Density of states and eigenvalue counts via approximation theory methods

Yousef Saad

Department of Computer Science and Engineering

University of Minnesota

PMAA-2014

Lugano, July 3, 2014

I. Eigenvalue counts [E. Polizzi, E. Di Napoli, YS]

The problem:

➤ Find an estimate of the number of eigenvalues of a matrix in a given interval $[a, \quad b]$.

Main motivation:

➤ Eigensolvers based on splitting the spectrum intervals and extracting eigenpairs from each interval independently.

➤ Contour integration-type methods, see, e,g.,:
  - FEAST approach [Polizzi 2011]
  - Sakurai-Sugiura - related method [2003, 2007, ..]

➤ Polynomial filtering, e.g.,:
  - Schofield, Chelikowsky, YS'2011.

Eigenvalue counts: Standard approach and an alternative

➤ Let A be a Hermitian matrix with eigenpairs (λ_i, u_i), where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and a, b such that $\lambda_1 \leq a \leq b \leq \lambda_n$.

➤ Want number $\mu_{[a,b]}$ of λ_i's $\in [a, b]$.

➤ Standard method: Use Sylvester inertia theorem. Requires two LDL^T factorizations $\rightarrow$ expensive!

➤ Alternative: Exploit trace of the eigen-projector:

$$P = \sum_{\lambda_i \in [a\ b]} u_i u_i^T.$$

➤ We know that :

$$\mathrm{Tr}\,(P) = \mu_{[a,b]}$$

➤ Goal: calculate an approximation to : $\mathrm{Tr}\,(P)$

➤ P is not available ... but can be approximated by

- ● a polynomial in A, or
- ● a rational function in A.

Approximation theory viewpoint:

➤ Interpret P as a step function of A, namely:

$$P = h(A) \quad \text{where} \quad h(t) = \begin{cases} 1 & \text{if } t \in [a \; b] \\ 0 & \text{otherwise} \end{cases}$$

➤ Approximate $h(t)$ by a polynomial ψ

➤ Then use statistical estimator for approximating $\text{Tr}(\psi(A))$

➤ Hutchinson's unbiased estimator uses only matrix-vector products to approximate the trace of a generic matrix A.

➤ Generate random unit vectors $v_k, k = 1, .., n_v$ with zero mean. Then

$$\mathrm{tr}(\psi(A)) \approx \frac{n}{n_v} \sum_{k=1}^{n_v} v_k^\top \psi(A) v_k.$$

➤ We use degree p Chebyshev polynomials, with Jackson damping (g_i^k) coefficients

$$\psi(t) = \sum_{i=0}^{k} g_i^k \gamma_i T_i(t)$$

➤ To compute 'moments' $v^T T_k(A) v$, let $v_k \equiv T_k(A) v$, and exploit 3-term recurrence $T_{k+1}(t) = 2t T_k(t) - T_{k-1}(t) \rightarrow$

$$v_{k+1} = 2A v_k - v_{k-1}$$

Computing the polynomials: Jackson-Chebyshev

$$\gamma_i = \frac{2 - \delta_{i0}}{\pi} \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x) dx \quad \delta_{i0} = \text{Kronecker symbol}$$

➤ The g_i^k's dampen high order terms in sum.

➤ Explicit expression known [L. O. Jay, et al CPC, 118:21–30 (1999)]

➤ Expansion coefficients γ_i also known

Mid−pass polynom. filter [−1 .3 .6 1]; Degree = 80

Standard Cheb.
Jackson−Cheb.
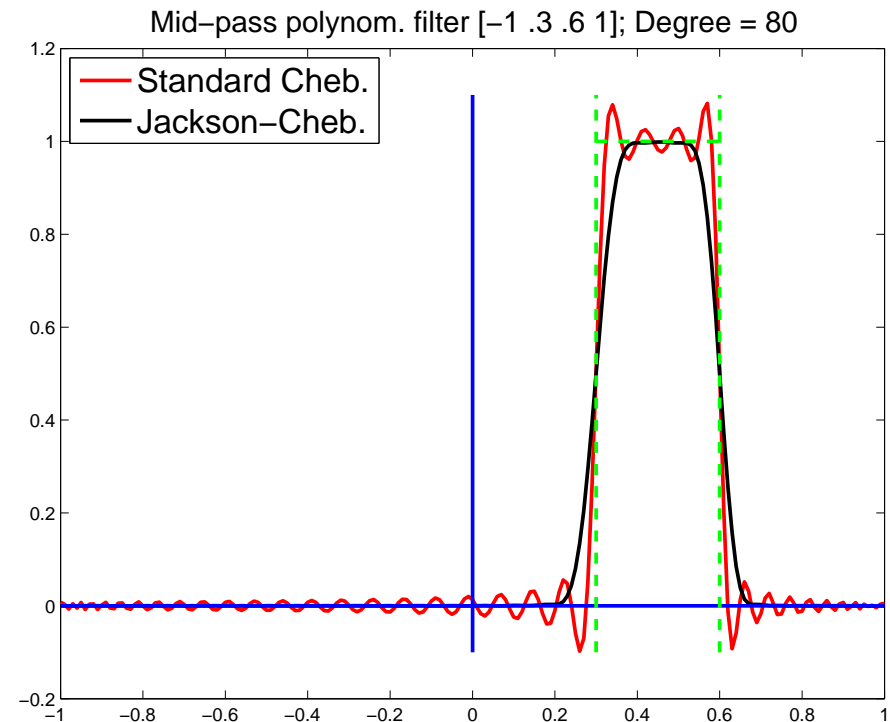
Generalized eigenvalue problems

$$Ax = \lambda Bx$$

➤ Matrices A and B are symmetric and B is positive definite.

