Preconditioned Low-Rank Riemannian Optimization for Symmetric Positive Definite Linear Matrix Equations

Ivan Bioli, Daniel Kressner, <u>Leonardo Robol</u>

Abstract

The topic of this work is the solution of multiterm linear matrix equations of the form

$$\mathcal{L}(X) := A_1 X B_1 + \ldots + A_\ell X B_\ell = F_{\ell}$$

under the assumption that the linear operator \mathcal{L} is positive definite. These equations often appear when discretizing elliptic PDEs, and in control problems. In this setting, it is often the case that F is either a low-rank matrix, or can be well-approximated in this way, i.e., it has fast decaying singular values. Throughout this work, we assume that F is a low-rank matrix.

A few special cases are easy to handle, and we can prove that the low-rank structure of F is inherited by X. If $\ell = 1$, X and F have the same rank; if $\ell = 2$ we have an elegant theory based on rational approximation that provides upper bounds for the singular values of X with a weak dependency on the condition number of the operator \mathcal{L} [2].

When $\ell > 2$, most theoretical guarantees are lost: except for a few special cases, none of the arguments used for $\ell = 2$ can be extended easily. Nevertheless, it can be experimentally verified that the structure is often present in X; sometimes, this can be justified via other means (such as the regularity of the solution of the discretized PDE).

Algorithmically, we face similar challenges: the cases $\ell = 1$ and $\ell = 2$ are well understood, and efficient low-rank solvers are available (examples include rational and extended Krylov methods, or the Alternative Direction Implicit method, known as ADI). When $\ell > 2$, fewer and less appealing options are available, mainly based on Krylov solvers for \mathcal{L} with low-rank truncation [5]. A major issue with this class of methods is that the intermediate iterates often have much higher ranks than the final solution, and aggressive truncations lead to degraded convergence properties.

We focus on the case $\ell > 2$, assuming that X is the discretization of the solution to an elliptic PDE. Assuming that X is well-approximated by a matrix of rank k, the problem can be recast into the optimization problem

Find
$$\min_{X} \Phi(X)$$
, subject to $\operatorname{rank}(X) = k$, $\Phi(X) := \frac{1}{2} \langle \mathcal{L}(X), X \rangle - \langle F, X \rangle$,

for some moderate k, where $\langle \cdot, \cdot \rangle$ is the standard scalar product inducing the Frobenius norm. The matrices of rank k have a Riemannian manifold structure, and thus the problem can be conveniently solved using Riemannian optimization [1, 4]. This idea has been successfully exploited in the past to solve positive definite Lyapunov equations, which correspond to $\ell = 2$ [6]. This strategy restricts our search to matrices of rank k, and avoids the rank-growth phenomenon often observed in Krylov methods. The convergence theory for Riemannian optimization schemes can be used to prove convergence.

However, a straightforward application of a vanilla Riemannian optimization method presents a major issue: when \mathcal{L} arises from an elliptic PDE, it is often ill-conditioned, and so is the Hessian of the objective function $\Phi(X)$. This leads to poor convergence for first-order optimization schemes.

We discuss how the idea of preconditioning can be extended to this setting, and show that one can design effective modifications to the problem that drastically improve the convergence. The key ingredient is finding a map \mathcal{P}_X on the tangent space at X which is positive definite and approximates the Hessian in a suitable sense. Then, the preconditioning can be described in two similar ways, which are only approximately equivalent:

- 1. We may take gradients with respect to the inner product induced by \mathcal{P}_X , instead of the canonical one; this turns the first-order method to be closer to a quasi-Newton scheme.
- 2. Alternatively, we may use \mathcal{P}_X to define a non-standard metric and inner product on the Riemannian manifold, and then apply the first-order scheme in this new setup.

We discuss differences, advantages, and disadvantages of these strategies; we show that is possible to employ both at the same time at once to design preconditioners for a large class of PDE problems discretize with finite element methods.

In the examples that we consider, the multiterm matrix equation is often obtained by non-constant diffusion coefficients in 2D diffusion problems; in this setting it is natural to choose \mathcal{P}_X as the projection of the operator obtained discretizing the same equation, but with constant diffusion coefficients; this typically is an operator of the same form, but with $\ell = 2$. Hence, we consider preconditioners of the form $\mathcal{P}_X = AXB$, $\mathcal{P}_X = AX + XB$, and $\mathcal{P}_X = AXB + CXD$. Applying these preconditioners in any of the sense described above require to solve an equation on the tangent space, which is low-dimensional. This is relatively easy to do in the first two cases, but it proves challenging in the latter, when $A, B, C, D \neq I$. We show how to this effectively by combining the two viewpoints discussed above, and interpreting this preconditioner as a composition of the first two options.

The proposed algorithm is competitive or faster than the state-of-the art in most cases, and in particular when the target rank k is moderate. When k is larger, solving the equations on the tangent space can become a bottleneck. For these cases, we propose yet another preconditioner, obtained generalizing the ADI iteration to an iteration on the tangent space of the manifold; our experiment show that it preserves the good convergence properties of the classical ADI scheme for matrix equations, while being much cheaper to compute than the previous approaches. A few steps of this "tangent ADI" iteration will prove to be a very good preconditioner for a large class of problems.

Time permitting, a few extra details essential to produce a robust implementation will be discussed, such as rank-adaptivity, and the use of randomized linear algebra to effectively estimate the residual; the latter helps in detectubg possible stagnation due to wrong estimates for the solution rank, and is essential for deciding when to perform rank changes inside the rank-adaptive scheme.

References

- [1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds.* Princeton University Press, 2008.
- [2] Bernhard Beckermann and Alex Townsend. On the singular values of matrices with displacement structure. SIAM Journal on Matrix Analysis and Applications, 38(4):1227–1248, 2017.
- [3] Ivan Bioli, Daniel Kressner, and Leonardo Robol. Preconditioned Low-Rank Riemannian Optimization for Symmetric Positive Definite Linear Matrix Equations. arXiv preprint arXiv:2408.16416, 2024.

- [4] Nicolas Boumal. An introduction to optimization on smooth manifolds. Cambridge University Press, 2023.
- [5] Davide Palitta and Patrick Kürschner. On the convergence of Krylov methods with low-rank truncations. *Numerical Algorithms*, 88(3):1383–1417, 2021.
- [6] Bart Vandereycken and Stefan Vandewalle. A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations. SIAM Journal on Matrix Analysis and Applications, 31(5):2553–2579, 2010.