Algorithms for Hermitian eigenproblems and their complexity

Aleksandros Sobczyk

Abstract

Hermitian eigenproblems, and, more broadly, Hermitian-definite pencils, arise naturally in many real world applications in Machine Learning, Scientific Computing, and Engineering. Given a Hermitian matrix **A** and a Hermitian positive-definite matrix **B**, the goal is to compute (a subset of) the eigenvalues λ and/or the eigenvectors **v**, which satisfy

$\mathbf{Av} = \lambda \mathbf{Bv}.$

In data science and machine learning, for example, they arise in Spectral Clustering [31, 39], Language Models [22], Image Processing [36, 1], Principal Components Analysis [11, 24, 37], and many others [12, 29, 14, 9, 28, 2]. A ubiquitous application in Scientific Computing is the computation of the density matrices and the electron densities in Density Functional Theory (DFT) [27].

Algorithms for eigenproblems have been studied since at least the nineteenth century, some early references being attributed to Jacobi [23, 16]. In this work we revisit algorithms from the computational complexity point of view, targeting (i) eigenvalue problems, which involve the approximation of eigenvalues, singular values, spectral gaps, and condition numbers, (ii) eigenspace problems, such as the approximation of eigenvectors, spectral projectors, and invariant subspaces, and (iii) diagonalization problems, which refer to the computation of all the eigenvalues and eigenvectors of a matrix, for example, a full spectral factorization, or the SVD.

This work is a summary of the results in [41, 40] for some variants of the aforementioned problems, which were part of the corresponding authors' doctoral dissertation. All of the algorithms and complexity bounds are in the following three *models of computation*:

- Exact real arithmetic (Real RAM), where a processor can execute the following operations: $\{+, -, \times, /, \sqrt{\cdot}, >\}$, on real numbers, in constant time, without any rounding errors;
- **Rational arithmetic**, where numbers are represented as rationals consisting of an integral numerator and denominator of finite (but variable bit length, and each arithmetic operation does not introduce errors but takes time proportional to the number of bits;
- Floating point, where any real number $\alpha \in \mathbb{R}$ is rounded to a floating point number $\mathbf{fl}(\alpha) = s \times 2^{e-t} \times m$, and each arithmetic operation $\odot \in \{+, -, \times, /, \sqrt{\cdot}\}$ introduces also some errors that satisfy $\mathbf{fl}(\alpha \odot \beta) = (1 + \theta)(\alpha \odot \beta)$, $\mathbf{fl}(\sqrt{a}) = (1 + \theta)\sqrt{a}$, where $|\theta| \leq \mathbf{u}$. Assuming a total of *b* bits for each number, every floating point operation costs $\mathcal{F}(b)$ bit operations, where typically it is assumed that $\mathcal{F}(b) \in \widetilde{O}(b)$ [19].

Algorithms in Real RAM. In the Real RAM model, we analyze the following problems:

- 1. Symmetric arrowhead/tridiagonal diagonalization,
- 2. Hermitian diagonalization,
- 3. Singular Value Decomposition.

We first provide an end-to-end complexity analysis of the divide-and-conquer algorithm of Gu and Eisenstat [18] for the diagonalization of tridiagonal and arrowhead matrices, when accelerated with the Fast Multipole Method [35]. By carefully analyzing all of the steps of the algorithm, we show that it provides provable approximation guarantees with the claimed nearly- $O(n^2)$ arithmetic complexity, which significantly improves classic (dense) eigensovers such as the QR algorithm.

