Interpolative Decomposition and its Applications in Quantum Chemistry

Lexing Ying
Department of Mathematics
Stanford University

Joint work with Anil Damle, Lin Lin, Jianfeng Lu

June 19-21, 2018, CSCAMM, UMD
Mathematical and Numerical Aspects of Quantum Dynamics
Outline

▶ What is interpolative decomposition?
  ▶ Some theory
  ▶ How to compute it
▶ Kohn-Sham density functional theory
▶ Applications
  ▶ Localization
  ▶ Electron repulsion integral tensor
Low-rank approximation

Given a matrix $A \in \mathbb{R}^{m \times n}$, a low-rank approximation of $A$ is

$$A \approx BC, \quad B \in \mathbb{R}^{m \times k}, C \in \mathbb{R}^{k \times n}, k \ll \min(m, n).$$

Example: singular value decomposition $A \approx (US)V^T$.

Applications
- Principal component analysis,
- Signal processing (compression, denoising, ...)
- Fast numerical linear algebra
- Sparse recovery (collaborative filtering)
Interpolative decomposition (ID)

- A low-rank approximation that uses $A$’s own columns.

- The picked columns are called the skeletons.
- Let $P$ be the permutation matrix moving the skeletons to the front.

$$AP \approx (AP)_{(:,1:k)} \begin{bmatrix} I & T \end{bmatrix}, \quad T \in \mathbb{R}^{k \times (n-k)}.$$

- Key advantages: keep the columns of $A$ in the approximation
  - Reuse the entries of $A$ (save space),
  - Keep structure of the columns.
Theory

Theorem (Gu+Eisenstat, Tyrtyshnikov)

For fixed $k$, there exists $P$ and $T$ s.t.

- $\|AP - (AP)(:,1:k) [I \ T]\|_2 \leq \sqrt{1 + k(n - k)\sigma_{k+1}(A)}$,  
- $|T_{ij}| \leq 1$.

Proof.

- Find $k$ columns of $A$ that span the maximal volume. This implies that $|T_{ij}| \leq 1$.
- Build a QR decomposition based on these columns to derive the error bound.

- $\sigma_{k+1}(A)$ is the best approximation result from SVD. ID has an extra $\sqrt{nk}$ factor (in the worst case).
- Complexity: combinatorial search, exponential cost.
Theory

Theorem (Gu+Eisenstat)

For fixed $k$ and fixed $f > 1$, there exists $P$ and $T$ s.t.

1. $\|AP - (AP)(:,:,1:k) [I \ T]\|_2 \leq \sqrt{1 + f^2 k(n-k)\sigma_{k+1}(A)},$
2. $|T_{ij}| \leq f,$

and it can be found in $O((m + n \log_f n)n^2)$ steps.

Proof.

1. Iteratively improve the column selection by finding the largest entry $T_{ij}$ with $|T_{ij}| > f$.
2. Number of iterations bounded by $O(\log_f n^{k/2}) = O(k \log_f n)$.
3. Efficient routines for updating the factorization once a new column is picked.

- Approx. error has an extra $\sqrt{f^2 nk}$ factor compared to SVD.
- Complexity: for $f = O(n^\alpha)$, the cost is cubic $O(mn \min(m, n))$. 
In practice

- QR with column pivoting (QRCP).
  - A greedy heuristic for maximizing the volume of the picked columns one by one.

\[
AP = QR = [Q_1 \quad Q_2] \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \approx Q_1 \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} \\
\approx (Q_1 R_{11}) \begin{bmatrix} I & R_{11}^{-1} R_{12} \end{bmatrix} := (AP)_{(:,1:k)} \begin{bmatrix} I & T \end{bmatrix}.
\]

- Almost the same cost of standard QR: \( O(mnk) \).
- No guarantee for a bound on \( |T_{ij}| \) but works well in most cases.
In practice

- Randomized approach if $\min(m,n) \gg k$
  - Project the columns (via randomized Fourier transform [Ailon-Chazelle-2009]) to a random $O(k)$ dimensional subspace.
  - Apply QRCP to the projected (fat) matrix.

- Benefits
  - Reuse the entries of $A$ (save space),
  - Inherit the structure of the columns: sparsity, locality, factorized form.
Extension

- Two-sided interpolative decomposition [Cheng et al.-2006]

\[ P^T AQ \approx \begin{bmatrix} I & T_1^T \end{bmatrix} (P^T AQ)_{(1:k,1:k)} \begin{bmatrix} I & T_2 \end{bmatrix}. \]

- Apply QRCP to both the rows and the columns.
- Can be combined with the randomization.

