

A hybrid method for computing a few singular triplets of very large sparse matrices

James Baglama, Jeniffer Picucci, Vasilije Perović

Abstract

Ability to efficiently compute a (partial) Singular Value Decomposition (SVD) of a matrix is essential in a wide range of problems, including data mining, genomics, machine learning, PCA, and signal processing. In such applications matrices tend to be very large and sparse and one is typically interested in computing only a few of the largest singular triplets. Over the last several decades this problem has led to a considerable amount of research and software development, see e.g., [4, 6, 7, 9, 10, 11, 15] and the references therein.

In this talk, for a large and sparse $A \in \mathbb{R}^{\ell \times n}$, we present a new hybrid restarted Lanczos bidiagonalization method for the computation of a small number, k , of its extreme singular triplets, i.e., we compute $\{\sigma_j, u_j, v_j\}_{j=1}^k$ such that

$$Av_j = \sigma_j u_j, \quad A^T u_j = \sigma_j v_j, \quad j = 1, 2, \dots, k.$$

At the core of our proposed algorithm [3], as in many of the above listed ones, is the Golub-Kahan-Lanczos (GKL) bidiagonalization procedure which at step m results in the m -GKL factorization

$$AP_m = Q_m B_m, \tag{1}$$

$$A^T Q_m = P_m B_m^T + f e_m^T = [P_m \ p_{m+1}] \begin{bmatrix} B_m^T \\ \beta_m e_m^T \end{bmatrix}, \tag{2}$$

where $P_m = [p_1, \dots, p_m] \in \mathbb{R}^{n \times m}$ and $Q_m = [q_1, \dots, q_m] \in \mathbb{R}^{\ell \times m}$ have orthonormal columns, the residual vector $f \in \mathbb{R}^n$ satisfies $P_m^T f = 0$, $\beta_m = \|f\|$, and $p_{m+1} = f/\beta_m$, and B_m is an $m \times m$ upper bidiagonal matrix.

Approximations of the singular triplets of A can then be obtained from the singular triplets of B_m , and in the case when the norm of the residual vector f is small, the singular values of B_m are close to the singular values of A . But for modest values of m these approximations are typically poor (assuming limited memory makes increasing m not an option) and thus leaving one with an option to modify, explicitly or implicitly, the starting vector p_1 and *restart the GKL process*. In [4] the authors exploited the mathematical equivalence for symmetric eigenvalue computations of the implicitly restarted Arnoldi (Lanczos) method of Sorensen [13] and the thick-restarting scheme of Wu and Simon [14], and applied it to a restarted GKL procedure. The resulting thick-restarted GKL routine, `irlba`, turns out to be a simple and computationally fast method for computing a few of the extreme singular triplets that is less sensitive to propagated round-off errors.

However, the `irlba` routine [4] often struggles when the dimension, m , of the Krylov subspaces is memory limited and kept relatively small in relationship to the number of desired singular triplets k . Very recently, in the context of symmetric eigenvalue computation, we were able to overcome this memory restriction by creating a hybrid restarted Lanczos method that combines thick-restarting with Ritz vectors with a new technique, *iteratively refined Ritz vectors* [1]. We recall that in [8] Jia proposed to use *refined Ritz* vectors in place of Ritz vectors as eigenvector approximations of a square matrix M [8]. More specifically, for a given approximate eigenvalue μ_j of M , Jia's method looks to minimize $\|Mz_j - \mu_j z_j\|$ for a unit vector z_j from a given subspace \mathcal{W} . Moreover, in [8] it was

shown that on the subspace \mathcal{W} an approximate eigenpair using the refined Ritz vector produced a “smaller” residual norm than an eigenpair approximation with the Ritz pair. More recently, in [1] we were able to extend this idea to *iterative refined Ritz* values/vectors for the symmetric eigenvalue problem that produces an even smaller residual norm than refined Ritz – this resulted in a better converging method that outperformed using Ritz or refined Ritz vectors. But this comes at the price as working with iteratively refined Ritz vectors is more challenging, in comparison to just Ritz vectors, due to the fact that the scheme of thick-restarting of Wu and Simon is not available [1]. However, not all is lost and in [1] we introduce an alternate scheme in which, based on the relationships first proposed by Sorensen [13] and later outlined in detail by Morgan [12], the iteratively refined Ritz vectors are linearly combined and then used to restart the process. We choose the constants in a way that the linear combination of the iteratively refined Ritz vectors resembles a restart, in a somewhat asymptotic sense, of thick-restarting, see [1] for details.

In our most recent work [3], we applied the earlier results from [1] by making a natural connection between the symmetric eigenvalue problem and the SVD of $A \in \mathbb{R}^{\ell \times n}$ and considered both the normal matrix $A^T A \in \mathbb{R}^{n \times n}$ and the augmented matrix $C = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \in \mathbb{R}^{(\ell+n) \times (\ell+n)}$. Multiplying (1) from the left by A^T produces the Lanczos tridiagonal decomposition of $A^T A$, namely

$$A^T A P_m = P_m B_m^T B_m + \alpha_m f_m e_m^T = \begin{bmatrix} P_m & p_{m+1} \end{bmatrix} \begin{bmatrix} B_m^T B_m \\ \alpha_m \beta_m e_m^T \end{bmatrix}. \quad (3)$$

