Decay Properties of Certain Matrix Functions Arising in Quantum Mechanics

Michele Benzi

Emory University
Department of Mathematics and Computer Science
Atlanta, GA 30322, USA

Acknowledgments

- Joint work with Nader Razouk (PhD thesis, Emory U., 2008)
- Thanks to Matt Challacombe (Los Alamos, T-12)
- NSF (Computational Mathematics)

Overview

1. Density matrices

Overview

1. Density matrices

2. Sparsity ("localization") in matrix functions

Overview

1. Density matrices

2. Sparsity ("localization") in matrix functions

3. $O(n)$ approximation of matrix functions

Overview

1. Density matrices
2. Sparsity ("localization") in matrix functions
3. $O(n)$ approximation of matrix functions
4. A few numerical experiments

Overview

1. Density matrices
2. Sparsity ("localization") in matrix functions
3. $O(n)$ approximation of matrix functions
4. A few numerical experiments
5. Some open problems

Overview

1. Density matrices
2. Sparsity ("localization") in matrix functions
3. $O(n)$ approximation of matrix functions
4. A few numerical experiments
5. Some open problems

Density Matrices

In quantum chemistry, one is interested in determining the electronic structure of (possibly large) molecules. In order to make the problem tractable, various approximations have been devised:

Density Matrices

In quantum chemistry, one is interested in determining the electronic structure of (possibly large) molecules. In order to make the problem tractable, various approximations have been devised:

- Wavefunction methods (e.g., Hartree-Fock)

Density Matrices

In quantum chemistry, one is interested in determining the electronic structure of (possibly large) molecules. In order to make the problem tractable, various approximations have been devised:

- Wavefunction methods (e.g., Hartree-Fock)
- Density Functional Theory (e.g., Kohn-Sham)

Density Matrices

In quantum chemistry, one is interested in determining the electronic structure of (possibly large) molecules. In order to make the problem tractable, various approximations have been devised:

- Wavefunction methods (e.g., Hartree-Fock)
- Density Functional Theory (e.g., Kohn-Sham)

In DFT, the electronic density ρ (a scalar field on $\mathbb{R}^3$) is sought, rather than the ground state wavefunction (a scalar field on $\mathbb{R}^{3N}$):

$$\rho(x) = N \int_{\mathbb{R}^{3(N-1)}} |\psi(x, x_2, \ldots, x_N)|^2 dx_2 \ldots dx_N$$

Density Matrices

In quantum chemistry, one is interested in determining the electronic structure of (possibly large) molecules. In order to make the problem tractable, various approximations have been devised:

- Wavefunction methods (e.g., Hartree-Fock)
- Density Functional Theory (e.g., Kohn-Sham)

In DFT, the electronic density ρ (a scalar field on $\mathbb{R}^3$) is sought, rather than the ground state wavefunction (a scalar field on $\mathbb{R}^{3N}$):

$$\rho(x) = N \int_{\mathbb{R}^{3(N-1)}} |\psi(x, x_2, \ldots, x_N)|^2 \mathrm{d}x_2 \ldots \mathrm{d}x_N$$

For large systems, further approximations are necessary. In the LDA (Local Density Approximation) framework, the problem is reduced to the computation of a sequence of density matrices of certain one-electron Hamiltonians (SCF iteration).

Density Matrices

All the statistical properties of a quantum-mechanical system in a given state can be described by a density matrix, i.e., a compact (in fact, trace class) operator P on a Hilbert space $\mathcal{H}$ such that:

1. $0 \preceq P = P^*$
2. $\mathrm{Trace}(P) = 1$ ($\Rightarrow 0 \preceq P \preceq I$)
3. For the ground state, $\mathrm{Trace}(PH) = \langle P, H \rangle_{HS} = \min$

where $H = H^*$ is the Hamiltonian and the minimization takes place over all trace class operators P satisfying conditions 1-2.

Density Matrices

All the statistical properties of a quantum-mechanical system in a given state can be described by a density matrix, i.e., a compact (in fact, trace class) operator P on a Hilbert space $\mathcal{H}$ such that:

1. $0 \preceq P = P^*$
2. $\text{Trace}(P) = 1$ ($\Rightarrow 0 \preceq P \preceq I$)
3. For the ground state, $\text{Trace}(PH) = \langle P, H \rangle_{HS} = \min$

where $H = H^*$ is the Hamiltonian and the minimization takes place over all trace class operators P satisfying conditions 1-2.

For systems in equilibrium, $[H, P] = 0$ and P is a function of the Hamiltonian: $P = f(H)$.

Density Matrices

In practice, the operators are replaced by matrices upon introduction of a set of basis functions $\{\phi\}_{i=1}^{n}$ into the Hilbert space $\mathcal{H}$, where n is a multiple of $N = \#$ of electrons. For simplicity, here we assume an orthonormal basis. The resulting matrices are 'sparse': their pattern/bandwidth is determined by the *range* of the interactions.

Density Matrices

In practice, the operators are replaced by matrices upon introduction of a set of basis functions $\{\phi\}_{i=1}^{n}$ into the Hilbert space $\mathcal{H}$, where n is a multiple of $N = \#$ of electrons. For simplicity, here we assume an orthonormal basis. The resulting matrices are 'sparse': their pattern/bandwidth is determined by the *range* of the interactions.

Once the density matrix is known, one can readily compute:

1. The probability of state ϕ_i , given by P_{ii}

Density Matrices

In practice, the operators are replaced by matrices upon introduction of a set of basis functions $\{\phi\}_{i=1}^{n}$ into the Hilbert space $\mathcal{H}$, where n is a multiple of $N = \#$ of electrons. For simplicity, here we assume an orthonormal basis. The resulting matrices are 'sparse': their pattern/bandwidth is determined by the *range* of the interactions.

