

# A SUBSPACE-CONJUGATE GRADIENT METHOD FOR LINEAR MATRIX EQUATIONS

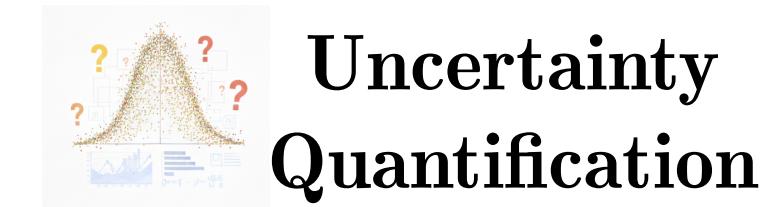
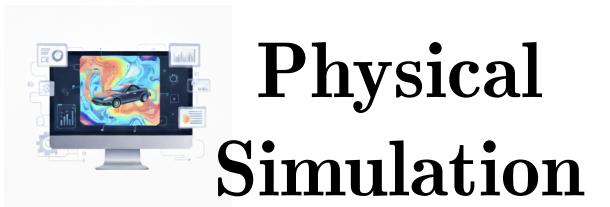
Davide Palitta, Martina Iannacito, Valeria Simoncini

## The problem

Consider a **multiterm Sylvester equation**

$$\mathbf{A}_1 \mathbf{X} \mathbf{B}_1 + \cdots + \mathbf{A}_m \mathbf{X} \mathbf{B}_m = \mathbf{C} \quad (1)$$

that appears in



Finding the solution  $\mathbf{X}$  of (1) is

- for  $m \leq 2$ : **easy** (projection methods, ADI, Riemannian optimization methods, etc...)
- for  $m > 2$ : **more challenging** (matrix-oriented Krylov methods, ad-hoc projection methods, Riemannian optimization)

## Matrix-oriented Conjugate Gradient

### Assumptions

- $\mathbf{A}_i$  and  $\mathbf{B}_i$  are symmetric matrices of size  $(n \times n)$ .
- Right-hand side  $\mathbf{C} = \mathbf{C}_1 \mathbf{C}_2^\top$  has low rank ( $s_C$ ).
- $\mathcal{L}(\mathbf{X}) = \sum_{i=1}^m \mathbf{A}_i \mathbf{X} \mathbf{B}_i$  is SPD w.r.t.  $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{trace}(\mathbf{X}^\top \mathbf{Y})$ .

Matrix-oriented CG is equivalent to standard CG, since

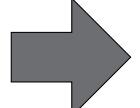
$$\sum_{i=1}^m \mathbf{A}_i \mathbf{X} \mathbf{B}_i = \mathbf{C} \iff \left( \sum_{i=1}^m \mathbf{B}_i^\top \otimes \mathbf{A}_i \right) \text{vec}(\mathbf{X}) = \text{vec}(\mathbf{C}).$$

Given an initial guess  $\mathbf{X}_0 \in \mathbb{R}^{n \times n}$ , the matrix-oriented CG iterates are

$$\begin{aligned} \mathbf{X}_{k+1} &= \mathbf{X}_k + \alpha_k \mathbf{P}_k && \text{where } \alpha_k \in \mathbb{R} \quad (\text{solution}) \\ \mathbf{R}_{k+1} &= \mathbf{C} - \mathcal{L}(\mathbf{X}_{k+1}) && (\text{residual}) \\ \mathbf{P}_{k+1} &= \mathbf{R}_{k+1} + \beta_k \mathbf{P}_k && \text{where } \beta_k \in \mathbb{R} \quad (\text{direction}) \end{aligned}$$

These matrices are in *factored form*, e.g.,  $\mathbf{P}_k = P_k P_k^\top$ , and throughout the iterations, blocks get larger accumulating redundant information.

