# Improved Spectral Density Estimation via Explicit and Implicit Deflation<sup>1</sup>

Rajarshi Bhattacharjee, Rajesh Jayaram, Cameron Musco, Christopher Musco, Archan Ray

#### Abstract

We study algorithms for approximating the spectral density of a symmetric matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  that is accessed through matrix-vector products. By combining an existing Chebyshev polynomial moment matching method with a deflation step that approximately projects off the largest magnitude eigendirections of  $\mathbf{A}$  before estimating the spectral density, we give an  $\epsilon \sigma_{\ell}(\mathbf{A})$  error approximation in the Wasserstein-1 metric using  $O(\ell \log n + 1/\epsilon)$  matrix-vector products, where  $\sigma_{\ell}(\mathbf{A})$  is the  $\ell^{th}$  largest singular value of  $\mathbf{A}$ . When  $\mathbf{A}$  exhibits fast singular value decay, this can be much stronger than prior work, which gives error  $\epsilon \sigma_1(\mathbf{A})$  using  $O(1/\epsilon)$  matrix-vector products are required to achieve error  $\epsilon \sigma_{\ell}(\mathbf{A})$ .

We further show that the popular Stochastic Lanczos Quadrature (SLQ) method matches the above bound, even though SLQ itself is parameter-free and performs no explicit deflation. This explains the strong practical performance of SLQ, and motivates a simple variant that achieves an even tighter error bound. Our error bound for SLQ leverages an analysis that views it as an implicit polynomial moment matching method, along with recent results on low-rank approximation with single-vector Krylov methods. We use these results to show that SLQ can perform implicit deflation as part of moment matching.

## **1** Introduction

Given a symmetric matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  with eigenvalues  $\lambda_1(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})$ , the spectral density of  $\mathbf{A}$  is defined as:

$$s_{\mathbf{A}}(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - \lambda_i(\mathbf{A})),$$

where  $\delta(.)$  is the Dirac delta function. The spectral density  $s_{\mathbf{A}}$  can be computed directly by performing a full eigendecomposition of  $\mathbf{A}$  in  $O(n^{\omega})$  time, for  $\omega \approx 2.37$  being the exponent of fast matrix multiplication. However, when  $\mathbf{A}$  is very large or where  $\mathbf{A}$  can only be accessed through a small number of queries, we often want to approximate  $s_{\mathbf{A}}$  by some  $\tilde{s}_{\mathbf{A}}$  such that  $\tilde{s}_{\mathbf{A}}$  and  $s_{\mathbf{A}}$  are close in some metric. Spectral density estimation is applied throughout the sciences [Ski89, SR94, STBB17, SRS20], network science [FDBV01, EG17, DBB19], machine learning and deep learning in particular [RL18, PSG18, MM19, GKX19], numerical linear algebra [DNPS16, LXES19], and beyond. In this work, we focus on the Wasserstein-1 (i.e., earth mover's) distance,  $W_1(s_{\mathbf{A}}, \tilde{s}_{\mathbf{A}})$ , which has been studied in a number of recent works giving formal approximation guarantees for SDE [CTU21, BKM22, CTU22]. Moreover,  $\mathbf{A}$  will be accessed only through matrix vector queries of the form  $\mathbf{Av}$  for any query vector  $\mathbf{v}$ . Most state-of-the-art matrix-vector query algorithms for SDE are based on Krylov subspace methods that fall into two general classes.

