A spectrum adaptive kernel polynomial method

View Online   Export Citation   CrossMark

Tyler Chen[a] (ID)

AFFILIATIONS

Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, New York 10012, USA and Department of Computer Science and Engineering, Tandon School of Engineering, New York University, 370 Jay Street, New York, New York 11201, USA

[a]Author to whom correspondence should be addressed: tyler.chen@nyu.edu. **URL:** https://research.chen.pw

ABSTRACT

The kernel polynomial method (KPM) is a powerful numerical method for approximating spectral densities. Typical implementations of the KPM require an *a prior* estimate for an interval containing the support of the target spectral density, and while such estimates can be obtained by classical techniques, this incurs addition computational costs. We propose a spectrum adaptive KPM based on the Lanczos algorithm without reorthogonalization, which allows the selection of KPM parameters to be deferred to after the expensive computation is finished. Theoretical results from numerical analysis are given to justify the suitability of the Lanczos algorithm for our approach, even in finite precision arithmetic. While conceptually simple, the paradigm of decoupling computation from approximation has a number of practical and pedagogical benefits, which we highlight with numerical examples.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0166678

I. INTRODUCTION

Over the past several decades, a number of moment-based schemes have been developed to approximate spectral densities of matrices.[1–5] These methods typically access the matrix of interest through matrix–vector products and are, therefore, especially well suited for applications where matrix–vector products can be performed quickly or in which the matrix of interest is so large that exact diagonalization techniques are infeasible. Perhaps, the most prominent of these methods is the Kernel Polynomial Method (KPM),[1,4,6–8] which has found widespread use in quantum physics/chemistry[1,4,6–13] and beyond.[14,15]

A standard implementation of the KPM[8] produces an approximate spectral density via a truncated Chebyshev polynomial expansion obtained using the low-degree Chebyshev moments of the target spectral density. These moments are computed using a Chebyshev recurrence shifted and scaled to an interval of approximation containing the support of the target spectral density. If this interval does not contain the support of the target spectral density, the KPM approximation is unlikely to converge, but if the interval is too large, then the KPM approximation will lose resolution. All implementations of the KPM that we are aware of require this interval of approximation to be determined ahead of time. Thus, as a preprocessing step, it is typical to run another algorithm to determine bounds for the support of the target spectral density.

The focus of this paper is an implementation of the KPM, which decouples computation from the choice of approximation method. For instance, our implementation allows many different intervals of approximation to be tested out at essentially no cost once the main computation is completed. In fact, and more importantly, approximations corresponding to different families of polynomials can also be efficiently obtained. The choice of polynomial family can significantly impact the qualitative properties of the resulting KPM approximation, but the use of non-standard orthogonal polynomial families has been limited in practice thus far, arguably due to the previous lack of a simple implementation.

Our spectrum adaptive KPM is based on the Lanczos algorithm *without reorthogonalization*. It is well-known that the Lanczos algorithm is unstable, and this has led to a general hesitance to use Lanczos-based methods for approximate spectral densities unless reorthogonalization is used.[2,7,8,16–18] Amazingly, this instability does not limit the usefulness of the Lanczos algorithm for many tasks. We use theoretical results from the numerical analysis literature to justify the validity of our approach in finite precision arithmetic. We believe that this commentary will be of general value to the computational physics/chemistry communities. Complimentary numerical experiments provide empirical evidence of the stability of the proposed algorithm.

Finally, we note that there are a number of related Lanczos-based algorithms, such as the Finite Temperature Lanczos Method

(FTLM);[2] see also stochastic Lanczos quadrature.[3,17] These methods are also widely used in practice and are viewed by some as preferable to KPM based methods in many settings.[19–21] We do not advocate the use of the KPM over these methods nor the use of any of these methods over KPM. Rather, our aim is to provide a new tool that allows practitioners to test out all of these algorithms for essentially free. For instance, users no longer need to make an *a priori* decision to use FTLM or KPM; they can simply output both approximations and then decide which to use later.

II. THE KERNEL POLYNOMIAL METHOD

We present the KPM from the perspective of orthogonal polynomials. Throughout, $H = \sum_{n=1}^{d} E_n |u_i\rangle\langle u_i|$ will be a Hermitian Hamiltonian of finite dimension $d < \infty$ with the corresponding Density of States (DOSs)

$$\rho(E) = \frac{1}{d}\sum_{n=0}^{d-1} \delta(E - E_n), \tag{1}$$

where $\delta(E)$ is a Dirac delta mass centered at zero. Given a state $|r\rangle$, the Local Density of State (LDOS)

$$\hat{\rho}(E) = \sum_{k=1}^{n-1} \langle r|u_n\rangle \delta(E - E_n) \tag{2}$$

is also of interest in many settings. For example, if $|r\rangle$ is a random state drawn from the uniform distribution on the unit hypersphere, then $\hat{\rho}(E)$ is an unbiased estimator for $\rho(E)$. In fact, *quantum typicality*[5,22] ensures that $\hat{\rho}(E)$ concentrates around $\rho(E)$. It therefore often suffices to use $\hat{\rho}(E)$ as a proxy for $\rho(E)$, but in the case that the variance of a single sample is too large, one can sample multiple random states independently and then average the corresponding LDOSs to reduce the variance.[1,2,8,23,24] In numerical analysis, this is called stochastic trace estimation and has been analyzed in detail.[25–30]

The aim of the KPM is to produce a "density" $\rho_{KPM}(E)$ approximating the LDOS[31] $\hat{\rho}(E)$. Toward this end, let $\sigma(E)$ be a fixed *reference density* and expand $\hat{\rho}(E)/\sigma(E)$ as a formal polynomial series,

$$\frac{\hat{\rho}(E)}{\sigma(E)} = \sum_{n=0}^{\infty} \mu_n p_n(E), \tag{3}$$

where $\{p_n\}$ are the orthonormal polynomials with respect to $\sigma(E)$. Using the orthonormality of $\{p_n\}$ with respect to $\sigma(E)$, we can compute μ_n by

$$\mu_n = \int \sigma(E)\frac{\hat{\rho}(E)}{\sigma(E)}p_n(E)\mathrm{d}E \tag{4}$$

$$= \int \hat{\rho}(E)p_n(E)\mathrm{d}E \tag{5}$$

$$= \langle r|p_n(H)|r\rangle. \tag{6}$$

Thus, we see that $\{\mu_n\}$ are the so-called (modified) moments of $\hat{\rho}(E)$ with respect to the orthogonal polynomials of $\sigma(E)$ and can be computed without explicit knowledge of $\hat{\rho}(E)$ using the expression in (6).

Computing the first s moments naturally gives an approximation $\rho_{KPM}(E)$ to $\hat{\rho}(E)$ defined by

$$\rho_{KPM}(E) = \sigma(E)\sum_{n=0}^{s} \mu_n p_n(E). \tag{7}$$

When the degree of the approximation $s \to \infty$,

$$\rho_{KPM}(E) \to \sigma(E)\sum_{n=0}^{\infty} \mu_n p_n(E) = \hat{\rho}(E), \tag{8}$$

and convergence is expected to be at a rate $O(s^{-1})$ in the Wasserstein distance.[8,21,32] A rigorous theoretical understanding of other types of convergence is important, particularly if a density approximation is desired. However, this is not entirely straightforward as $\hat{\rho}(E)$ is a linear combination of Dirac deltas and therefore is not even a density itself.

