

RANDOMIZED NUMERICAL LINEAR ALGEBRA FOR LARGE-SCALE MATRIX DATA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Kun Dong

August 2019

© 2019 Kun Dong

ALL RIGHTS RESERVED

RANDOMIZED NUMERICAL LINEAR ALGEBRA FOR LARGE-SCALE

MATRIX DATA

Kun Dong, Ph.D.

Cornell University 2019

Your abstract goes here. Make sure it sits inside the brackets. If not, your biosketch page may not be roman numeral iii, as required by the graduate school.

BIOGRAPHICAL SKETCH

Kun Dong was born in Shaoxing, Zhejiang Province, China on September 26, 1992 to Jian Dong and Chamei Sang. Kun became interested in puzzles and mathematics in early elementary school, and participated in mathematics competitions until the end of high school.

In 2008, Kun moved to Newmarket, Ontario, Canada and attended Sir William Mulock Secondary School. He spent the next two years learning the new language and culture.

After graduation, Kun attended the University of California, Los Angeles in 2010 to study applied mathematics. During the summers of 2012 and 2013, he participated in the UCLA Applied Math REU program to work on dynamical system and crime modeling. He received tremendous help and encouragement on his way to graduate school from his REU mentors, Scott McCalla and James von Brecht. In 2014, Kun completed the departmental scholar program in mathematics, earning his B.S. and M.A. concurrently. He graduated Summa Cum Laude, receiving the departmental highest honor in applied mathematics and the Daus memorial award for his achievement in mathematics as an undergraduate.

In 2014, Kun was admitted to the Ph.D. program at Center for Applied Mathematics in Cornell University. He soon started working with Professor David Bindel, who later became his advisor. During the summer of 2017, he interned at the Lawrence Berkeley National Laboratory, where he worked under the mentorship of Professor Lin Lin. For the last three years at Cornell, he was partially supported by a National Science Foundation grant. In September 2019, Kun will move to Seattle and become a research scientist at Facebook, Inc.

This document is dedicated to my parents and wife.

ACKNOWLEDGEMENTS

Your acknowledgements go here. Make sure it sits inside the brackets.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	vii
List of Figures	viii
1 Introduction	1
2 Network Density of States	2
3 Scalable Gaussian Processes	3
4 Robust Large-Vocabulary Topic Modeling	4
4.1 Abstract	4
4.2 Introduction	4
4.3 Background and Related Work	8
4.4 Low-rank Rectification and Compression	11
4.4.1 Epsilon Non-Negative Rectification (ENN)	11
4.4.2 Proximal Alternating Linearized Minimization Rectification (PALM)	13
4.5 Low-rank Anchor Word Algorithm	14
4.6 Experimental Results	18
4.7 Conclusion	22
5 Weighted K-Means for Electronic Structure Calculation	24
5.1 Abstract	24
5.2 Introduction	25
5.3 Interpolative Separable Density Fitting (ISDF) decomposition	29
5.4 Centroidal Voronoi Tessellation based ISDF decomposition	31
5.5 Numerical results	36
5.5.1 Accuracy: Si ₂₁₆ and Al ₁₇₆ Si ₂₄	38
5.5.2 Efficiency: Si ₁₀₀₀	40
5.5.3 Ab Initio Molecular Dynamics: Si ₆₄ and (H ₂ O) ₆₄	43
5.6 Conclusion	47
5.7 Acknowledgments	48
6 Conclusion	49

LIST OF TABLES

5.1	Accuracy of ACE-ISDF Based Hybrid Functional Calculations (HSE06) Obtained by Using the CVT method To Select Interpolation Points, with Varying Rank Parameter c for Semiconducting Si_{216} and Metallic $\text{Al}_{176}\text{Si}_{24}$ Systems ^a	41
5.2	Wall Clock Time (in seconds) Spent in the Components of the ACE-ISDF and ACE Enabled Hybrid DFT Calculations Related to the Exchange Operator, for Si_{1000} on 2002 Edison cores at Different E_{cut} Levels ^a	43

LIST OF FIGURES

4.1	Experiment on four datasets. ENN and LR-JSMF mostly agree with AP, whereas PALM has slight inconsistency. The general information of each dataset is above the corresponding row. Recovery, approximation, and runtimes are in \log_{10} scale. Note that ENN and LR-JSMF are almost two orders of magnitude faster than AP. The x -axis indicates the number of clusters K . Lower numbers in y -axis are better except Specificity and Dissimilarity.	19
4.2	As we increase the vocabulary size for four collections, anchor quality and sparsity improve, but running time is stable. The x -axis indicates the vocabulary size N in thousands. Values above 15k will not fit in memory on standard hardware with previous algorithms.	21
4.3	Losses or gains in topic words depending on the vocabulary size. Each row represents a topic from the NeurIPS dataset, with the top 6 topical words shown in the middle column. The red and green cells denote topic words that are lost or gained by shifting the vocabulary size from the default size 5000, respectively. The intensities of the colors indicate the words' contributions towards the specific topic.	23
5.1	Schematic illustration of the CVT procedure in a 2D domain, including (a) initial random choice of centroids and Voronoi tessellation and centroidal Voronoi tessellation generated by the weighted K-Means algorithm. The weight function is given by the linear superposition of 4 Gaussian functions.	35
5.2	The decomposition reaction process of BH_3NH_3 computed with hybrid functional (HSE06) calculations by using the CVT procedure to select interpolation points, including (a) the electron density (yellow isosurfaces), (b) the interpolation points (yellow squares) $\{\hat{\mathbf{r}}_\mu\}_{\mu=1}^{N_\mu}$ ($N_\mu = 8$) selected from the real space grid points $\{\mathbf{r}_i\}_{i=1}^{N_g}$ ($N_g = 100^3$ and $E_{\text{cut}} = 60$ Ha) when the BN distance respectively is 1.3, 1.7 and 2.8 Å and (c) the binding energy as a function of BN distance for BH_3NH_3 in a $10 \text{ \AA} \times 10 \text{ \AA} \times 10 \text{ \AA}$ box. The white, pink and blue pink balls denote hydrogen, boron and nitrogen atoms, respectively.	37
5.3	The accuracy of ACE-ISDF based hybrid functional calculations (HSE06) obtained by using the CVT and QRCP procedures to select the interpolation points, with varying rank parameter c from 4 to 20 for Si_{216} and $\text{Al}_{176}\text{Si}_{24}$, including the error of (a) Hartree-Fock exchange energy ΔE_{HF} (Ha/atom) and (b) total energy ΔE (Ha/atom).	42
5.4	Comparison of the ISDF-CVT method by using either random or QRCP initialization for hybrid DFT AIMD simulations on bulk silicon system Si_{64} and liquid water system $(\text{H}_2\text{O})_{64}$, including (a) the fraction of points what switch cluster in each K-Means iteration and (b) the number of K-Means iterations during each MD step.	45

5.5	Comparison of hybrid HSE06 DFT AIMD simulations by using the ISDF-CVT and ISDF-QRCP methods as well as exact nested two-level SCF iteration procedure as the reference on the bulk silicon Si ₆₄ , including (a) relatively energy drift and (b) potential energy during MD steps.	46
5.6	The oxygen-oxygen radial distribution functions $g_{OO}(r)$ of liquid water system (H ₂ O) ₆₄ at $T = 295$ K obtained from hybrid HSE06 + DFT-D2 AIMD simulations with the ISDF-CVT and ISDF-QRCP methods, exact nested two-level SCF iteration procedure (as the reference) as well as previous hybrid PBE0 + TS-vdW calculation [24].	47

CHAPTER 1

INTRODUCTION

CHAPTER 2

NETWORK DENSITY OF STATES

CHAPTER 3

SCALABLE GAUSSIAN PROCESSES

CHAPTER 4

ROBUST LARGE-VOCABULARY TOPIC MODELING

4.1 Abstract

Across many data domains, co-occurrence statistics about the joint appearance of objects are powerfully informative. By transforming unsupervised learning problems into decompositions of co-occurrence statistics, spectral inference provides transparent algorithms and optimality guarantees for non-linear dimensionality reduction or latent topic analysis. As object vocabularies grow, however, it becomes rapidly more expensive to store and run inference algorithms on co-occurrence statistics. Current rectification techniques, which preprocess real data in order to overcome model-data mismatch, are even more expensive, as they require iterative projections that destroy sparsity of the co-occurrence data. In this paper, we propose novel approaches that can simultaneously compress and rectify the co-occurrence statistics, scaling gracefully with the size of vocabulary and the dimension of latent space. We also present new algorithms that are capable of learning latent variables from the compressed statistics without losing visible precision, and verify that they perform comparably to previous approaches on both textual and non-textual data.

4.2 Introduction

Understanding the low-dimensional geometry of noisy and complex data is a fundamental problem of unsupervised learning. Probabilistic models explain data generation processes in terms of low-dimensional latent variables. Inferring a posterior distribution for

these latent variables provides us with a compact representation for various exploratory analyses and downstream tasks. However, exact inference is often intractable due to entangled interactions between the latent variables [16, 2, 1, 70]. Variational inference transforms the posterior approximation into an optimization problem over simpler distributions with independent parameters [43, 76, 17], while Markov Chain Monte Carlo enables users to sample from the desired posterior distribution [62, 63, 73]. However, these likelihood-based methods require numerous iterations without any guarantee beyond local improvement at each step [45].

When the data consists of collections of discrete objects, co-occurrence statistics summarize interactions between objects. Collaborative filtering learns low-dimensional representations of individual items, which are useful for recommendation systems, by explicitly decomposing the co -occurrence of items that are jointly consumed by certain users [47, 50]. Word-vector models learn low -dimensional embeddings of individual words, which encode useful linguistic biases for neural networks, by implicitly decomposing the co-occurrence of words that appear together [69, 49]. If co -occurrence provides a rich enough set of unbiased moments about an underlying generative model, spectral methods can provably learn posterior configurations from co-occurrence information alone, without iterating through individual training examples [8, 5, 37, 3].

However, two major limitations hinder users from taking advantage of spectral inference based on co-occurrence. First, the second-order co -occurrence matrix grows quadratically in the size of the objects (e.g. items, words, products). Pruning these objects is an option, but for a retailer selling millions of products, a low-dimensional representation of a small subset of the products is inadequate. Second, inference quality is poor in real data that does not necessarily follow our computational model. Whereas likelihood-based methods have an intrinsic capability to fit the data to the model despite

their mismatch, sample noise can destroy the performance of spectral methods even if the data is synthesized from the model [45]. *Rectification*, a process of projecting empirical co-occurrence onto a set consistent with the geometry of the model, can improve the performance of spectral inference in the face of model mismatch [47]. But running multiple projections dominates overall inference cost even when the vocabulary is small. In addition, the rectification process makes the co-occurrence dense, exacerbating storage costs when dealing with large vocabularies.

In this paper, we propose the Epsilon Non-Negative rectification (ENN) and the Low-rank Anchor Word algorithm (LAW). Given a vocabulary of N objects and the user-specified latent dimension K , ENN simultaneously compresses and rectifies the co-occurrence matrix $\mathbf{C} \in \mathbb{R}^{N \times N}$ into $\mathbf{Y}\mathbf{Y}^T$ with $\mathbf{Y} \in \mathbb{R}^{N \times K}$. Each entry of the decomposition $(\mathbf{Y}\mathbf{Y}^T)_{ij}$ tightly approximates the rectified co-occurrence C_{ij}^* but is allowed to be a tiny negative value above $-\epsilon$. Then LAW learns the latent clusters (e.g., topics of documents or genres of items) and their correlations provided only with \mathbf{Y} , guaranteeing the same performance as running the original Anchor Word algorithm on \mathbf{C}^* if $\mathbf{Y}\mathbf{Y}^T \geq 0$. In contrast, we also formulate the Proximal Alternating Linearized Minimization rectification algorithm (PALM) that approximates the rectified co-occurrence \mathbf{C}^* by $\mathbf{Y}\mathbf{Y}^T$ with $\mathbf{Y} \geq 0$. While the non-negativity of \mathbf{Y} in the PALM approach allows LAW to perform exact inference, the fact that PALM enforces more constraints than ENN means that PALM also provides a less faithful approximation to the original co-occurrence data. Our experiments on various textual and non-textual datasets show that ENN learns a high-quality factor \mathbf{Y} for which LAW provides results of quality comparable to those based on the full co-occurrence \mathbf{C}^* ; in contrast, while PALM works comparably in some settings, in others there is a visible loss of accuracy.

We also adopt a randomized algorithm that constructs a low-rank approximation of

the full co-occurrence C directly from the raw data. While PALM requires the full co-occurrence, ENN can work directly with the low-rank initialization, eliminating the need to ever store a full co-occurrence matrix. Note also that the second-order methods rely on the *separability assumption*,¹ which has been another criticism in theory despite their superior performance in practice [48]. Our analyses show that models based on large vocabularies find more separable anchor objects, learning stable latent clusters without much sensitivity to sample noise. Overall, we complete a robust and scalable pipeline: that efficiently performs quality posterior inference for unsupervised learning from co-occurrence information within time and space complexity linear in N .

The major contributions of our paper are:

- We introduce two efficient rectifications, ENN and PALM, that compress quadratic and noisy co-occurrence information on the fly with a linear space rectified representation.
- We develop a low-rank algorithm (LAW) for anchor-word-based topic modeling that works directly on the compressed rectified representations and provides near-exact performance.
- We propose a robust and scalable pipeline, LR-JSMF, that learns topic models with a small number of passes directly over the data. This new pipeline scales to large vocabularies that were previously intractable for spectral inference, and offers a $\sim 100 \times$ speedup over previous methods for general data sets.