The projector P becomes

$$P = \sum_{\lambda_i \in [a\ b]} u_i u_i^T B$$

➤ Again: Eigenvalue count $\mu_{[a,\ b]}$ equals the trace of P
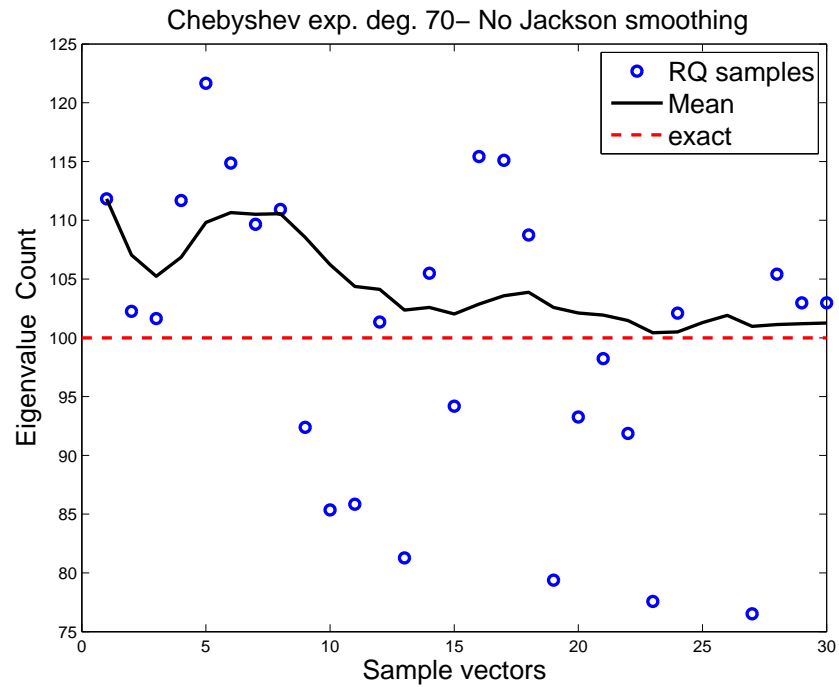
➤ Exploit relation:

$$\text{inertia}(A - \sigma B) = \text{inertia}(B^{-1}A - \sigma I)$$
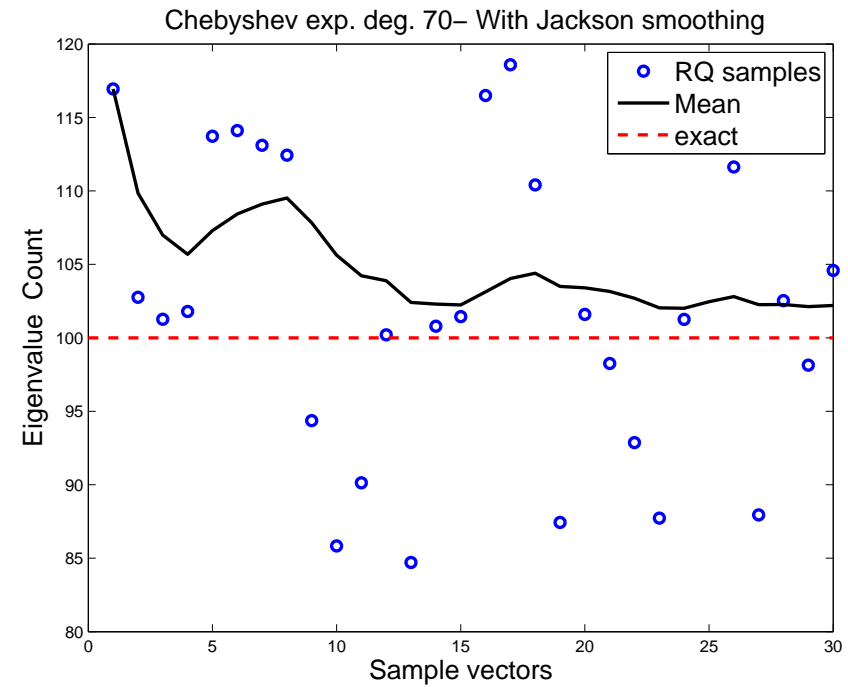
➤ No need to factor or to solve systems

An example

➤ Matrix 'Na5' from PARSEC [see U. Florida collection]

➤ $n = 5832$, $nnz = 305630$ nonzero entries.

➤ Obtain the eigenvalue count when $a = (\lambda_{100} + \lambda_{101})/2$ and $b = (\lambda_{200} + \lambda_{201})/2$ so $\mu_{[a,b]} = 100$.

➤ Use pol. of degree 70.