The tridiagonal diagonalization algorithm can be efficiently combined with the (rather overlooked) tridiagonal reduction algorithm of Schönhage [38], who proved that a Hermitian matrix can be reduced to tridiagonal form with unitary similarity transformations in $O(n^{\omega})$ arithmetic operations. Here $\omega \leq 2.371$ is the current best known upper bound for the matrix multiplication exponent. This way, we can diagonalize a Hermitian matrix in nearly matrix multiplication time, improving the $O(n^3)$ arithmetic complexity of classic algorithms [30, 17], as well as the more recent $O(n^{\omega} \log^2(\frac{n}{\epsilon}))$ complexity of the randomized algorithm of [3]. Similar bounds are obtained for the deterministic complexity of the SVD. Many theoretical works assume the exact computation of an SVD as "black-box" subroutine; see e.g. [34, 15, 13, 7, 9, 8], to name a few. However, its complexity and approximation guarantees are often unspecified. Our main results and comparisons with existing algorithms are outlined in Table 1.

Table 1: Complexities for diagonalization problems in the Real RAM model. The randomized algorithms succeed with high probability (at least $1 - 1/\operatorname{poly}(n)$). $O(n^{\omega(a,b,c)})$ denotes the complexity of multiplying two matrices with sizes $n^a \times n^b$ and $n^b \times n^c$, respectively.

	Arithmetic Complexity	Deterministic
Arrowhead/Tridiagonal diagonalization		
[10, 32, 18]	$O(n^3) + \widetilde{O}(n^2)$	\checkmark
Our results	$O\left(n^2\operatorname{polylog}(rac{n}{\epsilon}) ight)$	\checkmark
Hermitian diagonalization		
[3, 25]	$O\left(n^{\omega}\log^2(rac{n}{\epsilon}) ight)$	×
[4]	$\widetilde{O}\left(n^{\omega+1} ight)$	×
[30]	$O(n^3)$	\checkmark
Our results	$O\left(n^{\omega}\log(n) + n^2 \operatorname{polylog}(\frac{n}{\epsilon})\right)$	\checkmark
SVD		
Fast $MM + [30]$	$O\left(n^{\omega\left(1,k,1 ight)} ight)+\widetilde{O}\left(n^{3} ight)$	1
[3, 25]	$O\left(n^{\omega(1,k,1)} + n^{\omega}\log^2(\frac{n}{\epsilon})\right)$	×
Our results	$O\left(n^{\omega(1,k,1)} + n^{\omega}\log(n) + n^2\operatorname{polylog}(\frac{n\kappa(\mathbf{A})}{\epsilon})\right)$	1

Algorithms in finite precision. Similar deterministic complexity upper bounds are obtained for several problems in finite precision. We will present a stability analysis of the tridiagonal reduction algorithm of Schönhage, and its combination with the tridiagonal eigenvalue solver of [5, 6] (in rational arithmetic), to compute all the eigenvalues of a Hermitian matrix in nearly $O(n^{\omega})$ bit operations, deterministically.

Our main results, which are summarized in Table 2, improve several existing algorithms in different aspects. Pan and Chen [33] showed how to achieve additive errors in the eigenvalues in $O(n^{\omega})$ arithmetic operations, but this increases to $O(n^{\omega+1})$ boolean operations in rational arithmetic. The algorithm of [30] achieves $\tilde{O}(n^3)$ bit operations for all the eigenvalues. In the randomized setting, [3] also provides forward errors for the approximate eigenvalues in the Hermitian case, by exploiting a perturbation bound by Kahan [26, 41] in $O(n^{\omega} \operatorname{polylog}(n/\epsilon))$ boolean operations, but the logarithm power is fairly large. We also provide the analysis for several other useful subroutines related to eigenvalue computations, including singular values, condition numbers, Hermitian-definite pencil eigenvalues, spectral gaps, and spectral projectors.

Table 2: Boolean complexity comparison in finite precision. Here $\epsilon, \in (0, \frac{1}{2})$. FP stands for Floating Point and RA for Rational Arithmetic.

	Boolean Complexity	Success probability	Model
Tridiagonal Reduction			
[21, 20]	$O\left(n^{3}\mathcal{F}\left(\log(\frac{n}{\epsilon})\right)\right)$	Deterministic	\mathbf{FP}
Our results	$O\left(n^{\omega}\log(n)\mathcal{F}\left(\log(\frac{n}{\epsilon})\right)\right)$	Deterministic	\mathbf{FP}
Hermitian Eigenvalues			
[30]	$O\left(n^3 \mathcal{F}\Big(\log(rac{n}{\epsilon})\Big) ight)$	Deterministic	\mathbf{FP}
[3]+[26]	$O\left(n^{\omega}\log^2(\frac{n}{\epsilon})\mathcal{F}\left(\log^4(\frac{n}{\epsilon})\log(n)\right) ight)$	1 - O(1/n)	\mathbf{FP}
Our results	$O\left(n^{\omega}\mathcal{F}\left(\log(\frac{n}{\epsilon})\right) + n^2 \operatorname{polylog}(\frac{n}{\epsilon})\right)$	Deterministic	FP+RA

References

- Gökhan H Bakır, Jason Weston, and Bernhard Schölkopf. Learning to find pre-images. Advances in neural information processing systems, 16:449–456, 2004.
- [2] Pierre Baldi and Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima. Neural networks, 2(1):53–58, 1989.
- [3] Jess Banks, Jorge Garza-Vargas, Archit Kulkarni, and Nikhil Srivastava. Pseudospectral Shattering, the Sign Function, and Diagonalization in Nearly Matrix Multiplication Time. Foundations of Computational Mathematics, pages 1–89, 2022.
- [4] Michael Ben-Or and Lior Eldar. A Quasi-Random Approach to Matrix Spectral Analysis. In Proc. 9th Innovations in Theoretical Computer Science Conference, pages 6:1–6:22. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018.
- [5] Dario Bini and Victor Pan. Parallel complexity of tridiagonal symmetric eigenvalue problem. In Proc. of the Second Annual ACM-SIAM Symposium on Discrete Algorithms, pages 384–393, 1991.
- [6] Dario Bini and Victor Y Pan. Computing matrix eigenvalues and polynomial zeros where the output is real. SIAM Journal on Computing, 27(4):1099–1115, 1998.
- [7] Christos Boutsidis and David P Woodruff. Optimal cur matrix decompositions. In Proc. of the forty-sixth Symposium on Theory of Computing, pages 353–362, 2014.
- [8] Kenneth L Clarkson and David P Woodruff. Low-rank approximation and regression in input sparsity time. Journal of the ACM (JACM), 63(6):1–45, 2017.
- Michael B Cohen, Sam Elder, Cameron Musco, Christopher Musco, and Madalina Persu. Dimensionality reduction for k-means clustering and low rank approximation. In Proc. 47th Symposium on Theory of Computing, pages 163–172, 2015.
- [10] Jan JM Cuppen. A divide and conquer method for the symmetric tridiagonal eigenproblem. Numerische Mathematik, 36:177–195, 1980.
- [11] Konstantinos I Diamantaras and Sun Yuan Kung. Principal component neural networks: theory and applications. John Wiley & Sons, Inc., 1996.
- [12] Petros Drineas, Iordanis Kerenidis, and Prabhakar Raghavan. Competitive recommendation systems. In Proc. Thiry-Fourth Annual ACM Symposium on Theory of Computing, pages 82–90, 2002.
- [13] Petros Drineas, Michael W Mahoney, and Shan Muthukrishnan. Relative-error cur matrix decompositions. SIAM Journal on Matrix Analysis and Applications, 30(2):844–881, 2008.
- [14] Dan Feldman, Melanie Schmidt, and Christian Sohler. Turning big data into tiny data: Constant-size coresets for k-means, pca, and projective clustering. SIAM Journal on Computing, 49(3):601–657, 2020.
- [15] Alan Frieze, Ravi Kannan, and Santosh Vempala. Fast monte-carlo algorithms for finding low-rank approximations. Journal of the ACM (JACM), 51(6):1025–1041, 2004.