¹Each cluster has at least one anchor object which dedicates exclusively to that cluster, nothing else.

4.3 Background and Related Work

Our new algorithms build on the the **Joint-Stochastic Matrix Factorization (JSMF)** framework [47], which we now describe. Let $\mathbf{H} \in \mathbb{R}^{N \times M}$ be the object-example matrix whose m -th column vector \mathbf{h}_m counts the occurrences of each of the N objects in the vocabulary in example m . We denote the total number of objects in example m by n_m . Given a user-specified number of clusters K , we seek to learn an object-cluster matrix $\mathbf{B} \in \mathbb{R}^{N \times K}$ where B_{ik} is the conditional probability of observing object i given latent cluster k . Instead of learning \mathbf{B} directly from the sparse and noisy observations \mathbf{H} , JSMF begins with constructing the joint-stochastic co-occurrence $\mathbf{C} \in \mathbb{R}^{N \times N}$ by

$$\mathbf{C} = \hat{\mathbf{H}}\hat{\mathbf{H}}^T - \hat{\mathbf{H}}_{diag}, \quad \hat{\mathbf{h}}_m = \frac{\mathbf{h}_m}{\sqrt{n_m(n_m - 1)M}}, \quad \hat{\mathbf{H}}_{diag} = diag\left(\sum_{m=1}^M \frac{\mathbf{h}_m}{n_m(n_m - 1)M}\right). \quad (4.1)$$

Then the original Anchor Word algorithm decomposes \mathbf{C} into $\mathbf{B}\mathbf{A}\mathbf{B}^T$ by Algorithm 1, where $\mathbf{A} \in \mathbb{R}^{K \times K}$ is the cluster correlation matrix, whose entry A_{kl} captures the joint probability between two latent clusters k and l .² In the limit, using data generated according to the correct probabilistic models, \mathbf{A} must agree with the second-moment of the cluster proportions, which is given as a Bayesian prior in the models.

As with other spectral algorithms for latent variable models [37, 4], the decomposition as described so far may fail to learn high-quality clusters due to *model-data mismatch* [45]. Under the probabilistic model assumed to generate the data, the expected value of the co-occurrence should not only be normalized to sum to one (*NOR*) and be entry-wise non-negative (*NN*), but it should also be positive semi-definite with rank equal to the number of clusters K (\mathcal{PSD}_K) [47]. However, the empirical \mathbf{C} from real data is often indefinite and full-rank due to sample noise³ and the unbiased construction

² \mathbf{C} is proven to be a by far more robust estimator than \mathbf{H} in [7]. But actual construction of \mathbf{C} in [8] is slightly misleading without dividing by M . We report the full equations in this paper.

³Rectification still improves the quality of clusters on the synthetic data that is generated from our models.

Algorithm 1: Anchor Word algorithm (AW)

Input: Object co-occurrence $C \in \mathbb{R}^{N \times N}$
The number of clusters K

Output: Anchor objects $S = \{s_1, \dots, s_K\}$
Latent clusters $B \in \mathbb{R}^{N \times K}$
Cluster correlations $A \in \mathbb{R}^{K \times K}$

begin

L_1 -normalize the rows of C to form \bar{C} .
Find S via the column pivoted QR on \bar{C}^T .
Find \check{B} with $\check{B}_{ki} = p(\text{cluster } k \mid \text{object } i)$ by solving N simplex-constrained least squares in parallel to minimize $\|\bar{C} - \check{B}^T \bar{C}_{S*}\|_F$.
Recover B from \check{B} by the Bayes' rule.
Recover A by $B_{S*}^{-1} C_{SS} B_{S*}^{-1}$.

end

Algorithm 2: Rectified AW algorithm (RAW)

Input/Output: Same as Algorithm 1

begin

$C_0 \leftarrow C$
repeat with $t = 0, 1, 2, \dots$
 $(U, \Lambda_K) \leftarrow \text{Truncated-Eig}(C_t, K)$
 $\Lambda_K^+ \leftarrow \max(\Lambda_K, 0)$
 $C^{\mathcal{PSD}_K} \leftarrow U \Lambda_K^+ U^T$
 $C^{NOR} \leftarrow C^{\mathcal{PSD}_K} + \frac{1 - \sum_{ij} C_{ij}^{\mathcal{PSD}_K}}{N^2} ee^T$
 $C^{NN} \leftarrow \max(C^{NOR}, 0)$
 $C_{t+1} \leftarrow C^{NN}$
until converging to a certain C^*
 $(S, B, A) \leftarrow \text{AW}(C^*, K)$ (Algorithm 1)

end

of C in Equation (4.1), which penalizes all diagonal entries. The **Rectified Anchor Word (RAW)** algorithm has an additional rectification step that enforces that C should enjoy the expected structures before running the main algorithm. In [47], an Alternating Projection (AP) rectification as given in Algorithm 2 is used to overcome the gap between the underlying assumptions of our models and the actual data.

Rectification is also important for addressing the issue of *outlier bias*. Real data

often exhibits rare objects that are only present in a few examples. The corresponding co-occurrence of these objects are inevitably sparse with large variance, but the greedy anchor selection favors choosing these outliers due to L_1 normalization of \mathbf{C} . Previous work tried to bypass this problem by oversampling clusters by the number of outliers under some additional identifiability assumptions [27]. This approach is not always feasible, especially for a large vocabulary that introduces many low frequency objects. When synonyms and short documents cause undesirable sparsity to Latent Semantic Analysis [46], projection onto the leading eigen-subspace blurs sparse co-occurrences. Similarly, \mathcal{PSD}_K -projection turns out to significantly reduce outlier bias, and the remaining projections are useful for maintaining the probabilistic structures of \mathbf{C} , which then allow users to recover \mathbf{B} and \mathbf{A} in Algorithm 1.

Handling a *large vocabulary* is another major challenge for spectral methods. Even if we limit our focus only to second-order models, the space complexity of RAW is already $O(N^2)$, growing rapidly with increasing vocabulary. We are unable to exploit the high sparsity of \mathbf{C} as a single iteration of the AP-rectification makes \mathbf{C} significantly denser. The three projections in AP-rectification and the rest of the anchor word algorithm in Algorithm 1 have time complexities of $O(N^2K)$, $O(N^2)$, $O(N^2)$ and $O(N^2K)$, respectively, and so pose a difficulty when scaling to a large vocabulary size N . On the other hand, the *separability assumption* is crucial for the second-order models, and while there has been a line of research that tries to relax this assumption [13, 42], it has been formally shown that most topic models are indeed separable if their vocabulary sizes are sufficiently larger than the number of clusters [23], again emphasizing the urgency of an approach with better time and space scaling in the vocabulary size.

4.4 Low-rank Rectification and Compression

The rectified co-occurrence C^* in Section 4.3 must be of rank K and positive semidefinite, hinting at an opportunity to represent it in terms of an outer product $\mathbf{Y}\mathbf{Y}^T$ for some $\mathbf{Y} \in \mathbb{R}^{N \times K}$. One idea for achieving this structure is to use a low-rank representation $C_t = \mathbf{Y}_t\mathbf{Y}_t^T$ throughout the rectification in Algorithm 2. Another way to obtain this structure is to directly minimize $\|C - \mathbf{Y}\mathbf{Y}^T\|_F$ with the necessary constraints. In this section, we propose two new algorithms to simultaneously compress and rectify the input by representing $C \in \mathbb{R}^{N \times N}$ by a low-rank outer product $\mathbf{Y}\mathbf{Y}^T$.

Algorithm 3: ENN-rectification (ENN)

Input: Object co-occurrence $C \in \mathbb{R}^{N \times N}$
 The number of clusters K
Output: Rectified compression $\mathbf{Y} \in \mathbb{R}^{N \times K}$

begin

- $E \leftarrow \mathbf{0} \in \mathbb{R}^{N \times N}$ (sparse format)
- $C^{op} : \mathbf{x} \rightarrow C\mathbf{x}$ (Implicit operator)
- repeat** with $t = 0, 1, 2, \dots$
- $(U, \Lambda_K) \leftarrow \text{Truncated-Eig}(C^{op}, K)$
- $\Lambda_K^+ \leftarrow \max(\Lambda_K, 0)$
- $\mathbf{Y} \leftarrow U(\Lambda_K^+)^{1/2}$
- $E_{ij} \leftarrow \max(-\mathbf{Y}_{i*}\mathbf{Y}_{j*}^T, 0)$
- $r \leftarrow (1 - \|\mathbf{Y}^T e\|_2^2 - \sum_{ij} E_{ij})/N^2$
- $C^{op} : \mathbf{x} \rightarrow \mathbf{Y}(\mathbf{Y}^T \mathbf{x}) + E\mathbf{x} + r(e^T \mathbf{x})e$
- until** E converges

end

4.4.1 Epsilon Non-Negative Rectification (ENN)

The alternating projection iteration in Algorithm 2 produces low-rank semi-definite intermediate matrices in factored form at each iteration. By construction, the positive

Algorithm 4: PALM-rectification (PALM)

Input: Object co-occurrence $\mathbf{C} \in \mathbb{R}^{N \times N}$
 The number of clusters K
Output: Rectified compression $\mathbf{Y} \in \mathbb{R}^{N \times K}$

begin

- $(\mathbf{V}, \mathbf{D}) \leftarrow \text{Truncated-Eig}(\mathbf{C}, K)$
- $(\mathbf{X}_0, \mathbf{Y}_0) \leftarrow (\mathbf{V} \sqrt{\mathbf{D}}, \mathbf{V} \sqrt{\mathbf{D}})$
- repeat**

 - $c_t \leftarrow \gamma L_1(\mathbf{Y}_t)$
 - $\mathbf{X}'_{t+1} \leftarrow \mathbf{X}_t - (1/c_t) \nabla_{\mathbf{X}} J(\mathbf{X}_t, \mathbf{Y}_t)$
 - $\mathbf{X}_{t+1} \leftarrow \max(\mathbf{X}'_{t+1}, 0)$
 - $d_t \leftarrow \gamma L_2(\mathbf{X}_{t+1})$
 - $\mathbf{Y}'_{t+1} \leftarrow \mathbf{Y}_t - (1/d_t) \nabla_{\mathbf{Y}} J(\mathbf{X}_{t+1}, \mathbf{Y}_t)$
 - $\mathbf{Y}_{t+1} \leftarrow \max(\mathbf{Y}'_{t+1}, 0)$

- until** \mathbf{Y} converges

end

semi-definite projection (\mathcal{PSD}_K) and the normalization projection (\mathcal{NOR}) produce positive semi-definite matrices of rank K and $K + 1$, respectively. Unfortunately, the final projection to enforce elementwise non-negativity (\mathcal{NN}) destroys this low rank structure. However, the \mathcal{NN} projection only makes significant changes to a few elements; that is, the output of the \mathcal{NN} projection at step t is nearly rank $K + 1$ plus a sparse correction \mathbf{E}_t . The Epsilon Non-Negative Rectification algorithm (Algorithm 3) has the same structure as Algorithm 2, but with the key difference that it returns a sparse-plus-low-rank representation of the \mathcal{NN} projection rather than materializing a dense representation. Matrix-vector products with this sparse-plus-low-rank representation require $O(NK + \text{nnz}(\mathbf{E}_t))$ time, and $O(K)$ such matrix-vector products can be used in a Lanczos eigen-solver to compute the truncated eigen-decomposition at the start of the next iteration.

Maintaining a sparse correction matrix \mathbf{E}_t at each step lets the ENN approach avoid the storage overheads of the original alternating projection algorithm. To overcome the quadratic time cost at each iteration, though, we need to avoid explicitly computing ev-

every element of the intermediate $\mathbf{Y}\mathbf{Y}^T$ in the course of the \mathcal{NN} projection. However, we can bound the magnitude of many elements of $\mathbf{Y}\mathbf{Y}^T$ by the Cauchy-Schwartz inequality: $|C_{ij}| \leq \|\mathbf{y}_i\|_2 \|\mathbf{y}_j\|_2$ where \mathbf{y}_i and \mathbf{y}_j denote columns of \mathbf{Y}^T . Let I denote the index set indices $\{i : \|\mathbf{y}_i\|_2^2 > \epsilon\} \subseteq [N]$ for given ϵ ; then every large entry of \mathbf{C} belongs to either $\mathbf{Y}_{I*}\mathbf{Y}^T$ or $\mathbf{Y}(\mathbf{Y}^T)_{*I}$. As \mathbf{C} is symmetric, checking the negative entries in $\mathbf{Y}_{I*}\mathbf{Y}^T$ is sufficient to find a symmetric correction \mathbf{E} that guarantees $\mathbf{Y}\mathbf{Y}^T + \mathbf{E} \geq -\epsilon$. We refer to this property as *Epsilon Non-Negativity*: ϵ balances the trade-off between the effect of leaving small negative entries versus increasing the size of I to look up. We limit $|I|$ to be $O(K)$ based on the common sampling complexity of a suitable set of rows for a near-optimal rank-K approximation⁴.