**Moment Matching.** The first class of methods approximates  $s_{\mathbf{A}}$  by approximating its polynomial moments. I.e.,  $\mathbb{E}_{s_{\mathbf{A}}}[p(x)] = \frac{1}{n} \sum_{i=1}^{n} p(\lambda_i(\mathbf{A})) = \frac{1}{n} \operatorname{tr}(p(\mathbf{A}))$ , where p is a low-degree polynomial. We can employ stochastic trace estimation methods like Hutchinson's method [Gir87, Hut90] to approximate this trace using just a small number of matrix-vector products with  $p(\mathbf{A})$  and in turn

<sup>&</sup>lt;sup>1</sup>Based on paper to appear at SODA 2025. Preprint available at [BJM<sup>+</sup>24]

**A**, since if p has degree k, a single matrix-vector product with  $p(\mathbf{A})$  can be performed using k matrix vector products with  $\mathbf{A}$ . After approximating the moments for a set of low-degree polynomials (e.g., the first k monomials, or the first k Chebyshev polynomials), we can let  $\tilde{s}_{\mathbf{A}}$  be a distribution that matches these moments as closely as possible, and thus should closely match  $s_{\mathbf{A}}$ . Moment matching methods include the popular Kernel Polynomial Method (KPM) [SR94, Wan94, WWAF06] and its variants [CPB10, LSY16, BKM22, Che23]. Braverman et al. [BKM22] analyze a Chebyshev Moment Matching method, which can be thought of as a simple variant of KPM, showing that the method can compute  $\tilde{s}_{\mathbf{A}}$  satisfying  $W_1(s_{\mathbf{A}}, \tilde{s}_{\mathbf{A}}) \leq \epsilon \cdot \|\mathbf{A}\|_2$  with probability  $\geq 1 - \delta$  using just  $O(b/\epsilon)$  matrix vector products, where  $b = \max(1, \frac{1}{n\epsilon^2}\log^2 \frac{1}{\epsilon\delta}\log^2 \frac{1}{\epsilon})$ . Note that b = 1 in the common case when  $\epsilon = \tilde{\Omega}(1/\sqrt{n})$ . Here  $\|\mathbf{A}\|_2$  denotes the spectral norm of  $\mathbf{A}$  – i.e., its largest eigenvalue magnitude. They prove a similar guarantee for KPM itself, but with a worse dependence on  $\epsilon$ .

Lanczos-Based Methods. This class of methods computes a small number of approximate eigenvalues of **A** using the Lanczos method, and lets  $\tilde{s}_{\mathbf{A}}$  be a distribution supported on these eigenvalues, with appropriately chosen probability mass placed at each. The canonical method of this form is Stochastic Lanczos Quadrature (SLQ) [CTU21, GM09]. Many other variants have also been studied. Some place probability mass not just at the approximate eigenvalues, but on Gaussian or other simple distributions centered at these eigenvalues [LG82, BRP92, LSY16, HHK72]. Chen et al. [CTU21, CTU22] prove that the Lanczos-based SLQ method gives essentially the same approximation bound as [BKM22]: error  $\epsilon \cdot ||\mathbf{A}||_2$  using  $O(1/\epsilon)$  matrix-vector products when  $\epsilon = \tilde{\Omega}(1/\sqrt{n})^2$ .

## 2 Our Results

Our main contribution is to show that both moment matching and Lanczos based methods for SDE can achieve improved bounds on  $W_1(s_{\mathbf{A}}, \tilde{s}_{\mathbf{A}})$  that depend on  $\sigma_{l+1}(\mathbf{A})$ , the  $(l+1)^{st}$  largest singular value of  $\mathbf{A}$  for some parameter l, instead of  $\|\mathbf{A}\|_2$ . For matrices that exhibit spectral decay and thus have  $\sigma_{l+1}(\mathbf{A}) \ll \sigma_1(\mathbf{A}) = \|\mathbf{A}\|_2$ , our bounds can be much stronger than the bound  $W_1(s_{\mathbf{A}}, \tilde{s}_{\mathbf{A}}) \leq \epsilon \cdot \|\mathbf{A}\|_2$  achieved in prior work, which roughly corresponds to estimating each eigenvalue to average error  $\epsilon \cdot \|\mathbf{A}\|_2$ . We also provide a lower bound showing that our bounds are near optimal upto some logarithmic factors.