Once the density matrix is known, one can readily compute:

1. The probability of state ϕ_i , given by P_{ii}
2. The coherences between states, given by P_{ij} $(i \neq j)$

Density Matrices

In practice, the operators are replaced by matrices upon introduction of a set of basis functions $\{\phi\}_{i=1}^{n}$ into the Hilbert space $\mathcal{H}$, where n is a multiple of $N = \#$ of electrons. For simplicity, here we assume an orthonormal basis. The resulting matrices are 'sparse': their pattern/bandwidth is determined by the *range* of the interactions.

Once the density matrix is known, one can readily compute:

1. The probability of state ϕ_i , given by P_{ii}
2. The coherences between states, given by P_{ij} $(i \neq j)$
3. The expectation of a physical observable: $\langle A \rangle = \text{Trace}(AP)$

Density Matrices

In practice, the operators are replaced by matrices upon introduction of a set of basis functions $\{\phi\}_{i=1}^{n}$ into the Hilbert space $\mathcal{H}$, where n is a multiple of $N = \#$ of electrons. For simplicity, here we assume an orthonormal basis. The resulting matrices are 'sparse': their pattern/bandwidth is determined by the *range* of the interactions.

Once the density matrix is known, one can readily compute:

1. The probability of state ϕ_i , given by P_{ii}
2. The coherences between states, given by P_{ij} $(i \neq j)$
3. The expectation of a physical observable: $\langle A \rangle = \text{Trace}(AP)$
4. The uncertainty (dispersion) of an observable:
   $\Delta A = (\langle A^2 \rangle - \langle A \rangle^2)^{\frac{1}{2}} = [\text{Trace}(A^2 P) - \text{Trace}(AP)^2]^{\frac{1}{2}}$

Density Matrices

In (zero-temperature) electronic structure theory P is, up to a normalization factor, the spectral projector onto the subspace spanned by the N lowest eigenfunctions of H (occupied states):

$$P = \frac{1}{N}\left(\psi_1\psi_1^* + \cdots + \psi_N\psi_N^*\right)$$

where $H\psi_i = \lambda_i\,\psi_i,\ \ i = 1, \ldots, N$.

Density Matrices

In (zero-temperature) electronic structure theory P is, up to a normalization factor, the spectral projector onto the subspace spanned by the N lowest eigenfunctions of H (occupied states):

$$P = \frac{1}{N}\left(\psi_1\psi_1^* + \cdots + \psi_N\psi_N^*\right)$$

where $H\psi_i = \lambda_i\,\psi_i,\ \ i = 1,\ldots,N$. Note that $\mathrm{Trace}(PH) = \lambda_1 + \cdots + \lambda_N$.

Density Matrices

In (zero-temperature) electronic structure theory P is, up to a normalization factor, the spectral projector onto the subspace spanned by the N lowest eigenfunctions of H (occupied states):

$$P = \frac{1}{N}\left(\psi_1\psi_1^* + \cdots + \psi_N\psi_N^*\right)$$

where $H\psi_i = \lambda_i\,\psi_i,\ \ i = 1, \ldots, N$. Note that $\mathrm{Trace}(PH) = \lambda_1 + \cdots + \lambda_N$.

Ignoring the normalization factor, $P = f(H)$ where f is the step function

$$f(x) = \begin{cases} 1 & \text{if} \quad x \le \mu \\ 0 & \text{if} \quad x > \mu \end{cases}$$

with $\lambda_N \le \mu < \lambda_{N+1}$ ("Fermi level").

Density Matrices

If the spectral gap $\gamma = \lambda_{N+1} - \lambda_N$ is not too small, f can be well approximated by the Fermi-Dirac function

$$f(x) = \frac{1}{1 + e^{\beta(x-\mu)}}$$

which tends to a step function as the parameter β increases.

Density Matrices

If the spectral gap $\gamma = \lambda_{N+1} - \lambda_N$ is not too small, f can be well approximated by the Fermi-Dirac function

$$f(x) = \frac{1}{1 + e^{\beta(x-\mu)}}$$

which tends to a step function as the parameter β increases.

For systems at positive temperature ($T > 0$), the density matrix is given by the canonical (Boltzmann) distribution

$$P = e^{-\beta H}/Z, \quad Z = \mathrm{Trace}\left(e^{-\beta H}\right), \quad \text{where } \beta = (\kappa T)^{-1}.$$

Density Matrices

If the spectral gap $\gamma = \lambda_{N+1} - \lambda_N$ is not too small, f can be well approximated by the Fermi-Dirac function

$$f(x) = \frac{1}{1 + e^{\beta(x-\mu)}}$$

which tends to a step function as the parameter β increases.