4.4.2 Proximal Alternating Linearized Minimization Rectification (PALM)

To avoid small negative entries, we investigate another rectified compression algorithm that directly minimizes $\|\mathbf{C} - \mathbf{Y}\mathbf{Y}^T\|_F$ subject to the stronger \mathcal{NN} -constraint $\mathbf{Y} \geq 0$ and the usual \mathcal{NOR} -constraint $\|\mathbf{Y}^T \mathbf{e}\|_2 = 1$. Concretely, we try to

$$\text{minimize } J(\mathbf{X}, \mathbf{Y}) := \frac{1}{2} \|\mathbf{C} - \mathbf{X}\mathbf{Y}^T\|_F^2 + \frac{s}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2 \quad \text{subject to } \mathbf{X} \geq 0, \mathbf{Y} \geq 0. \quad (4.2)$$

\mathcal{PSD}_K - and \mathcal{NOR} -constraints are implicitly satisfied by jointly minimizing the two terms in the objective function J , whereas \mathcal{NN} -constraint is explicit in the formulation. Thus we can apply the Proximal Alternating Linearized Minimization [18] for learning \mathbf{Y} given \mathbf{C} ; the relevant proximal operator is \mathcal{NN} projection of \mathbf{Y} , which takes $O(NK)$ at most.

⁴This choice is standard in literature on low-rank approximation via column subset selection.

Note that J is semi-algebraic (as it is a real polynomial) with two partial derivatives: $\nabla_{\mathbf{X}} J = (\mathbf{X}\mathbf{Y}^T - \mathbf{C})\mathbf{Y} + s(\mathbf{X} - \mathbf{Y})$ and $\nabla_{\mathbf{Y}} J = (\mathbf{Y}\mathbf{X}^T - \mathbf{C})\mathbf{X} + s(\mathbf{Y} - \mathbf{X})$. So, the following lemma guarantees the global convergence.

Lemma 1. *For any fixed \mathbf{Y} , $\nabla_{\mathbf{X}} J(\mathbf{X}, \mathbf{Y})$ is globally Lipschitz continuous with the moduli $L_1(\mathbf{Y}) = \|\mathbf{Y}^T \mathbf{Y} + s\mathbf{I}_K\|_2$. So is $\nabla_{\mathbf{Y}} J(\mathbf{X}, \mathbf{Y})$ given any fixed \mathbf{X} with $L_2(\mathbf{X}) = \|\mathbf{X}^T \mathbf{X} + s\mathbf{I}_K\|_2$.*

Proof.

$$\begin{aligned} & \|\nabla_{\mathbf{X}} J(\mathbf{X}, \mathbf{Y}) - \nabla_{\mathbf{X}} J(\mathbf{X}', \mathbf{Y})\|_F \\ &= \|(\mathbf{Y}^T \mathbf{Y} + s\mathbf{I}_K)(\mathbf{X} - \mathbf{X}')\|_F \\ &\leq \|\mathbf{Y}^T \mathbf{Y} + s\mathbf{I}_K\|_2 \cdot \|\mathbf{X} - \mathbf{X}'\|_F \end{aligned}$$

The proof is symmetric for the other case with $L_2(\mathbf{X}) = \|\mathbf{X}^T \mathbf{X} + s\mathbf{I}_K\|_2$. \square

Algorithm 4 shows the PALM-rectification with the adaptive control of the learning rates based on the tight 2-norm Lipschitz modulis at each step t .

4.5 Low-rank Anchor Word Algorithm

The output for both methods in §4.4 is a compressed co-occurrence matrix $\mathbf{C} = \mathbf{Y}\mathbf{Y}^T$. In this section, we present the **Low-rank Anchor Word algorithm (LAW)** that reduces the time complexity of finding anchor objects from $O(N^2K)$ to $O(NK^2)$ by taking advantage of this form. We note that LAW applies whenever \mathbf{C} is in a low-rank representation, which does not have to be derived from our methods. Moreover, it is exact for non-negative \mathbf{C} , but it is robust in practice when we allow small negative entries in \mathbf{C} , as in the case with ENN.

Algorithm 5: Low-rank AW (LAW)

Input: Object co-occurrence $\mathbf{C} = \mathbf{Y}\mathbf{Y}^T$
Output: Anchor objects $\mathbf{S} = \{s_1, \dots, s_K\}$
Latent clusters $\mathbf{B} \in \mathbb{R}^{N \times K}$
Cluster correlations $\mathbf{A} \in \mathbb{R}^{K \times K}$

begin

- Calculate row sums $\mathbf{d} = \mathbf{Y}(\mathbf{Y}^T \mathbf{e})$.
- Normalize $\bar{\mathbf{Y}} \leftarrow \text{diag}(\mathbf{d})^{-1} \mathbf{Y}$.
- Compute QR decomposition of $\mathbf{Y} = \mathbf{Q}\mathbf{R}$.
- Form $\mathbf{X} = \bar{\mathbf{Y}}\mathbf{R}^T$.
- Select \mathbf{S} using column pivoted QR on \mathbf{X}^T .
- Solve n simplex-constrained least square problems to minimize $\|\mathbf{X} - \check{\mathbf{B}}\mathbf{X}_{S^*}\|_F$.
- Recover \mathbf{B} from $\check{\mathbf{B}}$ using Bayes' rule.
- Recover $\mathbf{A} = \mathbf{B}_{S^*}^{-1} \mathbf{Y}_{S^*} \mathbf{Y}_{S^*}^T \mathbf{B}_{S^*}^{-1}$.

end

Algorithm 6: Low-rank JSMF (LR-JSMF)

Input: Raw object-example $\mathbf{H} \in \mathbb{R}^{N \times M}$
Output: Anchor objects $\mathbf{S} = \{s_1, \dots, s_K\}$
Latent clusters $\mathbf{B} \in \mathbb{R}^{N \times K}$
Cluster correlations $\mathbf{A} \in \mathbb{R}^{K \times K}$

begin

- Get $\hat{\mathbf{H}}, \hat{\mathbf{H}}_{\text{diag}}$ from \mathbf{H} by (4.1).
- $C_{op} : \mathbf{x} \rightarrow \hat{\mathbf{H}}(\hat{\mathbf{H}}^T \mathbf{x}) - \hat{\mathbf{H}}_{\text{diag}} \mathbf{x}$
- $(\mathbf{U}, \Lambda_K) \leftarrow \text{Randomized-Eig}(C_{op}, K)$
- Initialize ENN with \mathbf{U}, Λ_K .
- $\mathbf{Y} \leftarrow \text{ENN-rectification}$
- $(\mathbf{S}, \mathbf{B}, \mathbf{A}) \leftarrow \text{LAW}(\mathbf{Y})$ (Algorithm 5)

end

The first step is to L_1 -normalize the rows of \mathbf{C} . Given $\mathbf{C} \geq 0$, the L_1 -norm of each row is simply the sum of all its entries, so we can calculate the row norms by $\mathbf{d} = \mathbf{Y}(\mathbf{Y}^T \mathbf{e})$. To obtain the normalized \mathbf{C} , we simply scale the rows of \mathbf{Y} , and $\overline{\mathbf{C}} = (\text{diag}(\mathbf{d})^{-1} \mathbf{Y}) \mathbf{Y}^T = \overline{\mathbf{Y}} \mathbf{Y}^T$. These steps cost $O(NK)$.

Next, we need to apply column pivoted QR to $\overline{\mathbf{C}}^T$ in order to identify the pivots as our anchor objects \mathbf{S} . By taking the QR decomposition $\mathbf{Y} = \mathbf{Q}\mathbf{R}$, $\overline{\mathbf{C}}^T$ can be further transformed into $\mathbf{Q}\mathbf{R}\overline{\mathbf{Y}}^T$. Notice that $\overline{\mathbf{C}}^T$ is an orthogonal embedding of $\mathbf{X}^T = \mathbf{R}\overline{\mathbf{Y}}^T$ onto a higher-dimensional space, which preserves the column L_2 -norms. Lemma 2 shows that column pivoted QR on $\overline{\mathbf{C}}^T$ and on $\mathbf{R}\overline{\mathbf{Y}}^T$ are equivalent, which allows us to lower the computation cost from $O(N^2K)$ to $O(NK^2)$.

Lemma 2. *Let \mathbf{S} be the set of pivots that have been selected by column pivoted QR on $\overline{\mathbf{C}}^T = \mathbf{Q}\mathbf{X}^T$. Given the QR decomposition, $\overline{\mathbf{X}}_{\mathbf{S}*}^T = \mathbf{P}\mathbf{T}$, then $\overline{\mathbf{C}}_{\mathbf{S}*}^T = (\mathbf{Q}\mathbf{P})\mathbf{T}$ is the corresponding QR decomposition for the columns of $\overline{\mathbf{C}}$. For any remaining row $i \in [N] \setminus \mathbf{S}$,*

$$\|(\mathbf{I} - \mathbf{P}\mathbf{P}^T)\overline{\mathbf{X}}_{i*}^T\|_2 = \|(\mathbf{I} - (\mathbf{Q}\mathbf{P})(\mathbf{Q}\mathbf{P})^T)\overline{\mathbf{C}}_{i*}^T\|_2 \quad (4.3)$$

Therefore, the next column pivot is identical for $\overline{\mathbf{C}}^T$ and $\overline{\mathbf{X}}^T$. By induction, column pivoted QR on $\overline{\mathbf{C}}^T$ and $\overline{\mathbf{X}}^T$ return the same pivots.

Proof. Because both \mathbf{Q} and \mathbf{P} have orthonormal columns,

$$(\mathbf{Q}\mathbf{P})^T(\mathbf{Q}\mathbf{P}) = \mathbf{P}^T(\mathbf{Q}^T\mathbf{Q})\mathbf{P} = \mathbf{P}^T\mathbf{P} = \mathbf{I} \quad (4.4)$$

Thus, $\mathbf{Q}\mathbf{P}$ and \mathbf{T} forms the QR decomposition of $\overline{\mathbf{C}}_{\mathbf{S}*}^T$. The residual of a remaining column $i \in [N] \setminus \mathbf{S}$ is $(\mathbf{I} - (\mathbf{Q}\mathbf{P})(\mathbf{Q}\mathbf{P})^T)\overline{\mathbf{C}}_{i*}^T$ and $(\mathbf{I} - \mathbf{P}\mathbf{P}^T)\overline{\mathbf{X}}_{i*}^T$ for $\overline{\mathbf{C}}^T$ and $\overline{\mathbf{X}}^T$, respectively. Simplify the former gives us

$$(\mathbf{I} - (\mathbf{Q}\mathbf{P})(\mathbf{Q}\mathbf{P})^T)\overline{\mathbf{C}}_{i*}^T$$

$$\begin{aligned}
&= (\mathbf{I} - (\mathbf{Q}\mathbf{P})(\mathbf{Q}\mathbf{P})^T)\mathbf{Q}\overline{\mathbf{X}}_{i*}^T \\
&= \mathbf{Q}\overline{\mathbf{X}}_{i*}^T - \mathbf{Q}\mathbf{P}\mathbf{P}^T\mathbf{Q}^T\mathbf{Q}\overline{\mathbf{X}}_{i*}^T \\
&= \mathbf{Q}(\mathbf{I} - \mathbf{P}\mathbf{P}^T)\overline{\mathbf{X}}_{i*}^T
\end{aligned}$$

Finally,

$$\|(\mathbf{I} - (\mathbf{Q}\mathbf{P})(\mathbf{Q}\mathbf{P})^T)\overline{\mathbf{C}}_{i*}^T\|_2^2 = \overline{\mathbf{X}}_{i*}(\mathbf{I} - \mathbf{P}\mathbf{P}^T)\mathbf{Q}^T\mathbf{Q}(\mathbf{I} - \mathbf{P}\mathbf{P}^T)\overline{\mathbf{X}}_{i*}^T = \|(\mathbf{I} - \mathbf{P}\mathbf{P}^T)\overline{\mathbf{X}}_{i*}^T\|_2^2$$
(4.5)

Because the next pivot is selected as the column whose residual has the largest L_2 -norm, Eq. 4.5 indicates that the same pivot will be selected for $\overline{\mathbf{C}}^T$ and $\overline{\mathbf{X}}^T$. Inductively, the anchors \mathbf{S} recovered by column pivoted QR on those matrices are equivalent. \square

Following the recovery of \mathbf{S} , AW solves N independent simplex-constrained least square problems $\|\overline{\mathbf{C}}_{i*} - \check{\mathbf{B}}_{i*}^T \overline{\mathbf{C}}_{S*}\|_2$. Again we can leverage the L_2 -norm preserving property,

$$\|\overline{\mathbf{C}}_{i*} - \check{\mathbf{B}}_{i*}^T \overline{\mathbf{C}}_{S*}\|_2 = \|\mathbf{X}_{i*}\mathbf{Q}^T - \check{\mathbf{B}}_{i*}^T \mathbf{X}_{S*}\mathbf{Q}^T\|_2 = \|\mathbf{X}_{i*} - \check{\mathbf{B}}_{i*}^T \mathbf{X}_{S*}\|_2$$
(4.6)

and reduce the dimension of the least-square problems from N to K , thereby the complexity from $O(N^2K)$ to $O(NK^2)$. The remaining part of the algorithm follows exactly as AW.

Low-rank Joint Stochastic Matrix Factorization (LR-JSMF) We complete our scalable framework of processing co-occurrence statistic by introducing a direct initialization method from the raw object-example data for ENN. This allows us to avoid creating and storing \mathbf{C} , which is a burden of memory when N becomes sufficiently large. In Algorithm 3, \mathbf{C} only appears in the initial truncated eigen-decomposition, after which we maintain the compressed operator \mathbf{C}_{op} independent of it. On the other hand, we just need the matrix-vector multiplication by \mathbf{C} for the iterative methods in

initialization. Using the generative formula in Eq. 4.1, we are able to implicitly apply \mathbf{C} to vectors as an outer-product plus diagonal operator, in terms of \mathbf{H} , at $O(NMK)$ computation cost. To further reduce the number of times the operator is applied, we adopt the one-pass randomized eigen-decomposition by Halko et al. [32]. This technique enables us to initialize with a single pass over the dataset, without concurrently storing the entire \mathbf{H} in memory. A limitation is when the number of clusters is large and the gap between the K -th eigenvalue and the ones below is small, we will have to incorporate a few power iterations for refinement, as suggested by the original paper. This will result in a multi-pass method, but still far more efficient on large object size and parallelization-friendly.