### 2.1 Improved SDE via Moment Matching with Explicit Deflation

Our first contribution is a modification of the moment matching method of [BKM22] that first 'deflates' off any eigenvalue of **A** with magnitude significantly larger than  $\sigma_{l+1}(\mathbf{A})$ , before estimating the spectral density. Eigenvalue deflation is widely applied throughout numerical linear algebra to problems like linear system solving [BEPW98, FV01, GOSS16, FTU23], trace estimation [GSO17, Lin17, MMMW21], norm estimation [MNS<sup>+</sup>18], and beyond [CS97]. Specifically, the method uses a block Krylov subspace method to first compute highly accurate approximations to the *p* largest magnitude eigenvalues of **A**, for some  $p \leq l$ , along with an orthonormal matrix  $\mathbf{Z} \in \mathbb{R}^{n \times p}$  with columns approximating the corresponding eigenvectors. It uses moment matching to estimate the spectral density of **A** projected away from these approximate eigendirections  $(\mathbf{I} - \mathbf{Z}\mathbf{Z}^T)\mathbf{A}(\mathbf{I} - \mathbf{Z}\mathbf{Z}^T)$ , achieving error  $\epsilon \sigma_{l+1}(\mathbf{A})$  since this matrix has spectral norm bounded by  $O(\sigma_{p+1}(\mathbf{A})) = O(\sigma_{l+1}(\mathbf{A}))$ if **Z** is sufficiently accurate. It then modifies this approximate density to account for the probability

<sup>&</sup>lt;sup>2</sup>The notations  $\tilde{O}$  and  $\tilde{\Omega}$  means some logarithmic factors are present.

mass at the top p eigenvalues. While block Krylov methods are well understood for related tasks like eigenvalue and eigenvector computation [Par98, Tro18], low-rank approximation [HMT11, MM15], singular value approximation [MM15, MNS<sup>+</sup>18, BN23], linear system solving [LSY98, Saa03], our work requires a careful analysis of eigenvalue/eigenvector approximation with these methods that may be of independent interest. Overall, the above approach gives the following result:

**Theorem 1** (SDE with Explicit Deflation). For any  $\epsilon \in (0, 1)$ ,  $l \in [n]$ , and any constants  $c_1, c_2 > 0$ , Algorithm 1 of  $[BJM^+24]$  performs  $O\left(l \log n + \frac{b}{\epsilon}\right)$  matrix-vector products with **A** where  $b = \max\left(1, \frac{1}{n\epsilon^2}\log^2\frac{n}{\epsilon}\log^2\frac{1}{\epsilon}\right)$  and computes a probability density function  $\tilde{s}_{\mathbf{A}}$  such that, with probability at least  $1 - \frac{1}{n^{c_1}}$ ,  $\|\mathbf{A}\|_2$ 

$$W_1(s_{\mathbf{A}}, \tilde{s}_{\mathbf{A}}) \le \epsilon \cdot \sigma_{l+1}(\mathbf{A}) + \frac{\|\mathbf{A}\|_2}{n^{c_2}}.$$

The additive error  $\frac{\|\mathbf{A}\|_2}{n^{c_2}}$  can be thought of as negligible – comparable e.g., to round-off error when directly computing  $s_{\mathbf{A}}$  using a full eigendecomposition in finite precision arithmetic [BGVKS22].

We further show that our algorithm is optimal amongst all matrix-vector query algorithms, up to logarithmic factors and the negligible additive error term. Our proof leverages an existing lower bound for distinguishing Wishart matrices of different ranks, previously used to give matrix-vector query lower bounds for the closely related problem of eigenvalue estimation [SW23]. Formally:

**Theorem 2** (SDE Lower Bound). Any (possibly randomized) algorithm that given symmetric  $\mathbf{A} \in \mathbb{R}^{n \times n}$  outputs  $\tilde{s}_{\mathbf{A}}$  such that, with probability at least 1/2,  $W_1(s_{\mathbf{A}}, \tilde{s}_{\mathbf{A}}) \leq \epsilon \sigma_{l+1}(\mathbf{A})$  for  $\epsilon \in (0, 1)$  and  $l \in [n]$  must make  $\Omega\left(l + \frac{1}{\epsilon}\right)$  (possibly adaptively chosen) matrix-vector queries to  $\mathbf{A}$ .