---

- In what follows, we assume that the columns are already in the correct order for ID (i.e., \( P = I \))

\[
A \approx A_{(:,1:k)} \begin{bmatrix} I & T \end{bmatrix} \\
A \approx \begin{bmatrix} I \\ T_1^T \end{bmatrix} A_{(1:k,1:k)} \begin{bmatrix} I & T_2 \end{bmatrix}.
\]
Related but different approaches

- Column/row sampling with leverage scores (Clarkson, Drineas, Kannan, Mahoney, Woodruff, ...)
- CUR decomposition (Tyrtyshnikov, Hackbusch, ...)
- Non-negative factorization (...)
- Nyström interpolation (...)

Outline

- What is interpolative decomposition?
  - Some theory
  - How to compute it
- Kohn-Sham density functional theory
- Applications in quantum chemistry
  - Localization
  - Electron repulsion integral tensor
Consider a quantum system with $N_e$ electrons. The many-body Schrödinger equation for the ground state

$$H \Psi = \left( \sum_{i=1}^{N_e} -\Delta x_i + \sum_{i<j} \frac{1}{|x_i - x_j|} - \sum_{i,\alpha} \frac{M_\alpha}{|x_i - z_\alpha|} \right) \Psi, \quad i\partial_t \Psi = H \Psi$$

where $\Psi = \Psi(x_1, \ldots, x_{N_e})$ and nuclei at $\{z_\alpha\}$ with charge $\{M_\alpha\}$.

**Ground state:** $H \Psi_0 = E_0 \Psi_0$

- $E_0$ is the lowest eigenvalue. $E_0 = \inf_{\|\Phi\|=1} \langle \Phi | H | \Phi \rangle$.
- $\Psi_0 = \Psi_0(x_1, \ldots, x_{N_e})$ is the lowest eigenfunction.

**Density** $\rho(x) = N_e \int |\Psi_0(x, x_2, \ldots, x_{N_e})|^2 dx_2,\ldots,N_e$.

**High-dimensional problem and curse of dimensionality.**
Kohn-Sham DFT

\[ H\Psi = \left( \sum_{i=1}^{N_e} -\Delta x_i + \sum_{i<j} \frac{1}{|x_i - x_j|} - \sum_{i,\alpha} \frac{M_\alpha}{|x_i - z_\alpha|} \right) \Psi, \quad i\partial_t \Psi = H\Psi \]

- Kohn-Sham density function theory: 3D nonlinear problem

\[( -\Delta + V[\rho]) \psi_i = \lambda_i \psi_i, \quad i = 1, \ldots, N_e.\]

\(\lambda_1, \ldots, \lambda_{N_e}\) are the smallest \(N_e\) eigenvalues.

\(\psi_1, \ldots, \psi_{N_e}\) are the Kohn-Sham (KS) orbitals (eigenfunctions)

\[\rho(x) = \sum_{i=1}^{N_e} |\psi_i(x)|^2.\]

- Nonlinear eigenvalue problem: self consistent iteration

\[\rho(x) \Rightarrow V[\rho](x) \Rightarrow \{\psi_i(x)\}_{1 \leq i \leq N_e} \Rightarrow \rho(x) \Rightarrow V[\rho](x) \Rightarrow \ldots\]

- At the end, we hold converged \(\{\psi_i(x)\}_{1 \leq i \leq N_e}\) and \(\rho(x)\).
Outline

- What is interpolative decomposition?
  - Some theory
  - How to compute it
- Kohn-Sham density functional theory
- Applications in quantum chemistry
  - Localization
  - Electron repulsion integral tensor
Kohn-Sham orbitals and subspace

\[
\begin{aligned}
\begin{cases}
(-\Delta + V[\rho]) \psi_i = \lambda_i \psi_i, & i = 1, \ldots, N_e \\
\rho(x) = \sum_{i=1}^{N_e} |\psi_i(x)|^2.
\end{cases}
\end{aligned}
\]

\[\rho(x) \Rightarrow V[\rho](x) \Rightarrow \{\psi_i(x)\}_{1 \leq i \leq N_e} \Rightarrow \rho(x) \Rightarrow V[\rho](x) \Rightarrow \ldots\]

- This is the starting point of computing other physical quantities.
- Many such quantities depend only on the subspace spanned by \(\{\psi_i(x)\}_{1 \leq i \leq N_e}\).
  
  e.g. \(\rho = \text{diag}(\Psi \Psi^T), \quad \Psi = [\psi_1 \ldots \psi_{N_e}], \quad \Psi \Psi^T = \text{proj. op.}\)

- Seek for a sparse and localized basis for span \(\Psi\)
  
  - Interpretability.
  
  - Computation and storage efficiency.
Localization [with A. Damle and L. Lin]

> Given $\Psi = [\psi_1, \ldots, \psi_{N_e}] \in \mathbb{R}^{n \times N_e}$, find $R \in \mathbb{R}^{N_e \times N_e}$ such that

$$\Phi = [\phi_1, \ldots, \phi_{N_e}] = \Psi R$$

has localized and well-conditioned columns.