4.6 Experimental Results

A good factorization should be accurate, meaningful, and fast. In two series of experiments, we show that LR-JSMF maintains model quality while running in a fraction of the space and time needed for the original JSMF method. Previous methods have required truncation of the vocabulary even to run on consumer-grade computers. We show not only that we are able to handle increasingly large vocabularies without loss of speed, but that using larger vocabularies measurably improves model quality relative to truncated vocabularies.

For the first series of experiments, we measure the accuracy of each rectification component as well as the entire pipeline of LR-JSMF. To produce a strong baseline, we begin with constructing the full co-occurrence \mathbf{C} from each of our datasets \mathbf{H} by (4.1), and produce the rectified \mathbf{C}_{AP} by running Alternating Projection (AP) on \mathbf{C} . Next we compress \mathbf{C} into \mathbf{Y}_{ENN} and \mathbf{Y}_{PALM} by running ENN (50 iterations, $|I| = 10K + 1000$) and

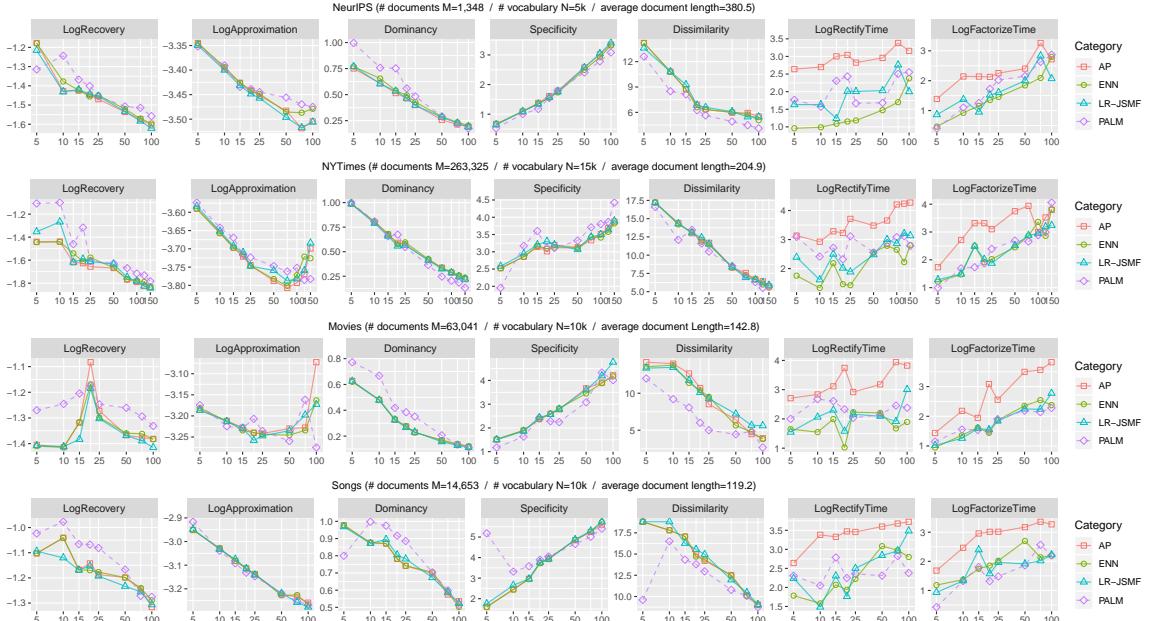


Figure 4.1: Experiment on four datasets. ENN and LR-JSMF mostly agree with AP, whereas PALM has slight inconsistency. The general information of each dataset is above the corresponding row. Recovery, approximation, and runtimes are in \log_{10} scale. Note that ENN and LR-JSMF are almost two orders of magnitude faster than AP. The x -axis indicates the number of clusters K . Lower numbers in y -axis are better except Specificity and Dissimilarity.

PALM (100 iterations, $s = 1e^{-4}$). For testing our complete low-rank pipeline, we also construct (\mathbf{V}, \mathbf{D}) directly from the raw data \mathbf{H} by the randomized eigen-decomposition in Algorithm 6, learning the compressed statistics $\mathbf{Y}_{LR-JSMF}$ again by running ENN initialized with $\mathbf{V} \sqrt{\mathbf{D}}$. Then we run the Anchor Word algorithm (AW) on \mathbf{C}_{AP} and the Low-rank Anchor Word algorithm (LAW) on each of \mathbf{Y}_{ENN} , \mathbf{Y}_{PALM} , and $\mathbf{Y}_{LR-JSMF}$.

The goal of rectification is to apply spectral inference to data that does not follow our modeling assumptions, so we evaluate on real data. In addition to two standard datasets from the UCI Machine Learning repository (NeurIPS papers and New York Times articles), we also use two non-textual datasets (Movies and Songs) previously used to demonstrate the performance of full algorithm with AP-rectification in [47]. Although our ultimate goal is to extend JSMF to large vocabularies, we use the same restricted vocabulary as [47] for a fair comparison in the first series of experiment.

Figure 4.1 shows the overall performance of the learned topic clusters from these four datasets with increasing number of clusters K . Low **Recovery** error $\frac{1}{N} \sum_i \|\overline{\mathbf{C}}_{i*} - \check{\mathbf{B}}_{i*} \overline{\mathbf{C}}_{S*}\|_2$ implies that the learned anchor objects successfully reconstruct the co-occurrence space of the entire objects. Low **Approximation** error $\|\mathbf{C} - \mathbf{BAB}^T\|_2$ means that our factorization captures most of information given in the unbiased co-occurrence statistics. In real data, low **Dominancy** $\frac{1}{K} \sum_k \mathbf{A}_{kk}$ implies that our models learn more correlations between clusters. High **Specificity** $\frac{1}{K} \sum_k \text{KL}(\mathbf{B}_{*k} \| \sum_i \mathbf{C}_{*i})$ indicates that the learned clusters are different enough from the corpus distribution, whereas high **Dissimilarity** counts the average number of objects in each cluster that do not occur among top 20 in other clusters, showing the interpretable difference across the learned clusters. We do not report the Cluster Coherence because it often measures deceptively [21]. The first five columns show that ENN and LR-JSMF learn approximately same clusters as JSMF with the full AP, showing no visible loss in accuracy across all settings. More importantly, the randomness we introduced into LR-JSMF results in a very low variance over a number of runs. This is important as the stability of spectral inference is a major advantage relative to MCMC or Variational Inference. Although PALM deviates a small amount from the other three methods in a few cases, it mostly achieves the same level of accuracy and follows the overall trend closely. In terms of runtimes, all of our methods have clear advantage over AP, gaining $1 \sim 2$ orders of magnitude speedup in most situations. Even for applications on relatively small vocabulary sizes, our algorithms shows a notable improvement in efficiency.

For the second series of experiments, we create eight corpora $\{\mathbf{H}_N, \mathbf{H}_{2N}, \dots, \mathbf{H}_{8N}\}$ for each dataset \mathbf{H} by tailoring their vocabulary sizes as multiples of a base vocabulary of N objects. In this case we are not able to compare LR-JSMF to previous methods because we cannot store the full co-occurrence matrices for the larger vocabulary: these models would be impossible. Figure 4.2 illustrates the overall performance of

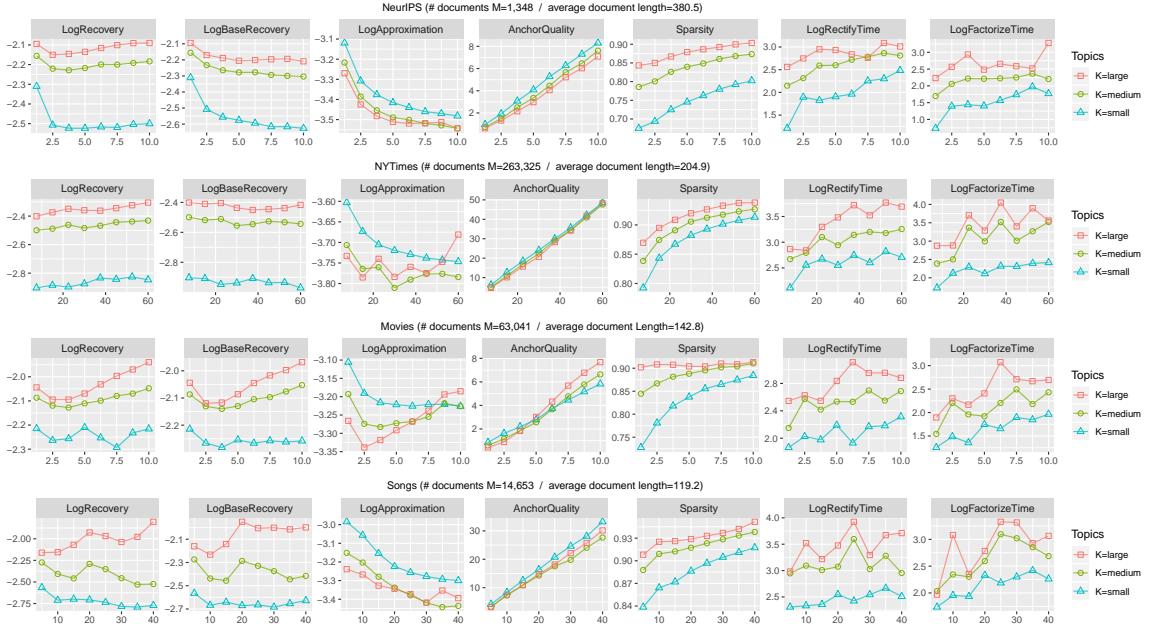


Figure 4.2: As we increase the vocabulary size for four collections, anchor quality and sparsity improve, but running time is stable. The x -axis indicates the vocabulary size N in thousands. Values above 15k will not fit in memory on standard hardware with previous algorithms.

the learned small/medium/large size clusters with increasing vocabulary size N . Low **BaseRecovery** means that the anchor objects from the models with larger vocabulary better reconstruct the objects in our base vocabulary (H_N). High **AnchorQuality** indicates that the average rank of the anchor objects s_k in every other topic clusters than k is high, implying the anchor objects rarely contribute to other clusters than their own. High **Sparsity** ($\frac{1}{K} \sum_k \frac{\sqrt{N} - (\|b_k\|_1 / \|b_k\|_2)}{\sqrt{N} - 1}$) [36] says that our topic clusters are more concentrated on specific objects.

We observe the quality of anchors increases with increasing vocabulary size, verifying that using larger vocabularies helps better satisfy the separability assumption. We also verify that a large vocabulary often better approximates the co-occurrence statistics and better reconstructs the co-occurrence space of the *base* vocabulary, but these patterns are not always consistent in non-textual datasets. In contrast, Sparsity consistently improves, increasing the interpretability of the learned clusters. Most excitingly, the run-

ning times of ENN and LAW show the scalability of our new rectification and low-rank algorithm, thereby demonstrating that LR-JSMF is an efficient and robust pipeline.

Finally, we have also inspected the qualitative behavior of the recovered clusters, as we increase the vocabulary size. The topic clusters become significantly more specific, while the clustering of objects is more conspicuous. Figure 4.3 shows how using a larger vocabulary size can lead to more distinguishable topics, especially as it allows us to make use of words that are relatively rare, but used in much narrower contexts. Going from left to right, we can observe that the set of topic words become more and more specific: for instance, the topic corresponding to the third row is slightly vague when observing just the left half of the row, while as we increase the vocabulary size beyond 5000, we gain access to highly topic-specific words such as `hjb` (Hamilton-Jacobi-Bellman equation) or `pid` (Proportional Integral Derivative), which signifies the row’s pertinence to dynamical and control systems. The flip side of the figure shows that words we would normally consider as non-topical can often be assigned high contributions towards certain topics. The strong red shade on the bottom left indicates that words such as “equivalent” or “cambridge” are strongly connected to the machine learning literature.

