Randomized matrix-free quadrature

Tyler Chen (joint with Tom Trogdon and Shashanka Ubaru)

https://chen.pw/slides.pdf

~~Randomized~~ matrix-free quadrature

Tyler Chen (joint with Tom Trogdon and Shashanka Ubaru)

https://chen.pw/slides.pdf

What is a matrix function?

An $n \times n$ symmetric matrix $\mathbf{A}$ has real eigenvalues and orthonormal eigenvectors:

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top.$$

The matrix function $f(\mathbf{A})$, induced by $f : \mathbb{R} \to \mathbb{R}$ and $\mathbf{A}$, is defined as

$$f(\mathbf{A}) := \sum_{i=1}^{n} f(\lambda_i) \mathbf{u}_i \mathbf{u}_i^\top.$$

Common functions are $1/x$, $\exp(-\beta x)$, $\sqrt{x}$, $\ln(x)$, etc.

Spectral sums and spectral measures

Spectral sums are integrals against a cumulative empirical spectral measure[1] (CSEM):

$$\operatorname{tr}(f(\mathbf{A})) = n \int f \, d\Phi, \qquad \Phi(x) = \sum_{i=1}^{n} n^{-1} \mathbb{1}[\lambda_i \leq x].$$

[1]also called density of states in physics

[2]Girard 1987; Hutchinson 1989; Skilling 1989, etc.

Spectral sums and spectral measures

Spectral sums are integrals against a cumulative empirical spectral measure[1] (CSEM):

$$\operatorname{tr}(f(\mathbf{A})) = n \int f \, \mathrm{d}\Phi, \qquad \Phi(x) = \sum_{i=1}^{n} n^{-1} \mathbb{1}[\lambda_i \leq x].$$

Quadratic forms of matrix functions are integrals against a weighted spectral measure (wCSEM):

$$\mathbf{v}^\mathsf{T} f(\mathbf{A}) \mathbf{v} = \int f \, \mathrm{d}\Psi, \qquad \Psi(x) = \sum_{i=1}^{n} |\mathbf{v}^\mathsf{T} \mathbf{u}_i|^2 \mathbb{1}[\lambda_i \leq x].$$

[1]also called density of states in physics

[2]Girard 1987; Hutchinson 1989; Skilling 1989, etc.

Spectral sums and spectral measures

Spectral sums are integrals against a cumulative empirical spectral measure[1] (CSEM):

$$\operatorname{tr}(f(\mathbf{A})) = n \int f \, d\Phi, \qquad \Phi(x) = \sum_{i=1}^{n} n^{-1} \mathbb{1}[\lambda_i \le x].$$

Quadratic forms of matrix functions are integrals against a weighted spectral measure (wCSEM):

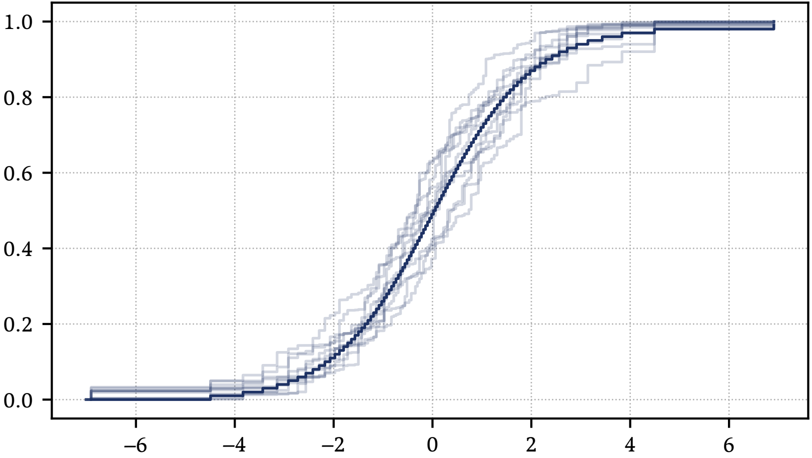
$$\mathbf{v}^\top f(\mathbf{A})\mathbf{v} = \int f \, d\Psi, \qquad \Psi(x) = \sum_{i=1}^{n} |\mathbf{v}^\top \mathbf{u}_i|^2 \mathbb{1}[\lambda_i \le x].$$