A. Damping

Strictly speaking, $\rho_{KPM}(E)$ need not be a proper density as it may be negative for some values of E. This effect is particularly noticeable if $\hat{\rho}(E)/\sigma(E)$ is very spiky so that polynomial approximations have large Gibbs oscillations. To ensure positivity, it is often suggested to use the so-called *damping kernels* that effectively results in an approximation,

$$\rho_{KPM}(E) = \sigma(E)\sum_{n=0}^{s} g_n\mu_n p_n(E), \tag{9}$$

where the damping coefficients $\{g_n\}$ are carefully chosen. The most common choice of coefficients corresponds to the so-called Jackson's damping kernel; see Ref. [8] for a detailed discussion on damping. It is also possible to simply apply a standard convolution against the resulting approximation, although this does not necessarily ensure positivity.

B. Evaluating orthogonal polynomials

Assuming that $\sigma(E)$ is a unit-mass (positive) density, the orthogonal polynomials satisfy a symmetric three-term recurrence,

$$p_{n+1}(E) = \frac{1}{\delta_n}\left(Ep_n(E) - \gamma_n p_n(E) - \delta_{n-1}p_{n-1}(E)\right) \tag{10}$$

with initial conditions $p_1(E) = (1/\delta_0)(Ep_0(E) - \gamma_0 p_0(E))$, $p_0(E) = 1$, for some set of recurrence coefficients $\{\gamma_n, \delta_n\}$ depending on $\sigma(E)$. Throughout, we will assume that these coefficients are known (or can be computed). Then, once the moments are known, the KPM approximation $\rho_{KPM}(E)$ defined in (7) can be obtained by evaluating the polynomials by this recurrence and then forming a linear combination of these polynomials.

C. Choice of reference density

The most common choice of reference density is

$$\sigma_{a,b}^{T}(E) = \frac{1}{\pi}\frac{1}{\sqrt{(b - E)(E - a)}}, \tag{11}$$

which is the orthogonality weight for the Chebyshev polynomials shifted and scaled from $[-1, 1]$ to $[a, b]$. For this choice of reference
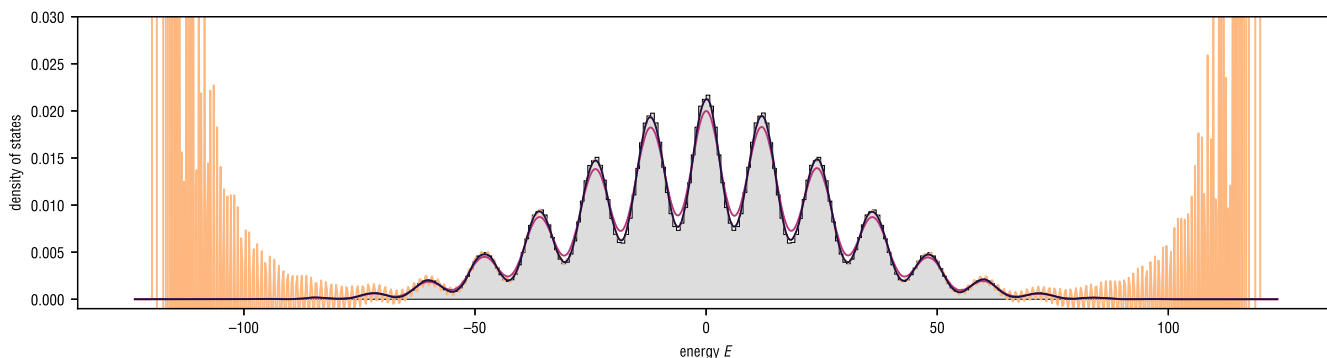
FIG. 1. Illustration showing the impact of the support $[a, b]$ for the standard Chebyshev KPM with $s = 500$ moments and Jackson's damping kernel. Here, $[a, b] = [E_{\min} - \eta, E_{\max} + \eta]$ for varying choices of η. *Legend*: $\eta = 0$ (——), $\eta = -0.00007(E_{\max} - E_{\min})$ (——), $\eta = 0.5(E_{\max} - E_{\min})$ (——), histogram of true eigenvalues (▭). *Takeaway*: Observe that even a slight underestimate of $[E_{\min}, E_{\max}]$ results in a loss of convergence, while an overestimate of $[E_{\min}, E_{\max}]$ results in a loss of resolution. Our spectrum adaptive KPM allows the reference density $\sigma(E)$ to be chosen after computation. In fact, many different $\sigma(E)$ can be efficiently obtained and compared.

density, there is an elegant and widely used algorithm for computing the moments, which we summarize in Sec. II D. In Fig. 1, we illustrate some of the impacts of the choice of $[a, b]$ on the KPM approximation when this reference density is used. In particular, it is very important that $[a, b]$ contains the spectrum of $\boldsymbol{H}$. Thus, it is often suggested to take $[a, b] = [E_{\min} - \eta, E_{\max} + \eta]$, where $\eta > 0$, to avoid the risk of $[a, b]$ not containing the entire spectrum.[8] However, using nonzero values of η reduces the resolution of the approach.[33]

The choice of reference density $\sigma(E)$ impacts the qualitative features of the KPM approximation $\rho_{\text{KPM}}(E)$, and in principle, any choice of unit-mass density with finite moments is possible.[6,8] First, observe that the KPM approximation is exact, if $\hat\rho(E)/\sigma(E)$ is a polynomial of degree at most s. Thus, one perspective is that $\rho(E)$ should be chosen to try and make $\hat\rho(E)/\sigma(E)$ as easy to approximate with polynomials as possible. In particular, the support of $\hat\rho(E)$ must contain the support of $\sigma(E)$. The difficulty of such an approach is that many properties of the spectrum of $\hat\rho(E)$ are not known ahead of time. Even the most basic properties such as an interval $[a, b]$ containing the support of $\hat\rho(E)$ are often unknown *a priori* and must be approximated numerically as a pre-processing step. The approach we describe in Sec. III addresses these difficulties by allowing $\sigma(E)$ to be chosen after the computation has completed.

In our experiments, we will make use of reference densities of the form

$$\sigma(E) = \sum_i w_i \sigma_{a_i,b_i}^T(E). \tag{12}$$

The recurrence coefficients for the orthogonal polynomials of distributions such as (12) are easily computed since integrals of polynomials against each term can be computed exactly using quadrature rules.[34,35]

D. Computing Chebyshev moments

In the case that $\sigma(E) = \sigma_{a,b}^T(E)$ is the Chebyshev density (11), the modified moments can be computed efficiently using properties of Chebyshev recurrences. Such an approach is described in detail in the literature,[1,6,8] but as this is by far the most common approach

to implementing the KPM, we provide a brief overview to put the contributions of this paper into context.