4.7 Conclusion

Spectral algorithms provide an appealing alternative for identifying interpretable low-rank subspaces by simple factorizations of higher-order moments. But this simplicity is also a weakness: violations of modeling assumptions destroy performance unless they are handled through rectification, and the size of the moment matrices limits us to small vocabularies. In this paper, we developed an efficient and scalable framework:

		Vocabulary Size						
		1250	2500	3750	5000 (base)	6250	7500	8750
refractory interconnection seen detail transmission considered	interspike	bursting	neuron	signalling	ipsp	tst		
	marder	stomatogastric	circuit	meilijson	stg	tam		
	abbott	konishi	synaptic	quiescent	substances	hyperpolarized		
	acad	axonal	cell	ryckebusch	inactivation	memorized		
	pyloric	modulatory	layer	leech	depolarized	transposed		
	bird	ionic	signal	silent	shapiro	tsividis		
san additional amount considered developed significant	male	kanji	recognition	radical	sicl	subword		
	henderson	phonemic	layer	npm	joe	phoneroes		
	jackel	subsystem	hidden	demi	chinese	otherfilter		
	recog	dtw	word	shikano	hanazawa	sdnn		
	dictionary	strokes	speech	letterform	lexicon	perplexity		
	ocr	gender	net	tebelskis	preprocessed	males		
cambridge pendulum requires con bellman plan	discounted	tutor	control	hjb	jacobi	ovi		
	bradtke	lqr	action	rein	forcement	pid		
	discount	disturbances	dynamic	biped	viscosity	idm		
	eligibility	disturbance	optimal	trol	sel	umass		
	indirect	hamilton	reinforcement	handicapped	bizzi	missile		
	amherst	smdp	controller	gullapalli	swinging	queueing		
directional seen dark neuroscience soc supported	transparent	luminance	cell	unoriented	aftereffect	moc		
	adelson	ruderman	field	heeger	blast	ori		
	geniculate	andersen	visual	thalamus	mae	taube		
	amacrine	bergen	motion	directionally	knierim	muller		
	deg	selectively	direction	hayashi	mexican	swindle		
	mennaughton	lond	image	werblin	specimen	skagg		
equivalent computing cambridge considered simply detail	fix	solvable	gaussian	birmingham	boxplot	pbr		
	satisfying	distribu	noise	kolmogorov	dependences	owi		
	royal	barber	approximation	gmm	cb2	danish		
	opper	parametrized	hidden	winther	trigonometric	ylz		
	leibler	const	bound	imation	colt	diabetes		
	treatment	eter	matrix	statist	minimizer	devroye		

Figure 4.3: Losses or gains in topic words depending on the vocabulary size. Each row represents a topic from the NeurIPS dataset, with the top 6 topical words shown in the middle column. The red and green cells denote topic words that are lost or gained by shifting the vocabulary size from the default size 5000, respectively. The intensities of the colors indicate the words’ contributions towards the specific topic.

Low-Rank Joint Stochastic Matrix Factorization. We provide theoretical advances in compressed matrix factorization, leading to high-quality low-rank non-negative approximations without quadratic blowup. The method provides orders of magnitude speedups for rectification even on small vocabularies. Perhaps most importantly, we can now apply reliable, high-quality factorizations of high-dimensional data sets on laptop-grade hardware, massively increasing the applicability and potential use of these algorithms.

CHAPTER 5

WEIGHTED K-MEANS FOR ELECTRONIC STRUCTURE CALCULATION

5.1 Abstract

The recently developed interpolative separable density fitting (ISDF) decomposition is a powerful way for compressing the redundant information in the set of orbital pairs, and has been used to accelerate quantum chemistry calculations in a number of contexts. The key ingredient of the ISDF decomposition is to select a set of non-uniform grid points, so that the values of the orbital pairs evaluated at such grid points can be used to accurately interpolate those evaluated at all grid points. The set of non-uniform grid points, called the interpolation points, can be automatically selected by a QR factorization with column pivoting (QRCP) procedure. This is the computationally most expensive step in the construction of the ISDF decomposition. In this work, we propose a new approach to find the interpolation points based on the centroidal Voronoi tessellation (CVT) method, which offers a much less expensive alternative to the QRCP procedure when ISDF is used in the context of hybrid functional electronic structure calculations. The CVT method only uses information from the electron density, and can be efficiently implemented using a K-Means algorithm. We find that this new method achieves comparable accuracy to the ISDF-QRCP method, at a cost that is negligible in the overall hybrid functional calculations. For instance, for a system containing 1000 silicon atoms simulated using the HSE06 hybrid functional on 2000 computational cores, the cost of QRCP-based method for finding the interpolation points costs 38.1 s, while the CVT procedure only takes 0.7 s. We also find that the ISDF-CVT method also enhances the smoothness of the potential energy surface in the context of *ab initio* molecular dynamics (AIMD) simulations with hybrid functionals.

5.2 Introduction

Orbital pairs of the form $\{\varphi_i(\mathbf{r})\psi_j(\mathbf{r})\}_{i,j=1}^N$, where φ_i, ψ_j are single particle orbitals, appear ubiquitously in quantum chemistry. A few examples include the Fock exchange operator, the MP2 amplitude, and the polarizability operator [74, 59]. When N is proportional to the number of electrons N_e in the system, the total number of orbital pairs is $N^2 \sim O(N_e^2)$. On the other hand, the number of degrees of freedom needed to resolve all orbital pairs on a dense grid is only $O(N_e)$. Hence as N_e becomes large, the set of all orbital pairs contains apparent redundant information. In order to compress the redundant information and to design more efficient numerical algorithms, many algorithms in the past few decades have been developed. Pseudospectral decomposition [61, 72], Cholesky decomposition [15, 44, 6, 58], density fitting (DF) or resolution of identity (RI) [71, 77], and tensor hypercontraction (THC) [67, 68] are only a few examples towards this goal. When the single particle orbitals φ_i, ψ_j are already localized functions, “local methods” or “linear scaling methods” [28, 19, 31, 64] can be applied to construct such decomposition with cost that scales linearly with respect to N_e . Otherwise, the storage cost of the matrix to represent all orbital pairs on a grid is already $O(N_e^3)$, and the computational cost of compressing the orbital pairs is then typically $O(N_e^4)$.

Recently, Lu and Ying developed a new decomposition called the interpolative separable density fitting (ISDF) [56], which takes the following form

$$\varphi_i(\mathbf{r})\psi_j(\mathbf{r}) \approx \sum_{\mu=1}^{N_\mu} \zeta_\mu(\mathbf{r}) (\varphi_i(\hat{\mathbf{r}}_\mu)\psi_j(\hat{\mathbf{r}}_\mu)). \quad (5.1)$$

For a given \mathbf{r} , if we view $[\psi_i(\mathbf{r})\psi_j(\mathbf{r})]$ as a row of the matrix $\{\psi_i\psi_j\}$ discretized on a dense grid, then the ISDF decomposition states that all such matrix rows can be approximately expanded using a linear combination of matrix rows with respect to a selected set of *interpolation points* $\{\hat{\mathbf{r}}_\mu\}_{\mu=1}^{N_\mu}$. The coefficients of such linear combination, or *inter-*

polating vectors, are denoted by $\{\zeta_\mu(\mathbf{r})\}_{\mu=1}^{N_\mu}$. Here N_μ can be interpreted as the numerical rank of the ISDF decomposition. Compared to the standard density fitting method, the three-tensor $\{\varphi_i(\hat{\mathbf{r}}_\mu)\psi_j(\hat{\mathbf{r}}_\mu)\}$ with three indices i, j, μ takes a separable form. This reduces the storage cost of the decomposed tensor from $\mathcal{O}(N_e^3)$ to $\mathcal{O}(N_e^2)$, and the computational cost from $\mathcal{O}(N_e^4)$ to $\mathcal{O}(N_e^3)$. Note that if the interpolation points $\{\hat{\mathbf{r}}_\mu\}_{\mu=1}^{N_\mu}$ are chosen to be on a uniform grid, then the ISDF decomposition reduces to the pseudospectral decomposition, where $N_\mu \sim \mathcal{O}(N_e)$ but with a large preconstant. For instance, the pseudospectral decomposition can be highly inefficient for molecular systems, where the grid points in the vacuum contribute nearly negligibly to the orbital pairs. On the other hand, by selecting the interpolation points carefully, e.g. through a randomized QR factorization with column pivoting (QRCP) procedure [29], the number of interpolation points can be significantly reduced. The QRCP based ISDF decomposition has been applied to accelerate a number of applications, at least in the context of pseudopotential approximation where the wavefunctions are smooth, including two-electron integral computation [56], correlation energy in the random phase approximation [55], density functional perturbation theory [52], and hybrid density functional calculations [40]. For example, when iterative solvers are used for hybrid density functional calculations, the Fock exchange operator V_X defined in terms of a set of orbitals $\{\varphi_i\}$ needs to be repeatedly applied to another set of Kohn-Sham orbitals $\{\varphi_j\}$

$$(V_X[\{\varphi_i\}]\psi_j)(\mathbf{r}) = - \sum_{i=1}^{N_e} \varphi_i(\mathbf{r}) \int K(\mathbf{r}, \mathbf{r}') \varphi_i(\mathbf{r}') \psi_j(\mathbf{r}') d\mathbf{r}'. \quad (5.2)$$

where $K(\mathbf{r}, \mathbf{r}')$ is the kernel for the Coulomb or the screened Coulomb operator. The integration in Eq. 5.2 is often carried out by solving Poisson-like equations, using e.g. a fast Fourier transform (FFT) method, and the computational cost is $\mathcal{O}(N_e^3)$ with a large preconstant. This is typically the most time consuming component in hybrid functional calculations, and can be accelerated by the ISDF decomposition for the orbital pairs $\{\varphi_i\psi_j\}$.

In Ref. [56], the interpolation points and the interpolation vectors are determined simultaneously through randomized QRCP applied to $\{\psi_i(\mathbf{r})\psi_j(\mathbf{r})\}$ directly. We recently found that the randomized QRCP procedure has $O(N_e^3)$ complexity but with a relatively large preconstant, and may not be competitive enough when used repeatedly. In order to overcome such difficulty, we proposed a different approach in Ref. [40] that determines the two parts separately and reduces the computational cost. We use the relatively expensive randomized QRCP procedure to find the interpolation points in advance, and only recompute the interpolation vectors whenever $\{\psi_i(\mathbf{r})\psi_j(\mathbf{r})\}$ has been updated using an efficient least squares procedure that exploits the separable nature of the matrix to be approximated. As a result, we can significantly accelerate hybrid functional calculations using the ISDF decomposition in all but the first SCF iteration.

In this work, we further remove the need of performing the QRCP decomposition completely and, hence, significantly reduce the computational cost. Note that an effective choice of the set of interpolation points should satisfy the following two conditions. (1) The distribution of the interpolation points should roughly follow the distribution of the electron density. In particular, there should be more points when the electron density is high, and less or even zero points if the electron density is very low. (2) The interpolation points should not be very close to each other. Otherwise, matrix rows represented by the interpolation points are nearly linearly dependent, and the matrix formed by the interpolation vectors will be highly ill-conditioned. The QRCP procedure satisfies both (1) and (2) simultaneously, and thus is an effective way for selecting the interpolation points. Here we demonstrate that (1) and (2) can also be satisfied through a much simpler centroidal Voronoi tessellation (CVT) procedure applied to a weight vector such as the electron density.

The Voronoi tessellation technique has been widely used in computer science [11],

and scientific and engineering applications such as image processing[25], pattern recognition [66], and numerical integration [14]. The concept of Voronoi tessellation can be simply understood as follows. Given a discrete set of weighted points, the CVT procedure divides a domain into a number of regions, each consisting of a collection of points that are closest to its weighted centroid. Here we choose the electron density as the weight, and the centroids as the interpolation points. The centroids must be located where the electron density is significant, and hence satisfy the requirement (1). The centroids are also mutually separated from each other by a finite distance due to the nearest neighbor principle, and hence satisfy the requirement (2). Although detailed analysis of the error stemming from such a choice of interpolation points is very difficult for general nonlinear functions, we find that the CVT procedure approximately minimizes the residual of the ISDF decomposition in Eq. 5.1. In practice, the CVT procedure only applies to one vector (the electron density) instead of $O(N_e^2)$ vectors and hence is very efficient.

We apply the ISDF-CVT method to accelerate hybrid functional calculations in a planewave basis set. We perform such calculations for different systems with insulating (liquid water), semiconducting (bulk silicon), and metallic (disordered silicon aluminum alloy) characters, as well as ab initio molecular dynamics (AIMD) simulations. We find that the ISDF-CVT method achieves similar accuracy to that obtained from the ISDF-QRCP method, with significantly improved efficiency. For instance, for a bulk silicon system containing 1000 silicon atoms computed on 2000 computational cores with kinetic energy cutoff being 10 Ha, the QRCP procedure finds the interpolation points with 38.1 s, while the CVT procedure only takes 0.7 s. Since the solution of the CVT procedure is continuous with respect to changes in the electron density, we also find that the CVT procedure produces a smoother potential energy surface than that by the QRCP procedure in the context of AIMD simulations.

The remainder of the paper is organized as follows. We briefly introduce the ISDF decomposition in §5.3. In §5.4 we describe the ISDF-CVT procedure and its implementation for hybrid functional calculations. We present numerical results of the ISDF-CVT method in §5.5, and conclude in §5.6.

5.3 Interpolative Separable Density Fitting (ISDF) decomposition

In this section, we briefly introduce the ISDF decomposition [56] evaluated using the method developed in Ref. [40], which employs a separate treatment of the interpolation points and interpolation vectors.