### 2.2 Implicit Deflation Bounds for Stochastic Lanczos Quadrature

Our second contribution is to show that the popular Stochastic Lanczos Quadrature (SLQ) method for SDE [LSY16, CTU21] nearly matches the improved error bound of Theorem 1 for any choice of l, even though SLQ is 'parameter-free' and performs no explicit deflation step. This result helps to justify the strong practical performance of SLQ and the observed 'spectrum adaptive' nature of this method as compared to standard moment matching-based methods like KPM [CTU21].

A key idea used to achieve this bound is to view SLQ as an implicit moment matching method as in [CTU21, CTU22], and to analyze it similarly to KPM and other explicit moment matching methods. We combine this analysis approach with recent work on low-rank approximation with single-vector (i.e., non-block) Krylov methods [MMM24] to show that SLQ can perform 'implicit deflation' as part of moment matching to achieve the improved error bound. Formally, we have:

**Theorem 3** (SDE with SLQ). Let  $l \in [n]$ , and  $\epsilon, \delta \in (0, 1)$ . Let  $g_{\min} = \min_{i \in [l]} \frac{\sigma_i(\mathbf{A}) - \sigma_{i+1}(\mathbf{A})}{\sigma_i(\mathbf{A})}$  and  $\kappa = \frac{\|\mathbf{A}\|_2}{\sigma_{l+1}(\mathbf{A})}$ . SLQ run for  $m = O(l \log \frac{1}{g_{\min}} + \frac{1}{\epsilon} \log \frac{n \cdot \kappa}{\delta})$  iterations performs m matrix vector products with  $\mathbf{A}$  and outputs a probability density function  $\tilde{s}_{\mathbf{A}}$  such that, with probability at least  $1 - \delta$ ,

$$W_1(s_{\mathbf{A}}, \tilde{s}_{\mathbf{A}}) \leq \tilde{O}\left(\epsilon \cdot \sigma_{l+1}(\mathbf{A}) + \frac{\sigma_{l+1}(\mathbf{A})}{\sqrt{n}} + \frac{l}{n} \|\mathbf{A}\|_2\right).$$

Theorem 3 essentially matches our result for moment matching with explicit deflation (Theorem 1) up to some small caveats, discussed below. First, the number of matrix vector products has a logarithmic dependence on the minimum gap  $g_{\min}$  amongst the top l singular values as well as the condition number  $\kappa = \frac{\|\mathbf{A}\|_2}{\sigma_{l+1}(\mathbf{A})}$ . The dependence on the minimum gap is inherent, as non-block

Krylov methods like SLQ require a dependence on  $g_{\min}$  in order to perform deflation/low-rank approximation [MMM24]. We note that, in practice,  $g_{\min}$  is generally not too small. Also, by adding a random perturbation to **A** with spectral norm bounded by  $\frac{\|\mathbf{A}\|_2}{\operatorname{poly}(n)}$ , one can ensure that both  $g_{\min} \geq \frac{1}{\operatorname{poly}(n)}$  and  $\kappa \leq \operatorname{poly}(n)$  with high probability, and thus replace the  $O(l \log \frac{1}{g_{\min}})$  term with an  $O(l \log n)$  and the  $O(\frac{\log(n\kappa)}{\epsilon})$  term with  $O(\frac{\log n}{\epsilon})$ , matching Theorem 1. See e.g., [MMM24]. Second, Theorem 3 has an additional error term of size  $\tilde{O}(\sigma_{l+1}(\mathbf{A})/\sqrt{n})$ . This term is lower order whenever  $\epsilon = \tilde{\Omega}(1/\sqrt{n})$ . Further, we believe that this term, along with the dependence on  $g_{\min}$  can be removed by using a variant on SLQ that is popular in practice, where the densities output by multiple independent runs of the method are averaged together to produce  $\tilde{s}(\mathbf{A})$ .