For systems at positive temperature ($T > 0$), the density matrix is given by the canonical (Boltzmann) distribution

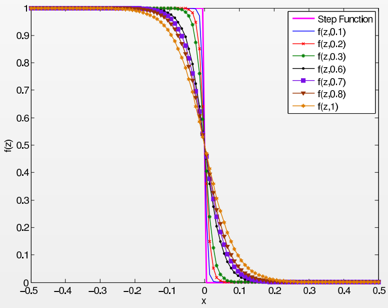
$$P = e^{-\beta H}/Z, \quad Z = \text{Trace}\left(e^{-\beta H}\right), \quad \text{where } \beta = (\kappa T)^{-1}.$$

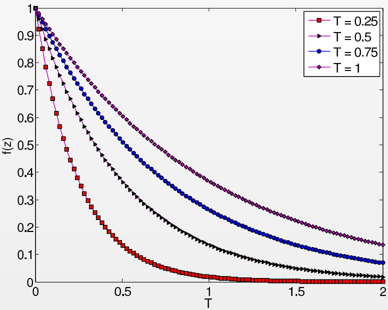
This expression for P is obtained by maximizing the 'von Neumann entropy' $\sigma = -\text{Trace}(P \log P)$ subject to $\text{Trace}(P) = 1$ and $\text{Trace}(HP) = \langle H \rangle$.

Approximations of P

$$f(x) = \frac{1}{1 + e^{\beta(x-\mu)}}$$

$$f(x) = e^{-\frac{x}{\kappa T}}$$





Locality of interactions ('Nearsightedness Principle')

- Physicists have observed long ago that for 'gapped systems' (like insulators), the entries of the density matrix P decay exponentially fast away from the main diagonal. For metallic systems, decay is only algebraic.

Locality of interactions ('Nearsightedness Principle')

- Physicists have observed long ago that for 'gapped systems' (like insulators), the entries of the density matrix P decay exponentially fast away from the main diagonal. For metallic systems, decay is only algebraic.
- This is closely related to the decay of the eigenfunctions corresponding to the occupied states. A classical example is the Anderson model

Locality of interactions ('Nearsightedness Principle')

- Physicists have observed long ago that for 'gapped systems' (like insulators), the entries of the density matrix P decay exponentially fast away from the main diagonal. For metallic systems, decay is only algebraic.
- This is closely related to the decay of the eigenfunctions corresponding to the occupied states. A classical example is the Anderson model
- In the last 10-15 years, this localization property has been exploited to develop 'linear scaling' algorithms for approximating P, i.e., algorithms that asymptotically require $O(n) = O(kN)$ work

Locality of interactions ('Nearsightedness Principle')

- Physicists have observed long ago that for 'gapped systems' (like insulators), the entries of the density matrix P decay exponentially fast away from the main diagonal. For metallic systems, decay is only algebraic.
- This is closely related to the decay of the eigenfunctions corresponding to the occupied states. A classical example is the Anderson model
- In the last 10-15 years, this localization property has been exploited to develop 'linear scaling' algorithms for approximating P, i.e., algorithms that asymptotically require $O(n) = O(kN)$ work
- There are also connections with random matrix theory and with the deflation phenomenon in the Divide-and-Conquer algorithm for the eigenvalues of symmetric tridiagonal matrices; see Trefethen and Bau's *Numerical Linear Algebra*, pp. 232–233 ▸ More

Overview

Sparse approximations of $f(A)$

Most of the literature deals with one of these two problems:

- Computing $f(A)$ for a general matrix A of moderate size

Sparse approximations of $f(A)$

Most of the literature deals with one of these two problems:

- Computing $f(A)$ for a general matrix A of moderate size
- Computing the product $f(A)v$ where A is large and sparse and v is a given vector

Sparse approximations of $f(A)$

Most of the literature deals with one of these two problems:

- Computing $f(A)$ for a general matrix A of moderate size
- Computing the product $f(A)v$ where A is large and sparse and v is a given vector

In some cases, however, we need to approximate $f(A)$ where A can be large and sparse (or banded).

Sparse approximations of $f(A)$

Most of the literature deals with one of these two problems:

- Computing $f(A)$ for a general matrix A of moderate size
- Computing the product $f(A)v$ where A is large and sparse and v is a given vector

In some cases, however, we need to approximate $f(A)$ where A can be large and sparse (or banded).

Example: density matrices $P = f(H)$.

Sparse approximations of $f(A)$

Most of the literature deals with one of these two problems:

- Computing $f(A)$ for a general matrix A of moderate size
- Computing the product $f(A)v$ where A is large and sparse and v is a given vector

In some cases, however, we need to approximate $f(A)$ where A can be large and sparse (or banded).

Example: density matrices $P = f(H)$.

Since we are interested in $\text{Trace}(PA)$ for different A, we need to compute P (to a certain accuracy). Diagonalization costs $O(n^3)$ work and $O(n^2)$ storage $\Rightarrow$ too expensive!

Sparse approximations of $f(A)$

In order to find viable alternatives to diagonalization, we first need to address a fundamental question:

Sparse approximations of $f(A)$

In order to find viable alternatives to diagonalization, we first need to address a fundamental question:

- Question
  If A is sparse, can we expect $f(A)$ to be sparse?

Sparse approximations of $f(A)$

In order to find viable alternatives to diagonalization, we first need to address a fundamental question:

- Question
  If A is sparse, can we expect $f(A)$ to be sparse?
- Answer
  Given an irreducible matrix A it is easy to show, under very mild assumptions on f, that $f(A)$ is structurally full, hence no sparsity is present in $f(A)$. This holds in particular for A^{-1} and $\exp(A)$

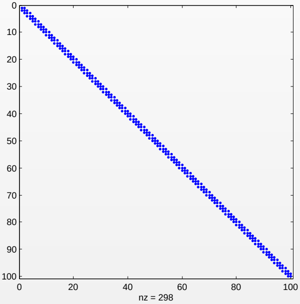
Sparse approximations of $f(A)$

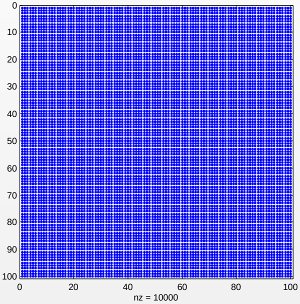
In order to find viable alternatives to diagonalization, we first need to address a fundamental question:

- Question
  If A is sparse, can we expect $f(A)$ to be sparse?
- Answer
  Given an irreducible matrix A it is easy to show, under very mild assumptions on f, that $f(A)$ is structurally full, hence no sparsity is present in $f(A)$. This holds in particular for A^{-1} and $\exp(A)$
- An important goal:
  To investigate the possibility of linear scaling algorithms to approximate $f(A)$ when A is sparse (or banded), and to develop such $O(n)$ methods when appropriate

An example for e^A with A tridiagonal

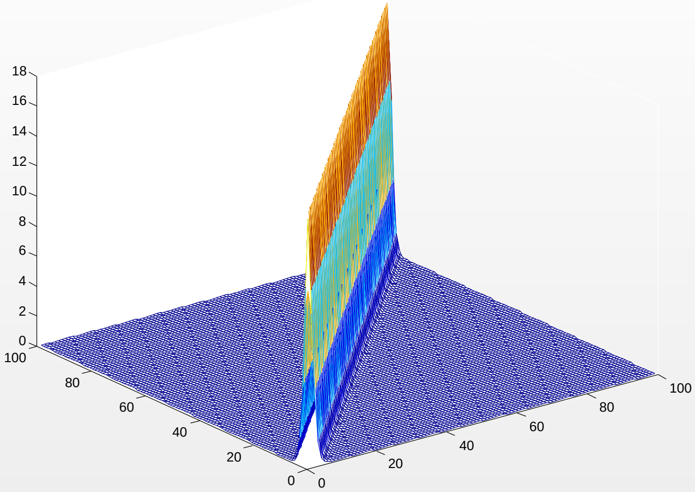
Sparsity pattern of $A = \text{trid}(-1, 2, -1)$ and $e^A = \texttt{expm}(A)$.





An example for e^A with tridiagonal A

$|[e^A]_{ij}|$



A decay result for functions of banded symmetric matrices

Theorem

Let A be a symmetric m-banded matrix and let f be a smooth function on the spectrum of A such that $f(x)$ is real for $x \in \mathbb{R}$. Then there exist $0 < \rho < 1$ and $K = K(f, A)$ such that $|[f(A)]_{ij}| \leq K\rho^{|i-j|}$.

Main ingredients of the proof: approximation theory (Bernstein's Thm.) and the Spectral Theorem.

Also valid for $A \in \mathcal{B}(\ell^2)$ if $f(A) \in \mathcal{B}(\ell^2)$.

M. B. & Gene Golub, Bounds for the entries of matrix functions with applications to preconditioning, BIT, 1999

Brief review of decay results

- In 1984, exponential decay bounds were proved for the inverse of banded, symmetric positive definite matrices by Demko, Moss & Smith; see also Jaffard (1991), Blatov (1996), et al.

Brief review of decay results

- In 1984, exponential decay bounds were proved for the inverse of banded, symmetric positive definite matrices by Demko, Moss & Smith; see also Jaffard (1991), Blatov (1996), et al.
- In 1999, B. & Golub proved the above-mentioned decay bound for $f(A)$ with A banded, symmetric

Brief review of decay results

- In 1984, exponential decay bounds were proved for the inverse of banded, symmetric positive definite matrices by Demko, Moss & Smith; see also Jaffard (1991), Blatov (1996), et al.
- In 1999, B. & Golub proved the above-mentioned decay bound for $f(A)$ with A banded, symmetric
- In 2000, Iserles proved decay results for the exponential of band matrices

Brief review of decay results

- In 1984, exponential decay bounds were proved for the inverse of banded, symmetric positive definite matrices by Demko, Moss & Smith; see also Jaffard (1991), Blatov (1996), et al.
- In 1999, B. & Golub proved the above-mentioned decay bound for $f(A)$ with A banded, symmetric
- In 2000, Iserles proved decay results for the exponential of band matrices
- In 2005, Del Buono, Lopez & Peluso proved decay bounds for functions of banded skew-symmetric matrices

Brief review of decay results

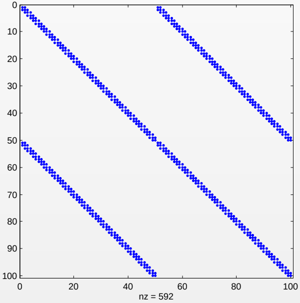
- In 1984, exponential decay bounds were proved for the inverse of banded, symmetric positive definite matrices by Demko, Moss & Smith; see also Jaffard (1991), Blatov (1996), et al.
- In 1999, B. & Golub proved the above-mentioned decay bound for $f(A)$ with A banded, symmetric
- In 2000, Iserles proved decay results for the exponential of band matrices
- In 2005, Del Buono, Lopez & Peluso proved decay bounds for functions of banded skew-symmetric matrices
- In 2006, extensions of the B.-Golub bounds to sparse Hermitian matrices appeared in the quantum computing literature

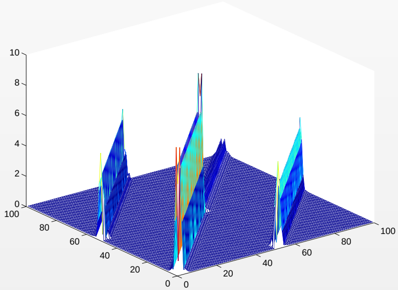
Brief review of decay results

- In 1984, exponential decay bounds were proved for the inverse of banded, symmetric positive definite matrices by Demko, Moss & Smith; see also Jaffard (1991), Blatov (1996), et al.
- In 1999, B. & Golub proved the above-mentioned decay bound for $f(A)$ with A banded, symmetric
- In 2000, Iserles proved decay results for the exponential of band matrices
- In 2005, Del Buono, Lopez & Peluso proved decay bounds for functions of banded skew-symmetric matrices
- In 2006, extensions of the B.-Golub bounds to sparse Hermitian matrices appeared in the quantum computing literature
- Further extension to non-normal matrices by B. & Razouk in 2007

Decay for exponential of a sparse Hamiltonian matrix

Sparsity pattern of a $2n \times 2n$ Hamiltonian matrix A and decay in $\exp(A)$.