First, assume the interpolation points $\{\hat{\mathbf{r}}_\mu\}_{\mu=1}^{N_\mu}$ are known, then the interpolation vectors can be efficiently evaluated using a least squares method as follows. Using a linear algebra notation, Eq. 5.1 can be written as

$$\mathbf{Z} \approx \Theta \mathbf{C}, \quad (5.3)$$

where each column of \mathbf{Z} is given by $Z_{ij}(\mathbf{r}) = \varphi_i(\mathbf{r})\psi_j(\mathbf{r})$ sampled on a dense real space grids $\{\mathbf{r}_i\}_{i=1}^{N_g}$, and $\Theta = [\zeta_1, \zeta_2, \dots, \zeta_{N_\mu}]$ contains the interpolating vectors. Each column of \mathbf{C} indexed by (i, j) is given by

$$[\varphi_i(\hat{\mathbf{r}}_1)\psi_j(\hat{\mathbf{r}}_1), \dots, \varphi_i(\hat{\mathbf{r}}_\mu)\psi_j(\hat{\mathbf{r}}_\mu), \dots, \varphi_i(\hat{\mathbf{r}}_{N_\mu})\psi_j(\hat{\mathbf{r}}_{N_\mu})]^T. \quad (5.4)$$

Eq. 5.3 is an overdetermined linear system with respect to the interpolation vectors Θ . The least squares approximation to the solution is given by

$$\Theta = \mathbf{Z} \mathbf{C}^T (\mathbf{C} \mathbf{C}^T)^{-1}. \quad (5.5)$$

It may appear that the matrix-matrix multiplications $\mathbf{Z} \mathbf{C}^T$ and $\mathbf{C} \mathbf{C}^T$ take $O(N_e^4)$ operations because the size of \mathbf{Z} is $N_g \times N^2$ and the size of \mathbf{C} is $N_\mu \times N^2$. However, both

multiplications can be carried out with fewer operations due to the separable structure of \mathbf{Z} and \mathbf{C} . The computational complexity for computing the interpolation vectors is $O(N_e^3)$, and numerical results indicate that the preconstant is also much smaller than that involved in hybrid functional calculations [40]. Hence the interpolation vectors can be obtained efficiently using the least squares procedure.

The problem for finding a suitable set of interpolation points $\{\hat{r}_\mu\}_{\mu=1}^{N_\mu}$ can be formulated as the following linear algebra problem. Consider the discretized matrix \mathbf{Z} of size $N_g \times N^2$, and find N_μ rows of \mathbf{Z} so that the rest of the rows of \mathbf{Z} can be approximated by the linear combination of the selected N_μ rows. This is called an interpolative decomposition [20], and a standard method to achieve such a decomposition is the QR factorization with column pivoting (QRCP) procedure [20] as

$$\mathbf{Z}^T \boldsymbol{\Pi} = \mathbf{Q} \mathbf{R}. \quad (5.6)$$

Here \mathbf{Z}^T is the transpose of \mathbf{Z} , \mathbf{Q} is an $N^2 \times N_g$ matrix that has orthonormal columns, \mathbf{R} is an upper triangular matrix, and $\boldsymbol{\Pi}$ is a permutation matrix chosen so that the magnitude of the diagonal elements of \mathbf{R} form a non-increasing sequence. The magnitude of each diagonal element of \mathbf{R} indicates how important the corresponding column of the permuted \mathbf{Z}^T is, and whether the corresponding grid point should be chosen as an interpolation point. The QRCP factorization can be terminated when the $(N_\mu + 1)$ -th diagonal element of \mathbf{R} becomes less than a predetermined threshold. The leading N_μ columns of the permuted \mathbf{Z}^T are considered to be linearly independent numerically. The corresponding grid points are chosen as the interpolation points. The indices for the chosen interpolation points $\{\hat{r}_\mu\}$ can be obtained from indices of the nonzero entries of the first N_μ columns of the permutation matrix $\boldsymbol{\Pi}$.

The QRCP decomposition satisfies the requirements (1) and (2) discussed in §5.2. First, QRCP permutes matrix columns of \mathbf{Z}^T with large norms to the front, and pushes

matrix columns of \mathbf{Z}^T with small norms to the back. Note that the square of the vector 2-norm of the column of \mathbf{Z}^T labeled by \mathbf{r} is just

$$\sum_{i,j=1}^N \varphi_i^2(\mathbf{r})\varphi_j^2(\mathbf{r}) = \left(\sum_{i=1}^N \varphi_i^2(\mathbf{r}) \right) \left(\sum_{j=1}^N \psi_j^2(\mathbf{r}) \right). \quad (5.7)$$

In the case when φ_i, ψ_j are the set of occupied orbitals, the norm of each column of \mathbf{Z}^T is simply the electron density. Hence the interpolation points chosen by QRCP will occur where the electron density is significant. Second, once a column is selected, all other columns are immediately orthogonalized with respect to the chosen column. Hence nearly linearly dependent matrix columns will not be selected repeatedly. As a result, the interpolation points chosen by QRCP are well separated spatially.

It turns out that the direct application of the QRCP procedure (Eq. 5.6) still requires $O(N_e^4)$ computational complexity. The key idea used in Ref. [56] to lower the cost is to randomly subsample columns of the matrix \mathbf{Z} to form a smaller matrix $\tilde{\mathbf{Z}}$ of size $N_g \times \tilde{N}_\mu$, where \tilde{N}_μ is only slightly larger than N_μ . Applying the QRCP procedure to this subsampled matrix $\tilde{\mathbf{Z}}$ approximately yields the choice of interpolation points, but the computational complexity is reduced to $O(N_e^3)$. In the context of hybrid density functional calculations, the cost of the randomized QRCP method can be comparable to that of applying the exchange operator in the planewave basis set [40]. However, the ISDF decomposition can still significantly reduce the computational cost, since the interpolation points only need to be performed once for a fixed geometric configuration.

5.4 Centroidal Voronoi Tessellation based ISDF decomposition

In this section, we demonstrate that the interpolation points can also be selected from a Voronoi tessellation procedure. For a d -dimensional space, the Voronoi tessellation partitions a set of points $\{\mathbf{r}_i\}_{i=1}^{N_g}$ in \mathbb{R}^d into a number of disjoint cells. The partition is

based on the distance of each point to a finite set of points, called its generators. In our context, let $\{\hat{\mathbf{r}}_\mu\}_{\mu=1}^{N_\mu}$ denote such a set of generators, and the corresponding cell, C_μ , of a given generator $\hat{\mathbf{r}}_\mu$ is defined as a cluster of points

$$C_\mu = \{\mathbf{r}_i \mid \text{dist}(\mathbf{r}_i, \hat{\mathbf{r}}_\mu) < \text{dist}(\mathbf{r}_i, \hat{\mathbf{r}}_\nu) \text{ for all } \nu \neq \mu\}. \quad (5.8)$$

The distance can be chosen to be any metric, e.g. the L_2 distance as $\text{dist}(\mathbf{r}, \mathbf{r}') = \|\mathbf{r} - \mathbf{r}'\|_2$. In the case when the distances of a point \mathbf{r} to $\hat{\mathbf{r}}_\mu, \hat{\mathbf{r}}_\nu$ are exactly the same, we may arbitrarily assign \mathbf{r} to one of the clusters.

The Centroidal Voronoi tessellation (CVT) is a specific type of Voronoi tessellation in which the generator $\hat{\mathbf{r}}_\mu$ is chosen to be the centroid of its cell. Given a weight function $\rho(\mathbf{r})$ (such as the electron density), the centroid of a cluster C_μ is defined as

$$\mathbf{c}(C_\mu) = \frac{\sum_{\mathbf{r}_j \in C_\mu} \mathbf{r}_j \rho(\mathbf{r}_j)}{\sum_{\mathbf{r}_j \in C_\mu} \rho(\mathbf{r}_j)}. \quad (5.9)$$

Combined with the L_2 distance, CVT can be viewed as a minimization problem over both all possible partition of the cells and the centroids as [57]

$$\{C_\mu^*, \mathbf{c}_\mu^*\} = \arg \min_{\{C_\mu, \mathbf{c}_\mu\}} \sum_{\mu=1}^{N_\mu} \sum_{\mathbf{r}_k \in C_\mu} \rho(\mathbf{r}_k) \|\mathbf{r}_k - \mathbf{c}_\mu\|^2, \quad (5.10)$$

and the interpolation points are then chosen to be the minimizers $\hat{\mathbf{r}}_\mu = \mathbf{c}_\mu(C_\mu^*) = \mathbf{c}_\mu^*$. Following the discussion in §5.2, the electron density as the weight function (5.10) enforces that the interpolation points should locate at points where the electron density is significant and hence satisfies the requirement (1). Since the cells C_μ^* are disjoint, the centroids \mathbf{c}_μ^* are also separated by a finite distance away from each other and hence satisfies the requirement (2). Because the ISDF decomposition is a highly nonlinear process, in general we cannot expect the choice of interpolation points from CVT decomposition to maximally reduce the error of the decomposition. Instead, we demonstrate that the choice of the interpolation points from CVT approximately minimizes the residual

for the ISDF decomposition, and hence provides a heuristic solution to the problem of finding interpolation points.

Theorem 3. *When the set of electron orbitals $\{\varphi_i\}$ are Lipschitz continuous, CVT method approximately minimizes the residual error of the ISDF decomposition.*

Proof. For simplicity we assume the limiting case where $\varphi_i = \psi_i$, and hence each row of Z is $Z(\mathbf{r}) = [\varphi_i(\mathbf{r})\varphi_j(\mathbf{r})]_{i,j=1}^N$.

Now suppose we cluster all matrix rows of Z into sub-collections $\{C_\mu\}_{\mu=1}^{N_\mu}$, and for each C_μ we choose a representative matrix row $Z(r_\mu)$. Then the error of the ISDF can be approximately characterized as

$$R = \sum_{\mu=1}^{N_\mu} \sum_{\mathbf{r}_k \in C_\mu} \|Z(\mathbf{r}_k) - \text{Proj}_{\text{Span}\{Z(r_\mu)\}} Z(\mathbf{r}_k)\|^2, \quad (5.11)$$

where the projection is defined according to the L_2 inner product as

$$\text{Proj}_{\text{Span}\{Z(r_\mu)\}} Z(\mathbf{r}_k) = \frac{Z(\mathbf{r}_k) \cdot Z(r_\mu)}{Z(r_\mu) \cdot Z(r_\mu)} Z(r_\mu). \quad (5.12)$$

Let Φ be the $N_g \times N$ matrix with each row $\Phi(\mathbf{r}) = [\varphi_i(\mathbf{r})]_{i=1}^N$, then the electron density $\rho(\mathbf{r})$ is equal to $\Phi(\mathbf{r}) \cdot \Phi(\mathbf{r})$. Using the relation

$$Z(r_\mu) \cdot Z(r_\mu) = (\Phi(r_\mu) \cdot \Phi(r_\mu))^2 = \rho(r_\mu)^2, \quad (5.13)$$

we have

$$R = \sum_{\mu=1}^{N_\mu} \sum_{\mathbf{r}_k \in C_\mu} \rho(\mathbf{r}_k)^2 \left(1 - \frac{(\Phi(\mathbf{r}_k) \cdot \Phi(r_\mu))^4}{\rho(\mathbf{r}_k)^2 \rho(r_\mu)^2} \right) \quad (5.14)$$

$$= \sum_{\mu=1}^{N_\mu} \sum_{\mathbf{r}_k \in C_\mu} \rho(\mathbf{r}_k)^2 [1 - \cos^4(\theta(\mathbf{r}_k, r_\mu))]. \quad (5.15)$$

Here $\theta(\mathbf{r}_k, r_\mu)$ is the angle between the vectors $\Phi(\mathbf{r}_k)$ and $\Phi(r_\mu)$. Since

$$\rho(\mathbf{r}_k)[1 - \cos^4(\theta(\mathbf{r}_k, r_\mu))] \leq 2\Phi(\mathbf{r}_k) \cdot \Phi(\mathbf{r}_k) \sin^2(\theta(\mathbf{r}_k, r_\mu)) \quad (5.16)$$

$$\leq 2 \|\Phi(\mathbf{r}_k) - \Phi(r_\mu)\|^2, \quad (5.17)$$

we have

$$R \leq 2 \sum_{\mu=1}^{N_\mu} \sum_{r_k \in C_\mu} \rho(r_k) \|\Phi(r_k) - \Phi(r_\mu)\|^2 \quad (5.18)$$

$$\approx 2 \sum_{\mu=1}^{N_\mu} \sum_{r_k \in C_\mu} \rho(r_k) \|\nabla_r \Phi(r_\mu)\|^2 \|r_k - r_\mu\|^2. \quad (5.19)$$

If we bound the gradient of $\Phi(\mathbf{r})$ by its Lipschitz constant, or simply neglect the spatial inhomogeneity in the electron orbitals, we arrive at the minimization criterion for the centroidal Voronoi tessellation decomposition. \square

Many algorithms have been developed to efficiently compute the Voronoi tessellation [60]. One most widely used method is the Llyod's algorithm [54], which in discrete case is equivalent to the K-Means algorithm [57]. The K-Means algorithm is an iterative method that greedily minimizes the objective by taking alternating steps between $\{C_\mu\}$ and $\{\mathbf{c}_\mu\}$. In this work, we adopt a weighted version of the K-Means algorithm, which is demonstrated in Algorithm 7. Note that the K-Means algorithm can be straightforwardly parallelized. We distribute the grid points evenly at the beginning. The classification step is the most time consuming step, and can be locally computed for each group of grid points. After this step, the weighted sum and total weight of all clusters can be reduced from and broadcast to all processors for the next iteration.

In order to demonstrate the CVT procedure, we consider the weight function $\rho(\mathbf{r})$ given by the summation of four Gaussian functions in a 2D domain. The initial choice of centroids, given by 40 uniformly distributed random points, together with its associated Voronoi tessellation are plotted in Figure 5.1 (a). Figure 5.1 (b) demonstrates the converged centroids and the associated Voronoi tessellation using the weighted K-Means algorithm. We observe that the centroids concentrate on where the weight function is significant, and are well-separated.