- [16] Gene H Golub and Henk A Van der Vorst. Eigenvalue computation in the 20th century. Journal of Computational and Applied Mathematics, 123(1-2):35–65, 2000.
- [17] Gene H Golub and Charles F Van Loan. Matrix Computations. Johns Hopkins University Press, 2013.
- [18] Ming Gu and Stanley C Eisenstat. A divide-and-conquer algorithm for the symmetric tridiagonal eigenproblem. SIAM Journal on Matrix Analysis and Applications, 16(1):172–191, 1995.
- [19] David Harvey and Joris Van Der Hoeven. Integer multiplication in time O(nlogn). Annals of Mathematics, 193(2):563–617, 2021.
- [20] Nicholas J Higham. Accuracy and stability of numerical algorithms. SIAM, 2002.
- [21] Alston S Householder. Unitary triangularization of a nonsymmetric matrix. Journal of the ACM, 5(4):339–342, 1958.
- [22] Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.
- [23] C.G.J. Jacobi. Über ein leichtes verfahren die in der theorie der säcularstörungen vorkommenden gleichungen numerisch aufzulösen. Journal für die reine und angewandte Mathematik, 30:51–94, 1846.
- [24] Ian T Jolliffe. Principal component analysis for special types of data. Springer, 2002.
- [25] Praneeth Kacham and David P Woodruff. Faster algorithms for schatten-p low rank approximation. In Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2024). Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2024.
- [26] W Kahan. Spectra of nearly hermitian matrices. Proc. of the American Mathematical Society, 48(1):11–17, 1975.
- [27] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. Physical Review, 140(4A):A1133, 1965.
- [28] Arnaz Malhi and Robert X Gao. Pca-based feature selection scheme for machine defect classification. IEEE transactions on instrumentation and measurement, 53(6):1517–1525, 2004.
- [29] Olvi L Mangasarian and Edward W Wild. Multisurface proximal support vector machine classification via generalized eigenvalues. *IEEE transactions on pattern analysis and machine intelligence*, 28(1):69–74, 2005.
- [30] Yuji Nakatsukasa and Nicholas J Higham. Stable and efficient spectral divide and conquer algorithms for the symmetric eigenvalue decomposition and the svd. SIAM Journal on Scientific Computing, 35(3):A1325–A1349, 2013.
- [31] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. Advances in Neural Information Processing Systems, 14, 2001.
- [32] Dianne P O'Leary and Gilbert W Stewart. Computing the eigenvalues and eigenvectors of symmetric arrowhead matrices. Journal of Computational Physics, 90(2):497–505, 1990.
- [33] Victor Y Pan and Zhao Q Chen. The complexity of the matrix eigenproblem. In Proc. 31st Annual ACM Symposium on Theory of Computing, pages 507–516, 1999.
- [34] Christos H Papadimitriou, Hisao Tamaki, Prabhakar Raghavan, and Santosh Vempala. Latent semantic indexing: A probabilistic analysis. In Proc. of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems, pages 159–168, 1998.
- [35] Vladimir Rokhlin. Rapid solution of integral equations of classical potential theory. Journal of computational physics, 60(2):187–207, 1985.
- [36] Awwal Mohammed Rufai, Gholamreza Anbarjafari, and Hasan Demirel. Lossy image compression using singular value decomposition and wavelet difference reduction. *Digital signal processing*, 24:117–123, 2014.
- [37] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. Neural computation, 10(5):1299–1319, 1998.
- [38] Arnold Schönhage and Volker Strassen. Fast multiplication of large numbers. Computing, 7:281–292, 1971.
- [39] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. IEEE Transactions on pattern analysis and machine intelligence, 22(8):888–905, 2000.
- [40] Aleksandros Sobczyk. Deterministic complexity analysis of Hermitian eigenproblems. arXiv preprint arXiv:2410.21550, Under review, 2024.
- [41] Aleksandros Sobczyk, Marko Mladenović, and Mathieu Luisier. Invariant subspaces and PCA in nearly matrix multiplication time, 2024.