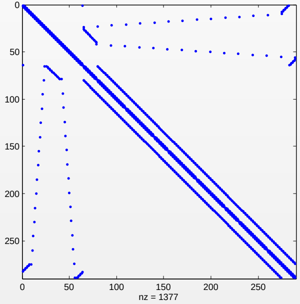


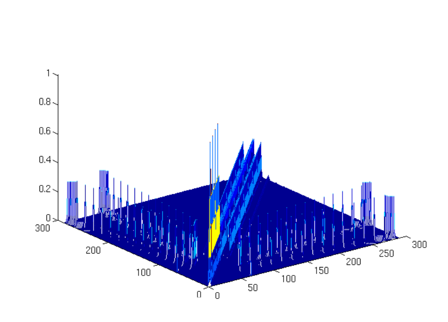


Note that $\exp(A)$ is symplectic. Also, here A is non-normal.

Decay for logarithm of a sparse matrix

Sparsity pattern of $A =$ `mesh3e1` (from NASA) and decay in $\log(A)$.

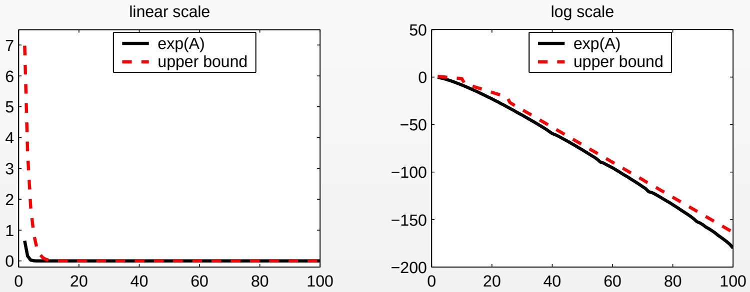




Here A is symmetric positive definite.

Assessment of the bound for A banded Hermitian

Upper bounds vs. $|[e^A]_{ij}|$ first row.



Overview

Sufficient conditions for $O(n)$ approximation of $f(A)$

- Let $\{A_n\}$ be a sequence of $n \times n$ Hermitian matrices such that there is a closed interval $\mathcal{I}$ with the property that $\sigma(A_n) \subset \mathcal{I}$ for all n

Sufficient conditions for $O(n)$ approximation of $f(A)$

- Let $\{A_n\}$ be a sequence of $n \times n$ Hermitian matrices such that there is a closed interval $\mathcal{I}$ with the property that $\sigma(A_n) \subset \mathcal{I}$ for all n
- Assume that $\{A_n\}$ has bandwidth m independent of n

Sufficient conditions for $O(n)$ approximation of $f(A)$

- Let $\{A_n\}$ be a sequence of $n \times n$ Hermitian matrices such that there is a closed interval $\mathcal{I}$ with the property that $\sigma(A_n) \subset \mathcal{I}$ for all n
- Assume that $\{A_n\}$ has bandwidth m independent of n
- Let f be a function analytic on a neighborhood of $\mathcal{I}$

Sufficient conditions for $O(n)$ approximation of $f(A)$

- Let $\{A_n\}$ be a sequence of $n \times n$ Hermitian matrices such that there is a closed interval $\mathcal{I}$ with the property that $\sigma(A_n) \subset \mathcal{I}$ for all n
- Assume that $\{A_n\}$ has bandwidth m independent of n
- Let f be a function analytic on a neighborhood of $\mathcal{I}$
- Assume furthermore that the spectrum of $\{A_n\}$ remains bounded away from the singularities of f as $n \to \infty$

Sufficient conditions for $O(n)$ approximation of $f(A)$

- Let $\{A_n\}$ be a sequence of $n \times n$ Hermitian matrices such that there is a closed interval $\mathcal{I}$ with the property that $\sigma(A_n) \subset \mathcal{I}$ for all n
- Assume that $\{A_n\}$ has bandwidth m independent of n
- Let f be a function analytic on a neighborhood of $\mathcal{I}$
- Assume furthermore that the spectrum of $\{A_n\}$ remains bounded away from the singularities of f as $n \to \infty$
- Then there is an $\hat{m}$ such that $f(A_n)$ can be uniformly approximated by the truncated matrix $[f(A_n)]^{(\hat{m})}$ for all n

Sufficient conditions for $O(n)$ approximation of $f(A)$

- Let $\{A_n\}$ be a sequence of $n \times n$ Hermitian matrices such that there is a closed interval $\mathcal{I}$ with the property that $\sigma(A_n) \subset \mathcal{I}$ for all n
- Assume that $\{A_n\}$ has bandwidth m independent of n
- Let f be a function analytic on a neighborhood of $\mathcal{I}$
- Assume furthermore that the spectrum of $\{A_n\}$ remains bounded away from the singularities of f as $n \to \infty$
- Then there is an $\hat{m}$ such that $f(A_n)$ can be uniformly approximated by the truncated matrix $[f(A_n)]^{(\hat{m})}$ for all n