Algorithm 7: Weighted K-Means Algorithm to Find Interpolation Points for Density Fitting

Input : Grid points $\{\mathbf{r}_i\}_{i=1}^{N_g}$, Weight function $\rho(\mathbf{r})$, Initial centroids $\{\mathbf{c}_\mu^{(0)}\}$

Output: Interpolation points $\{\hat{\mathbf{r}}_\mu\}_{\mu=1}^{N_\mu}$

Set $t \leftarrow 0$

do

Classification step:

for $i = 1$ **to** N_g

| Assign point \mathbf{r}_i to the cluster $C_\mu^{(t)}$ **if** $\mathbf{c}_\mu^{(t)}$ is the closest centroid to \mathbf{r}_i

end

Update step:

for $\mu = 1$ **to** N_μ

| $\mathbf{c}_\mu^{(t+1)} \leftarrow \sum_{\mathbf{r}_j \in C_\mu^{(t)}} \mathbf{r}_j \rho(\mathbf{r}_j) / \sum_{\mathbf{r}_j \in C_\mu^{(t)}} \rho(\mathbf{r}_j)$

end

Set $t \leftarrow t + 1$

while $\{\mathbf{c}_\mu^{(t)}\}$ not converged and maximum steps not reached

for $\mu = 1$ **to** N_μ

| Set $\hat{\mathbf{r}}_\mu \leftarrow \mathbf{c}_\mu^{(t)}$

end

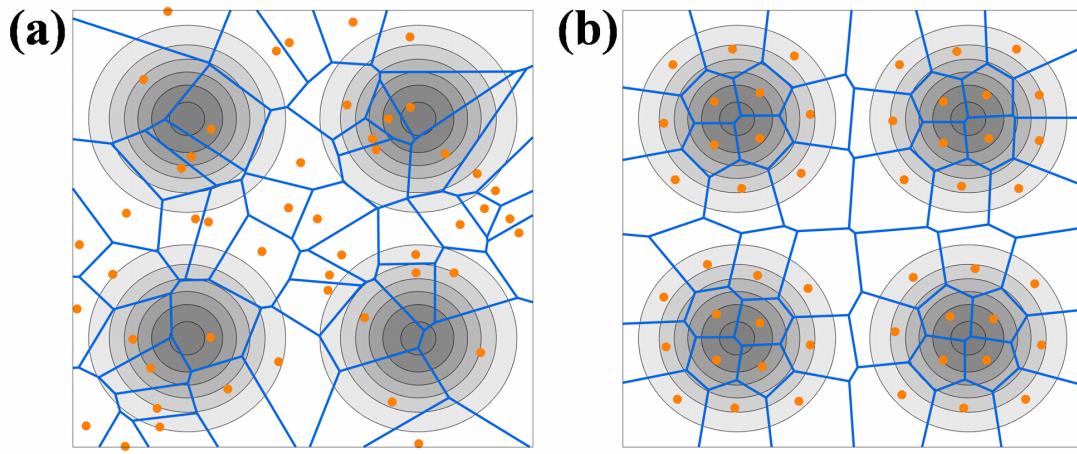


Figure 5.1: Schematic illustration of the CVT procedure in a 2D domain, including (a) initial random choice of centroids and Voronoi tessellation and centroidal Voronoi tessellation generated by the weighted K-Means algorithm. The weight function is given by the linear superposition of 4 Gaussian functions.

We also show how the interpolation points are placed and moved in real chemical systems, i.e. the ammonia-borane (BH_3NH_3) decomposition reaction process. Figure 5.2 (a) shows the electron density of the molecule at the compressed, equilibrium, and dissociated configurations, respectively, according to the energy landscape in Fig. 5.2 (c). We plot the interpolation points found by the weighted K-Means algorithm in Fig. 5.2 (b). At the compressed configuration, all the interpolation points are distributed evenly around the molecule. As the bond length increases, some interpolation points are transferred from BH_3 to NH_3 . Finally at the dissociated configuration, NH_3 has more interpolation points around the molecule, since there are more electrons in NH_3 than BH_3 . Along the decomposition reaction process, both the transfer of the interpolation points and the potential energy landscape are smooth with respect to the change of the bond length.

5.5 Numerical results

We demonstrate the accuracy and efficiency of the ISDF-CVT method for hybrid functional calculations by using the DGDFT (Discontinuous Galerkin Density Functional Theory) software package [51, 38, 39, 12, 78]. DGDFT is a massively parallel electronic structure software package designed for large scale DFT calculations involving up to tens of thousands of atoms. It includes a self-contained module called PWDFT for performing planewave based electronic structure calculations (mostly for benchmark and validation purposes). We implemented the ISDF-CVT method in PWDFT. We use the Message Passing Interface (MPI) to handle data communication. We use the Hartwigsen-Goedecker-Hutter (HGH) norm-conserving pseudopotential [33]. The atomic valence electron configuration is $1s^1$ for the H atom, $2s^22p^1$ for the B atom, $2s^22p^3$ for the N atom, $2s^22p^4$ for the O atom, $3s^23p^2$ for the Si atom in our DFT calcu-

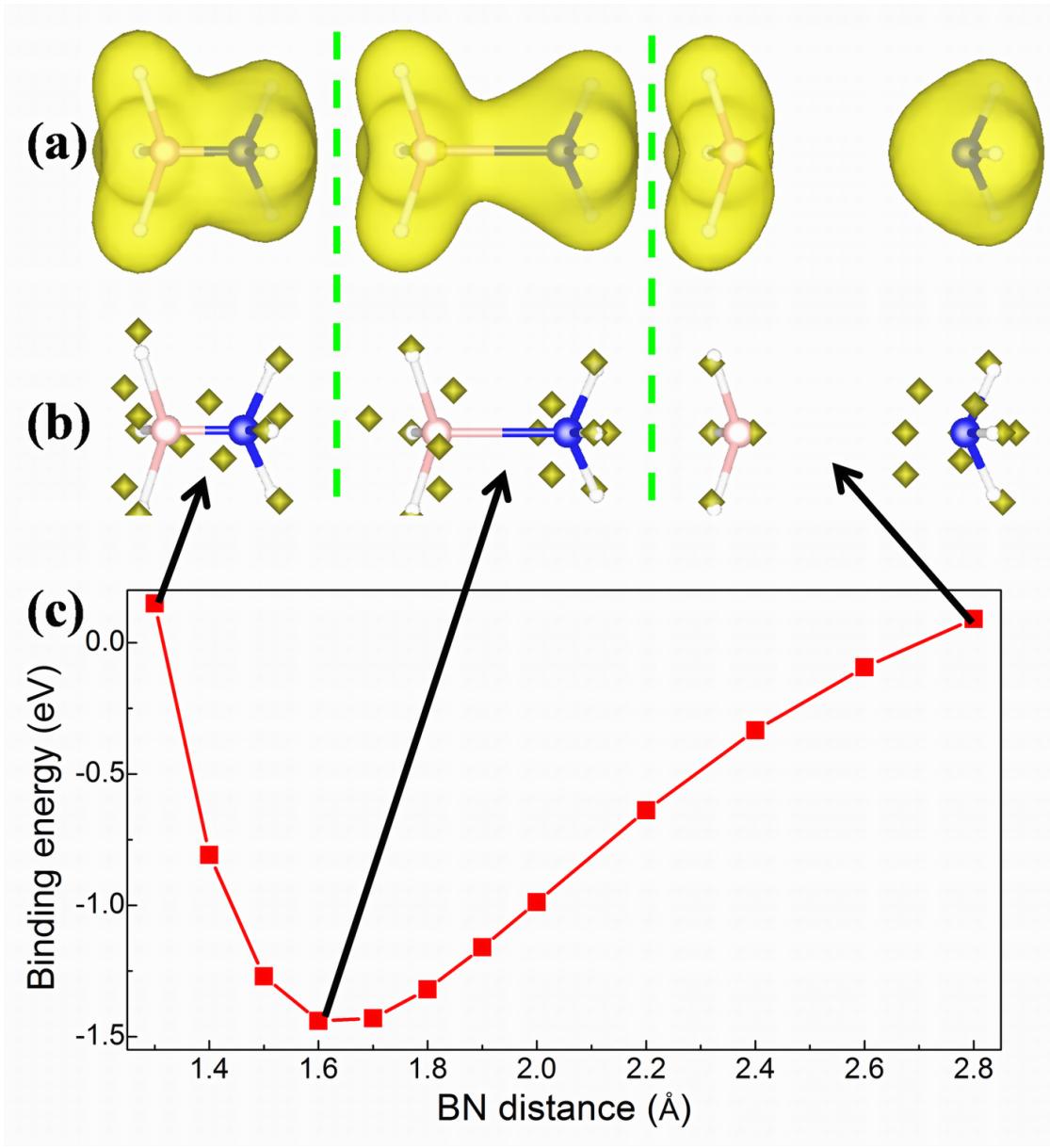


Figure 5.2: The decomposition reaction process of BH_3NH_3 computed with hybrid functional (HSE06) calculations by using the CVT procedure to select interpolation points, including (a) the electron density (yellow isosurfaces), (b) the interpolation points (yellow squares) $\{\hat{r}_\mu\}_{\mu=1}^{N_\mu}$ ($N_\mu = 8$) selected from the real space grid points $\{r_i\}_{i=1}^{N_g}$ ($N_g = 100^3$ and $E_{\text{cut}} = 60$ Ha) when the BN distance respectively is 1.3, 1.7 and 2.8 Å and (c) the binding energy as a function of BN distance for BH_3NH_3 in a $10 \text{ \AA} \times 10 \text{ \AA} \times 10 \text{ \AA}$ box. The white, pink and blue pink balls denote hydrogen, boron and nitrogen atoms, respectively.

lations, respectively. All calculations use the HSE06 functional [34], carried out on the Edison systems at the National Energy Research Scientific Computing Center (NERSC). Each node consists of two Intel “Ivy Bridge” processors with 24 cores in total and 64 gigabyte (GB) of memory. Our implementation only uses MPI. The number of cores is equal to the number of MPI ranks used in the simulation.

In this section, we demonstrate the performance of the ISDF-CVT method for accelerating hybrid functional calculations by using three types of systems [41]. They consist of bulk silicon systems (Si_{64} , Si_{216} , and Si_{1000}), a bulk water system with 64 molecules $((\text{H}_2\text{O})_{64})$, and a disordered silicon aluminum alloy system ($\text{Al}_{176}\text{Si}_{24}$). Bulk silicon systems (Si_{64} , Si_{216} and Si_{1000}) and bulk water system $((\text{H}_2\text{O})_{64})$ are semiconducting with a relatively large energy gap $E_{\text{gap}} > 1.0$ eV, and the $\text{Al}_{176}\text{Si}_{24}$ system is metallic with a small energy gap $E_{\text{gap}} < 0.1$ eV. All systems are closed shell systems, and the number of occupied bands is $N_{\text{band}} = N_e/2$, where N_e is the number of valence electrons. In order to compute the energy gap in the systems, we also include two unoccupied bands in all calculations.

5.5.1 Accuracy: Si_{216} and $\text{Al}_{176}\text{Si}_{24}$

We demonstrate the accuracy of the CVT-based ISDF decomposition in the hybrid functional calculation for semiconducting Si_{216} and metallic $\text{Al}_{176}\text{Si}_{24}$ systems, respectively. Although there is no general theoretical guarantee for the convergence of the K-Means algorithm and the convergence can depend sensitively on the initialization [9, 10], we find that, in the current context, initialization to have little impact on the final accuracy of the approximation. Hence we use random initialization for the K-Means algorithm. In all calculations, the adaptively compressed exchange (ACE) technique is used to acceler-

ate hybrid functional calculations without loss of accuracy [53]. The results obtained in this work are labeled as ACE-ISDF (CVT), which are compared against those obtained from the previous work based on the QRCP decomposition [40] labeled as ACE-ISDF (QRCP). In both cases, we introduce a rank parameter c to control the trade-off between efficiency and accuracy, by setting the number of interpolation points $N_\mu = cN_e$. We measure the error using the valence band maximum (VBM) energy level, the conduction band minimum (CBM) energy level, the energy gap, the Hartree-Fock exchange energy, the total energy, and the atomic forces, respectively. We remark that, in ISDF-CVT and ISDF-QRCP, the atomic force is computed directly using the Hellmann-Feynman formula, thereby neglects the Pulay force contribution from the change of the interpolation points. On the other hand, there is no Pulay contribution in the ACE formulation, and the Hellmann-Feynman force F_I^{ACE} can be used as the reference solution.

The last three quantities are defined as

$$\begin{aligned}\Delta E_{\text{HF}} &= \left| E_{\text{HF}}^{\text{ACE-ISDF (CVT)}} - E_{\text{HF}}^{\text{ACE}} \right| / N_A \\ \Delta E &= \left| E^{\text{ACE-ISDF (CVT)}} - E^{\text{ACE}} \right| / N_A \\ \Delta F &= \max_I \| F_I^{\text{ACE-ISDF (CVT)}} - F_I^{\text{ACE}} \| \end{aligned}$$

where N_A is the number of atoms and I is the atom index.

Table 5.1 shows that the accuracy of the ACE-ISDF (CVT) method can systematically improve as the rank parameter c increases. When the rank parameter is large enough (≥ 20.0), the results from ACE-ISDF (CVT) are fully comparable (the energy error is below 10^{-6} Ha/atom and the force error is below 10^{-5} Ha/Bohr) to those obtained from the benchmark calculations. Furthermore, for a moderate choice of the rank parameter $c = 6.0$, the error of the energy per atom reaches below the chemical accuracy of 1 kcal/mol (1.6×10^{-3} Ha/atom), and the error of the force is around 10^{-3} Ha/Bohr. This is comparable to the accuracy obtained from ACE-ISDF (QRCP), and to e.g. linear

scaling methods for insulating systems with reasonable amount of truncation needed to achieve significant speedup [22]. In fact, when compared with ACE-ISDF (QRCP) in Figure 5.3, we find that the CVT based ISDF decomposition achieves slightly higher accuracy, though there is no theoretical guarantee for this to hold in general. The last column of Table 5.1 shows the runtime of the K-Means algorithm. As c increases, the number of interpolation points as well as the number of cells increases proportionally. Hence we observe that the runtime of K-Means scales linearly with respect to c .

