{A}^{-1}\mathcal{U}(I + \mathcal{V}^T\mathcal{A}^{-1}\mathcal{U})^{-1}\mathcal{V}^T\mathcal{A}^{-1}c$$

(joint work with Yue Hao, Lanzhou University, China)

Sherman-Morrison-Woodbury (vector) formula

$$\boldsymbol{x} = \mathcal{A}^{-1}c - \mathcal{A}^{-1}\mathcal{U}(I + \mathcal{V}^T\mathcal{A}^{-1}\mathcal{U})^{-1}\mathcal{V}^T\mathcal{A}^{-1}c$$

Algorithm 0.

1. Solve $\mathcal{A}w = c$
2. Solve $\mathcal{A}\mathrm{p}_j = \mathrm{u}_j$ where $\mathcal{U} = [\mathrm{u}_1, \ldots, \mathrm{u}_{s^2}]$ to give $\mathcal{P} = [\mathrm{p}_1, \ldots, \mathrm{p}_{s^2}]$;
3. Compute $H = I + \mathcal{V}^T\mathcal{P} \in \mathbb{R}^{s^2 \times s^2}$
4. Solve $Hg = \mathcal{V}^T w$
5. Compute $x = w - \mathcal{P}g$

Transforming into matrix-matrix operations:

- Step 1: $w = \mathcal{A}^{-1}c \quad \Leftrightarrow \quad AW + WA^T = C$ with $w = \mathrm{vec}(W)$
- Step 2: $\mathcal{U} = [\mathrm{u}_1, \ldots, \mathrm{u}_{s^2}]$ with $\mathrm{u}_t = u_k \otimes u_i$ where $t = (k-1)s + i$, hence

$$\mathrm{p}_j = \mathcal{A}^{-1}\mathrm{u}_j \quad \Leftrightarrow \quad AP_j + P_jA^T = u_iu_k^T, \quad \mathrm{p}_j = \mathrm{vec}(P_j).$$

- Step 3: $H = I + \mathcal{V}^T[\mathrm{p}_1, \ldots, \mathrm{p}_{s^2}]$ can be written by using

$$\mathrm{v}_j^T\mathcal{A}^{-1}\mathrm{u}_t = v_i^T P_t v_k, \quad j = (k-1)s + i.$$

- Step 4:

$$\mathcal{V}^T\mathcal{A}^{-1}c = \begin{bmatrix} v_1^T W v_1 \\ v_2^T W v_1 \\ \vdots \\ v_s^T W v_s \end{bmatrix}.$$

Generalizations to multiterm low-rank equations

$$A\boldsymbol{X} + \boldsymbol{X}A^T + U_1V_1^T\boldsymbol{X}(U_2V_2^T)^T + U_3V_3^T\boldsymbol{X}(U_4V_4^T)^T = C$$

With U_i, V_i, $i = 1, ..., 4$ low rank (not necessarily the same)

Setting

$$\mathcal{U} = [U_2 \otimes U_1, U_4 \otimes U_3] \quad \text{and} \quad \mathcal{V} = [V_2 \otimes V_1, V_4 \otimes V_3],$$

we obtain again

$$(\mathcal{A} + \mathcal{U}\mathcal{V}^T)\boldsymbol{x} = c$$

An experiment with A symmetric and dense

"Direct": Solve $(\mathcal{A} + \mathcal{U}\mathcal{V}^T)\boldsymbol{x} = c$ using \