Recall that the Chebyshev polynomials are defined by the recurrence

$$T_{n+1}(E) = 2xT_n(E) - T_{n-1}(E) \tag{13}$$

with initial conditions $T_1(E) = E$ and $T_0(E) = 1$. These polynomials are orthogonal with respect to the weight $1/\sqrt{1 - E^2}$. The Chebyshev polynomials also satisfy the useful identities

$$T_{2n}(E) = 2T_n(E)^2 - 1, \tag{14}$$

$$T_{2n+1}(E) = 2T_{n+1}(E)T_n(E) - T_1(E). \tag{15}$$

The Chebyshev polynomials shifted and scaled from $[-1, 1]$ to $[a, b]$ are defined by

$$\tilde{T}_n(E) = T_n((E - \alpha)/\beta), \tag{16}$$

where

$$\alpha = (b + a)/2, \qquad \beta = (b - a)/2. \tag{17}$$

It is straightforward to see that the orthonormal polynomials of $\sigma_{a,b}^T(E)$ are

$$p_n(E) = \sqrt{2 - \delta_{0,n}}\,\tilde{T}_n(E), \tag{18}$$

where $\delta_{0,n}$ is the Kronecker delta.

In order to compute the moments, one can run the matrix version of the Chebyshev recurrence (16),

$$|\boldsymbol{v}_{n+1}\rangle = (2/\beta)(\boldsymbol{H} - \alpha\boldsymbol{I})|\boldsymbol{v}_n\rangle - |\boldsymbol{v}_{n-1}\rangle, \tag{19}$$

with initial conditions $|\boldsymbol{v}_1\rangle = (1/\beta)\,(\boldsymbol{H} - \alpha\boldsymbol{I})|\boldsymbol{v}_0\rangle$ and $|\boldsymbol{v}_0\rangle = |\boldsymbol{r}\rangle$. Then, at step n, we have $|\boldsymbol{v}_n\rangle = \tilde{T}_n(\boldsymbol{H})|\boldsymbol{r}\rangle$. Using (14) and (15), we see that the moments can then be computed by $\mu_0 = 1$, $\mu_1 = \sqrt{2}\langle\boldsymbol{r}|\boldsymbol{v}_1\rangle$, and

$$\mu_{2n} = 2\sqrt{2}\langle\boldsymbol{v}_n|\boldsymbol{v}_n\rangle - \sqrt{2}\mu_0, \tag{20}$$

ALGORITHM 1. Chebyshev moments.

1: **procedure** CHEB-MOMENTS$(H, |r\rangle, k, a, b)$
2:     $\alpha = (b - a)/2, \beta = (b + a)/2$
3:     $|v_0\rangle = |r\rangle, |v_1\rangle = (1/\beta) (H - \alpha I)|v_0\rangle$
4:     $\mu_0 = 1, \mu_1 = \sqrt{2}\langle v_1 | v_0 \rangle$
5:     **for** $n = 1, 2, \ldots, k - 1$ **do**
6:         $|v_{n+1}\rangle = (2/\beta) (H - \alpha I)|v_n\rangle - |v_{n-1}\rangle$
7:         $\mu_{2n} = 2\sqrt{2}\langle v_n | v_n \rangle - \sqrt{2}\mu_0$
8:         $\mu_{2n+1} = 2\sqrt{2}\langle v_{n+1} | v_n \rangle - \mu_1$
9:     **return** $\mu_0, \mu_1, \ldots, \mu_{2k}$

$$\mu_{2n+1} = 2\sqrt{2}\langle v_{n+1} | v_n \rangle - \mu_1. \tag{21}$$

Note that the factors of $\sqrt{2}$ are due to the fact that we are working with the orthonormal Chebyshev polynomials (18) rather than the typical Chebyshev polynomials.

This approach is summarized in Algorithm 1 and clearly requires a and b to be specified ahead of time. If $[a, b]$ does not contain all of the energies of H, then the algorithm is *exponentially unstable*. Meanwhile, if $[a, b]$ is much wider than the energies of H, then the convergence may be slowed.

III. A LANCZOS-BASED SPECTRUM ADAPTIVE KPM

We now describe our proposed algorithm, which allows $\sigma(E)$ (including its support) to be chosen after the expensive aspects of the algorithm have been carried out. This allows the KPM approximation generated to adapt to the energy spectrum of the Hamiltonian H. A related and very general approach to obtaining quadrature approximations from moment data has recently been described.[21] The approach in this paper is more focused/straightforward, as we focus only on implementing the KPM using Lanczos. Indeed, our approach can be summarized in one sentence: *use the output of Lanczos to compute the KPM moments.*

A. The Lanczos algorithm

When run on H and $|r\rangle$ for k iterations, the Lanczos algorithm (Algorithm 2) iteratively produces an orthonormal basis $|v_n\rangle$ for the Krylov subspace,

$$\text{span}\{|r\rangle, H|r\rangle, \ldots, H^k|r\rangle\}. \tag{22}$$

This is performed via a symmetric three-term recurrence

$$|v_{n+1}\rangle = \frac{1}{\beta_n} (H|v_n\rangle - \alpha_n|v_n\rangle - \beta_{n-1}|v_{n-1}\rangle) \tag{23}$$

with initial conditions $|v_1\rangle = (1/\beta_0)(H|v_0\rangle - \alpha_0|v_0\rangle)$ and $|v_0\rangle = |r\rangle$. At each step, α_n is chosen so that $\langle v_{n+1}|v_n\rangle = 0$ and then β_n is chosen so that $\langle v_{n+1}|v_{n+1}\rangle = 1$. In exact arithmetic, $|v_{n+1}\rangle$ is automatically orthogonal to $|v_i\rangle$ for all $i \leq n - 2$ by symmetry. However, those familiar with the Lanczos algorithm in finite precision arithmetic may be skeptical that we have omitted any form of reorthogonalization. We discuss the stability of our approach in finite precision arithmetic in Sec. III C and argue that reorthogonalization is not needed.

ALGORITHM 2. Lanczos.

1: **procedure** LANCZOS$(H, |r\rangle, k)$
2:     $|v_0\rangle = |r\rangle$
3:     $|\tilde{v}_1\rangle = H|v_0\rangle$
4:     $|\hat{v}_1\rangle = |\tilde{v}_1\rangle - \alpha_0|v_0\rangle, \alpha_0 = \langle v_0|\tilde{v}_1\rangle$
5:     $|v_1\rangle = \hat{v}_1/\beta_0, \beta_0 = \langle \hat{v}_1|\hat{v}_1\rangle$
6:     **for** $n = 1, 2, \ldots, k - 1$ **do**
7:         $|\tilde{v}_{n+1}\rangle = H|v_n\rangle - \beta_{n-1}|v_{n-1}\rangle$
8:         $\alpha_n = \langle v_n|\tilde{v}_n\rangle$
9:         $|\hat{v}_{n+1}\rangle = |\tilde{v}_{n+1}\rangle - \alpha_n|v_n\rangle$
10:         $\beta_n = \langle \hat{v}_{n+1}|\hat{v}_{n+1}\rangle$
11:         $|v_{n+1}\rangle = |\hat{v}_{n+1}\rangle/\beta_n$
12:     **return** $\alpha_0, \alpha_1, \ldots, \alpha_{k-1}, \beta_0, \beta_1, \ldots, \beta_{k-1}$

After k iterations of the Lanczos iteration, the recurrence coefficients form a $(k + 1) \times (k + 1)$ symmetric tridiagonal matrix,

$$H_k = \text{tridiag}\begin{pmatrix} \beta_0 & \beta_1 & \cdots & \beta_{k-1} \\ \alpha_0 & \alpha_1 & \cdots & \alpha_{k-1} & 0 \\ \beta_0 & \beta_1 & \cdots & \beta_{k-1} \end{pmatrix}. \tag{24}$$

If we write the Lanczos basis as $V = \sum_{n=0}^{k} |v_n\rangle\langle e_n|$, where $|e_n\rangle$ is the all-zeros vector with a one in index n, it is not hard to see that (23) implies

$$HV = VH_k + |v\rangle\langle e_k| \tag{25}$$

for some vector $|v\rangle$. Note that it is somewhat more common to write such a recurrence with the upper-left $k \times k$ principle submatrix of H_k, which is closely related to Gaussian quadrature.[36] However, as will become apparent in Sec. III B, using H_k as defined in (24) will provide slightly more approximation power in our Lanczos-based KPM.