If $\mathbb{E}[\mathbf{v}\mathbf{v}^\top] = n^{-1}\mathbf{I}$, then $\Psi(x)$ is an unbiased estimator for $\Phi(x)$; see also quadratic trace estimation[2]: $\mathbb{E}[\mathbf{v}^\top \mathbf{B}\mathbf{v}] = n^{-1}\operatorname{tr}(\mathbf{B})$.

[1] also called density of states in physics

[2] Girard 1987; Hutchinson 1989; Skilling 1989, etc.

Example: CSEM vs wCESM



Legend: CESM Φ (——), samples of weighted CESM Ψ corresponding to random $\mathbf{v}$ (——).

A prototypical algorithm for randomized matrix free quadrature

Many standard algorithms approximate the CESM Φ in two stages:

1. approximate Φ by weighted CESM Ψ by sampling $\mathbf{v}$
2. approximate Ψ by a polynomial quadrature $[\Psi]_s^{\circ q}$

A prototypical algorithm for randomized matrix free quadrature

Many standard algorithms approximate the CESM Φ in two stages:

1. approximate Φ by weighted CESM Ψ by sampling $\mathbf{v}$
2. approximate Ψ by a polynomial quadrature $[\Psi]_s^{\circ q}$

We need to enforce that low-degree polynomials are integrated exactly. This can be done with knowledge of polynomial moments

$$m_i = \int p_i \mathrm{d}\Psi = \mathbf{v}^\mathsf{T} p_i(\mathbf{A})\mathbf{v}.$$

A prototypical algorithm for randomized matrix free quadrature

Many standard algorithms approximate the CESM Φ in two stages:

1. approximate Φ by weighted CESM Ψ by sampling $\mathbf{v}$
2. approximate Ψ by a polynomial quadrature $[\Psi]_s^{\circ q}$

We need to enforce that low-degree polynomials are integrated exactly. This can be done with knowledge of polynomial moments

$$m_i = \int p_i \mathrm{d}\Psi = \mathbf{v}^\top p_i(\mathbf{A})\mathbf{v}.$$

Moments $m_0, m_1, \ldots, m_{2k}$ can be computed from the Krylov subspace

$$\mathcal{K}_k(\mathbf{A}, \mathbf{v}) := \operatorname{span}\{\mathbf{v}, \mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^k\mathbf{v}\}.$$

Polynomial quadrature

Fix a reference measure μ.

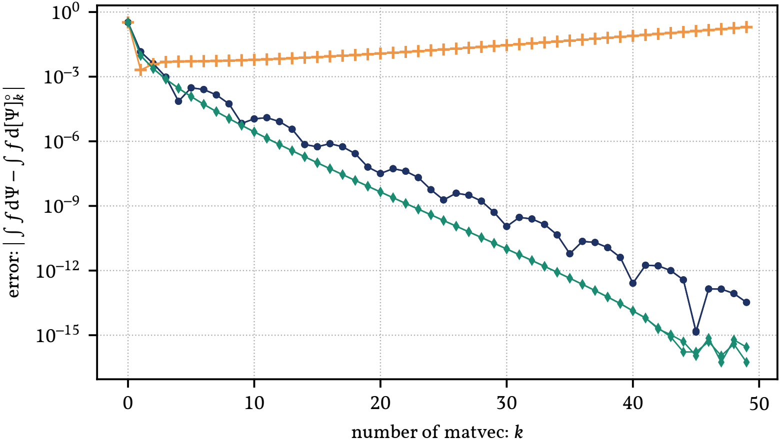
Examples of choices of $[f]_s^{\circ p}$:

- truncated μ-orthogonal polynomial series of f
  - **Kernel polynomial method**[3]: μ fixed (e.g. arcsin), possibly apply damping kernel
- polynomial interpolate at roots of an orthogonal polynomial of μ
  - **Stochastic Lanczos quadrature**[4]: $\mu = \Psi$ (Gaussian quadrature)

KPM and SLQ are probably the most widely used[5] algorithms for spectrum and spectral sum approximation.

