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 [EFLSV02], solving differential equations in applied math [Saa92; HL97], Gaussian process sampling in statistics [PJEDG20], principle component projection and regression in data science [JS19], and a range of other applications [Hig08].

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 [Hut89; MMMW21] to approximate the trace of a matrix function [UCS17; CTU21a]. Applications of the trace of matrix functions include characterizing the degree of protein folding in biology [Est00], maximum likelihood estimation in statistics [BP99; PL04], designing better public transit in urban planning [BS21; WSMB21] and finding triangle counts and other motifs in network science [Avr10]. The trace of matrix functions is intimately related to spectral density estimation, which is used in electronic structure computations [DCL70; WB72; HHK72] and other tasks in physics [WWAF06], probing the behavior of neural networks in machine learning [GKX19; Pap19; YGKM20]. and load balancing modern parallel eigensolvers in numerical linear algebra [Pol09; LXES19].

PERSONAL BACKGROUND/ACOMPLISHMENTS

Thesis. My thesis work is centered on the Lanczos method for matrix function approximation (Lanczos-FA) [HS52; Saa92]. 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) [CTU21a; CTU21b; CGMM21a; CGMM21b; CT21] and (ii) designing more efficient implementations of Lanczos-FA for use in modern computing environments (e.g. distributed memory supercomputers) [CC20; CGMM21b].

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.

OVERVIEW OF PROPOSAL

If the mathematical science postdoctoral research fellowship is awarded, 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.** Towards this end, the sponsoring scientist ████████████ and intended co-mentor ████████████ both of whom are professors at New York University (NYU), have been carefully chosen for their expertise in theoretical computer science, optimization, and numerical linear algebra. **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. Extend stochastic gradient methods for linear systems to the more general task of applying matrix functions to vectors.

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 $\lambda_{\max} = \lambda_1 \geq \cdots \geq \lambda_n = \lambda_{\min}$ and corresponding (orthonormal) eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Thus, $\mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^\top$ and $\mathbf{u}_i^\top \mathbf{u}_j = 0$ 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 $f[\mathbf{A}] := \sum_{i=1}^n f(\lambda_i) \mathbf{u}_i \mathbf{u}_i^\top$. 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 $\sqrt{\mathbf{y}^\top \mathbf{A} \mathbf{y}}$.

Research Objective 1: Compare conjugate gradient and fast stochastic gradient methods. When $\mathbf{A}$ is positive definite, the conjugate gradient algorithm (CG) [HS52], 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 $\kappa = \lambda_{\max}/\lambda_{\min}$ is the condition number of $\mathbf{A}$ [Gre97]. 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 $\kappa^* = \lambda_{\text{ave}}/\lambda_{\min}$ is the "smoothed condition number" of $\mathbf{A}$ [LS13; JZ13; AKKLNS17].

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 [Gre89; Gre97]. During my PhD I extended many existing bounds for CG to bounds for Lanczos-FA and related algorithms [CGMM21a; CTU21a; CGMM21b]. 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 [DT20; PT20] whereas in others it is known that it is loose [BK01; DT21]. Gradient descent and accelerated gradient descent have also been studied in this setting [PMPP20; PLPP21]. 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: stochastic methods for $f[\mathbf{A}]\mathbf{b}$. If $f(x) = 1/x$, then $f[\mathbf{A}]\mathbf{b} = \mathbf{A}^{-1}\mathbf{b}$ is the solution to the linear system of equation $\mathbf{A}\mathbf{x} = \mathbf{b}$. In this case, a range of algorithms can be used to compute approximations to $f[\mathbf{A}]\mathbf{b}$, including CG and the fast SGMs described above. For general f, however, current techniques are far more limited. The most common approaches are closely related to Lanczos-FA, which is a generalization of CG. Deriving a generalization of SGMs for this task remains open, but has the potential to yield algorithms which inherit the fast runtime gurantees of SGMs.