Without Jackson Damping

With Jackson Damping

Chebyshev exp. deg. 70– No Jackson smoothing

Chebyshev exp. deg. 70– With Jackson smoothing

An example from FEAST

➤ FEAST developed by Eric Polizzi (U. Mass. Amherst)..

➤ Based on a form of subspace iteration with a rational function of A

➤ Also works for generalized problems $Au = \lambda B$.

➤ Example: a small generalized problem ($n = 12,450, nnz = 86,808$).

➤ Result with standard Chebyshev shown. Deg=100, $nv = 70$.

Case: Gen2D; deg = 100; $n_v = 70$
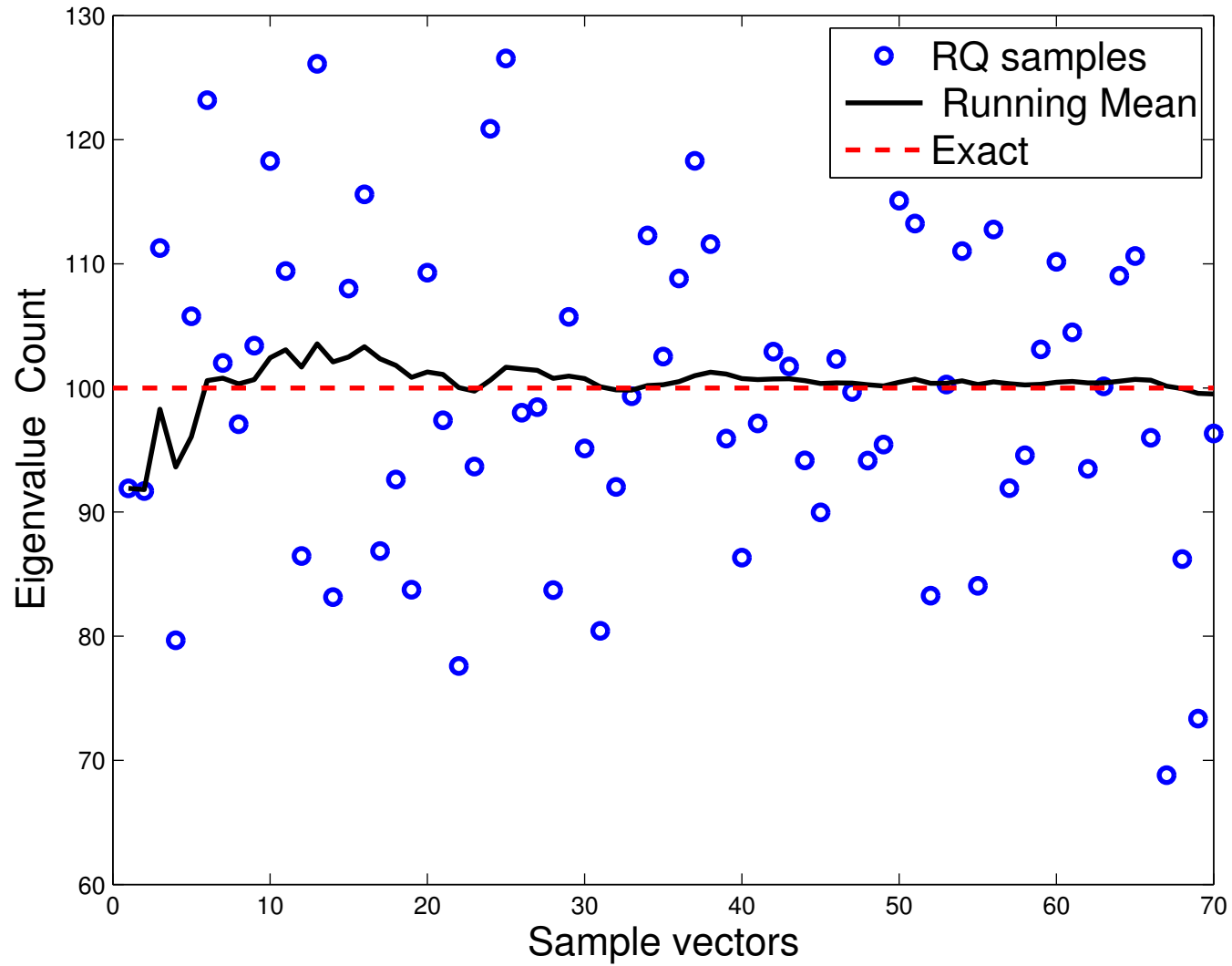
➤ A few more comments:

• Method also works with rational approximations ...

• .. and it works for nonsymmetric problems (eigenvalues inside a given contour).

• For details see paper:

E. Di Napoli, E, Polizzi, and YS. Efficient estimation of eigenvalue counts in an interval. Preprint:
arXiv: http://arxiv.org/abs/1308.4275.

II. Computing Densities of States [Lin-Lin, Chao Yang, YS]

➤ Formally, the Density Of States (DOS) of a matrix A is

$$\phi(t) = \frac{1}{n} \sum_{j=1}^{n} \delta(t - \lambda_j),$$

where

- δ is the Dirac δ-function or Dirac distribution
- $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ are the eigenvalues of A

➤ Note: $\mu_{[ab]}$ can be obtained from ϕ

➤ $\phi(t)$ == a probability distribution function == probability of finding eigenvalues of A in a given infinitesimal interval near t.