[3] Skilling 1989; Silver and Röder 1994; Weiße, Wellein, Alvermann, and Fehske 2006.
[4] Bai, Fahey, and Golub 1996; Golub and Meurant 2009.
[5] Weiße, Wellein, Alvermann, and Fehske 2006; Lin, Saad, and Yang 2016; Ubaru, Chen, and Saad 2017; Martinsson and Tropp 2020; Murray et al. 2023.

Choosing the reference measure/approximation



Legend: KPM with correct support ( ), 5% too large ( ), 5% too small ( ).

Computing moments

Let p_i be the orthogonal polynomials of μ with three-term recurrence:

$$xp_i(x) = \beta_{i-1}p_{i-1}(x) + \alpha_i p_i(x) + \beta_i p_{i+1}(x).$$

[6] Skilling 1989; Weiße, Wellein, Alvermann, and Fehske 2006.

Computing moments

Let p_i be the orthogonal polynomials of μ with three-term recurrence:

$$xp_i(x) = \beta_{i-1}p_{i-1}(x) + \alpha_i p_i(x) + \beta_i p_{i+1}(x).$$

We can run a matrix version of the recurrence to compute $p_i(\mathbf{A})\mathbf{v}$. Then, to get moments:

- Compute $m_i = \mathbf{v}^\top p_i(\mathbf{A})\mathbf{v}$ as you go.
  - This works fine, but we only get degree k not $2k$.
- Instead store basis $\mathbf{B} = [p_0(\mathbf{A})\mathbf{v}, \ldots, p_k(\mathbf{A})\mathbf{v}]$ and compute $\mathbf{B}^\top\mathbf{B}$.
  - This gets degree $2k$, but requires high memory.

For Chebyshev polynomials, can get both from[6]:

$$T_{2i}(x) = 2T_i(x)^2 - 1, \qquad T_{2i+1}(x) = 2T_i(x)T_{i+1}(x) - x.$$

[6] Skilling 1989; Weiße, Wellein, Alvermann, and Fehske 2006.

Connection coefficients for more modified moments

The connection coefficient matrix $\mathbf{C} = \mathbf{C}_{\mu\to\nu}$ is the upper triangular matrix representing a change of basis between the orthogonal polynomials $\{p_i\}_{i=1}^{\infty}$ with respect to μ and the orthogonal polynomials $\{q_i\}_{i=1}^{\infty}$ with respect to ν, whose entries satisfy,

$$p_s(x) = [\mathbf{C}]_{0,s} q_0(x) + [\mathbf{C}]_{1,s} q_1(x) + \cdots + [\mathbf{C}]_{s,s} q_s(x).$$

- Connection coefficient matrix can be computed recursively[7] from recurrence formulas for orthogonal polynomials of μ and ν.
- If we have moments with respect to ν, we can get moments with respect to μ.

[7] Sack and Donovan 1971; Wheeler 1974; Webb and Olver 2021.

The Lanczos algorithm

The Lanczos algorithm (efficiently) computes an orthonormal basis for the Krylov subspace $\mathcal{K}_k(\mathbf{A}, \mathbf{v})$.

Equivalently, Lanczos computes the orthogonal polynomials of Ψ! Resulting Gaussian quadrature integrates polynomials of degree $2k - 1$ exactly.