B. Getting the KPM moments

It is well-known that the Lanczos tridiagonal matrix H_k contains information suitable for computing the polynomial moments of $\hat{\rho}(E)$ through degree $2k$.

Theorem 1. *Let p be any polynomial of degree at most $2k$. Then,*

$$\langle r|p(H)|r\rangle = \int \hat{\rho}(E)p(E)\,dE = \langle e_0|p(H_k)|e_0\rangle. \tag{26}$$

Proof. The first equality is by definition of the LDOS $\hat{\rho}(E)$. Due to the linearity of polynomials, it suffices to show

$$\langle r|H^n|r\rangle = \langle e_0|H_k^n|e_0\rangle \tag{27}$$

for $n = 0, 1, \ldots, 2k$. We will first show that $H^n|r\rangle = VH_k^n|e_0\rangle$ for all $n \leq k$. Since $V^\dagger V = I$, this immediately implies the desired result.

Suppose $H^{n-1}|r\rangle = VH_k^{n-1}|e_0\rangle$. Then, since $|r\rangle = V|e_0\rangle$, we can use (25) to write

$$H^n|r\rangle = HVH_k^{n-1}|e_0\rangle = VH_k^n|e_0\rangle + |v\rangle\langle e_k|H_k^n|e_0\rangle. \tag{28}$$

ALGORITHM 3. Get KPM moments from Lanczos.

1: **procedure** MOMENTS-FROM-LANCZOS($H_k, \sigma(E)$)
2:     Obtain recurrence coefficients $\{\gamma_n, \delta_n\}$ for the orthogonal polynomials of $\sigma(E)$
3:     $|u_0\rangle = |e_0\rangle$, $|u_1\rangle = (1/\delta_0)(H_k|u_0\rangle - \gamma_0 u_0\rangle)$
4:     $\mu_0 = \langle e_0|u_0\rangle$, $\mu_1 = \langle e_0|u_1\rangle$
5:     **for** $n = 1, 2, \ldots, 2k - 1$ **do**
6:         $|u_{n+1}\rangle = (1/\delta_n)(H_k|u_n\rangle - \gamma_n|u_n\rangle - \delta_{n-1}|u_{n-1}\rangle)$
7:         $\mu_{n+1} = \langle e_0|u_{n+1}\rangle$
8::   **return** $\mu_0, \mu_1, \ldots, \mu_{2k}$

Since H_k is tridiagonal, H_k^n has a bandwidth $2n + 1$ and $\langle e_k|H_k^n|e_0\rangle = 0$ provided $n \leq k$. The base case $|r\rangle = V|e_0\rangle$ is trivial. $\quad\square$

The critical observation is that this allows us to obtain the KPM moments $\{\mu_n\}$ with respect to a reference density $\sigma(E)$, which we can choose *after* we have run the Lanczos computation. In fact, we can cheaply produce approximations corresponding to various different reference measures as this process no longer involves computations with H or any vectors of length d.

If we choose $\sigma(E) = \sigma_{a,b}^T(E)$, we can compute the moments by applying Algorithm 1 to H_k and $|e_0\rangle$. Of note is the fact that the Lanczos algorithm produces high accuracy estimates of extremal eigenvalues when started on a random vector.[37–39] In particular, the largest and smallest eigenvalues of the top-left $k \times k$ sub-matrix of H_k approximate those of H from the interior. We can use this to help guide our choice of a and b.

For other choices of $\sigma(E)$, we can compute the moments directly via the three-term recurrence for the orthogonal polynomials. In particular, we use the matrix version of (10),

$$|u_{n+1}\rangle = \frac{1}{\delta_n}(H_k|u_n\rangle - \gamma_n|u_n\rangle - \delta_{n-1}|u_{n-1}\rangle) \qquad (29)$$

with initial conditions $|u_1\rangle = (1/\delta_0)(H_k|u_0\rangle - \gamma_0|u_0\rangle)$ and $|u_0\rangle = |e_0\rangle$. Then, $|u_n\rangle = |p_n(H_k)|e_0\rangle$, so we can compute $\mu_n = \langle e_0|u_n\rangle$ for $n \leq 2k$. This is summarized in Algorithm 3. The cost of this process depends only on k and not on d, so tricks for halving the number of matrix–vector products with H_k are not needed.

C. Stability of the Lanczos algorithm

The Lanczos algorithm is unstable in the sense that the tridiagonal matrix H_k and basis vectors $\{|v_n\rangle\}$ produced in finite precision arithmetic may be nothing like what would be obtained in exact arithmetic. In particular, by symmetry, $|v_{n+1}\rangle$ is automatically orthogonal to $|v_0\rangle, \ldots, |v_{n-1}\rangle$ in exact arithmetic. However, in finite precision arithmetic, this is no longer even approximately true and the Lanczos basis vectors produced can completely lose orthogonality and even linear independence. To fix this, it is common to use *reorthogonalization*, that is, to explicitly orthogonalize $|\hat{v}_{n+1}\rangle$ against $|v_0\rangle, \ldots, |v_{n-1}\rangle$ before normalizing in Line 10 of Algorithm 2. This, of course, drastically increases the computational requirements to be able to run the algorithm; in particular, the memory required for reorthogonalization scales as $O(dk)$ and the arithmetic cost as $O(dk^2)$.

Despite the instabilities of the Lanczos algorithm without reorthogonalization, there is significant theoretical[40–45] and empirical[19,46,47] evidence that the Lanczos algorithm is highly effective for tasks related to the density of state approximation. In fact, while not widely known, the Lanczos algorithm is *forward stable* for the tasks of computing Chebyshev polynomials and moments.[45,48] We summarize the high-level ideas behind these works, the results of which we believe are relevant to the computational physics and chemistry communities.

In finite precision arithmetic, the outputs of the Lanczos algorithm satisfy a perturbed version of (25),

$$HV = VH_k + |v\rangle\langle e_k| + F, \qquad (30)$$

where F accounts for local errors, which can be expected to be on the size of machine precision. In addition, while V need not be orthonormal, $|1 - \langle v_{n+1}|v_{n+1}\rangle|$ and $|\langle v_{n+1}|v_n\rangle|$ are on the order of machine precision. This (and much more) is analyzed in detail in Refs. 40–42. These assumptions form the basis of essentially all analyses of the behavior of Lanczos-based methods in finite precision arithmetic.

Following the Proof of Theorem 1, we might try to use (30) to understand the difference between $H^n|r\rangle$ and $VH_k^n|e_0\rangle$; i.e., how closely (27) holds in finite precision arithmetic. However, it is not hard to see that this results in an error term with an exponential dependence on k. This is fundamentally because the monomial basis is very poorly conditioned and therefore not a good choice to work with numerically. Indeed, if, instead, we use a Chebyshev basis, then it can be shown the error term itself satisfies Chebyshev-like three-term recurrence and grows only polynomially with k. This yields the following bound.

Theorem 2 (Druskin and Knizhnerman,[45] informal). *Suppose $[a, b] = [E_{\min} - \eta, E_{\max} + \eta]$ for some $\eta = O(\epsilon_{\text{mach}}\text{poly}(k))$. Let H_k be the output of the Lanczos algorithm run on H and $|r\rangle$ for k iterations in finite precision arithmetic without reorthogonalization. Then, for any $n \leq 2k$,*

$$\left\| \tilde{T}_n(H)|r\rangle - V\tilde{T}_n(H_k)|e_0\rangle \right\| = O(\epsilon_{\text{mach}}\,\text{poly}(k)), \qquad (31)$$

where $\tilde{T}_n$ is as in (16) and the big-O hides mild dimensional constants and dependencies on a and b. The actual statement of Ref. 45 (Theorem 1) is more precise and gives explicit bounds for the degree of the $\text{poly}(k)$ terms. We remark that in numerical analysis, the precise numerical value of bounds is often much less important than the intuition conveyed by the bound.

With some additional knowledge of properties satisfied by $V^\dagger V$ in finite precision arithmetic (which can be very far from the identity), a similar result can be shown for the Chebyshev moments:

Theorem 3 (Knizhnerman,[48] informal). *Under similar assumptions to Theorem 2, for any $n \leq 2k$,*

$$\left| \langle r|\tilde{T}_n(H)|r\rangle - \langle e_0|\tilde{T}_n(H_k)|e_0\rangle \right| = O(\epsilon_{\text{mach}}\,\text{poly}(k)). \qquad (32)$$

Theorem 3 implies that the (appropriately scaled) Chebyshev moments can be obtained from the matrix H_k, even when Lanczos was carried out in finite precision arithmetic. Thus, the KPM

with $\sigma_{a,b}^T(E)$ can be implemented from Lanczos after $\boldsymbol{H}_k$ has been obtained, even in finite precision arithmetic. As with Theorem 2,[48] Theorem 1 is much more precise than what is stated in Theorem 3.

Note that any polynomial $p(x)$ of degree $2k$ can be decomposed in a Chebyshev series

$$p(E) = \sum_{n=0}^{2k} c_n \tilde{T}_n(E). \qquad (33)$$

Moreover, if $|p(E)| \le M$ for all $E \in [a, b]$, then

$$|c_i| = \left| (2 - \delta_{0,n}) \int \sigma_{a,b}^T(E) p(E) \tilde{T}_n(E) dE \right| \le 2M. \qquad (34)$$

Applying the triangle inequality gives the bound

$$|\langle \boldsymbol{r}|p(\boldsymbol{H})|\boldsymbol{r}\rangle - \langle \boldsymbol{e}_0|p(\boldsymbol{H}_k)|\boldsymbol{e}_0\rangle| = O(M\epsilon_{\text{mach}} \operatorname{poly}(k)). \qquad (35)$$

In other words, Theorem 3 can be upgraded to hold for any bounded polynomial. Thus, the KPM can also be implemented for choices of $\sigma(E)$ whose orthogonal polynomials are well-behaved.

We remark this that a similar argument also implies that, even in finite precision arithmetic, the FTLM approximation to $\langle \boldsymbol{r}|f(\boldsymbol{H})|\boldsymbol{r}\rangle$ is accurate provided $f(E)$ has a good polynomial approximation, i.e., the same sort of result as would be expected in exact arithmetic. A more detailed analysis is found in Ref. 48. This provides a theoretical justification for the observation that the FTLM works well, even without reorthogonalization.[19]

D. Computational costs

The overall computational costs of our energy adaptive KPM and the standard KPM (assuming $[a, b]$ is known) are almost identical. In addition to the storage required for $\boldsymbol{H}$, both algorithms require storage for three vectors of length d. At iteration n, the Lanczos algorithm (without reorthogonalization) and a standard implementation of the KPM require one matrix–vector product, several vector updates, and two inner products. The algorithms also have additional lower-order arithmetic and storage costs, which depend only on the maximum number of iterations k but not the dimension d. Assuming $k \ll d$, these costs are negligible.

As noted above, the standard KPM typically requires a preprocessing step in which a suitable interval $[a, b]$ is determined, often via a few iterations of Lanczos. While our Lanczos-based KPM avoids the need for such a pre-processing step, this pre-processing step is often cheap relative to the overall computation. In such cases, the fact that our algorithm avoids this step is not particularly significant from a runtime perspective.

Finally, we note one situation in which the runtimes of the two algorithms may differ is on high-performance clusters where the time spent on communication for inner products can dominant the time spent on arithmetic computation. Indeed, in the case of the KPM, the two inner products are used to compute μ_n and μ_{n+1} and do not prevent the algorithm from proceeding. Meanwhile, the two inner products in Lanczos are used to compute α_n and β_n for blocking and therefore must be completed before the algorithm can proceed.

In such settings, if the cost of the pre-processing step is significant, one could run the energy-adaptive KPM suggested here for several iterations to determine good choices of $[a, b]$ and $\sigma(E)$.

Assuming that the Lanczos basis vectors are stored, $p_n(\boldsymbol{H})|\boldsymbol{r}\rangle$ and $p_{n-1}(\boldsymbol{H})|\boldsymbol{r}\rangle$ can be computed without any additional matrix-vector products at which point an explicit three-term recurrence can be continued without the need for blocking inner products. We leave such implementation details to practitioners who have better knowledge of their individual computing environments.

IV. NUMERICAL EXPERIMENTS

Here, we provide several numerical examples to demonstrate the potential usefulness of our spectrum adaptive KPM. Rather than focusing on any single domain area, we aim to provide a diverse collection of examples from a variety of applications. Each of these examples demonstrates a particular aspect of the paradigm of decoupling computation from approximation, which may prove useful in practical settings. Unless stated otherwise, all KPM approximations are computed using our Lanczos-based approach.

A. Spin systems

One of the main uses of KPM and related algorithms is in the study of thermodynamic properties of Heisenberg spin systems.[19,20] Here, we consider the simplest example: the 1D XX spin chain of length m with Hamiltonian

$$\boldsymbol{H} = J \sum_{i=1}^{m-1} \left(\boldsymbol{\sigma}_i^{\text{x}} \boldsymbol{\sigma}_{i+1}^{\text{x}} + \boldsymbol{\sigma}_i^{\text{y}} \boldsymbol{\sigma}_{i+1}^{\text{y}} \right) + h \sum_{i=1}^{m} \boldsymbol{\sigma}_i^{\text{z}}.$$

This system is exactly solvable, meaning that the true spectrum can be computed analytically.[49] For our numerical experiments, we set $m = 20$ so that $d = 2^{20} \approx 10^6$ and use $J = 1/6$ and $h = 6$.

Figure 1 shows a histogram of the exact spectrum, along with the degree $s = 500$ KPM approximation of the LDOS corresponding to a single random state $|\boldsymbol{r}\rangle$. The support $[a, b]$ of the KPM approximation is varied to study the impact of estimating $[E_{\min}, E_{\max}]$. If $E_n \notin [a, b]$ for some index n, then the KPM approximation deteriorates, but if $[a, b]$ is taken too large, then convergence is slowed. Thus, reasonably accurate estimates of $E_{\min}$ and $E_{\max}$ are required. Our spectrum adaptive KPM allows these estimates to be determined after computation with $\boldsymbol{H}$ has occurred.

In Fig. 2, we study the impact of finite precision arithmetic. We first consider how accurately the moments μ_n can be computed. In particular, we compare our Lanczos-based algorithm and a standard implementation of the KPM and observe that the computed moments agree with essentially machine precision. This is expected due to Theorem 3. For reference, we also show the orthogonality of the Lanczos vector without reorthogonalization as well as the difference between the Lanczos recurrence coefficients $\{\alpha_n, \beta_n\}$ and what would have been obtained with reorthogonalization $\{\alpha_n^*, \beta_n^*\}$. There is a complete loss of orthogonality, and the coefficients obtained with and without reorthogonalization are vastly different. This implies that the moments agreeing are not simply due to the outputs with and without reorthogonalization being similar.

B. Tight binding models

Another common use of the KPM is in tight binding models. Here, we consider a cuboid section of a zinc blende crystal with nearest neighbor hopping. The Hamiltonian and visualization of the zinc blende crystal in Fig. 3 were generated using the Kwant code.[50]
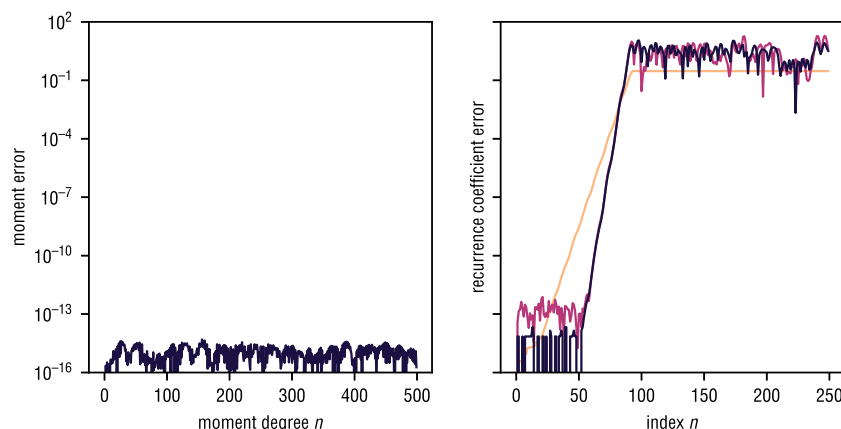
FIG. 2. Study of relevant quantities in finite precision arithmetic. *Legend*: Left: error between Chebyshev moments computed directly and using the finite precision Lanczos recurrence and Theorem 1. Right: level of orthogonality of the Lanczos basis vectors $\max_{i<j\leq n}|\langle \boldsymbol{v}_i|\boldsymbol{v}_j\rangle|$ (——) and error in Lanczos coefficients with and without reorthogonalization $|\alpha_n - \alpha_n^*|$ (——) and $|\beta_n - \beta_n^*|$ (——). *Takeaway*: Even though the Lanczos algorithm completely lost orthogonality, the Chebyshev moments are computed stably! This enables us to stably implement the Chebyshev KPM with Lanczos, even without reorthogonalization.

The resulting Hamiltonian is of dimension $d = 56\,000$, and we output the average of 10 LDOSs corresponding to random independent samples of $|\boldsymbol{r}\rangle$.

The DOS has a large spike at zero, and this spike causes an undamped Chebyshev KPM to exhibit massive Gibbs oscillations. These oscillations can be mitigated somewhat through the use of a damping kernel. However, as shown in Fig. 3, the resolution is only $O(1/k)$, and therefore, s must be taken large to get high resolution.

Instead, we can construct a reference density, which is adaptive to this spike. In particular, we define

$$\sigma(E) = 0.05\sigma_{1-\eta,1+\eta}^T(E) + 0.95\sigma_{a,b}^T(E), \qquad (36)$$

where $\eta = 10^{-2}$. Note that the relative weighting of the spike and the bulk spectrum, as well as the width of the spike, can be tuned using

our energy adaptive KPM. When properly chosen, this in higher resolution in other parts of the spectrum as the KPM approximation does not have to use as much of its approximation power on approximating the spike in $\hat{\rho}(E)/\sigma(E)$. Moreover, even without damping, Gibbs oscillations are relatively minor.

C. Density functional theory

In this example, we consider a matrix obtained in the study of $\mathrm{Ga_{41}As_{41}H_{72}}$ with the pseudo-potential algorithm for real-space electronic structure calculation (PARSEC).[51] The matrix $\mathrm{Ga_{41}As_{41}H_{72}}$ can be obtained from the sparse matrix collection[52] and has been used as a test matrix in a past numerical study.[38,53] This matrix is of dimension $d = 268\,096$ and has many low-lying eigenvalues $|E_n| < 100$ and a cluster of 123 large eigenvalues
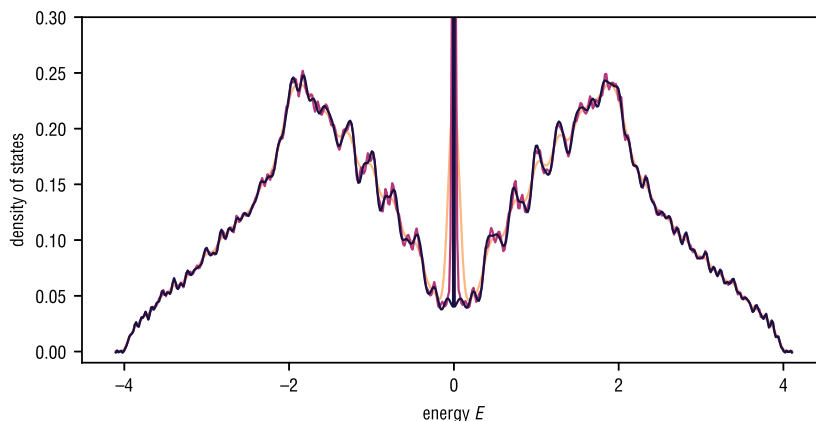

FIG. 3. Approximation of DOS of a cubic zinc blende crystal with nearest neighbor hopping. *Legend*: reference density (36) with $s = 240$ (——) and Chebyshev KPM with Jackson's damping with $s = 240$ (——) and $s = 800$ (——). *Takeaway*: Large spikes in the spectrum are hard to resolve with regular Chebyshev KPM but can be resolved with a suitable choice of reference density $\sigma(E)$. Our spectrum adaptive KPM allows such an approximation to be computed without prior knowledge of the spectrum.