> Assumption: working with insulators so such a basis exists.

> Requirements
  > 2. Orthogonal (at least well-conditioned).
Previous work

- Maximally localized Wannier functions [Marzari+Vanderbilt].
\[
\min_{R \in SO_k} \sum_{i=1}^{N_e} \left( \int r^2 |\psi_i^R(r)|^2 dr - \left( \int r |\psi_i^R(r)|^2 dr \right)^2 \right).
\]

- Challenges
  - Non-convex optimization problem.
  - Needs smart initial guess.
Density matrix

- Requirement 1: sparse/localized
- Idea: consider the density matrix (projector) \( Z = \Psi \Psi^T \).
- For insulators, \( Z \) has localized and sparse columns.

Instead of using arbitrary columns for \( R \), only look for selected columns of the density matrix (SCDM).
Interpolative decomposition

- Requirement 2: orthogonal or at least well-conditioned
- Apply ID to $\Psi^T$: (let $C$ be the picked columns)

$$R := (\Psi^T)_{(:,C)}, \quad \Phi = \Psi R = \Psi (\Psi^T)_{(:,C)}.$$  

- For orthonormal $\Phi$, the QRCP gives $\Psi^T P = Q \begin{bmatrix} R_{11} & R_{12} \end{bmatrix}$:

$$\Phi = \Psi Q.$$
SCDM example: silicon and water

Yellow: Silicon crystal structure
Red: isosurface of localized orbital

Red and white: water molecule structure. Blue and yellow: isosurface of localized orbital

Si

Water

\[(\text{average } n_z^2)^{1/3}\]

- Orthogonalized SCDM
- SCDM

Relative truncation value, \(\log_{10} \varepsilon\)

- 0.35
- 0.3
- 0.25
- 0.2
- 0.15
- 0.1
- 0.05

- 0.3
- 0.25
- 0.2
- 0.15
- 0.1
- 0.05
Outline

- What is interpolative decomposition?
  - Some theory
  - How to compute it
- Kohn-Sham density functional theory
- Applications in quantum chemistry
  - Localization
  - Electron repulsion integral tensor
Electron repulsion integral tensor [with J. Lu]

- Given a set \( \{\psi_i(x)\}_{1 \leq i \leq N_e} \) of basis functions.
- Electron repulsion integral tensor \( R \in \mathbb{R}^{N_e^2 \times N_e^2} \)

\[
R_{ij,kl} = \int \int \psi_i(x)\psi_j(x) \frac{1}{|x-y|} \psi_k(y)\psi_l(y) \, dx \, dy.
\]

(e.g. \( N_e = 10^3 \) and \( n = 10^6 \)).

- Introduce \( W \in \mathbb{R}^{N_e^2 \times n} \) and \( G \in \mathbb{R}^{n \times n} \) with
  \[
  W_{ij,x} = \psi_i(x)\psi_j(x) \quad \text{and} \quad G_{x,y} = \frac{1}{|x-y|}
  \]

\[
R = WGW^T.
\]

Goal: compute and represent \( R = (R_{i,j,k,l}) \) efficiently.
Interpolative separable density fitting (ISDF)

- Each column of $W$ is separable (i.e. an outer-product)

$$W_{ij,x} = \psi_i(x)\psi_j(x)$$

- In most cases $W$ is numerically low-rank.
- Consider $\psi_i(x) = \exp\left(2\pi \sqrt{-1}ix\right)$,

$$\psi_i(x)\psi_j(x) = \exp\left(2\pi \sqrt{-1}(i+j)x\right).$$

- Only $O(N_e)$ choices for $(i+j)$ for $N_e^2$ combinations of $i$ and $j$.
- Idea: $W$ has rank $O(N_e)$. Apply ID to compress $W$. 

![Diagram](image)
Interpolative separable density fitting (ISDF)

- Given an interpolative decomposition of $W$

- Then $R = WGW^T$ has approximation

Middle matrix products done with FFTs.

- Costs: $O(N_e^2)$ storage and $O(N_e^2 n)$ time
How to compute the ID of $W$

- Size of $W$: $N_e^2 \times n$
- Naive randomized ID costs at lest $O(N_e^3 n)$
- Idea: use the separable (outer-product) structure of $W$’s columns
- Reshape $W$ from $N_e^2 \times n$ to $N_e \times N_e \times n$

- Randomized ID with random project in each dimension of size $N_e$
- Cost: $O(N_e^2 n)$ time

- So total cost of ISDF: $O(N_e^2)$ storage and $O(N_e^2 n \log n)$ time
Thank you

- Email: lexing@stanford.edu
- URL: http://web.stanford.edu/~lexing
- Research supported by NSF and DOE