This can be done efficiently with a three term recurrence:

$$\mathbf{A}\mathbf{q}_i = \beta_{i-1}\mathbf{q}_{i-1} + \alpha_i\mathbf{q}_i + \beta_i\mathbf{q}_{i+1}.$$

Compared with explicit polynomials: we already know the modified moments, but need to compute the recurrence coefficients.

Example: instability of Lanczos

In finite precision arithmetic, the Lanczos algorithm behaves extremely differently than in exact arithmetic.

Toy example[8]:

$$\mathbf{A} = \begin{bmatrix} 0 & & & & & \\ & 0.00025 & & & & \\ & & 0.0005 & & & \\ & & & 0.00075 & & \\ & & & & 0.001 & \\ & & & & & 10 \end{bmatrix}, \qquad \mathbf{v} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

[8] Parlet and Scott 1979.

Example: instability of Lanczos

Denote by $\mathbf{T}, \mathbf{Q}$ the finite precision arithmetic output of Lanczos and $\tilde{\mathbf{T}}, \tilde{\mathbf{Q}}$ the "exact" arithmetic output. How many digits of accuracy do we have for the following quantities:

Example: instability of Lanczos

Denote by $\mathbf{T}, \mathbf{Q}$ the finite precision arithmetic output of Lanczos and $\tilde{\mathbf{T}}, \tilde{\mathbf{Q}}$ the "exact" arithmetic output. How many digits of accuracy do we have for the following quantities:

$$\tilde{\mathbf{Q}} - \mathbf{Q} \qquad \tilde{\mathbf{T}} - \mathbf{T} \qquad \mathbf{Q}^\mathsf{T}\mathbf{Q} - \mathbf{I}$$

$$\begin{bmatrix} - & - & 12 & 7 & 1 \\ - & - & 12 & 7 & 0 \\ - & 17 & 13 & 11 & 0 \\ - & - & 12 & 7 & 0 \\ - & - & 12 & 7 & 1 \\ - & 17 & 8 & 3 & 0 \end{bmatrix} \quad \begin{bmatrix} - & - & & & & \\ - & - & - & & & \\ & - & - & 19 & & \\ & & 19 & 14 & 10 & \\ & & & 10 & 5 & 2 \\ & & & & 2 & 0 \end{bmatrix} \quad \begin{bmatrix} 16 & 16 & 17 & 8 & 4 & 0 \\ 16 & 16 & 12 & 8 & 3 & 0 \\ 17 & 12 & 16 & 15 & 7 & 4 \\ 8 & 8 & 15 & 15 & 15 & 9 \\ 4 & 3 & 7 & 15 & - & 17 \\ 0 & 0 & 4 & 9 & 17 & - \end{bmatrix}$$

Stability of matrix-free quadrature

Practitioners (and theorists) are wary of using Lanczos-based methods ($\mu = \Psi$), at least without reorthogonalization[9] (expensive)!

Instead, they prefer methods based on explicit polynomails (μ fixed) such as the Chevyshev polynomails.

[9]Jaklič and Prelovšek 1994; Aichhorn, Daghofer, Evertz, and Linden 2003; Weiße, Wellein, Alvermann, and Fehske 2006; Ubaru, Chen, and Saad 2017; Granziol, Wan, and Garipov 2019.

Practitioners (and theorists) are wary of using Lanczos-based methods ($\mu = \Psi$), at least without reorthogonalization[9] (expensive)!

Instead, they prefer methods based on explicit polynomails (μ fixed) such as the Chevyshev polynomails.

However...

[9] Jaklič and Prelovšek 1994; Aichhorn, Daghofer, Evertz, and Linden 2003; Weiße, Wellein, Alvermann, and Fehske 2006; Ubaru, Chen, and Saad 2017; Granziol, Wan, and Garipov 2019.

Stability of matrix-free quadrature

Practitioners (and theorists) are wary of using Lanczos-based methods ($\mu = \Psi$), at least without reorthogonalization[9] (expensive)!

Instead, they prefer methods based on explicit polynomails (μ fixed) such as the Chevyshev polynomails.

However...

- Explicit methods are not adaptive to the spectrum
- Explicit methods are exponentialy unstable unless certain hyperparemeters are selected properly

[9] Jaklič and Prelovšek 1994; Aichhorn, Daghofer, Evertz, and Linden 2003; Weiße, Wellein, Alvermann, and Fehske 2006; Ubaru, Chen, and Saad 2017; Granziol, Wan, and Garipov 2019.

Lanczos in finite precision arithmetic

A lot is known: Perturbed Lanczos recurrence[10], CG/Backwards stability[11], Matrix functions[12].

[10] Paige 1970; Paige 1972; Paige 1976; Paige 1980.
[11] Greenbaum 1989.
[12] Druskin and Knizhnerman 1992; Knizhnerman 1996; Musco, Musco, and Sidford 2018.
[13] Unfortunately this paper is hard to find, so we included similar proofs in Chen and Trogdon 2023.
[14] technically, it just shows the Chebyshev moments can still be obtained accurately

Lanczos in finite precision arithmetic

A lot is known: Perturbed Lanczos recurrence[10], CG/Backwards stability[11], Matrix functions[12].

Knizhnerman 1996[13] shows that finite precision Lanczos approximates Chebyshev moments accurately:

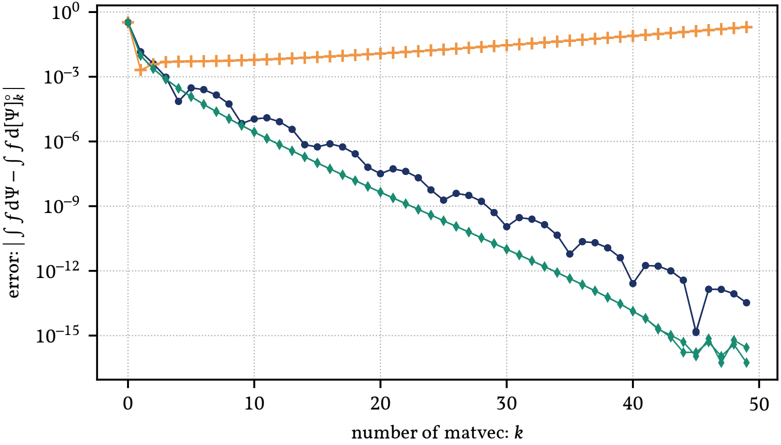
$$\| \underbrace{\mathbf{v}^\top T_i(\mathbf{A})\mathbf{v}}_{\text{true moment}} - \underbrace{\mathbf{e}_1^\top T_i(\mathbf{T})\mathbf{e}_1}_{\text{Lanczos version}} \| \leq \epsilon_{\text{mach}} \cdot \text{poly}(k).$$

Proofs straightforward given Paige 1976 and Paige 1980.

Knizhnerman 1996 implies[14] that KPM can be implemented stably using Lanczos.

[10] Paige 1970; Paige 1972; Paige 1976; Paige 1980.
[11] Greenbaum 1989.
[12] Druskin and Knizhnerman 1992; Knizhnerman 1996; Musco, Musco, and Sidford 2018.
[13] Unfortunately this paper is hard to find, so we included similar proofs in Chen and Trogdon 2023.
[14] technically, it just shows the Chebyshev moments can still be obtained accurately

Choosing the reference measure/approximation revisited



Legend: KPM with correct support ( ), 5% too large ( ), 5% too small ( ).

The big picture

The ideas we described here are old[15]

[15]Gautschi 1970; Sack and Donovan 1971; Wheeler 1974; Golub and Meurant 1994; Gautschi 2006; Golub and Meurant 2009.

The big picture

The ideas we described here are old[15], so what's the point?

More interaction with application domains is needed.

- Practitioners have lots of good algorithms (that we'll re-discover in 10 years)
- We have the tools to improve their algorihms

[15]Gautschi 1970; Sack and Donovan 1971; Wheeler 1974; Golub and Meurant 1994; Gautschi 2006; Golub and Meurant 2009.

The big picture

The ideas we described here are old[15], so what's the point?

More interaction with application domains is needed.

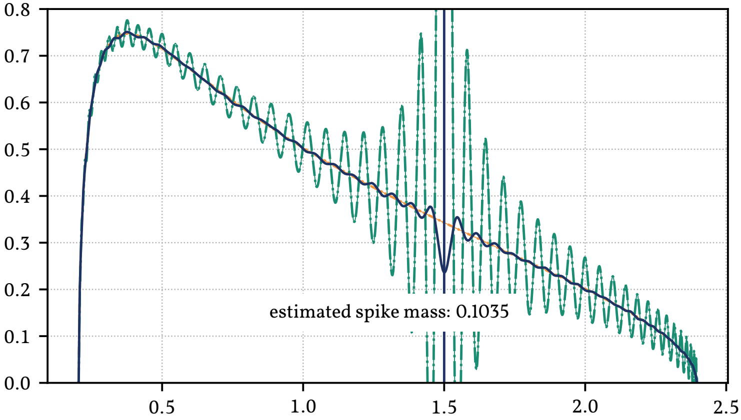
- Practitioners have lots of good algorithms (that we'll re-discover in 10 years)
- We have the tools to improve their algorihms

This talk:

- We can cheaply try out lots of different quadrature rules (decouple computation from approximation) once we've run Lanczos.
  - This allows variants of KPM which are spectrum adaptive
  - We do not need to know hyperparemeters ahead of time!
  - This avoids potential instabilities of KPM with bad parameter choices
- Better explanation of stability of Lanczos-based methods

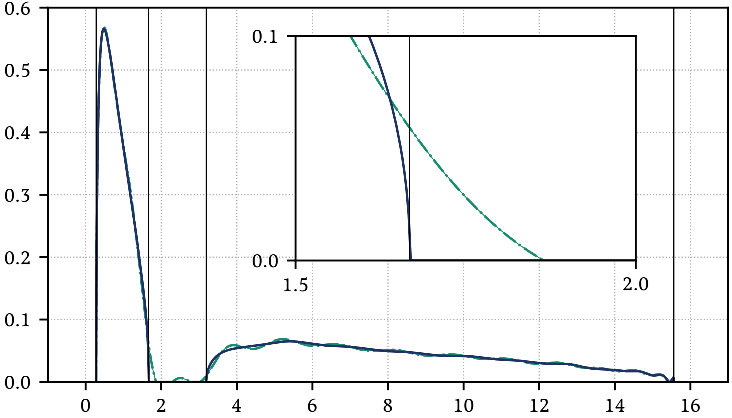
[15] Gautschi 1970; Sack and Donovan 1971; Wheeler 1974; Golub and Meurant 1994; Gautschi 2006; Golub and Meurant 2009.

Example: smooth spectrum with spike



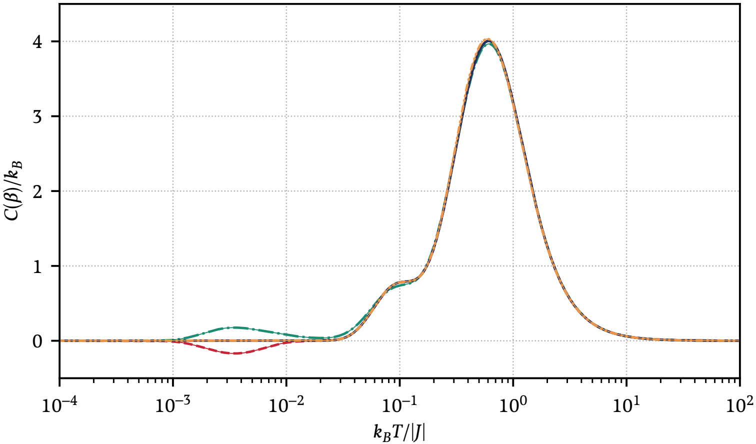
Legend: limiting density ( ), kernel polynomial method: $\mu = (1-p)\mu^U_{a,b} + p\,\delta(x-z)$ ( ), kernel polynomial method: $\mu = \mu^U_{a,b}$ ( ).