➤ Also known as the spectral density

➤ Very important uses in Solid-State physics

The Kernel Polynomial Method

➤ Used by Chemists to calculate the DOS – see Silver and Röder'94 , Wang '94, Drabold-Sankey'93, + others

➤ Basic idea: expand DOS into Chebyshev polynomials

➤ Coefficients γ_k lead to evaluating $\mathrm{Tr}\,(T_k(A))$

➤ Use trace estimators [discovered independently] to get traces

➤ Next: A few details

➤ Assume change of variable done so eigenvalues lie in $[-1,\ 1]$.

➤ Include the weight function in the expansion so expand:

$$\hat{\phi}(t) = \sqrt{1 - t^2}\phi(t) = \sqrt{1 - t^2} \times \frac{1}{n}\sum_{j=1}^{n}\delta(t - \lambda_j).$$

➤ Then, (full) expansion is: $\hat{\phi}(t) = \sum_{k=0}^{\infty} \mu_k T_k(t)$.

➤ Expansion coefficients μ_k are formally defined by:

$$
\begin{aligned}
\mu_k &= \frac{2 - \delta_{k0}}{\pi} \int_{-1}^{1} \frac{1}{\sqrt{1 - t^2}} T_k(t) \hat{\phi}(t) dt \\
&= \frac{2 - \delta_{k0}}{\pi} \int_{-1}^{1} \frac{1}{\sqrt{1 - t^2}} T_k(t) \sqrt{1 - t^2} \phi(t) dt \\
&= \frac{2 - \delta_{k0}}{n\pi} \sum_{j=1}^{n} T_k(\lambda_j). \qquad \text{with} \quad \delta_{ij} = \text{Dirac symbol}
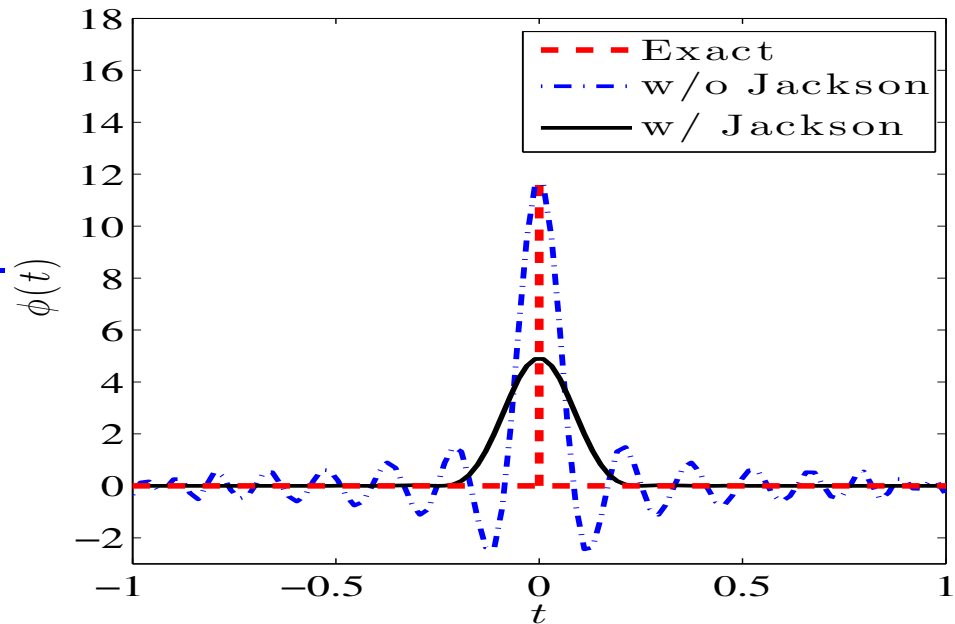\end{aligned}
$$