Low rank truncation



- delayed or stagnating convergence
- challenging to control the rank

## Subspace-Conjugate Gradient

### Key idea

Replace  $\alpha, \beta$  scalars by  $\alpha, \beta$  matrices

As in matrix-oriented CG, define  $\Phi : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$  such that

$$\Phi(\mathbf{X}) = \frac{1}{2} \langle \mathcal{L}(\mathbf{X}), \mathbf{X} \rangle - \langle \mathbf{C}, \mathbf{X} \rangle,$$

so that  $\mathbf{X}^*$ , the exact solution of (1), satisfies  $\mathbf{X}^* = \arg \min_{\mathbf{X} \in \mathbb{R}^{n \times n}} \Phi(\mathbf{X})$ .

The new iterate for the solution is

$$\mathbf{X}_{k+1} = \mathbf{X}_k + P_k \alpha_k P_k^\top \quad \text{where } \alpha_k \in \mathbb{R}^{s_k \times s_k}.$$

Let  $\phi : \mathbb{R}^{s_k \times s_k} \rightarrow \mathbb{R}$  be  $\phi(\alpha) = \Phi(\mathbf{X}_k + P_k \alpha P_k^\top)$ , then

$$\alpha_k = \arg \min_{\alpha \in \mathbb{R}^{s_k \times s_k}} \phi(\alpha) \quad (2)$$

### Proposition

The matrix  $\alpha_k \in \mathbb{R}^{s_k \times s_k}$  minimizer of (2) is the unique solution of

$$P_k^\top \mathcal{L}(P_k \alpha_k P_k^\top) P_k = P_k^\top \mathbf{R}_k P_k. \quad (3)$$

$\alpha_k$  solution of (3) is equivalent to imposing the **local orthogonality condition**

$$\text{vec}(\mathbf{R}_{k+1}) \perp \text{range}(P_k \otimes P_k). \quad (4)$$

The new iterate for the direction matrix is

$$\mathbf{P}_{k+1} = \mathbf{R}_{k+1} + P_k \beta_k P_k^\top \quad \text{where } \beta_k \in \mathbb{R}^{s_k \times s_k}.$$

### Proposition

The matrix  $\beta_k \in \mathbb{R}^{s_k \times s_k}$  is obtained imposing the  **$\mathcal{L}$ -orthogonality condition**

$$\text{vec}(\mathcal{L}(\mathbf{P}_{k+1})) \perp \text{range}(P_k \otimes P_k). \quad (5)$$

$\beta_k$  satisfying (5) is the unique solution of

$$P_k^\top \mathcal{L}(P_k \beta_k P_k^\top) P_k = -P_k^\top \mathcal{L}(\mathbf{R}_{k+1}) P_k. \quad (6)$$

Notice that (3) and (6) are smaller size multiterm Sylvester equations, that is

$$\sum_{i=1}^m \tilde{\mathbf{A}}_i \alpha \tilde{\mathbf{B}}_i = P_k^\top \mathbf{R}_k P_k \quad \text{and} \quad \sum_{i=1}^m \tilde{\mathbf{A}}_i \beta \tilde{\mathbf{B}}_i = -P_k^\top \mathcal{L}(\mathbf{R}_{k+1}) P_k$$

where  $\tilde{\mathbf{A}}_i = P_k^\top \mathbf{A}_i P_k$  and  $\tilde{\mathbf{B}}_i = P_k^\top \mathbf{B}_i P_k$  have size  $(s_k \times s_k)$ .

## Further details

D. Palitta, M. Iannacito, and V. Simoncini. A Subspace-Conjugate Gradient Method for Linear Matrix Equations. *SIAM J. Matrix Anal. & Appl.*, 46(4):2197–2225, 2025.



## References

- [1] I. Bioli, D. Kressner, and L. Robol. Preconditioned Low-Rank Riemannian Optimization for Symmetric Positive Definite Linear Matrix Equations. *SIAM J. Sci. Comput.*, 47(2):A1091–A1116, 2025.
- [2] D. Kressner and C. Tobler. Low-Rank Tensor Krylov Subspace Methods for Parametrized Linear Systems. *SIAM J. Matrix Anal. & Appl.*, 32(4):1288–1316, 2011.
- [3] V. Simoncini and Y. Hao. Analysis of the Truncated Conjugate Gradient Method for Linear Matrix Equations. *SIAM J. Matrix Anal. & Appl.*, 44(1):359–381, 2023.



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA