Tyler Chen
chentyl@uw.edu

Research Statement

https://chen.pw

INTRODUCTION

Computational approaches to the pressing and world-changing questions of today are reliant on sub-routines for fundamental linear algebraic tasks. One important set of such tasks is computing matrix functions and products of matrix functions with vectors. Matrix functions transform the eigenvalues of a symmetric (or Hermitian) matrix according to some scalar function while leaving the eigenvectors untouched. For example, the matrix inverse, which corresponds to the inverse function $f(x) = 1/x$, inverts each of the eigenvalues of the given matrix. Other common matrix functions include the matrix sign, logarithm, exponential, square root, and inverse square root.

In many situations, it is desirable to compute a vector equal to the product of a matrix function with a fixed vector (rather than the matrix function itself). For instance, the matrix inverse applied to a vector corresponds to the solution of a linear system of equations, which is useful, even without knowing the inverse matrix. Beyond the multitude of applications of linear systems, matrix functions applied to vectors are used for computing the overlap operator in quantum chromodynamics [7], solving differential equations in applied math [8, 9], Gaussian process sampling in statistics [10], principle component projection and regression in data science [11], and a range of other applications [12].

Computing a scalar equal to the quadratic form of a matrix function and fixed vector is similarly desirable and is often combined with stochastic trace estimation [13, 14] to approximate the trace of a matrix function [15, 3]. Applications of the trace of matrix functions include characterizing the degree of protein folding in biology [16], maximum likelihood estimation in statistics [17, 18], designing better public transit in urban planning [19, 20] and finding triangle counts and other motifs in network science [21]. The trace of matrix functions is intimately related to spectral density estimation, which is used in electronic structure computations [22, 23, 24] and other tasks in physics [25], probing the behavior of neural networks in machine learning [26, 27, 28], and load balancing modern parallel eigensolvers in numerical linear algebra [29, 30].

PERSONAL BACKGROUND/ACOMPLISHMENTS

Thesis. My thesis work is centered on the Lanczos method for matrix function approximation (Lanczos-FA) [31, 8]. Lanczos-FA is an algorithm for applying matrix functions to vectors and computing quadratic forms involving matrix functions and is among the most efficient and widely used method for all of the applications listed above. Critically, like many methods for solving linear systems, Lanczos-FA does not require computing a full matrix function in order to compute the product of a matrix function with a vector. Changing computing goals and environments necessitate continued research, and my thesis aims to address this need by (i) analyzing existing Lanczos-FA based algorithms for tasks which have emerged in recent years (e.g. studying large neural networks) [3, 32, 5, 6, 33] and (ii) designing more efficient implementations of Lanczos-FA for use in modern computing environments (e.g. distributed memory supercomputers) [1, 6].

Accomplishments. During my PhD, I was fortunate to be supported by the NSF Graduate Research Fellowship Program (GRFP). The GRFP provided me the freedom to work on problems which interested me individually and which aligned with my career goal of supporting basic science. As a result, over the past several years, my research has repeatedly received special recognition. In particular:

- I won best student paper award at the 16th Copper Mountain Conference on Iterative Methods.
- I gave a long presentation at the International Conference on Machine Learning (ICML) 2021 ($\approx 3\%$ of submissions were selected for long presentations).
- I was awarded Boeing research award by my department.

Collaboration.

Throughout my PhD, I sought out collaborators from a range of distinct disciplines. I have found such collaborations extremely effective in stimulating the development of new ideas, and this experience motivates me to pursue a similarly diverse research environment as preparation for a career of collaborative and cross-disciplinary work.

OVERVIEW OF PROPOSAL

During my postdoc, I will build on my PhD work by **combining the power of recent randomization techniques from theoretical computer science and optimization with the practicality of Krylov subspace methods (like Lanczos-FA) from applied math.** The chosen research objectives require combining ideas and techniques from a range of backgrounds, but, as a result, have the potential to significantly benefit the advancement and sharing of knowledge across the sciences. Simultaneously, my outreach objective of organizing equitable conference and seminar sessions will contribute to building a more inclusive scientific community.

At a high level, my research and outreach objectives are:

RO 1. Compare the convergence guarantees and practicality of the conjugate gradient algorithm with recently developed fast stochastic gradient methods for the task of solving linear systems.

RO 2. Derive sharper, spectrum dependent, bounds for the convergence of Krylov subspace methods used to approximate a matrix function applied to a vector and quadratic forms of matrix functions.

RO 3. Design refined estimates for matrix function trace estimation and develop new randomized sketching based approaches for computing general matrix functions from the ground up, rather than as ad-hoc combinations of existing techniques.

OO. Organize conference and seminar sessions with diverse speaker lineups and research focuses.

RESEARCH OBJECTIVES

Throughout this section, $\mathbf{A}$ will be a $n \times n$ symmetric matrix with eigenvalues $\lambda_{\text{max}} = \lambda_1 \geq \cdots \geq \lambda_n = \lambda_{\text{min}}$ and corresponding (orthonormal) eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Thus, $\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^\mathsf{T}$ and $\mathbf{u}_i^\mathsf{T} \mathbf{u}_j = 0$ if $i \neq j$ and 1 if $i = j$. The average eigenvalue is $\lambda_{\text{ave}} :=$

$(\lambda_1 + \cdots + \lambda_n)/n = \text{tr}(\mathbf{A})/n$. Given a function $f : \mathbb{R} \to \mathbb{R}$, we define the matrix function $f[\mathbf{A}]$ in the usual way by $f[\mathbf{A}] := \sum_{i=1}^{n} f(\lambda_i)\mathbf{u}_i\mathbf{u}_i^\mathsf{T}$. Thus, $f[\mathbf{A}]$ is a matrix with the same eigenvectors as $\mathbf{A}$ but whose eigenvalues are transformed by f (again, the proptotypical matrix function is the matrix inverse which corresponds to $f(x) = 1/x$). Finally, when $\mathbf{A}$ is positive definite (i.e. $\lambda_{\min} > 0$), the $\mathbf{A}$-norm of a vector $\mathbf{y}$, denoted $\|\mathbf{y}\|_\mathbf{A}$, is given by $\sqrt{\mathbf{y}^\mathsf{T}\mathbf{A}\mathbf{y}}$.

Research Objective 1: Compare conjugate gradient and fast stochastic gradient methods. When $\mathbf{A}$ is positive definite, the conjugate gradient algorithm (CG) [31], which can be viewed as a special case of Lanczos-FA, is of the most used methods for approximating $\mathbf{A}^{-1}\mathbf{b}$, the solution to the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$. CG outputs an estimate $\tilde{\mathbf{x}}$ satisfying $\|\mathbf{A}^{-1}\mathbf{b} - \tilde{\mathbf{x}}\|_\mathbf{A} < \epsilon$ in runtime $O(ns\sqrt{\kappa}\log(1/\epsilon))$, where $s = s(\mathbf{A})$ is the maximum number of nonzero entries in a row of $\mathbf{A}$ and $\kappa = \lambda_{\text{max}}/\lambda_{\text{min}}$ is the condition number of $\mathbf{A}$ [34]. However, in the past few years, a collection of fast stochastic gradient methods (SGMs), such as accelerated coordinate descent and stochastic variance reduced gradient, offer runtime guarantees of $O(ns\sqrt{\kappa^*}\log(1/\epsilon))$, where $\kappa^* = \lambda_{\text{ave}}/\lambda_{\text{min}}$ is the "smoothed condition number" of $\mathbf{A}$ [35, 36, 37].

It's always true that $\kappa^* \leq \kappa$, and in practice, it's often the case that $\kappa^* \ll \kappa$. Thus, in settings where CG's $\sqrt{\kappa}$ runtime bound is tight, fast SGMs provably outperform CG. However, CG often performs far better than the $\sqrt{\kappa}$ bound, and it's unclear whether SGMs still outperform CG in such settings. Moreover, while CG accesses $\mathbf{A}$ through matrix-vector products, SGMs, which are based on randomized estimates to the gradient, access $\mathbf{A}$ through sequential inner products with individual rows of $\mathbf{A}$. Since a matrix-vector products is equivalent to n inner products, the theoretical runtimes of the algorithms are comparable. However, the inner products in a matrix-vector product can be optimized through parallelization or other hardware-aware approaches, so the real-world runtime of a matrix-vector product is often far shorter than if the constituent inner products were computed sequentially.

Spectrum dependent bounds. The optimality of CG over Krylov subspace means there exist a range of "spectrum dependent" bounds which account for more information about the spectrum of $\mathbf{A}$ than just the largest and smallest eigenvalues [38, 34]. During my PhD I extended many existing bounds for CG to bounds for Lanczos-FA and related algorithms [5, 3, 6]. However, existing bounds are not easily compared directly to the $\sqrt{\kappa^*}$ bound for SGMs described above. Thus, in order to make a more direct comparison between CG and SGMs, I will seek spectrum dependent bounds for CG depending on the smoothed condition number κ^*.

Average case analysis. My second approach, complimentary to the first, will be to consider the average case behavior of CG and fast SGMs when applied to a large random system. CG has been studied extensively in this setting, and in certain cases it is known that the $\sqrt{\kappa}$ bound is tight [39, 40] whereas in others it is known that it is loose [41, 42]. Gradient descent and accelerated gradient descent have also been studied in this setting [43, 44]. The range of existing works will provide the groundwork for a similar analysis of fast gradient methods. However, fast SGMs access the matrix $\mathbf{A}$ in fundamentally different ways than CG and (accelerated) gradient descent; specifically, the fact SGMs use randomized estimators of the gradient rather than the full gradient. Therefore, analyzing these algorithms in this setting is non-trivial and will require the development of new analysis techniques.

Research Objective 2: Spectrum dependent error bounds for Krylov subspace based matrix function approximation.

The most general bounds for Krylov subspace methods (KSMs) used to approximate $f[\mathbf{A}]\mathbf{b}$ or $\mathbf{b}^\mathsf{T} f[\mathbf{A}]\mathbf{b}$ are based on polynomial approximation theory, and in particular, on approximation of f over a single fixed interval such as $[\lambda_{\min}, \lambda_{\max}]$. Such bounds are useful in that they provide simple convergence guarantees which require only minimal information about $\mathbf{A}$ and $\mathbf{b}$. Necessarily, however, this type of bound does not account for fine-grained properties of the spectrum of $\mathbf{A}$ such as isolated or clustered eigenvalues.

In the case that $f(x) = 1/x$, algorithms such as the conjugate gradient algorithm (CG) and minimum residual algorithm (MINRES) have instance optimality guarantees; i.e. they provide optimal (in a certain norm) approximations to $f[\mathbf{A}]\mathbf{b} = \mathbf{A}^{-1}\mathbf{b}$ over Krylov subspace. As a result, error bounds depending on the best approximation to on the eigenvalues of $\mathbf{A}$ can be obtained for such algorithms. These error bounds can be significantly better than bounds based on the best uniform approximation over $[\lambda_{\min}, \lambda_{\max}]$ when the spectrum of $\mathbf{A}$ has favorable properties.

It's often possible to approximate $f[\mathbf{A}]\mathbf{b}$ by a proxy rational matrix function of the form $\sum_i (\mathbf{A} + c_i\mathbf{I})^{-1}\mathbf{b}$ or $\sum_i (\mathbf{A}^2 + c_i\mathbf{I})^{-1}\mathbf{b}$ [45]; i.e. by solving a series of shifted linear systems. Such approximations can be derived by districting an integral representation of f, and when Lanczos-FA is used to approximate each system, the resulting approximations converge to the Lanczos-FA approximation to $f[\mathbf{A}]\mathbf{b}$ (as the numerical integral approximation to f is refined). In [5] we take advantage of this fact to leverage existing fine grained convergence guarantees for Lanczos-FA on linear systems to provide refined error bounds for analytic or piecewise analytic functions f. Our work extends past work [46, 47, 48, 49, 50] in that it applies to a much broader class of functions and that the impact of a perturbed Lanczos recurrence in finite precision is considered.

Interestingly, existing KSMs, such as Lanczos-FA, often perform nearly optimally in numerical experiments. While [5] provides spectrum dependent bounds for general functions, it does not explain this near optimality. In [6] we take a step towards understanding this phenomenon by describing memory-efficient algorithms to compute nearly optimal approximations to $f[\mathbf{A}]\mathbf{b}$ for *any* rational matrix function f. The algorithms we describe are closely related to Lanczos-FA and therefore help to illuminate convergence properties of Lanczos-FA. Even so, the question of optimality for general f remains open.

If my application is successful, I will work towards a better understanding of the convergence of Lanczos-FA and other KSMs in terms of optimality over Krylov subspace. Because we already have stronger guarantees for rational functions, a natural approach is to consider functions "well approximated" by rational functions. As preliminary steps, there are a interesting technical problems to explore. For instance, it would be interesting to study properties of the generalized "harmonic Ritz values" induced by the optimal algorithms from [6].

Research Objective 3: trace and low-rank approximation of $f[\mathbf{A}]$. The most widely used algorithms for estimating $\operatorname{tr}(f[\mathbf{A}]) = f(\lambda_1) + \cdots + f(\lambda_n)$ are closely related to the task of computing $\mathbf{b}^\mathsf{T} f[\mathbf{A}]\mathbf{b}$ for a suitible choice of random vector $\mathbf{b}$; i.e. using stochastic trace estimation [13]. While a range of existing analyses balance the errors of stochastic trace estimation with

the convergence guarantees of Krylov subspace methods [15, 3, 32], these analyses do not take advantage of the specific properties of the randomized estimator, and instead apply general worst-case bounds for KSMs. Thus, I will design (high probability) bounds which take advantage of the full random structure of $\mathbf{b}$. This is closely related to the average case analysis approach in Objective 1, although here the randomness is due to $\mathbf{b}$ rather than $\mathbf{A}$.

Following the successful analysis of stochastic matrix function trace estimation, I will study the more general task of low-rank approximation via matrix sketching. Low rank approximations to a matrix are widely useful because they they can easily speed up essentially any downstream applications involving the original matrix. A variety of algorithms for obtaining a low rank approximations have been developed, and a large number of such algorithms are based on an algorithmic technique called sketching [51, 52]. A core step of sketching involves computing the product of the matrix in question with a set of suitably chosen random vectors. If this matrix is a matrix-function $f[\mathbf{A}]$, then each of these products can be computed approximately using Lanczos-FA or the proposed methods from Objective 2. This naturally yields algorithms for obtaining low rank approximations to $f[\mathbf{A}]$, and balancing the error from the sketching step with the error from Lanczos-FA is a reasonable and fairly straightforward task. However, as with matrix function stochastic trace estimation, there is significant potential for redundancies, and accounting for this has the potential to lead to better algorithms.

Extensions to non-normal matrices. Objectives 2 and 3 have been framed in the setting of symmetric matrices (which are normal). The Cauchy Integral Formula yields a natural way of defining a matrix function of a non-normal matrices, and the same questions are still relevant in the non-normal setting. However, the behavior of non-normal matrices is far more subtle and difficult to analyze than the behavior of their normal counterparts. Therefore, the extension of these objectives to the non-normal setting provides a set of new non-trivial problems.

A range of the work from my PhD focuses on algorithms for symmetric matrices, but in many cases, the ideas can be naturally generalized to non-normal matrices. In fact, in several cases we have preliminary results in this direction. Generalizing these analyses to the non-normal setting is more straightforward than generalizing some of the objectives in this section since we have already addressed the symmetric/normal case. Working on generalizing the work from my PhD will provide a natural entry point to non-normal numerical linear algebra which I will use to familiarize myself with the tools and techniques needed to address the proposed research objective in the non-normal setting.

OUTREACH OBJECTIVE

My outreach objective is to continue organizing conference and seminars sessions with the aim of ensuring speaker lineups are representative of broader society, not just the current state of science or academia. **Finding potential speakers from underrepresented groups is inherently difficult, but during my PhD I have gained experience doing exactly this.** For example, in the [redacted] Achieving this level of representation took a deliberate

and extended effort on my part, not only to find and invite a diverse set of speakers, but also to design a topic for the session which would maximize the number of potential speakers from underrepresented groups as well as provide a range of research backgrounds and perspectives.

I also aim to invite speakers who will provide a diversity of perspectives, especially perspectives which are directly relevant to increasing equity within the field. As an example, when organizing the department's numerical analysis seminar I brought in a data-ethicist to speak about her research, which addresses important questions such as the role of data and algorithms in perpetuating system injustices along the lines of rage, gender, and ability. Such topics are obviously of great importance, but are not formally taught in most STEM programs, necessitating the need for other means of dissemination.

Other Service. Beyond research, I will continue to work to empower the next generation of scientists by advocating for increasing student representation in academia, by creating inclusive and engaging classroom environments, and by direct mentorship.

I have demonstrated a commitment to this type of outreach during grad school. For instance, I served as my department's graduate student representative (GSR) for the 2020-2021 academic year during which I (i) organized the student body to (successfully) petition the department for more transparency and representation in future faculty hirings, and (ii) requested, and helped implement, the use of gender-neutral phrasing on the department's website. I have also helped organize and served on many student panels related to grad student well-being, including panels on mental health and department climate.

Tyler Chen
chentyl@uw.edu

Bibliography

https://chen.pw

[1] T. Chen and E. C. Carson. "Predict-and-recompute conjugate gradient variants". In: *SIAM Journal on Scientific Computing* 42.5 (2020). DOI: 10.1137/19m1276856. arXiv: 1905.01549 [cs.NA].

[2] T. Chen. "Non-asymptotic moment bounds for random variables rounded to non-uniformly spaced sets". In: *Stat* (2021). DOI: 10.1002/STA4.395. arXiv: 2007.11041 [math.ST].

[3] T. Chen, T. Trogdon, and S. Ubaru. "Analysis of stochastic Lanczos quadrature for spectrum approximation". In: *Proceedings of the 37th International Conference on Machine Learning*. Proceedings of Machine Learning Research. PMLR, 2021. arXiv: 2105.06595 [cs.DS].

[4] A. Greenbaum, H. Liu, and T. Chen. "On the Convergence Rate of Variants of the Conjugate Gradient Algorithm in Finite Precision Arithmetic". In: 43.5 (Jan. 2021). DOI: 10.1137/20m1346249. URL: https://doi.org/10.1137/20m1346249.

[5] T. Chen, A. Greenbaum, C. Musco, and C. Musco. *Error bounds for Lanczos-based matrix function approximation*. 2021. arXiv: 2106.09806 [math.NA].

[6] T. Chen, A. Greenbaum, C. Musco, and C. Musco. *Optimal low-memory rational matrix function approximation*. 2021. Preprint available upon request.

[7] J. van den Eshof, A. Frommer, T. Lippert, K. Schilling, and H. van der Vorst. "Numerical methods for the QCDd overlap operator. I. Sign-function and error bounds". In: *Computer Physics Communications* 146.2 (2002). ISSN: 0010-4655. DOI: https://doi.org/10.1016/S0010-4655(02)00455-1. URL: http://www.sciencedirect.com/science/article/pii/S0010465502004551.

[8] Y. Saad. "Analysis of Some Krylov Subspace Approximations to the Matrix Exponential Operator". In: *SIAM Journal on Numerical Analysis* 29.1 (1992). DOI: 10.1137/0729014. eprint: https://doi.org/10.1137/0729014. URL: https://doi.org/10.1137/0729014.

[9] M. Hochbruck and C. Lubich. "On Krylov Subspace Approximations to the Matrix Exponential Operator". In: *SIAM Journal on Numerical Analysis* 5 (Oct. 1997). DOI: 10.1137/s0036142995280572. URL: https://doi.org/10.1137/s0036142995280572.

[10] G. Pleiss, M. Jankowiak, D. Eriksson, A. Damle, and J. R. Gardner. *Fast Matrix Square Roots with Applications to Gaussian Processes and Bayesian Optimization*. 2020. arXiv: 2006.11267 [cs.LG].

[11] Y. Jin and A. Sidford. "Principal Component Projection and Regression in Nearly Linear Time through Asymmetric SVRG". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett. Curran Associates, Inc., 2019. URL: http://papers.nips.cc/paper/8642-principal-component-projection-and-regression-in-nearly-linear-time-through-asymmetric-svrg.pdf.

[12] N. J. Higham. *Functions of Matrices*. Society for Industrial and Applied Mathematics, 2008.

[13] M. Hutchinson. "A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines". In: *Communications in Statistics - Simulation and Computation* 18.3 (Jan. 1989). DOI: 10.1080/03610918908812806. URL: https://doi.org/10.1080/03610918908812806.

[14] R. A. Meyer, C. Musco, C. Musco, and D. P. Woodruff. *Hutch++: Optimal Stochastic Trace Estimation*. 2021. arXiv: 2010.09649 [cs.DS].

[15] S. Ubaru, J. Chen, and Y. Saad. "Fast Estimation of $tr(f(A))$ via Stochastic Lanczos Quadrature". In: *SIAM Journal on Matrix Analysis and Applications* 38.4 (2017).

[16] E. Estrada. "Characterization of 3D molecular structure". In: *Chemical Physics Letters* 319.5-6 (Mar. 2000). DOI: 10.1016/s0009-2614(00)00158-5. URL: https://doi.org/10.1016/s0009-2614(00)00158-5.

[17] R. P. Barry and R. K. Pace. "Monte Carlo estimates of the log determinant of large sparse matrices". In: *Linear Algebra and its Applications* 289.1-3 (Mar. 1999). DOI: 10.1016/s0024-3795(97)10009-x. URL: https://doi.org/10.1016/s0024-3795(97)10009-x.

[18] R. Pace and J. P. LeSage. "Chebyshev approximation of log-determinants of spatial weight matrices". In: *Computational Statistics & Data Analysis* 45.2 (Mar. 2004). DOI: 10.1016/s0167-9473(02)00321-3. URL: https://doi.org/10.1016/s0167-9473(02)00321-3.

[19] K. Bergermann and M. Stoll. *Orientations and matrix function-based centralities in multiplex network analysis of urban public transport*. 2021. arXiv: 2107.12695 [physics.soc-ph].

[20] S. Wang, Y. Sun, C. Musco, and Z. Bao. "Public Transport Planning". In: *Proceedings of the 2021 International Conference on Management of Data*. ACM, June 2021. DOI: 10.1145/3448016.3457247. URL: https://doi.org/10.1145/3448016.3457247.

[21] H. Avron. "Counting Triangles in Large Graphs using Randomized Matrix Trace Estimation". In: *Proceedings of KDD-LDMTA*. 2010.

[22] F. Ducastelle and F. Cyrot-Lackmann. "Moments developments and their application to the electronic charge distribution of d bands". In: *Journal of Physics and Chemistry of Solids* 31.6 (June 1970). DOI: 10.1016/0022-3697(70)90134-4. URL: https://doi.org/10.1016/0022-3697(70)90134-4.

[23] J. C. Wheeler and C. Blumstein. "Modified Moments for Harmonic Solids". In: *Physical Review B* 6.12 (Dec. 1972). DOI: 10.1103/physrevb.6.4380. URL: https://doi.org/10.1103/physrevb.6.4380.

[24] R Haydock, V Heine, and M. J. Kelly. "Electronic structure based on the local atomic environment for tight-binding bands". In: *Journal of Physics C: Solid State Physics* 5.20 (Oct. 1972). DOI: 10.1088/0022-3719/5/20/004. URL: https://doi.org/10.1088/0022-3719/5/20/004.

[25] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske. "The kernel polynomial method". In: *Reviews of Modern Physics* 78.1 (Mar. 2006). DOI: 10.1103/revmodphys.78.275. URL: https://doi.org/10.1103/revmodphys.78.275.

[26] B. Ghorbani, S. Krishnan, and Y. Xiao. *An Investigation into Neural Net Optimization via Hessian Eigenvalue Density*. 2019. arXiv: 1901.10159 [cs.LG].

[27] V. Papyan. *The Full Spectrum of Deepnet Hessians at Scale: Dynamics with SGD Training and Sample Size*. 2019. arXiv: 1811.07062 [cs.LG].

[28] Z. Yao, A. Gholami, K. Keutzer, and M. Mahoney. *PyHessian: Neural Networks Through the Lens of the Hessian*. 2020. arXiv: 1912.07145 [cs.LG].

[29] E. Polizzi. "Density-matrix-based algorithm for solving eigenvalue problems". In: *Physical Review B* 79.11 (Mar. 2009). DOI: 10.1103/physrevb.79.115112. URL: https://doi.org/10.1103/physrevb.79.115112.

[30] R. Li, Y. Xi, L. Erlandson, and Y. Saad. "The Eigenvalues Slicing Library (EVSL): Algorithms, Implementation, and Software". In: *SIAM Journal on Scientific Computing* 41.4 (Jan. 2019). DOI: 10.1137/18m1170935. URL: https://doi.org/10.1137/18m1170935.

[31] M. R. Hestenes and E. Stiefel. *Methods of conjugate gradients for solving linear systems*. Vol. 49. NBS Washington, DC, 1952.

[32] T. Chen, T. Trogdon, and S. Ubaru. *Randomized matrix-free quadrature*. 2021.

[33] T. Chen and T. Trogdon. *Average case behavior of the Lanczos algorithm in finite precision arithmetic*. 2021.

[34] A. Greenbaum. *Iterative Methods for Solving Linear Systems*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1997.

[35] Y. T. Lee and A. Sidford. "Efficient accelerated coordinate descent methods and faster algorithms for solving linear systems". In: *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE. 2013.

[36] R. Johnson and T. Zhang. "Accelerating stochastic gradient descent using predictive variance reduction". In: *Advances in neural information processing systems* 26 (2013).