- The result holds for any sparsity pattern of $\{A_n\}$ (independent of n)

Approximation of $f(A)$ by polynomials

Algorithm ▸ More

- We compute approximations of $f(A)$ using Chebyshev polynomials
  - The degree of the polynomial can be estimated a priori
  - The coefficients of the polynomial can be pre-computed (indep. of n)
  - Estimates for the extreme eigenvalues of A are required
- The polynomial expansion is combined with a procedure that a priori determines a bandwidth or sparsity pattern for $f(A)$ outside which the elements are so small that they can be neglected

Approximation of $f(A)$ by polynomials

Algorithm ▸ More

- We compute approximations of $f(A)$ using Chebyshev polynomials
  - The degree of the polynomial can be estimated a priori
  - The coefficients of the polynomial can be pre-computed (indep. of n)
  - Estimates for the extreme eigenvalues of A are required
- The polynomial expansion is combined with a procedure that a priori determines a bandwidth or sparsity pattern for $f(A)$ outside which the elements are so small that they can be neglected

Cost

This method is multiplication-rich; the matrices are kept sparse throughout the computation, hence $O(n)$ arithmetic and storage requirements. Matrix polynomials are evaluated with the classical Paterson-Stockmeyer algorithm.

Decay bounds for the Fermi-Dirac approximation

Assume that H is m-banded and has spectrum in $[-1, 1]$, then

$$\left| \left[\left(I + e^{\beta(H-\mu I)} \right)^{-1} \right]_{ij} \right| \le K(\gamma)\rho(\gamma)^{\frac{2|i-j|}{m}}.$$

Decay bounds for the Fermi-Dirac approximation

Assume that H is m-banded and has spectrum in $[-1, 1]$, then

$$\left| \left[\left(I + e^{\beta(H-\mu I)} \right)^{-1} \right]_{ij} \right| \leq K(\gamma)\rho(\gamma)^{\frac{2|i-j|}{m}}.$$

Note that β depends on γ and on the desired accuracy. Furthermore, if

$$\gamma \to 0 \quad \text{then} \quad \rho(\gamma) \to 1$$

and if

$$\gamma \to 1 \quad \text{then} \quad \rho(\gamma) \to 0.872.$$

We choose β and $\hat{m}$ so as to guarantee an accuracy $\|P - f(H)\|_2 < 10^{-6}$.

Decay bounds for the Fermi-Dirac approximation

Assume that H is m-banded and has spectrum in $[-1, 1]$, then

$$\left| \left[\left(I + e^{\beta(H - \mu I)} \right)^{-1} \right]_{ij} \right| \le K(\gamma) \rho(\gamma)^{\frac{2|i-j|}{m}}.$$

Note that β depends on γ and on the desired accuracy. Furthermore, if

$$\gamma \to 0 \quad \text{then} \quad \rho(\gamma) \to 1$$

and if

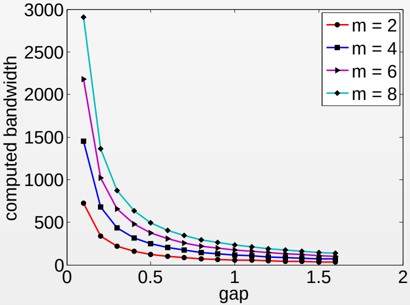
$$\gamma \to 1 \quad \text{then} \quad \rho(\gamma) \to 0.872.$$

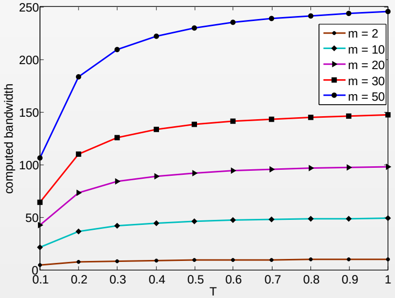
We choose β and $\hat{m}$ so as to guarantee an accuracy $\|P - f(H)\|_2 < 10^{-6}$.

Remark: The above bound only depends on m and γ.

Computed bandwidth for approximations of P

$$f(x) = \frac{1}{1 + e^{\beta(x-\mu)}} \qquad\qquad f(x) = e^{-\frac{x}{\kappa T}}$$





Significance of our decay bounds for $O(N)$ scaling

We quote a passage from Claude Le Bris (2005):

Significance of our decay bounds for $O(N)$ scaling

We quote a passage from Claude Le Bris (2005):

> *In order to make alternatives to diagonalization practical (...) an algorithm is constructed, which might scale cubically in the whole generality, but scales linearly when H is sparse and when the density matrix P to be determined is assumed to be sparse. The latter assumption is in some sense an* a posteriori *assumption, and not easy to analyse... It is to be emphasized that the numerical analysis of the linear scaling methods overviewed above that would account for cut-off rules and locality assumptions, is not yet available.*

Significance of our decay bounds for $O(N)$ scaling

We quote a passage from Claude Le Bris (2005):

> *In order to make alternatives to diagonalization practical (...) an algorithm is constructed, which might scale cubically in the whole generality, but scales linearly when H is sparse and when the density matrix P to be determined is assumed to be sparse. The latter assumption is in some sense an* a posteriori *assumption, and not easy to analyse... It is to be emphasized that the numerical analysis of the linear scaling methods overviewed above that would account for cut-off rules and locality assumptions, is not yet available.*

Our bounds, depending only on the interaction range m and on the spectral gap γ, are *a priori* and provide a justification of linear scaling algorithms. However, some estimate of γ is needed.