"Matrix form": Use the matrix-oriented Sherman-Morrison formula

"Vectorized form": Use the Kronecker form of the Sherman-Morrison formula

		Direct			Matrix form			Vectorized Form		
n	s_1/s_3	CPU	Res	Err	CPU	Res	Err	CPU	Res	Err
40	3/5	0.08	1.2e-15	3.7e-12	0.02	1.2e-14	2.1e-11	0.23	1.2e-13	3.5e-11
	4/6	0.08	2.2e-15	5.2e-12	0.02	6.5e-15	1.8e-11	0.29	1.0e-12	4.8e-11
	5/7	0.10	1.1e-15	5.7e-11	0.02	1.1e-14	1.2e-10	0.37	4.9e-12	2.2e-10
80	3/5	2.01	3.4e-15	3.3e-11	0.02	5.7e-15	4.6e-11	6.14	8.3e-13	9.4e-10
	4/6	2.05	2.3e-15	2.8e-10	0.02	3.3e-15	1.4e-10	8.19	3.9e-12	6.6e-10
	5/7	2.00	2.7e-15	7.2e-11	0.03	5.7e-14	2.0e-10	10.6	1.5e-12	1.8e-09
160	3/5	85.6	9.2e-15	1.5e-10	0.04	2.4e-14	3.9e-10	168	1.6e-13	9.6e-09
	4/6	86.4	9.0e-15	1.4e-09	0.05	7.4e-14	5.3e-09	211	3.1e-11	7.9e-08
	5/7	86.9	6.4e-15	3.0e-10	0.08	8.6e-14	2.3e-09	257	2.7e-12	2.0e-07

An experiment with A symmetric and pentadiagonal

"Direct": Solve $(\mathcal{A}+\mathcal{U}\mathcal{V}^T)\boldsymbol{x} = c$ using \

"Matrix form": Use the matrix-oriented Sherman-Morrison formula

"Vectorized form": Use the Kronecker form of the Sherman-Morrison formula

		Direct			Matrix form			Vectorized Form		
U_1, U_3	n	CPU	Res	Err	CPU	Res	Err	CPU	Res	Err
non-orth	40	0.08	9.1e-16	3.8e-12	0.02	4.8e-14	1.9e-11	0.04	1.5e-12	1.1e-10
	80	1.94	1.1e-15	5.8e-10	0.02	3.0e-14	2.3e-10	0.22	1.7e-11	2.6e-09
	160	85.1	5.9e-15	9.2e-09	0.04	2.5e-13	1.9e-08	1.24	1.5e-10	7.4e-08
	320	–	–	–	0.08	1.4e-12	5.0e-08	6.64	8.3e-11	4.8e-06
orth	40	0.08	3.2e-15	8.9e-14	0.02	1.8e-15	1.1e-14	0.04	2.6e-16	9.9e-15
	80	1.88	6.0e-15	3.2e-13	0.02	1.7e-15	2.1e-14	0.18	1.8e-16	2.2e-14
	160	84.8	1.7e-14	2.7e-12	0.03	1.7e-15	1.6e-13	1.32	2.5e-15	3.2e-12
	320	–	–	–	0.07	2.0e-15	3.9e-13	6.81	1.6e-15	1.3e-12

More generalizations

$$AX + XA^T + M \circ X = C, \quad M = UV^T$$

with $U = [u_1, \ldots, u_s]$, $V = [v_1, \ldots, v_s]$ sparse, $\circ$ Hadamard (element-wise) product

♣ Using the property: $\quad (UV^T) \circ X = \sum_{i=1}^{s} \mathrm{diag}(u_i) X \mathrm{diag}(v_i)$

we obtain

$$AX + XA^T + \sum_{i=1}^{s} \mathrm{diag}(u_i) X \mathrm{diag}(v_i) = C$$

with $\mathrm{diag}(u_i), \mathrm{diag}(v_i)$ low rank. The previous procedures can be applied

More generalizations

$$AX + XA^T + M \circ X = C, \quad M = UV^T$$

with $U = [u_1, \dots, u_s]$, $V = [v_1, \dots, v_s]$ sparse, $\circ$ Hadamard (element-wise) product

♣ Using the property: $$(UV^T) \circ X = \sum_{i=1}^{s} \mathrm{diag}(u_i) X \mathrm{diag}(v_i)$$

we obtain

$$AX + XA^T + \sum_{i=1}^{s} \mathrm{diag}(u_i) X \mathrm{diag}(v_i) = C$$

with $\mathrm{diag}(u_i), \mathrm{diag}(v_i)$ low rank. The previous procedures can be applied

Considerations and conclusions

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

- Many scientific applications now lead to multiterm linear matrix equations
- Multiterm equations give new challenges and are a rich source of solution-brainstorming
- Small scale and large scale require different perspectives, though "small" has a different scale than in the past

REFERENCE

Yue Hao and V. Simoncini, *The Sherman-Morrison-Woodbury formula for generalized linear matrix equations and applications*, To appear in Numer. Linear Algebra w/Appl. DOI: 10.1002/nla.2384