➤ Note: $\sum T_k(\lambda_i) = \text{Trace}[T_k(A)]$
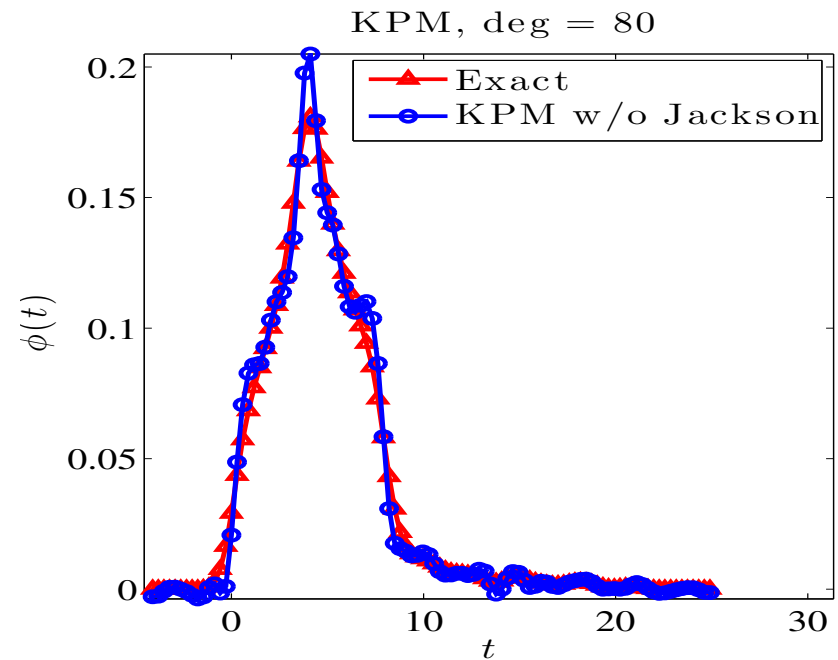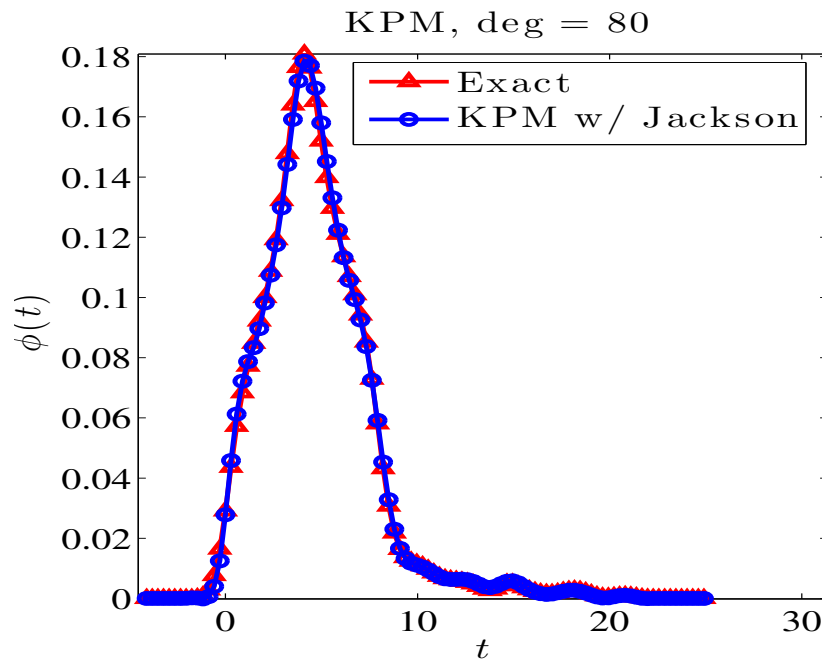
➤ Estimate this, e.g., via stochastic estimator

$$
\text{Trace}(T_k(A)) \approx \frac{1}{n_{\text{vec}}} \sum_{l=1}^{n_{\text{vec}}} \left(v^{(l)} \right)^T T_k(A) v^{(l)}.
$$

➤ To compute scalars of the form $v^T T_k(A) v$, exploit again 3-term recurrence of the Chebyshev polynomial ...

➤ Same Jackson smoothing as before can be used

An example with degree 80 polynomials

Left: Jackson damping; right: without Jackson damping.

Issue: How to deal with Distributions

➤ Highly discontinuous nature – not easy to handle

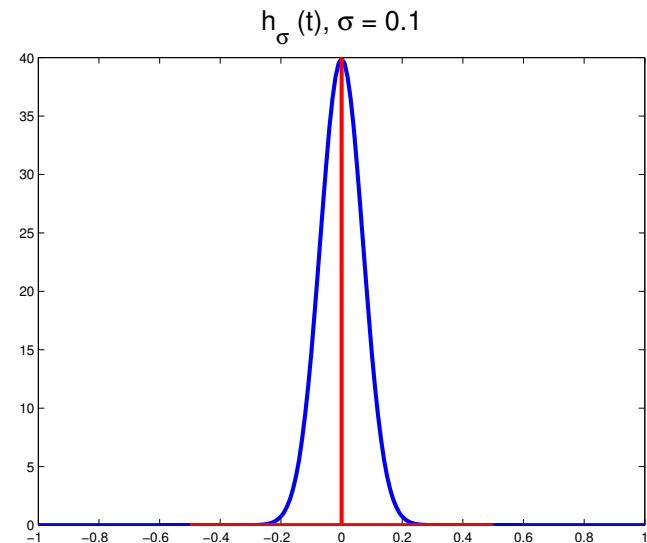➤ Solution for practical and theoretical purposes: replace ϕ by a 'blurred" (continuous) version ϕ_σ:

$$\phi_\sigma(t) = \frac{1}{n} \sum_{j=1}^{n} h_\sigma(t - \lambda_j),$$

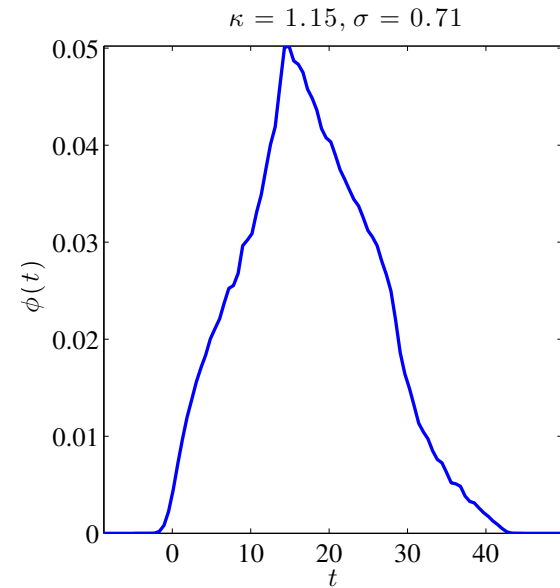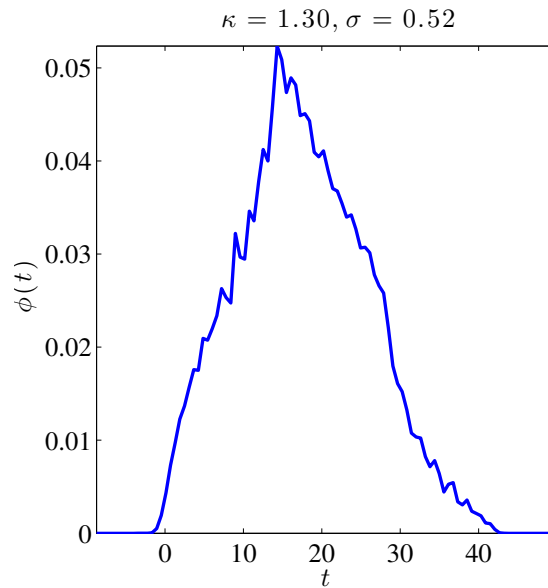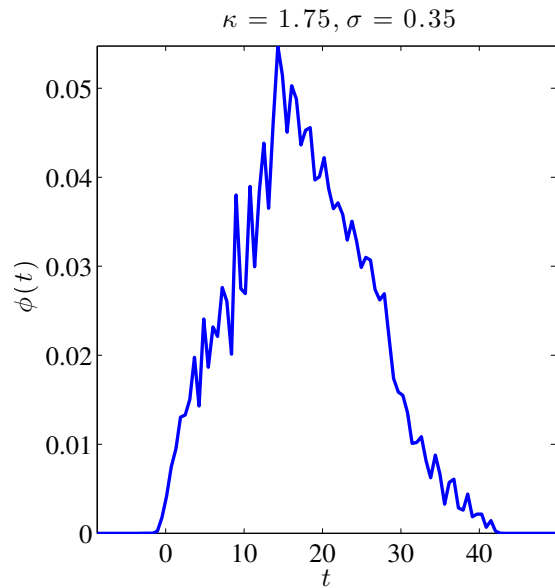where $h_\sigma(t)$ = any $\mathcal{C}^\infty$ function s.t.:
- $\int_{-\infty}^{+\infty} h_\sigma(s)ds = 1$
- h_σ has a peak at zero

➤ An example is the Gaussian:

$$h_\sigma(t) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{t^2}{2\sigma^2}}.$$

$h_\sigma(t)$, σ = 0.1

How to select σ? Example for Si_2

$\kappa = 1.75, \sigma = 0.35$ $\kappa = 1.30, \sigma = 0.52$ $\kappa = 1.15, \sigma = 0.71$

$\kappa = 1.08, \sigma = 0.96$

➤ Higher $\sigma \rightarrow$ smoother curve

➤ But loss of detail ..

➤ Compromise: $\sigma = \dfrac{h}{2\sqrt{2\log(\kappa)}}$,

➤ $h =$ resolution, $\kappa =$ parameter > 1
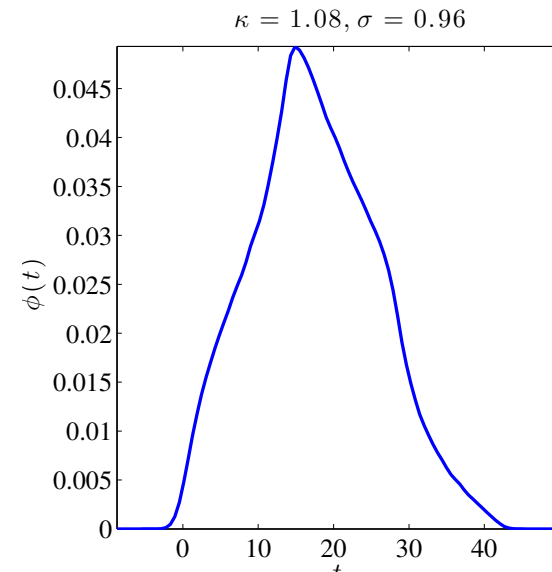
Delta-Gauss Legendre

➤ Idea: Instead of approximating ϕ directly, first select a representative ϕ_σ of ϕ for a given σ and then approximate ϕ_σ.

➤ ϕ_σ is a 'surrogate' for ϕ. Obtained by replacing δ_λ by :

$$h_\sigma(\lambda - t) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{(\lambda - t)^2}{2\sigma^2}\right].$$

➤ Goal: to expand into Legendre polynomials $L_k(\lambda)$

➤ With normalization factor expansion is written as:

$$h_\sigma(\lambda - t) = \frac{1}{(2\pi\sigma^2)^{1/2}} \sum_{k=0}^{\infty} \left(k + \frac{1}{2}\right) \gamma_k L_k(\lambda).$$

➤ To determine the γ_k's we will also need to compute:

$$\psi_k = \int_{-1}^{1} L_k'(s) e^{-\frac{1}{2}((s-t)/\sigma)^2} ds.$$

Set $\zeta_k = e^{-\frac{1}{2}((1-t)/\sigma)^2} - (-1)^k e^{-\frac{1}{2}((1+t)/\sigma)^2}$.