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 $\sum_i (\mathbf{A}^2 + c_i \mathbf{I})^{-1}\mathbf{b}$; i.e. by solving a series of shifted linear systems [HHT08]. CG/Lancos-FA can solve all of the systems for essentially the cost of solving a single system by taking advantage of redundant information [GS21; CGMM21b]. It's known that SGMs can be used to efficiently solve individual systems from both classes of approximation [JS19], but how to efficiently solve the sequence of systems by taking advantage of the shift structure is less clear. Understanding this will serve as a preliminary step towards deriving new methods for computing $f[\mathbf{A}]\mathbf{b}$.

Research Objective 3: trace and low-rank approximation of $f[\mathbf{A}]$. The most widely used algorithms for estimating $\text{tr}(f[\mathbf{A}]) = f(\lambda_1) + \cdots + f(\lambda_n)$ 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 [Hut89]. While a range of existing analyses balance the errors of stochastic trace estimation with the convergence guarantees of Krylov subspace methods [UCS17; CTU21a; CTU21b], 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 $\mathbf{A}$.
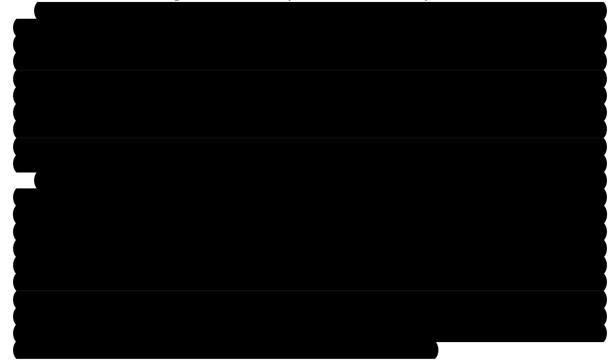
Following the successful analysis of stochastic matrix function trace estimation, I will study the more general task of low-rank approximation via matrix sketching. Low rank approximations to a matrix are widely useful because they they can easily speed up essentially any downstream applications involving 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 [MT20; MM15]. 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.

CAREER DEVELOPMENT AND CHOICE OF SPONSORING SCIENTIST/HOST INSTITUTION

The MSPRF offers a once-in-career opportunity to have near complete control over the research environment I will be in during the next few years. During my PhD I worked closely with co-advisors and collaborators from a range of distinct research backgrounds. This experience motivates me to pursue a similarly diverse environment as preparation for a career of collaborative and cross-disciplinary work. **The sponsoring scientist▮▮▮▮▮▮▮▮▮ and intended co-mentor ▮▮▮▮▮▮▮▮▮ have backgrounds complimentary to my own** which are well suited to support me in successfully carrying out the research objectives of this proposal. In addition, **the research community at the host institution, New York University, and throughout the greater New York City area provides a unique opportunity to collaborate with researchers from a huge number of nearby academic and industry institutions.**

▮▮▮

Beyond mentorship, NYU and it's proximity to other research institutions will directly support my career development by providing easy access to a wide range of researchers including ▮▮▮▮▮▮▮▮▮▮▮▮▮
▮▮▮

[REDACTED]

BROADER IMPACTS

Working closely with ███████████ will help strengthen ties between theoretical computer science, optimization, and numerical linear algebra, and ████████████████████████████████ ████████████████████ This, in and of itself, is a meaningful contribution – the intersection of these fields contains a wealth of important problems, and engaging in such collaborations serves as a stepping stone for future collaborations between researchers in these fields.

Research Objectives. The proposed research objectives support basic science by ensuring that fundamental linear algebraic tasks can be solved quickly and with sufficient accuracy. Since algorithms for these tasks are often used in a modular fashion in downstream applications, any new algorithms developed have the potential to immediately speed up these applications. All research code I write is released to the public domain at the same time the corresponding manuscript is posted. The code is formatted such that all figures can be generated by executing a single script to help ensure the accuracy and reproducibility of the figures and experiments in my papers. Moreover, because the code is open source, it can easily used by others.

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 ██████████████████ ███ 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, and 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 wellbeing, including panels on mental health and department culture/climate.