FIG. 4. Approximation of DOS with a large gap in the spectrum. *Legend*: reference density (37) with a convolutional filter and $s = 200$ (——), Chebyshev KPM with Jackson's damping $s = 200$ (——) and $s = 800$ (——), and histogram of top eigenvalues (▭). *Takeaway*: Even with four times the computation, the standard Chebyshev KPM does match the resolution of KPM with a more suitable choice of reference density $\sigma(E)$. Our spectrum adaptive KPM allows such an approximation to be computed without prior knowledge of the spectrum.
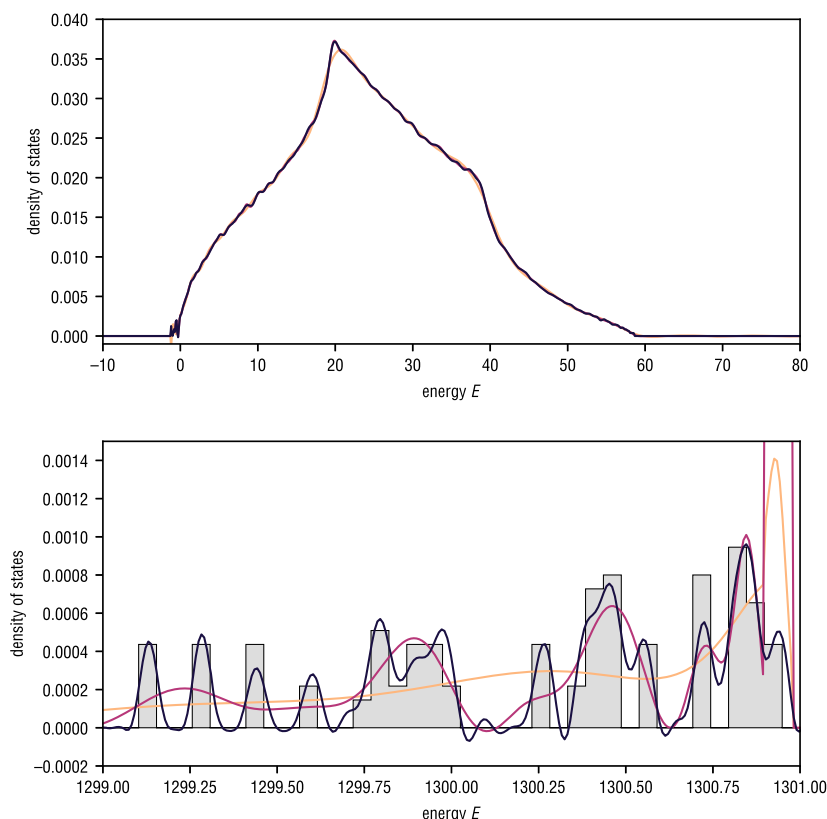
$E_n \in [1299, 1301]$. Thus, while the spectrum can be contained in two reasonably sized intervals, any single interval containing the spectrum must be large. In this experiment, we output the average of 10 random LDOSs.

If the standard Chebyshev KPM is used, the zero part of $\hat{\rho}(E)/\sigma(E)$ in the gap must be approximated with a polynomial.[54] This significantly slows convergence, and even with $s = 800$ moments, the fine-grained structure of the upper spectrum is not resolved. To avoid this delay of convergence, we can take $\sigma(E)$ as a density supported on two disjoint intervals $[a_1, b_1]$ and $[a_2, b_2]$ containing the spectrum of $\boldsymbol{H}$. In particular, we take

$$\sigma(E) = 0.95\sigma_{a_1,b_1}^T(E) + 0.05\sigma_{a_2,b_2}^T(E), \tag{37}$$

where the intervals $[a_1, b_1]$ and $[a_2, b_2]$ are computed based on the eigenvalues of $\boldsymbol{H}_k$. As shown in Fig. 4, this provides a higher resolution in each interval and the structure of the upper cluster is visible. Here, we have applied a simple convolutional filter to the KPM approximation of the upper eigenvalues to reduce Gibbs oscillations.

V. CONCLUSION

We have described an energy adaptive kernel polynomial method based on the Lanczos algorithm. Our approach allows many different KPM approximations to be tested out for close to zero cost, after computation with $\boldsymbol{H}$ has finished. Experiments demonstrate situations in which this allows the reference density $\sigma(E)$ to be chosen in such a way to improve the resolution of the approximation. It is our belief that the paradigm of separating computation from the desired approximation is beneficial in most settings in which the KPM is used and that our algorithm has the potential to improve the usability of the KPM.

AUTHOR DECLARATIONS

Conflict of Interest

The author has no conflicts to disclose.

Author Contributions

Tyler Chen: Conceptualization (equal); Formal analysis (equal); Visualization (equal); Writing – original draft (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

[1] J. Skilling, "The eigenvalues of mega-dimensional matrices," in *Maximum Entropy and Bayesian Methods* (Springer, Netherlands, 1989), pp. 455–466.

[2] J. Jaklič and P. Prelovšek, "Lanczos method for the calculation of finite-temperature quantities in correlated systems," Phys. Rev. B **49**, 5065–5068 (1994).

[3] Z. Bai, G. Fahey, and G. Golub, "Some large-scale matrix computation problems," J. Comput. Appl. Math. **74**, 71–89 (1996).

[4] L. Lin, Y. Saad, and C. Yang, "Approximating spectral densities of large matrices," SIAM Rev. **58**, 34–65 (2016).

[5] F. Jin, D. Willsch, M. Willsch, H. Lagemann, K. Michielsen, and H. De Raedt, "Random state technology," J. Phys. Soc. Jpn. **90**, 012001 (2021).

[6] R. Silver and H. Röder, "Densities of states of mega-dimensional Hamiltonian matrices," Int. J. Mod. Phys. C **05**, 735–753 (1994).

[7] R. Silver, H. Roeder, A. Voter, and J. Kress, "Kernel polynomial approximations for densities of states and spectral functions," J. Comput. Phys. **124**, 115–130 (1996).

[8] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, "The kernel polynomial method," Rev. Mod. Phys. **78**, 275–306 (2006).

[9] S. Ganeshan, J. Pixley, and S. Das Sarma, "Nearest neighbor tight binding models with an exact mobility edge in one dimension," Phys. Rev. Lett. **114**, 146601 (2015).

[10] J. H. García, L. Covaci, and T. G. Rappoport, "Real-space calculation of the conductivity tensor for disordered topological matter," Phys. Rev. Lett. **114**, 116602 (2015).

[11] S. Carr, D. Massatt, S. Fang, P. Cazeaux, M. Luskin, and E. Kaxiras, "Twistronics: Manipulating the electronic properties of two-dimensional layered materials through their twist angle," Phys. Rev. B **95**, 075420 (2017).

[12] D. Carvalho, N. A. García-Martínez, J. L. Lado, and J. Fernández-Rossier, "Real-space mapping of topological invariants using artificial neural networks," Phys. Rev. B **97**, 115453 (2018).

[13] D. Varjas, M. Fruchart, A. R. Akhmerov, and P. M. Perez-Piskunow, "Computation of topological phase diagram of disordered Pb$_{1-x}$Sn$_x$Te using the kernel polynomial method," Phys. Rev. Res. **2**, 013229 (2020).

[14] I. Han, D. Malioutov, H. Avron, and J. Shin, "Approximating spectral sums of large-scale matrices using stochastic Chebyshev approximations," SIAM J. Sci. Comput. **39**, A1558–A1585 (2017).

[15] K. Dong, A. R. Benson, and D. Bindel, "Network density of states," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2019).

[16] M. Aichhorn, M. Daghofer, H. G. Evertz, and W. von der Linden, "Low-temperature Lanczos method for strongly correlated systems," Phys. Rev. B **67**, 161103 (2003).