➤ Then, for $k = 0, 1, \cdots ,$:

$$\begin{cases} \gamma_{k+1} = \frac{2k+1}{k+1}\left[\sigma^2(\psi_k - \zeta_k) + t\gamma_k\right] - \frac{k}{k+1}\gamma_{k-1} \\ \psi_{k+1} = (2k+1)\gamma_k + \psi_{k-1}. \end{cases}$$

Initiialization: set $\gamma_{-1} = \psi_{-1} = 0 \; \psi_1 = \gamma_0$, and $\psi_0 = 0$ and:

$$\gamma_0 = \sigma \sqrt{\frac{\pi}{2}} \left[\text{erf}\left(\frac{1-t}{\sqrt{2}\sigma}\right) + \text{erf}\left(\frac{1+t}{\sqrt{2}\sigma}\right)\right],$$

Use of the Lanczos Algorithm

➤ Let θ_i, $i = 1\cdots, m$ be the Ritz values obtained from Lanczos with starting vector v

➤ y_i's associated eigenvectors; Ritz vectors: $\{V_m y_i\}_{i=1:m}$

➤ Ritz values approximate eigenvalues [from 'outside in']

➤ Could compute θ_i's then get approximate DOS from these

➤ Problem: θ_i not good enough approximations – especially inside the spectrum.

➤ Better idea: exploit relation of Lanczos with (discrete) orthogonal polynomials and related Gaussian quadrature:

$$\int p(t)dt \approx \sum_{i=1}^{m} a_i p(\theta_i) \quad a_i = \left[e_1^T y_i\right]^2$$

➤ See, e.g., Golub & Meurant '93, and also Gautschi'81, Golub and Welsch '69.

➤ Formula exact when p is a polynomial of degree $\leq 2m+1$

➤ Consider now $\int p(t)dt =$ discrete integral $\equiv$

$$(p(A)v, v) = \sum \beta_i^2 p(\lambda_i) \equiv\ <\phi_v, p>$$

➤ Then $\langle\phi_v, p\rangle \approx \sum a_i p(\theta_i) = \sum a_i \langle\delta_{\theta_i}, p\rangle \rightarrow$

$$\phi_v \approx \sum a_i \delta_{\theta_i}$$

➤ To mimick the effect of $\beta_i = 1, \forall i$, use several vectors v and average the result of the above formula over them..

Experiments

➤ Goal: to compare errors for similar number of matrix-vector products

➤ Example: Kohn-Sham Hamiltonian associated with a benzene molecule generated from PARSEC; size $n = 8,219$

➤ In all cases, we use 10 sampling vectors

➤ General observation: DGL, Lanczos, and KPM are best,

➤ Spectroscopic method does OK

➤ Haydock's method [another method based on the Lanczos algorithm] not as good

Method	L^1 error	L^2 error	L^∞ error
KPM w/ Jackson, deg=80	2.592e-02	5.032e-03	2.785e-03
KPM w/o Jackson, deg=80	2.634e-02	4.454e-03	2.002e-03
KPM Legendre, deg=80	2.504e-02	3.788e-03	1.174e-03
Spectroscopic, deg=40	5.589e-02	8.652e-03	2.871e-03
Spectroscopic, deg=100	4.624e-02	7.582e-03	2.447e-03
DGL, deg=80	1.998e-02	3.379e-03	1.149e-03
Lanczos, deg=80	2.755e-02	4.178e-03	1.599e-03
Haydock, deg=40	6.951e-01	1.302e-01	6.176e-02
Haydock, deg=100	2.581e-01	4.653e-02	1.420e-02