Overview

Chebyshev expansion

Some results for A_n tridiagonal, SPD

	$A \log(A)$	$\text{Trace}[A \log(A)]$		
n	rel. error	error	$\hat{m}$	k
100	$5e{-}07$	$3e{-}04$	20	9
200	$6e{-}07$	$8e{-}04$	20	9
300	$1e{-}07$	$3e{-}04$	20	10
500	$2e{-}07$	$5e{-}04$	20	10

In the Table, $\hat{m}$ is the estimated bandwidth and k is the number of terms in the Chebyshev expansion. Note the $O(n)$ behavior in terms of cost.

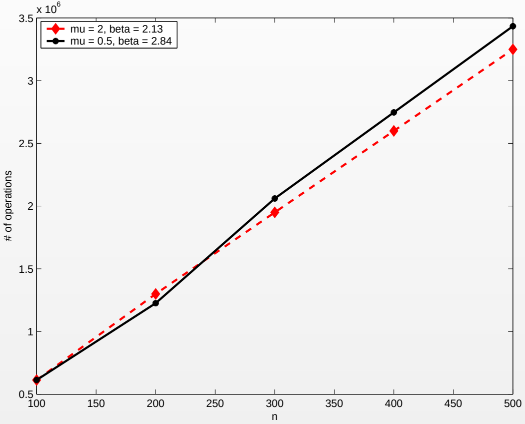
Density matrix computation (toy example)

The bandwidth was computed prior to the calculation to be ≈ 20; here H is tridiagonal (1D Anderson model).

Table: Results for $f(x) = \frac{1}{1+\mathrm{e}^{(\beta(x-\mu))}}$

	$\mu = 2,\ \beta = 2.13$			$\mu = 0.5,\ \beta = 1.84$		
n	error	k	$\hat{m}$	error	k	$\hat{m}$
100	$9e{-}06$	18	20	$6e{-}06$	18	22
200	$4e{-}06$	19	20	$9e{-}06$	18	22
300	$4e{-}06$	19	20	$5e{-}06$	20	22
400	$6e{-}06$	19	20	$8e{-}06$	20	22
500	$8e{-}06$	19	20	$8e{-}06$	20	22

Density matrix computation



The $O(n)$ behavior of Chebyshev's approximation to the Fermi–Dirac function $f(H) = (\exp(\beta(H - \mu I)) + I)^{-1}$.

Summary

- 'Gapped' systems, like insulators, exhibit strong localization

Summary

- 'Gapped' systems, like insulators, exhibit strong localization
- Localization in $f(A)$, when present, can lead to fast approximation algorithms

Summary

- 'Gapped' systems, like insulators, exhibit strong localization
- Localization in $f(A)$, when present, can lead to fast approximation algorithms
- Our exponential decay bounds for density matrices depend only on the parameters m and γ

Summary

- 'Gapped' systems, like insulators, exhibit strong localization
- Localization in $f(A)$, when present, can lead to fast approximation algorithms
- Our exponential decay bounds for density matrices depend only on the parameters m and γ
- These bounds can be useful in determining appropriate sparsity patterns (or bandwidths) that capture the 'important' entries in $f(A)$

Summary

- 'Gapped' systems, like insulators, exhibit strong localization
- Localization in $f(A)$, when present, can lead to fast approximation algorithms
- Our exponential decay bounds for density matrices depend only on the parameters m and γ
- These bounds can be useful in determining appropriate sparsity patterns (or bandwidths) that capture the 'important' entries in $f(A)$
- Chebyshev approximations need estimates of the extremal eigenvalues

Summary

- 'Gapped' systems, like insulators, exhibit strong localization
- Localization in $f(A)$, when present, can lead to fast approximation algorithms
- Our exponential decay bounds for density matrices depend only on the parameters m and γ
- These bounds can be useful in determining appropriate sparsity patterns (or bandwidths) that capture the 'important' entries in $f(A)$
- Chebyshev approximations need estimates of the extremal eigenvalues
- Extension to non-normal case possible

Overview

Some open problems

- Tighter bounds?

Some open problems

- Tighter bounds?
- Better $O(n)$ algorithms?

Some open problems

- Tighter bounds?
- Better $O(n)$ algorithms?
- How to deal with metallic systems ($\gamma \to 0$ as $n \to \infty$)?

Some open problems

- Tighter bounds?
- Better $O(n)$ algorithms?
- How to deal with metallic systems ($\gamma \to 0$ as $n \to \infty$)?
  1. Wavelets?

Some open problems

- Tighter bounds?
- Better $O(n)$ algorithms?
- How to deal with metallic systems ($\gamma \to 0$ as $n \to \infty$)?
  1. Wavelets?
  2. Hierarchical matrices? Semiseparable?

Some open problems

- Tighter bounds?
- Better $O(n)$ algorithms?
- How to deal with metallic systems ($\gamma \to 0$ as $n \to \infty$)?
  1. Wavelets?
  2. Hierarchical matrices? Semiseparable?
  3. Other bases?

Some open problems

- Tighter bounds?
- Better $O(n)$ algorithms?
- How to deal with metallic systems ($\gamma \to 0$ as $n \to \infty$)?
  1. Wavelets?
  2. Hierarchical matrices? Semiseparable?
  3. Other bases?
- How to exploit structure? Lie group/algebra, Toeplitz, etc.