Finally, Theorem 3 has an additional error term of size  $O(\|\mathbf{A}\|_2 \cdot l/n)$ . In the natural case when we run for  $m \ll n$  iterations and thus  $l \ll n$ , this term will be small. However, it cannot be avoided: even for a matrix with rank  $\leq l$  with well-separated eigenvalues, while the Lanczos method will converge to near-exact approximations to these eigenvalues (with error bounded by  $\frac{\|\mathbf{A}\|_2}{n^c}$ ), the probability distribution output by SLQ will not place mass exactly 1/n at these approximate eigenvalues and thus will not achieve SDE error  $O(\frac{\|\mathbf{A}\|_2}{n^c})$ .

This limitation motivates us to introduce a simple variant of SLQ, which we call variance reduced SLQ, which places mass exactly 1/n at any eigenvalue computed by Lanczos that has converged to sufficiently small error. This variant gives the following stronger error bound:

**Theorem 4** (SDE with Variance Reduced SLQ). Let  $l \in [n]$ , and  $\epsilon, \delta \in (0, 1)$ . Let  $g_{\min} = \min_{i \in [l]} \frac{\sigma_i(\mathbf{A}) - \sigma_{i+1}(\mathbf{A})}{\sigma_i(\mathbf{A})}$  and  $\kappa = \frac{\|\mathbf{A}\|_2}{\sigma_{l+1}(\mathbf{A})}$ . Algorithm 5 of  $[BJM^+ 24]$  run for  $m = O(l \log \frac{1}{g_{\min}} + \frac{1}{\epsilon} \log \frac{n \cdot \kappa}{\delta})$  iterations performs m matrix vector products with  $\mathbf{A}$  and outputs a probability density function  $\tilde{s}_{\mathbf{A}}$  such that, with probability at least  $1 - \delta$ , for some fixed constant c > 0,

$$W_1(s_{\mathbf{A}}, \tilde{s}_{\mathbf{A}}) \le \tilde{O}\left(\epsilon \cdot \sigma_{l+1}(\mathbf{A}) + \frac{\sigma_{l+1}(\mathbf{A})}{\sqrt{n}} + \frac{l}{n}\sigma_{l+1}(\mathbf{A})\right) + \frac{\|\mathbf{A}\|_2}{n^c}$$

## 3 Future Work

There are a number of directions inspired by our work which can be pursued in the future.

Lanczos based Matrix Function Approximation. Variants of SLQ and Lanczos have been used to obtain algorithms for estimating general functions of the trace of  $\mathbf{A}$ , tr(f(A)) [UCS17, CTU22, CH23]. The Lanczos method itself can approximate different matrix functions like rational functions very accurately [ACG<sup>+</sup>24]. Our deflation based analysis, particularly that of the variance reduced SLQ, could be used to give improved spectrum adaptive bounds for all these methods.

Numerical stability. The Lanczos algorithm is known to suffer from numerical stability issues when implemented in finite precision arithmetic [Che24]. A more careful analysis of how the algorithms perform under finite precision arithmetic will be interesting. However, we note that our experiments (Section 6 of  $[BJM^+24]$ ) show that our algorithms work pretty well in practice.

### References

[ACG<sup>+</sup>24] Noah Amsel, Tyler Chen, Anne Greenbaum, Cameron Musco, and Chris Musco. Nearly optimal approximation of matrix functions by the lanczos method. In NeuRIPS, 2024.