L^1, L^2, and L^∞ error compared with the normalized "surrogate" DOS for benzene matrix
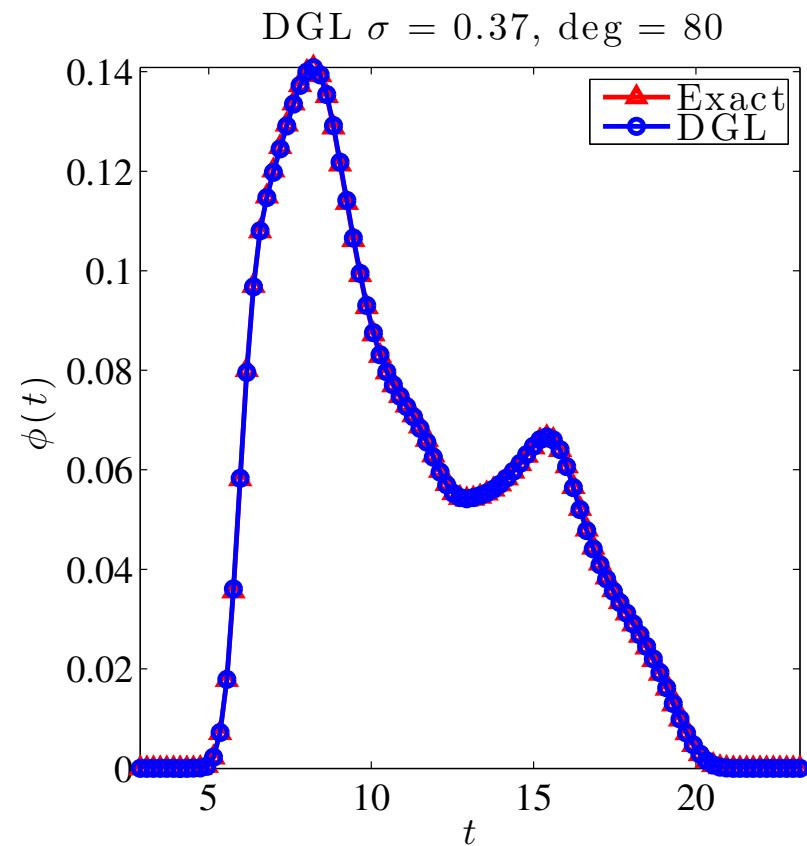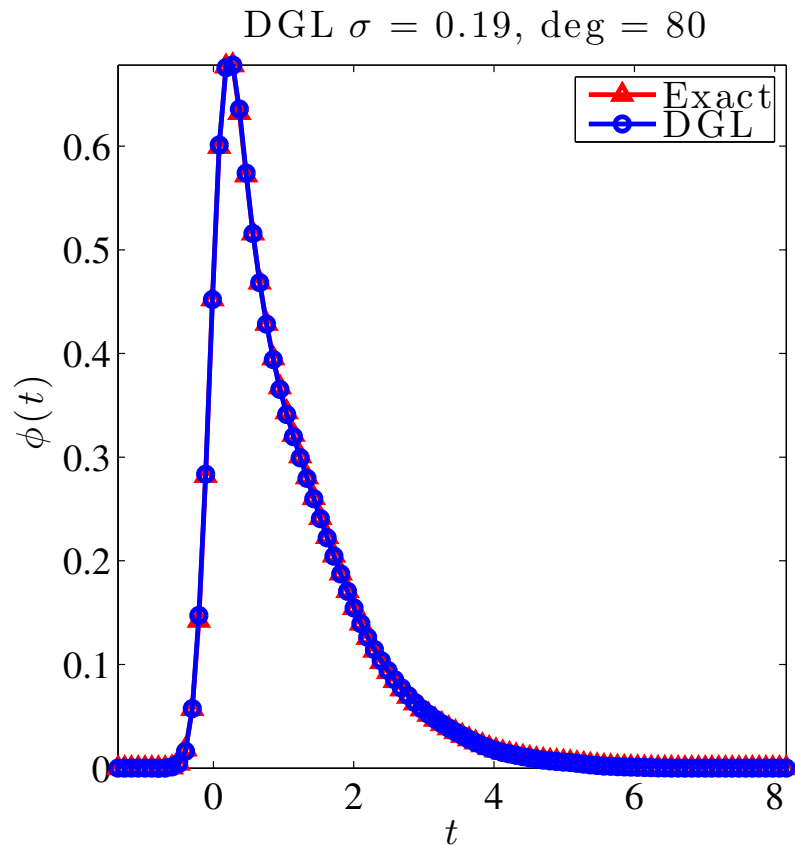
Other matrices

Matrix	n	λ_1	λ_n
$Ga_{10}As_{10}H_{30}$	113,081	-1.2	1.3×10^3
PE3K	9,000	8.1×10^{-6}	1.3×10^2
CFD1	70,656	2.0×10^{-5}	6.8
SHWATER	81,920	5.8	2.0×10^1

Description of the size and the spectrum range of the test matrices.

Matrix	Method	L^1 error	L^2 error	L^∞ error
$\mathbf{Ga_{10}As_{10}H_{30}}$	DGL	3.937e-03	3.214e-04	4.301e-05
	Lanczos	4.828e-03	3.940e-04	5.452e-05
PE3K	DGL	4.562e-03	7.368e-04	3.143e-04
	Lanczos	5.459e-03	7.372e-04	3.294e-04
CFD1	DGL	2.276e-03	1.299e-03	1.746e-03
	Lanczos	2.024e-03	1.286e-03	2.478e-03
SHWATER	DGL	3.779e-03	1.282e-03	9.328e-04
	Lanczos	3.047e-03	9.829e-04	6.100e-04

L^1, L^2, and L^∞ error associated with the approximate spectral densities produced by the DGL and Lanczos methods for different test matrices.

Approximate spectral densities of CFD1 and SHWATER matrices obtained by DGL along with exact smoothed ones

Conclusion

➤ Probabilistic algorithms provide powerful tools for solving various problems: eigenvalue counts, DOS, Diag $(f(A))$..

➤ Most of the algorithms we discussed rely on estimating trace of $f(A)$.

Q: Can we do better than random sampling [e.g., probing,..]?

➤ Physicists are interested in modified forms of the density of states. $\rightarrow$ Explore extentions of what we did.