[CTU21a]   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].

[CGMM21a]   T. Chen, A. Greenbaum, C. Musco, and C. Musco. *Error bounds for Lanczos-based matrix function approximation*. 2021. arXiv: 2106.09806 [math.NA].

[CC20]   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].

[GLC21]   A. Greenbaum, H. Liu, and T. Chen. "On the Convergence Rate of Variants of the Conjugate Gradient Algorithm in Finite Precision Arithmetic". In: *SIAM Journal on Scientific Computing* (July 2021). DOI: 10.1137/20m1346249. arXiv: 1905.05874 [cs.NA].

[PET]   PETSc. *PIPEPRCG method*. URL: https://gitlab.com/petsc/petsc/-/blob/main/src/ksp/ksp/impls/cg/pipeprcg/pipeprcg.c.

[Chea]   T. Chen. *An Introduction to Modern Analysis of the Conjugate Gradient Algorithm in Exact and Finite Precision*. URL: https://chen.pw/research/cg/.

[Che21]   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].

[Cheb]   T. Chen. *Code accompanying: Analysis of stochastic Lanczos quadrature for spectrum approximation*. URL: https://github.com/chentyl/SLQ_analysis.

[Chec]   T. Chen. *Code accompanying: Analysis of stochastic Lanczos quadrature for spectrum approximation*. URL: https://github.com/chentyl/lanczos_function_CIF.

[Ched]   T. Chen. *Code accompanying: Predict-and-recompute conjugate gradient variants*. URL: https://github.com/tchen01/new_cg_variants/tree/master/predict_and_recompute.

[EFLSV02]   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.

[Saa92]   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.

[HL97]   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.

[PJEDG20]   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].

[JS19]   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.

[Hig08]   N. J. Higham. *Functions of Matrices*. Society for Industrial and Applied Mathematics, 2008.

[Hut89]   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.

[MMMW21] R. A. Meyer, C. Musco, C. Musco, and D. P. Woodruff. *Hutch++: Optimal Stochastic Trace Estimation*. 2021. arXiv: 2010.09649 [cs.DS].

[UCS17]   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).

[Est00]   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.

[BP99]   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.

[PL04]   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.

[BS21]   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].

[WSMB21] 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.

[Avr10]   H. Avron. "Counting Triangles in Large Graphs using Randomized Matrix Trace Estimation". In: *Proceedings of KDD-LDMTA*. 2010.

[DCL70]   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.

[WB72]   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.

[HHK72]   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.

[WWAF06]   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.

[GKX19]   B. Ghorbani, S. Krishnan, and Y. Xiao. *An Investigation into Neural Net Optimization via Hessian Eigenvalue Density*. 2019. arXiv: 1901.10159 [cs.LG].

[Pap19]   V. Papyan. *The Full Spectrum of Deepnet Hessians at Scale: Dynamics with SGD Training and Sample Size*. 2019. arXiv: 1811.07062 [cs.LG].

[YGKM20]   Z. Yao, A. Gholami, K. Keutzer, and M. Mahoney. *PyHessian: Neural Networks Through the Lens of the Hessian*. 2020. arXiv: 1912.07145 [cs.LG].

[Pol09]   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.

[LXES19]   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.

[HS52]   M. R. Hestenes and E. Stiefel. *Methods of conjugate gradients for solving linear systems*. Vol. 49. NBS Washington, DC, 1952.

[CTU21b]   T. Chen, T. Trogdon, and S. Ubaru. *Randomized matrix-free quadrature*. 2021.

[CGMM21b]   T. Chen, A. Greenbaum, C. Musco, and C. Musco. *Optimal low-memory rational matrix function approximation*. 2021.

[CT21]   T. Chen and T. Trogdon. *Average case behavior of the Lanczos algorithm in finite precision arithmetic*. 2021.

[Gre97]   A. Greenbaum. *Iterative Methods for Solving Linear Systems*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1997.

[LS13]   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.