- [BEPW98] Kevin Burrage, Jocelyne Erhel, Bert Pohl, and Alan Williams. A deflation technique for linear systems of equations. *SIAM Journal on Scientific Computing*, 1998.
- [BGVKS22] 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*, 2022.
- [BJM<sup>+</sup>24] Rajarshi Bhattacharjee, Rajesh Jayaram, Cameron Musco, Christopher Musco, and Archan Ray. Improved spectral density estimation via explicit and implicit deflation. https://arxiv.org/abs/2410.21690, 2024.
- [BKM22] Vladimir Braverman, Aditya Krishnan, and Christopher Musco. Sublinear time spectral density estimation. In *Proceedings of the 54th Annual ACM Symposium on Theory* of Computing (STOC), 2022.
- [BN23] Ainesh Bakshi and Shyam Narayanan. Krylov methods are (nearly) optimal for lowrank approximation. In *Proceedings of the 64th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2023.
- [BRP92] C Benoit, E Royer, and G Poussigue. The spectral moments method. Journal of *Physics: Condensed Matter*, 1992.
- [CH23] Tyler Chen and Eric Hallman. Krylov-aware stochastic trace estimation. SIAM Journal on Matrix Analysis and Applications, 44(3):1218–1244, 2023.
- [Che23] Tyler Chen. A spectrum adaptive kernel polynomial method. *The Journal of Chemical Physics*, 2023.
- [Che24] Tyler Chen. The lanczos algorithm for matrix functions: a handbook for scientists, 2024.
- [CPB10] L Covaci, FM Peeters, and M Berciu. Efficient numerical approach to inhomogeneous superconductivity: The Chebyshev-Bogoliubov–de Gennes method. *Physical review letters*, 2010.
- [CS97] Andrew Chapman and Yousef Saad. Deflated and augmented Krylov subspace techniques. *Numerical linear algebra with applications*, 1997.
- [CTU21] Tyler Chen, Thomas Trogdon, and Shashanka Ubaru. Analysis of stochastic Lanczos quadrature for spectrum approximation. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.
- [CTU22] Tyler Chen, Thomas Trogdon, and Shashanka Ubaru. Randomized matrix-free quadrature for spectrum and spectral sum approximation. *arXiv:2204.01941*, 2022.
- [DBB19] Kun Dong, Austin R Benson, and David Bindel. Network density of states. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2019.
- [DNPS16] Edoardo Di Napoli, Eric Polizzi, and Yousef Saad. Efficient estimation of eigenvalue counts in an interval. *Numerical Linear Algebra with Applications*, 2016.

- [EG17] Nicole Eikmeier and David F Gleich. Revisiting power-law distributions in spectra of real world networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2017.
- [FDBV01] Illés J Farkas, Imre Derényi, Albert-László Barabási, and Tamas Vicsek. Spectra of "real-world" graphs: Beyond the semicircle law. *Physical Review E*, 2001.
- [FTU23] Zachary Frangella, Joel A Tropp, and Madeleine Udell. Randomized Nyström preconditioning. SIAM Journal on Matrix Analysis and Applications, 2023.
- [FV01] Jason Frank and Cornelis Vuik. On the construction of deflation-based preconditioners. SIAM Journal on Scientific Computing, 2001.
- [Gir87] Didier Girard. Un algorithme simple et rapide pour la validation croisee géenéralisée sur des problémes de grande taille. Technical report, 1987.
- [GKX19] Behrooz Ghorbani, Shankar Krishnan, and Ying Xiao. An investigation into neural net optimization via Hessian eigenvalue density. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019.
- [GOSS16] Alon Gonen, Francesco Orabona, and Shai Shalev-Shwartz. Solving ridge regression using sketched preconditioned SVRG. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 2016.
- [GSO17] Arjun Singh Gambhir, Andreas Stathopoulos, and Kostas Orginos. Deflation as a method of variance reduction for estimating the trace of a matrix inverse. *SIAM Journal on Scientific Computing*, 2017.
- [HHK72] R Haydock, V Heine, and M J Kelly. Electronic structure based on the local atomic environment for tight-binding bands. *Journal of Physics C: Solid State Physics*, 1972.
- [HMT11] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 2011.
- [Hut90] Michael F. Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 1990.
- [LG82] P Lambin and J-P. Gaspard. Continued-fraction technique for tight-binding systems. A generalized-moments method. *Physical Review B*, 1982.
- [Lin17] Lin Lin. Randomized estimation of spectral densities of large matrices made accurate. Numerische Mathematik, 2017.
- [LSY98] Richard B Lehoucq, Danny C Sorensen, and Chao Yang. ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods. SIAM, 1998.
- [LSY16] Lin Lin, Yousef Saad, and Chao Yang. Approximating spectral densities of large matrices. SIAM Review, 2016.