Example: spectrum with disjoint support



Legend: kernel polynomial method: $\mu = \mu^U_{a_1,b_2}$ ( ), damped kernel polynomial method: $\mu = \frac{1}{2}\mu^U_{a_1,b_1} + \frac{1}{2}\mu^U_{a_2,b_2}$ ( ).

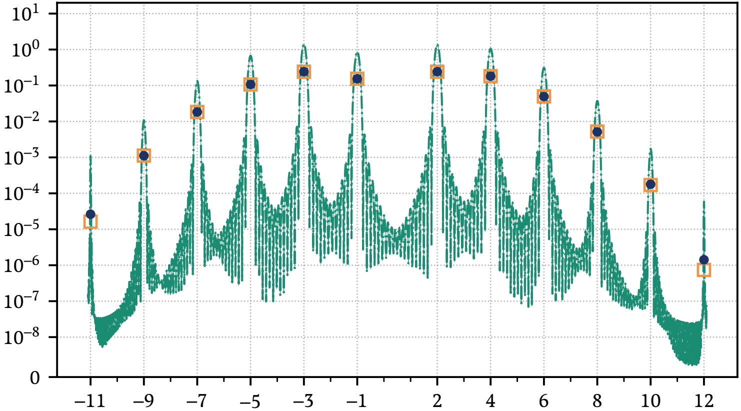
Example: heat capacity of quantum spin system[16]



Legend: exact diagonalization ( ), stochastic Lanczos quadrature ( ), kernel polynomial method ( ), and damped kernel polynomial method ( ).

[16]Schlüter, Gayk, Schmidt, Honecker, and Schnack 2021.

Example: a sparse spectrum



Legend: true spectrum (□), stochastic Lanczos quadrature $k = 12$ (•), kernel polynomial method $k = 250$ (—·—)

References I

Aichhorn, Markus et al. (Apr. 2003). “Low-temperature Lanczos method for strongly correlated systems”. In: *Physical Review B* 67.16.

Bai, Zhaojun, Gark Fahey, and Gene Golub (Nov. 1996). “Some large-scale matrix computation problems”. In: *Journal of Computational and Applied Mathematics* 74.1-2, pp. 71–89.

Chen, Tyler and Thomas Trogdon (2023). *Stability of the Lanczos algorithm on matrices with regular spectral distributions*.

Druskin, Vladimir and Leonid Knizhnerman (July 1992). “Error Bounds in the Simple Lanczos Procedure for Computing Functions of Symmetric Matrices and Eigenvalues”. In: *Comput. Math. Math. Phys.* 31.7, pp. 20–30.

Gautschi, Walter (Apr. 1970). “On the Construction of Gaussian Quadrature Rules from Modified Moments”. In: *Mathematics of Computation* 24.110, p. 245.

— (2006). “Orthogonal Polynomials, Quadrature, and Approximation: Computational Methods and Software (in Matlab)”. In: *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, pp. 1–77.

Girard, Didier (1987). *Un algorithme simple et rapide pour la validation croisée généralisée sur des problèmes de grande taille*.

Golub, Gene H and Gérard Meurant (2009). *Matrices, moments and quadrature with applications*. Vol. 30. Princeton University Press.

— (1994). “Matrices, moments and quadrature”. In.

Granziol, Diego, Xingchen Wan, and Timur Garipov (2019). *Deep Curvature Suite*.

Greenbaum, Anne (1989). “Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences”. In: *Linear Algebra and its Applications* 113, pp. 7–63.