[JZ13]   R. Johnson and T. Zhang. "Accelerating stochastic gradient descent using predictive variance reduction". In: *Advances in neural information processing systems* 26 (2013).

[AKKLNS17] 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).

[Gre89]   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.

[DT20]   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.

[PT20]   E. Paquette and T. Trogdon. *Universality for the conjugate gradient and MINRES algorithms on sample covariance matrices*. 2020. arXiv: 2007.00640 [math.NA].

[BK01]   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.

[DT21]   X. Ding and T. Trogdon. *The conjugate gradient algorithm on a general class of spiked
           covariance matrices.* 2021. arXiv: 2106.13902 [math.NA].

[PMPP20]   C. Paquette, B. van MerriÃńnboer, E. Paquette, and F. Pedregosa. *Halting Time is Pre-
           dictable for Large Models: A Universality Property and Average-case Analysis.* 2020. arXiv:
           2006.04299 [math.OC].

[PLPP21]   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].

[HHT08]   N. Hale, N. J. Higham, and L. N. Trefethen. "Computing A^{α}, $\log(A)$, 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.

[GS21]   S. Güttel and M. Schweitzer. "A Comparison of Limited-memory Krylov Methods for
           Stieltjes Functions of Hermitian Matrices". In: *SIAM Journal on Matrix Analysis and
           Applications* 42.1 (Jan. 2021). DOI: 10.1137/20m1351072. URL: https://doi.org/10.
           1137/20m1351072.

[MT20]   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.

[MM15]   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.

[BO01]   J. V. Burke and M. L. Overton. "Variational analysis of non-Lipschitz spectral functions".
           In: *Math. Program.* 90.2, Ser. A (2001). ISSN: 0025-5610. DOI: 10.1007/s101070100225.
           URL: https://doi.org/10.1007/s101070100225.

[BLO05]   J. V. Burke, A. S. Lewis, and M. L. Overton. "A robust gradient sampling algorithm for
           nonsmooth, nonconvex optimization". In: *SIAM J. Optim.* 15.3 (2005). ISSN: 1052-6234.
           DOI: 10.1137/030601296. URL: https://doi.org/10.1137/030601296.

[BHLO06]   J. V. Burke, D. Henrion, A. S. Lewis, and M. L. Overton. "Stabilization via nonsmooth,
           nonconvex optimization". In: *IEEE Trans. Automat. Control* 51.11 (2006). ISSN: 0018-
           9286. DOI: 10.1109/TAC.2006.884944. URL: https://doi.org/10.1109/TAC.
           2006.884944.

[LO13]   A. S. Lewis and M. L. Overton. "Nonsmooth optimization via quasi-Newton methods". In:
           Math. Program. 141.1-2, Ser. A (2013). ISSN: 0025-5610. DOI: 10.1007/s10107-012-
           0514-2. URL: https://doi.org/10.1007/s10107-012-0514-2.

[ABBO11]   R. Alam, S. Bora, R. Byers, and M. L. Overton. "Characterization and construction of
           the nearest defective matrix via coalescence of pseudospectral components". In: *Linear
           Algebra Appl.* 435.3 (2011). ISSN: 0024-3795. DOI: 10.1016/j.laa.2010.09.022. URL:
           https://doi.org/10.1016/j.laa.2010.09.022.

[GLO17]   A. Greenbaum, A. S. Lewis, and M. L. Overton. "Variational analysis of the Crouzeix ratio".
            In: *Math. Program.* 164.1-2, Ser. A (2017). ISSN: 0025-5610. DOI: `10.1007/s10107-016-`
            `1083-6`. URL: `https://doi.org/10.1007/s10107-016-1083-6`.

[DM21]   P. Dharangutte and C. Musco. "Dynamic trace estimation". In: *Proceedings of the 35th*
            International Conference on Neural Information Processing Systems. Dec. 2021.

[BKM21]   V. Braverman, A. Krishnan, and C. Musco. *Linear and Sublinear Time Spectral Density*
            Estimation. 2021. arXiv: `2104.03461` [`cs.DS`].