- [LXES19] Ruipeng Li, Yuanzhe Xi, Lucas Erlandson, and Yousef Saad. The eigenvalues slicing library (EVSL): Algorithms, implementation, and software. *SIAM Journal on Scientific Computing*, 2019.
- [MM15] Cameron Musco and Christopher Musco. Randomized block Krylov methods for stronger and faster approximate singular value decomposition. In Advances in Neural Information Processing Systems 28 (NIPS), 2015.
- [MM19] Michael Mahoney and Charles Martin. Traditional and heavy tailed self regularization in neural network models. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019.
- [MMM24] Raphael Meyer, Cameron Musco, and Christopher Musco. On the unreasonable effectiveness of single vector Krylov methods for low-rank approximation. In *Proceedings* of the 35th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2024.
- [MMMW21] Raphael A Meyer, Cameron Musco, Christopher Musco, and David P Woodruff. Hutch++: Optimal stochastic trace estimation. In Symposium on Simplicity in Algorithms (SOSA), 2021.
- [MNS<sup>+</sup>18] Cameron Musco, Praneeth Netrapalli, Aaron Sidford, Shashanka Ubaru, and David P. Woodruff. Spectrum approximation beyond fast matrix multiplication: Algorithms and hardness. In *Proceedings of the 9th Conference on Innovations in Theoretical Computer Science (ITCS)*, 2018.
- [Par98] Beresford N Parlett. The symmetric eigenvalue problem. SIAM, 1998.
- [PSG18] Jeffrey Pennington, Samuel Schoenholz, and Surya Ganguli. The emergence of spectral universality in deep networks. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2018.
- [RL18] Aditya Ramesh and Yann LeCun. Backpropagation for implicit spectral densities. *arXiv:1806.00499*, 2018.
- [Saa03] Yousef Saad. Iterative methods for sparse linear systems. SIAM, 2003.
- [Ski89] John Skilling. The eigenvalues of mega-dimensional matrices. Maximum Entropy and Bayesian Methods, 1989.
- [SR94] RN Silver and H Röder. Densities of states of mega-dimensional Hamiltonian matrices. International Journal of Modern Physics C, 1994.
- [SRS20] Jürgen Schnack, Johannes Richter, and Robin Steinigeweg. Accuracy of the finitetemperature Lanczos method compared to simple typicality-based estimates. *Physical Review Research*, 2020.
- [STBB17] Björn Sbierski, Maximilian Trescher, Emil J Bergholtz, and Piet W Brouwer. Disordered double Weyl node: Comparison of transport and density of states calculations. *Physical Review B*, 2017.
- [SW23] William Swartworth and David P Woodruff. Optimal eigenvalue approximation via sketching. In Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC), 2023.

- [Tro18] Joel A. Tropp. Analysis of randomized block Krylov methods. Technical report, California Institute of Technology, Pasadena, CA, 2018.
- [UCS17] Shashanka Ubaru, Jie Chen, and Yousef Saad. Fast estimation of tr(f(a)) via stochastic Lanczos quadrature. SIAM Journal on Matrix Analysis and Applications, 2017.
- [Wan94] Lin-Wang Wang. Calculating the density of states and optical-absorption spectra of large quantum systems by the plane-wave moments method. *Physical Review B*, 1994.
- [WWAF06] Alexander Weiße, Gerhard Wellein, Andreas Alvermann, and Holger Fehske. The kernel polynomial method. *Reviews of Modern Physics*, 2006.
- [GM09] Gene H Golub and Gérard Meurant. *Matrices, moments and quadrature with applications.* Princeton University Press, 2009.