

Lanczos with compression for symmetric matrix functions

Monday, 20 January 2025 09:00 (20 minutes)

In this work we present a low-memory method for the approximation of the action of a symmetric matrix function $f(A) \in \mathbb{R}^{n \times n}$ on a vector $\mathbf{b} \in \mathbb{R}^n$, where the matrix A is large and sparse. A popular approach for approximating $f(A)\mathbf{b}$ is the Lanczos algorithm. Given an orthonormal basis $Q_M \in \mathbb{R}^{n \times M}$ of the Krylov subspace $\mathcal{K}_M(A, \mathbf{b}) = \text{span}\{\mathbf{b}, A\mathbf{b}, \dots, A^{M-1}\mathbf{b}\}$, the Lanczos approximation of $f(A)\mathbf{b}$ is given by $\mathbf{f}_M = Q_M f(T_M) Q_M^T \mathbf{b}$, where $T_M = Q_M^T A Q_M$. Although the basis Q_M can be constructed with a short-term recurrence in a setting with limited memory, the computation of \mathbf{f}_M still requires the storage of the whole Krylov basis. A simple way to overcome this issue and obtain a low-memory algorithm is to use the two-pass Lanczos approach, which however requires double the number of matrix vector products with A . Several other techniques have been proposed in the literature.

In this talk we propose a novel algorithm to approximate $f(A)\mathbf{b}$ in a limited memory setting, by combining the Lanczos method with a basis compression procedure involving rational Krylov subspaces, which is employed whenever the basis grows beyond a certain size. The method that we propose has essentially the same convergence behaviour as Lanczos, with an additional error term that depends on rational approximation of the function f and is typically negligible. The cost of the basis compression procedure is also negligible with respect to the cost of the Lanczos algorithm. In particular, the rational Krylov subspaces used for the compression of the Lanczos basis are built using small projected matrices, so their construction is cheap and does not involve the solution of any linear systems with the matrix A .

The key theoretical result used in the development of our algorithm is a formula for the computation of low-rank updates of matrix functions using rational Krylov subspaces, derived in [1]. This update formula is applied iteratively to the tridiagonal projected matrix T_M , by splitting it into a block diagonal matrix and a rank-two off-diagonal part. The resulting algorithm is composed of a sequence of cycles: in each cycle, we first perform a fixed number of Lanczos iterations to expand the Krylov basis, and then we construct an appropriate rational Krylov subspace associated with the current projected matrix in order to compress the basis back to a fixed size, while also updating the approximation to $f(A)\mathbf{b}$.

Numerical experiments demonstrate that our algorithm exhibits competitive performance when compared against other low-memory methods for $f(A)\mathbf{b}$. Further details can be found in the preprint [2].

[1] B. Beckermann, A. Cortinovis, D. Kressner, and M. Schweitzer, *Low-rank updates of matrix functions II: rational Krylov methods*, SIAM J. Numer. Anal., 59 (2021), pp. 1325-1347.

[2] A. A. Casulli and I. Simunec, *A low-memory Lanczos method with rational Krylov compression for matrix functions*, arXiv:2403.04390, 2024.

Primary authors: CASULLI, Angelo Alberto (Gran Sasso Science Institute); SIMUNEC, Igor (École Polytechnique Fédérale de Lausanne)

Presenter: SIMUNEC, Igor (École Polytechnique Fédérale de Lausanne)

Session Classification: Morning Session I