Some open problems

- Tighter bounds?
- Better $O(n)$ algorithms?
- How to deal with metallic systems ($\gamma \to 0$ as $n \to \infty$)?
  1. Wavelets?
  2. Hierarchical matrices? Semiseparable?
  3. Other bases?
- How to exploit structure? Lie group/algebra, Toeplitz, etc.
- Rigorous error analysis? What if spectrum is not well-estimated?

Some open problems

- Tighter bounds?
- Better $O(n)$ algorithms?
- How to deal with metallic systems ($\gamma \to 0$ as $n \to \infty$)?
  1. Wavelets?
  2. Hierarchical matrices? Semiseparable?
  3. Other bases?
- How to exploit structure? Lie group/algebra, Toeplitz, etc.
- Rigorous error analysis? What if spectrum is not well-estimated?
- Rational approximations? See Sidje & Saad (2008)

Some open problems

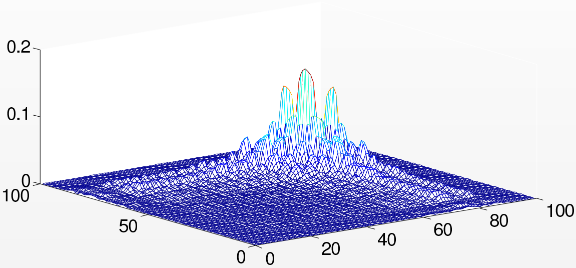
- Tighter bounds?
- Better $O(n)$ algorithms?
- How to deal with metallic systems ($\gamma \to 0$ as $n \to \infty$)?
  1. Wavelets?
  2. Hierarchical matrices? Semiseparable?
  3. Other bases?
- How to exploit structure? Lie group/algebra, Toeplitz, etc.
- Rigorous error analysis? What if spectrum is not well-estimated?
- Rational approximations? See Sidje & Saad (2008)
- Software for $O(n)$ approximations?

Some open problems

- Tighter bounds?
- Better $O(n)$ algorithms?
- How to deal with metallic systems ($\gamma \to 0$ as $n \to \infty$)?
  1. Wavelets?
  2. Hierarchical matrices? Semiseparable?
  3. Other bases?
- How to exploit structure? Lie group/algebra, Toeplitz, etc.
- Rigorous error analysis? What if spectrum is not well-estimated?
- Rational approximations? See Sidje & Saad (2008)
- Software for $O(n)$ approximations?

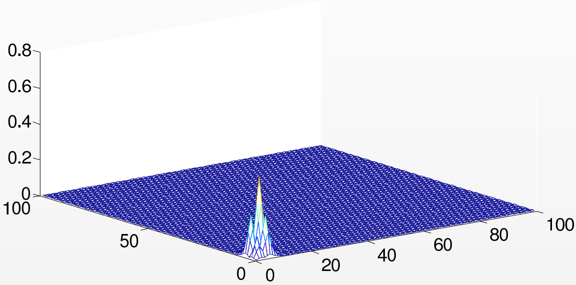
An excellent reference: C. LeBris, *Computational Chemistry from the Perspective of Numerical Analysis*, Acta Numerica 14 (2005), 363-444.

Localization in spectral projectors: small gap



Rank-one spectral projector for $A = A^T$ tridiagonal. Relative gap $\gamma = 10^{-3}$. Note the slow decay and oscillatory behavior.

Localization in spectral projectors: large gap



Rank-one spectral projector for $A = A^T$ tridiagonal. Relative gap $\gamma = 0.5$.

Back

Chebyshev approximation

For A with $\sigma(A) \subset [-1, 1]$ the Chebyshev polynomials are given by

$$T_{k+1}(A) = 2AT_k(A) - T_{k-1}(A),\ T_1(A) = A,\ T_0(A) = I.$$

Then $f(A)$ can be represented in a series of the form

$$f(A) = \sum_{k=0}^{\infty} c_k T_k(A).$$

The coefficients of the expansion are given by

$$c_k \approx \frac{2}{M} \sum_{j=1}^{M} f(\cos(\theta_j)) \cos((k-1)\theta_j),$$

where $\theta_j = \pi(j - \frac{1}{2})/M$. Back

The n-independence of the error

The Nth truncation error without dropping can be written as

$$\|e_N(A)\| = \|f(A) - \sum_{k=0}^{N} c_k T_k(A)\|.$$

For x in $[-1,1]$ we have that $|T_k(x)| \leq 1$ for $k = 1, 2, \ldots$. Then

$$\|e_N(A)\| = \| \sum_{k=N+1}^{\infty} c_k T_k(A)\| \leq \sum_{k=N+1}^{\infty} |c_k|.$$

Back

A Theorem of Bernstein

The set of Faber polynomials can be used to obtain a uniform approximation to an analytic function f with a sequence of polynomials of bounded degree, i.e.,

$$|f(z) - \Pi_N(z)| < cq^N \quad (0 < q < 1)$$

for all $z \in F$, where c and q depend on the analytic properties of f.

A Theorem of Bernstein

The set of Faber polynomials can be used to obtain a uniform approximation to an analytic function f with a sequence of polynomials of bounded degree, i.e.,

$$|f(z) - \Pi_N(z)| < cq^N \quad (0 < q < 1)$$

for all $z \in F$, where c and q depend on the analytic properties of f.

Example – Disk

If the region is a disk of radius ρ centered at z_0, then for any function f analytic on the disk of radius ρ/q centered at z_0, where $0 < q < 1$, there exists a polynomial Π_N of degree at most N and a positive constant c such that

$$|f(z) - \Pi_N(z)| < cq^N,$$

for all $z \in F$.

Back