Hutchinson, M.F. (Jan. 1989). “A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines”. In: *Communications in Statistics - Simulation and Computation* 18.3, pp. 1059–1076.

Jaklič, J. and P. Prelovšek (Feb. 1994). “Lanczos method for the calculation of finite-temperature quantities in correlated systems”. In: *Physical Review B* 49.7, pp. 5065–5068.

Knizhnerman, L. A. (Jan. 1996). “The Simple Lanczos Procedure: Estimates of the Error of the Gauss Quadrature Formula and Their Applications”. In: *Comput. Math. Math. Phys.* 36.11, pp. 1481–1492.

References II

Lin, Lin, Yousef Saad, and Chao Yang (Jan. 2016). “Approximating Spectral Densities of Large Matrices”. In: *SIAM Review* 58.1, pp. 34–65.

Martinsson, Per-Gunnar and Joel A. Tropp (May 2020). “Randomized numerical linear algebra: Foundations and algorithms”. In: *Acta Numerica* 29, pp. 403–572.

Murray, Riley et al. (2023). *Randomized Numerical Linear Algebra : A Perspective on the Field With an Eye to Software*.

Musco, Cameron, Christopher Musco, and Aaron Sidford (2018). “Stability of the Lanczos Method for Matrix Function Approximation”. In: *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA ’18. New Orleans, Louisiana: Society for Industrial and Applied Mathematics, pp. 1605–1624.

Paige, C. C. (June 1970). “Practical use of the symmetric Lanczos process with re-orthogonalization”. In: *BIT* 10.2, pp. 183–195.

— (1972). “Computational Variants of the Lanczos Method for the Eigenproblem”. In: *IMA Journal of Applied Mathematics* 10.3, pp. 373–381.

Paige, Christopher Conway (Dec. 1976). “Error Analysis of the Lanczos Algorithm for Tridiagonalizing a Symmetric Matrix”. In: *IMA Journal of Applied Mathematics* 18.3, pp. 341–349.

— (1980). “Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem”. In: *Linear Algebra and its Applications* 34, pp. 235–258.

Parlet, Beresford Neill and David St. Clair Scott (Jan. 1979). “The Lanczos algorithm with selective orthogonalization”. In: *Mathematics of Computation* 33.145, pp. 217–238.

Sack, R. A. and A. F. Donovan (Oct. 1971). “An algorithm for Gaussian quadrature given modified moments”. In: *Numerische Mathematik* 18.5, pp. 465–478.

Schlüter, Henrik et al. (June 2021). “Accuracy of the typicality approach using Chebyshev polynomials”. In: *Zeitschrift für Naturforschung A* 76.9, pp. 823–834.

Silver, R.N. and H. Röder (Aug. 1994). “Densities of states of mega-dimensional Hamiltonian matrices”. In: *International Journal of Modern Physics C* 05.04, pp. 735–753.

References III

Skilling, John (1989). "The Eigenvalues of Mega-dimensional Matrices". In: *Maximum Entropy and Bayesian Methods*. Springer Netherlands, pp. 455–466.

Ubaru, Shashanka, Jie Chen, and Yousef Saad (2017). "Fast Estimation of $tr(f(A))$ via Stochastic Lanczos Quadrature". In: *SIAM Journal on Matrix Analysis and Applications* 38.4, pp. 1075–1099.

Webb, Marcus and Sheehan Olver (Feb. 2021). "Spectra of Jacobi Operators via Connection Coefficient Matrices". In: *Communications in Mathematical Physics* 382.2, pp. 657–707.

Weiße, Alexander et al. (Mar. 2006). "The kernel polynomial method". In: *Reviews of Modern Physics* 78.1, pp. 275–306.

Wheeler, John C. (June 1974). "Modified moments and Gaussian quadratures". In: *Rocky Mountain Journal of Mathematics* 4.2.