[17] S. Ubaru, J. Chen, and Y. Saad, "Fast estimation of $tr(f(a))$ via stochastic Lanczos quadrature," SIAM J. Matrix Anal. Appl. **38**, 1075–1099 (2017).

[18] D. Granziol, X. Wan, and T. Garipov, "Deep curvature suite," arXiv:1912.09656 [stat.ML] (2019).

[19] J. Schnack, J. Richter, and R. Steinigeweg, "Accuracy of the finite-temperature Lanczos method compared to simple typicality-based estimates," Phys. Rev. Res. **2**, 013186 (2020).

[20] K. Morita and T. Tohyama, "Finite-temperature properties of the Kitaev–Heisenberg models on kagome and triangular lattices studied by improved finite-temperature Lanczos methods," Phys. Rev. Res. **2**, 013205 (2020).

[21] T. Chen, T. Trogdon, and S. Ubaru, "Randomized matrix-free quadrature for spectrum and spectral sum approximation," arXiv:2204.01941 [math.NA] (2022).

[22] S. Goldstein, J. L. Lebowitz, C. Mastrodonato, R. Tumulka, and N. Zanghì, "Normal typicality and von Neumann's quantum ergodic theorem," Proc. R. Soc. A **466**, 3203–3224 (2010).

[23] R. Alben, M. Blume, H. Krakauer, and L. Schwartz, "Exact results for a three-dimensional alloy with site diagonal disorder: Comparison with the coherent potential approximation," Phys. Rev. B **12**, 4090–4094 (1975).

[24] J. Schnack, J. Richter, T. Heitmann, J. Richter, and R. Steinigeweg, "Finite-size scaling of typicality-based estimates," Z. Naturforsch. A **75**, 465–473 (2020).

[25] D. Girard, "Un algorithme simple et rapide pour la validation croisée généralisée sur des problèmes de grande taille," *RR 669-M, TIM3, Informatique et Mathématiques Appliquées de Grenoble* (1987).

[26] A. Girard, "A fast 'Monte-Carlo cross-validation' procedure for large least squares problems with noisy data," Numer. Math. **56**, 1–23 (1989).

[27] M. Hutchinson, "A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines," Commun. Stat. Simul. Comput. **18**, 1059–1076 (1989).

[28] H. Avron and S. Toledo, "Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix," J. ACM **58**, 1–34 (2011).

[29] F. Roosta-Khorasani and U. Ascher, "Improved bounds on sample size for implicit matrix trace estimators," Found. Comput. Math. **15**, 1187–1212 (2014).

[30] A. Cortinovis and D. Kressner, "On randomized trace estimates for indefinite matrices with an application to determinants," Found. Comput. Math. **22**, 875 (2021).

[31] If we can compute the moments of the DOS $\rho(E)$, then we can apply the KPM to the DOS directly.

[32] V. Braverman, A. Krishnan, and C. Musco, "Sublinear time spectral density estimation," in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing* (ACM, 2022); arXiv:2104.03461 [cs.DS].

[33] H. Schlüter, F. Gayk, H.-J. Schmidt, A. Honecker, and J. Schnack, "Accuracy of the typicality approach using Chebyshev polynomials," Z. Naturforsch. A **76**, 823–834 (2021).

[34] Y. Saad, "Iterative solution of indefinite symmetric linear systems by methods using orthogonal polynomials over two disjoint intervals," SIAM J. Numer. Anal. **20**, 784–811 (1983).

[35] J. S. Geronimo and W. Van Assche, "Orthogonal polynomials on several intervals via a polynomial mapping," Trans. Am. Math. Soc. **308**, 559–581 (1988).

[36] G. H. Golub and G. Meurant, *Matrices, Moments and Quadrature with Applications*, Princeton Series in Applied Mathematics Vol. 30 (Princeton University Press, 2009).

[37] J. Kuczyński and H. Woźniakowski, "Estimating the largest eigenvalue by the power and Lanczos algorithms with a random start," SIAM J. Matrix Anal. Appl. **13**, 1094–1122 (1992).

[38] Y. Zhou and R.-C. Li, "Bounding the spectrum of large Hermitian matrices," Linear Algebra Appl. **435**, 480–493 (2011).

[39] P.-G. Martinsson and J. A. Tropp, "Randomized numerical linear algebra: Foundations and algorithms," Acta Numer. **29**, 403–572 (2020).

[40] C. C. Paige, "Practical use of the symmetric Lanczos process with re-orthogonalization," BIT Numer. Math. **10**, 183–195 (1970).

[41] C. C. Paige, "Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix," IMA J. Appl. Math. **18**, 341–349 (1976).

[42] C. C. Paige, "Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem," Linear Algebra Appl. **34**, 235–258 (1980).

[43] A. Greenbaum, "Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences," Linear Algebra Appl. **113**, 7–63 (1989).

[44] Z. Strakos and A. Greenbaum, Open questions in the convergence analysis of the Lanczos process for the real symmetric eigenvalue problem, University of Minnesota, 1992, https://conservancy.umn.edu/handle/11299/1838.

[45] V. L. Druskin and L. A. Knizhnerman, "Error bounds in the simple Lanczos procedure for computing functions of symmetric matrices and eigenvalues," Comput. Math. Math. Phys. **31**, 20–30 (1991).

[46] M. W. Long, P. Prelovšek, S. El Shawish, J. Karadamoglou, and X. Zotos, "Finite-temperature dynamical correlations using the microcanonical ensemble and the Lanczos algorithm," Phys. Rev. B **68**, 235106 (2003).

[47] T. Chen, T. Trogdon, and S. Ubaru, "Analysis of stochastic Lanczos quadrature for spectrum approximation," in *Proceedings of the 38th International Conference on Machine Learning, Proceedings of the Machine Learning Research* (PMLR, 2021), Vol. 139, pp. 1728–1739; arXiv:2105.06595 [cs.DS].

[48] L. A. Knizhnerman, "The simple Lanczos procedure: Estimates of the error of the Gauss quadrature formula and their applications," Comput. Math. Math. Phys. **36**, 1481–1492 (1996).

[49] M. Karabach, G. Müller, H. Gould, and J. Tobochnik, "Introduction to the Bethe ansatz I," Comput. Phys. **11**, 36 (1997).

[50] C. W. Groth, M. Wimmer, A. R. Akhmerov, and X. Waintal, "Kwant: A software package for quantum transport," New J. Phys. **16**, 063065 (2014).

[51] L. Kronik, A. Makmal, M. L. Tiago, M. M. G. Alemany, M. Jain, X. Huang, Y. Saad, and J. R. Chelikowsky, "PARSEC—The pseudopotential algorithm for real-space electronic structure calculations: Recent advances and novel applications to nano-structures," Phys. Status Solidi B **243**, 1063–1079 (2006).

[52] T. A. Davis and Y. Hu, "The university of Florida sparse matrix collection," ACM Trans. Math. Softw. **38**, 1–25 (2011).

[53] R. Li, Y. Xi, L. Erlandson, and Y. Saad, "The eigenvalues slicing library (EVSL): Algorithms, implementation, and software," SIAM J. Sci. Comput. **41**, C393–C415 (2019).

[54] If the number of outlying eigenvalues is known, it is possible to deflate them.[8,20] However, this requires additional computation, including the storage of the eigenvectors (which may be intractable).

25 March 2024 02:32:44