Similarly, in the case of matrix C , after performing $2m$ steps of the standard Lanczos algorithm with the starting vector $[0; p_1] \in \mathbb{R}^{\ell+n}$ we have a $2m \times 2m$ tridiagonal projection matrix, which when followed by an odd-even permutation gives the following Lanczos factorization [10]

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} Q_m & 0 \\ 0 & P_m \end{bmatrix} = \begin{bmatrix} Q_m & 0 & 0 \\ 0 & P_m & p_{m+1} \end{bmatrix} \begin{bmatrix} 0 & B_m \\ B_m^T & 0 \\ \beta_m e_m^T & 0 \end{bmatrix}. \quad (4)$$

With the two Lanczos factorization relationships (3) and (4), the theoretical results and properties related to the hybrid iterative refined Ritz (eigenvalue) scheme in [1] are carried over resulting into two hybrid routines capable of computing few extreme singular triplets A based on either $A^T A$ or $C = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$. While this extension seems straightforward its implementation on (4), as well as the new development of iterative refined Ritz working on the normal system, is nontrivial.

Through numerical examples, we have observed in [1, 3] that when memory was limited and only iterative refined Ritz vectors were used to restart the method there was potential for either slow or no convergence. In order to overcome this challenge, we developed a hybrid method that, based on extensive numerical tests and existing heuristics, switches between thick-restarted with Ritz vectors and under certain criteria it restarts with a linear combination of iterative refined Ritz vectors. We note that a careful balance is needed here, since on the one side the iterative refined Ritz vectors can give a better approximation but with possible stagnation, while on the other side thick-restarted is a more efficient restarting scheme, but with not as good of approximations.

In the rest of the talk we discuss multiple criteria we used to determine when to make a switch and provide some justification along with several useful heuristics. We also describe a simple yet powerful variant of our proposed hybrid algorithm (≈ 100 lines of MATLAB code) where the Krylov basis size is $m = 2$ and which requires a nontrivial purging of converged vectors. This simplified code, with potential for extensions [2], does not require calls to any LAPACK routines, thus making

it portable and at the same time very competitive on a number of large matrices. For instance, in [3] we computed the largest singular triplet of a 214 million \times 214 million matrix (1.2GB) from the `kmerV1r` dataset on a laptop in just 31 minutes, whereas the `irlba` method [4] took 75 minutes, and MATLAB’s internal `svds` function required 4 hours. We conclude with several additional examples that illustrate the proposed hybrid scheme is competitive with other publicly available code when there are limited memory requirements.

References

- [1] Baglama, J., Bella, T., Picucci, J.: Hybrid iterative refined method for computing a few extreme eigenpairs of a symmetric matrix. *SIAM J. Sci. Comput.* **43**(5), S200–S224 (2021)
- [2] Baglama, J., Perović, V.: Explicit deflation in Golub-Kahan-Lanczos bidiagonalization methods. *ETNA*, **58** (2023)
- [3] Baglama, J., Perović, V., Picucci, J.: Hybrid iterative refined restarted Lanczos bidiagonalization methods. *Numer. Algorithms*, **92**(2) (2023)
- [4] Baglama, J., Reichel, L.: Augmented implicitly restarted Lanczos bidiagonalization methods. *SIAM J. Sci. Comput.* **27**(1), 19–42 (2005)
- [5] Baglama, J., Reichel, L.: An implicitly restarted block Lanczos bidiagonalization method using Leja shifts. *BIT Numer. Math.* **53**(2), 285–310 (2013)
- [6] Goldenberg, S., Stathopoulos, A., Romero, E.: A Golub–Kahan Davidson method for accurately computing a few singular triplets of large sparse matrices. *SIAM J. Sci. Comput.* **41**(4), A2172–A2192 (2019)
- [7] Hochstenbach, M.E.: Harmonic and refined extraction methods for the singular value problem, with applications in least squares problems. *BIT Numer. Math.* **44**(4), 721–754 (2004)
- [8] Jia, Z.: Refined iterative algorithms based on Arnoldi’s process for large unsymmetric eigenproblems. *Linear Algebra Appl.* **259**, 1–23 (1997)
- [9] Jia, Z., Niu, D.: An implicitly restarted refined bidiagonalization Lanczos method for computing a partial SVD. *SIAM J. Matrix Anal. Appl.* **25**(1), 246–265 (2003)
- [10] Kokiopoulou, E., Bekas, C., Gallopoulos, E.: Computing smallest singular triplets with implicitly restarted Lanczos bidiagonalization. *Applied Numer. Math.* **49**, 39–61 (2004).
- [11] Lehoucq, R.B., Sorensen, D.C., Yang, C.: ARPACK users’ guide: Solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods. *SIAM* (1998)
- [12] Morgan, R.B.: On restarting the Arnoldi method for large nonsymmetric eigenvalue problems. *Math. Comp.* **65**(215), 1213–1230 (1996)
- [13] Sorensen, D.C.: Implicit application of polynomial filters in a k -step Arnoldi method. *SIAM J. Matrix Anal. Appl.* **13**(1), 357–385 (1992)
- [14] Wu, K., Simon, H.: Thick-restart Lanczos method for large symmetric eigenvalue problems. *SIAM J. Matrix Anal. Appl.* **22**(2), 602–616 (2000)
- [15] Wu, L., Romero, E., Stathopoulos, A.: `Primme_svds`: A high-performance preconditioned svd solver for accurate large-scale computations. *SIAM J. Sci. Comput.* **39**(5), S248–S271 (2017)