[37] N. Agarwal, S. Kakade, R. Kidambi, Y. T. Lee, P. Netrapalli, and A. Sidford. "Leverage score sampling for faster accelerated regression and erm". In: *arXiv preprint arXiv:1711.08426* (2017).

[38] A. Greenbaum. "Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences". In: *Linear Algebra and its Applications* 113 (1989). DOI: https://doi.org/10.1016/0024-3795(89)90285-1. URL: http://www.sciencedirect.com/science/article/pii/0024379589902851.

[39] P. Deift and T. Trogdon. "The conjugate gradient algorithm on well-conditioned Wishart matrices is almost deterministic". In: *Quarterly of Applied Mathematics* (July 2020). DOI: 10.1090/qam/1574. URL: https://doi.org/10.1090/qam/1574.

[40] E. Paquette and T. Trogdon. *Universality for the conjugate gradient and MINRES algorithms on sample covariance matrices*. 2020. arXiv: 2007.00640 [math.NA].

[41] B. Beckermann and A. B. J. Kuijlaars. "Superlinear Convergence of Conjugate Gradients". In: *SIAM Journal on Numerical Analysis* 39.1 (2001). DOI: 10.1137/S0036142999363188. eprint: https://doi.org/10.1137/S0036142999363188. URL: https://doi.org/10.1137/S0036142999363188.

[42] X. Ding and T. Trogdon. *The conjugate gradient algorithm on a general class of spiked covariance matrices*. 2021. arXiv: 2106.13902 [math.NA].

[43] C. Paquette, B. van Merriënboer, E. Paquette, and F. Pedregosa. *Halting Time is Predictable for Large Models: A Universality Property and Average-case Analysis*. 2020. arXiv: 2006.04299 [math.OC].

[44] C. Paquette, K. Lee, F. Pedregosa, and E. Paquette. *SGD in the Large: Average-case Analysis, Asymptotics, and Stepsize Criticality*. 2021. arXiv: 2102.04396 [math.OC].

[45] N. Hale, N. J. Higham, and L. N. Trefethen. "Computing $A^\alpha, \log(A)$, and Related Matrix Functions by Contour Integrals". In: *SIAM Journal on Numerical Analysis* 46.5 (2008). DOI: 10.1137/070700607. eprint: https://doi.org/10.1137/070700607. URL: https://doi.org/10.1137/070700607.

[46] M. Ilic, I. W. Turner, and D. P. Simpson. "A restarted Lanczos approximation to functions of a symmetric matrix". In: *IMA Journal of Numerical Analysis* 30.4 (June 2009). DOI: 10.1093/imanum/drp003. URL: https://doi.org/10.1093/imanum/drp003.

[47] A. Frommer, S. Güttel, and M. Schweitzer. "Efficient and Stable Arnoldi Restarts for Matrix Functions Based on Quadrature". In: *SIAM Journal on Matrix Analysis and Applications* 35.2 (Jan. 2014). DOI: 10.1137/13093491x. URL: https://doi.org/10.1137/13093491x.

[48] A. Frommer and M. Schweitzer. "Error bounds and estimates for Krylov subspace approximations of Stieltjes matrix functions". In: *BIT Numerical Mathematics* 56.3 (Dec. 2015). DOI: 10.1007/s10543-015-0596-3. URL: https://doi.org/10.1007/s10543-015-0596-3.

[49] A. Frommer and V. Simoncini. "Error Bounds for Lanczos Approximations of Rational Functions of Matrices". In: *Numerical Validation in Current Hardware Architectures*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. ISBN: 978-3-642-01591-5.

[50] A. Frommer, K. Kahl, T. Lippert, and H. Rittich. "2-Norm Error Bounds and Estimates for Lanczos Approximations to Linear Systems and Rational Matrix Functions". In: *SIAM Journal on Matrix Analysis and Applications* 34.3 (2013). DOI: 10.1137/110859749. eprint: https://doi.org/10.1137/110859749. URL: https://doi.org/10.1137/110859749.

[51] P.-G. Martinsson and J. A. Tropp. "Randomized numerical linear algebra: Foundations and algorithms". In: *Acta Numerica* 29 (May 2020). DOI: 10.1017/s0962492920000021. URL: https://doi.org/10.1017/s0962492920000021.

[52] C. Musco and C. Musco. "Randomized Block Krylov Methods for Stronger and Faster Approximate Singular Value Decomposition". In: *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*. NIPS'15. Montreal, Canada: MIT Press, 2015.