5.5.2 Efficiency: Si₁₀₀₀

We report the efficiency of the ISDF-CVT method by performing hybrid DFT calculations for a bulk silicon system with 1000 atoms ($N_{\text{band}} = 2000$) on 2000 computational cores as shown in Table 5.2, with respect to various choices of the kinetic energy cutoff (E_{cut}). With the number of interpolation points fixed at $N_\mu = 12000$, both QRCP and K-Means scales linearly with the number of grid points N_g . Yet the runtime of K-Means is around two orders of magnitude faster than QRCP. The determination of interpolation vectors, which consists of solving a least-square problem, previously costs a fifth of the ISDF runtime but now becomes the dominating component in CVT-based ISDF decomposition. Notice that the ISDF method allows us to reduce the number of Poisson-like equations from $N_e^2 = 4 \times 10^6$ to $N_\mu = 12000$, which results in a significant speedup in terms of the cost of the FFT operations.

Table 5.1: Accuracy of ACE-ISDF Based Hybrid Functional Calculations (HSE06) Obtained by Using the CVT method To Select Interpolation Points, with Varying Rank Parameter c for Semiconducting Si_{216} and Metallic $\text{Al}_{176}\text{Si}_{24}$ Systems^a.

c	E_{VBM}	E_{CBM}	E_{gap}	ΔE_{HF}	ΔE	ΔF	T_{KMEANS}
ACE-ISDF: Semiconducting Si_{216} ($N_{\text{band}} = 432$)							
4.0	6.7467	8.3433	-1.5967	2.69E-03	3.08E-03	5.04E-03	0.228
5.0	6.6852	8.2231	-1.5379	9.46E-04	1.12E-03	2.29E-03	0.248
6.0	6.6640	8.1522	-1.4882	3.76E-04	4.62E-04	1.05E-03	0.301
7.0	6.6550	8.1163	-1.4613	1.55E-04	1.98E-04	6.49E-04	0.312
8.0	6.6510	8.1030	-1.4520	7.33E-05	9.55E-05	3.07E-04	0.349
9.0	6.6490	8.0980	-1.4490	3.60E-05	4.96E-05	2.30E-04	0.398
10.0	6.6479	8.0959	-1.4480	1.78E-05	2.64E-05	1.30E-04	0.477
12.0	6.6472	8.0945	-1.4473	4.46E-06	8.91E-06	8.37E-05	0.530
16.0	6.6469	8.0937	-1.4468	1.51E-07	1.41E-06	3.20E-05	0.773
20.0	6.6468	8.0935	-1.4467	4.06E-07	3.33E-07	1.20E-05	0.830
24.0	6.6468	8.0935	-1.4467	2.99E-07	1.06E-07	5.18E-06	0.931
ACE	6.6468	8.0934	-1.4466	0.00E+00	0.00E+00	0.00E+00	-
ACE-ISDF: Metallic $\text{Al}_{176}\text{Si}_{24}$ ($N_{\text{band}} = 312$)							
4.0	7.9258	8.0335	-0.1076	3.80E-03	4.03E-03	8.01E-03	0.430
5.0	7.8537	7.9596	-0.1059	1.60E-03	1.69E-03	3.18E-03	0.535
6.0	7.8071	7.9127	-0.1056	6.07E-04	6.39E-04	1.48E-03	0.611
7.0	7.7843	7.8860	-0.1017	2.07E-04	2.17E-04	1.03E-03	0.731
8.0	7.7749	7.8749	-0.1000	7.43E-05	7.77E-05	4.40E-04	0.948
9.0	7.7718	7.8710	-0.0992	3.02E-05	3.20E-05	1.98E-04	0.947
10.0	7.7709	7.8697	-0.0989	1.48E-05	1.60E-05	1.80E-04	1.096
12.0	7.7703	7.8690	-0.0987	4.64E-06	5.60E-06	8.51E-05	1.305
16.0	7.7702	7.8688	-0.0986	6.35E-07	1.41E-06	3.24E-05	1.646
20.0	7.7701	7.8687	-0.0986	1.70E-08	5.30E-07	1.91E-05	2.037
ACE	7.7701	7.8687	-0.0986	0.00E+00	0.00E+00	0.00E+00	-

^a The unit for VBM (E_{VBM}), CBM (E_{CBM}) and the energy gap E_{gap} is eV. The unit for the error in the Hartree-Fock exchange energy ΔE_{HF} and the total energy ΔE is Ha/atom, and the unit for the error in atomic forces ΔF is Ha/Bohr. We use the results from the ACE-enabled hybrid functional calculations as the reference. The last column shows the time (in seconds) for K-Means with different c values, with 434 cores for Si_{216} and 314 cores for $\text{Al}_{176}\text{Si}_{24}$ on Edison.

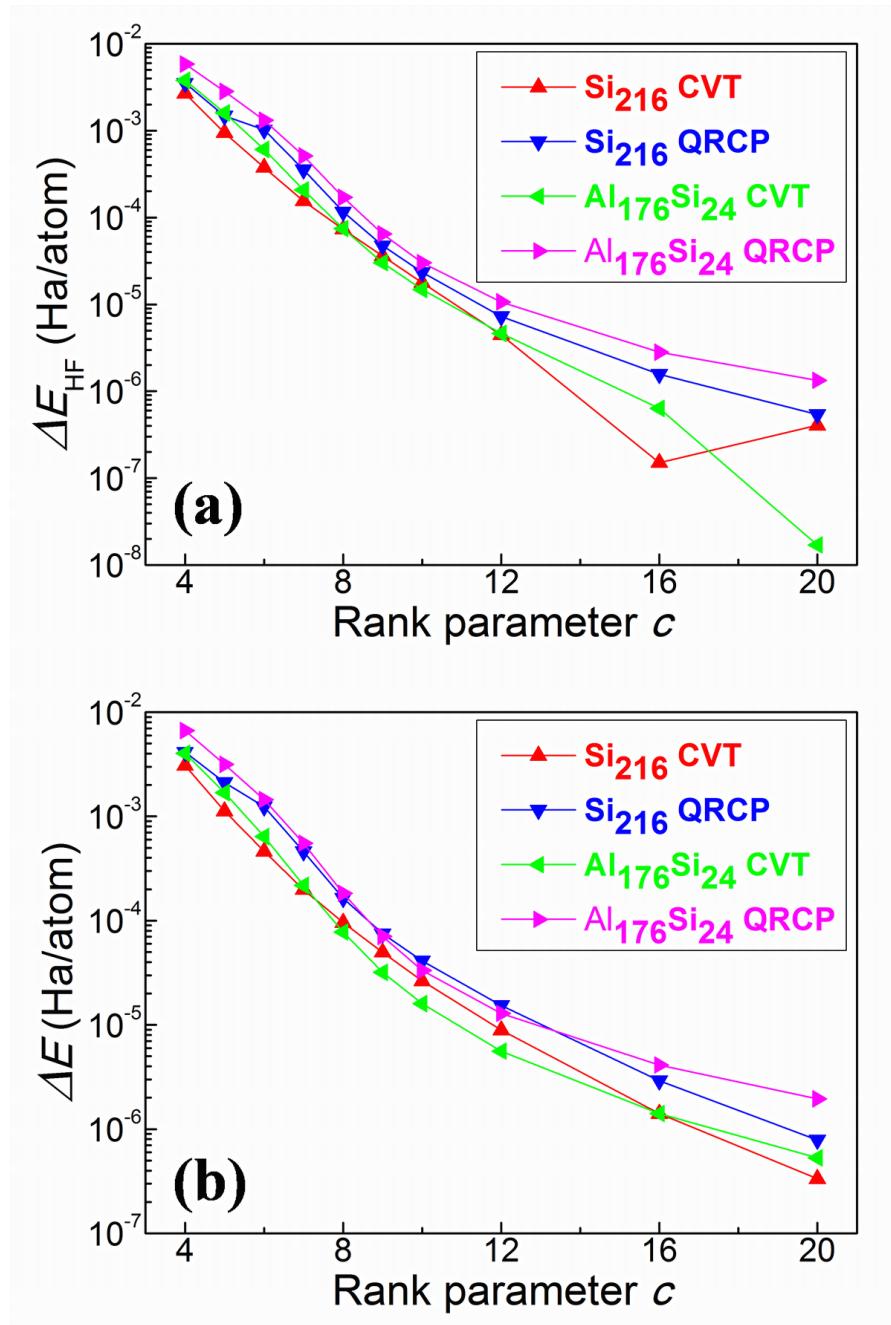


Figure 5.3: The accuracy of ACE-ISDF based hybrid functional calculations (HSE06) obtained by using the CVT and QRCP procedures to select the interpolation points, with varying rank parameter c from 4 to 20 for Si_{216} and $Al_{176}Si_{24}$, including the error of (a) Hartree-Fock exchange energy ΔE_{HF} (Ha/atom) and (b) total energy ΔE (Ha/atom).

Table 5.2: Wall Clock Time (in seconds) Spent in the Components of the ACE-ISDF and ACE Enabled Hybrid DFT Calculations Related to the Exchange Operator, for Si_{1000} on 2002 Edison cores at Different E_{cut} Levels^a.

Si_{1000}		ACE-ISDF		ACE
E_{cut}	N_g	IP_{QRCP}	$\text{IP}_{\text{KMEANS}}$	IV (FFT)
10	74^3	38.06	0.70	12.48 (0.33)
20	104^3	126.39	1.24	36.48 (0.71)
30	128^3	240.87	2.03	68.50 (1.43)
40	148^3	434.16	3.26	108.18 (3.10)

^a Interpolation points are selected via either the QRCP or CVT procedure with the same rank parameter $c = 6.0$. N_g is the number of grid points in real space.

5.5.3 Ab Initio Molecular Dynamics: Si_{64} and $(\text{H}_2\text{O})_{64}$

In this section, we demonstrate the accuracy of the ACE-ISDF (CVT) method in the context of AIMD simulations for a bulk silicon system Si_{64} under the NVE ensemble [26], and a liquid water system $(\text{H}_2\text{O})_{64}$ under the NVT ensemble [26], respectively. For the Si_{64} system, the initial MD structure (initial temperature $T = 300 \text{ K}$) is optimized by hybrid DFT calculations, and we perform the simulation ($E_{\text{cut}} = 20 \text{ Ha}$) for 1.0 ps with a MD time step of 1.0 fs. For the $(\text{H}_2\text{O})_{64}$ system, we perform the simulation ($E_{\text{cut}} = 60 \text{ Ha}$) for 2.0 ps with a MD time step of 0.5 fs to sample the radial distribution function after equilibrating the system starting from a prepared initial guess [24]. In this case, the Van der Waals (VdW) interaction is modeled at the level of the DFT-D2 method [30]. We use a single level Nose-Hoover thermostat [65, 35] at $T = 295 \text{ K}$, and the choice of mass of the Nose-Hoover thermostat is 85000 au.

In the AIMD simulation, the interpolation points need to be recomputed for each atomic configuration. At the initial MD step, although the initialization strategy does not impact the accuracy of the physical observable, it can affect the convergence rate of the K-Means algorithm. We measure the convergence in terms of the fraction of

points that switch clusters during two consecutive iterations. Figure 5.4 (a) shows the convergence of the K-Means algorithm with interpolation points initially chosen from a random distribution and from the QRCP solution, respectively. We find that the K-Means algorithm spends around half the number of iterations to wait for 0.1% of the points to settle on the respective clusters. However, these points often belong to the boundary of the clusters and have little effect on the positions of the centroids (interpolation points). Therefore, we decide to terminate K-Means algorithm whenever the fraction of points that switch clusters falls below the 0.1% threshold. It is evident that QRCP initialization leads to faster convergence than random sampling. However, in the AIMD simulation, a very good initial guess of the interpolation points can be simply obtained from those from the previous MD step. Figure 5.4 (b) shows that the number of K-Means iterations in the MD simulation can be very small, which demonstrates the effectiveness of this initialization strategy.

Figure 5.5 (a) shows that both the CVT-based and QRCP-based ISDF decomposition lead to controlled energy drift, defined as $E_{\text{drift}}(t) = (E_{\text{tot}}(t) - E_{\text{tot}}(0))/E_{\text{tot}}(0)$. In the NVE simulation on bulk silicon system Si_{64} , the energy drift per atom is 6.6×10^{-5} , 7.5×10^{-5} and 2.5×10^{-5} Ha/ps respectively for the ISDF-CVT, ISDF-QRCP, and the conventional nested two-level SCF iteration procedure, indicating that ISDF is a promising method for reducing the cost of hybrid functional calculations with controllable loss of accuracy. Figure 5.5 (b) shows the total potential energy obtained by the three methods along the MD trajectory, and the difference among the three methods is more noticeable. This is due to the fact that ISDF decomposition is a low rank decomposition for the pair product of orbitals, which leads to error in the Fock exchange energy and hence the total potential energy. Nonetheless, we find that such difference mainly results in a shift of the potential energy surface along the MD trajectory, and hence has little affect on physical observables defined via relative potential energy differences. Furthermore, the

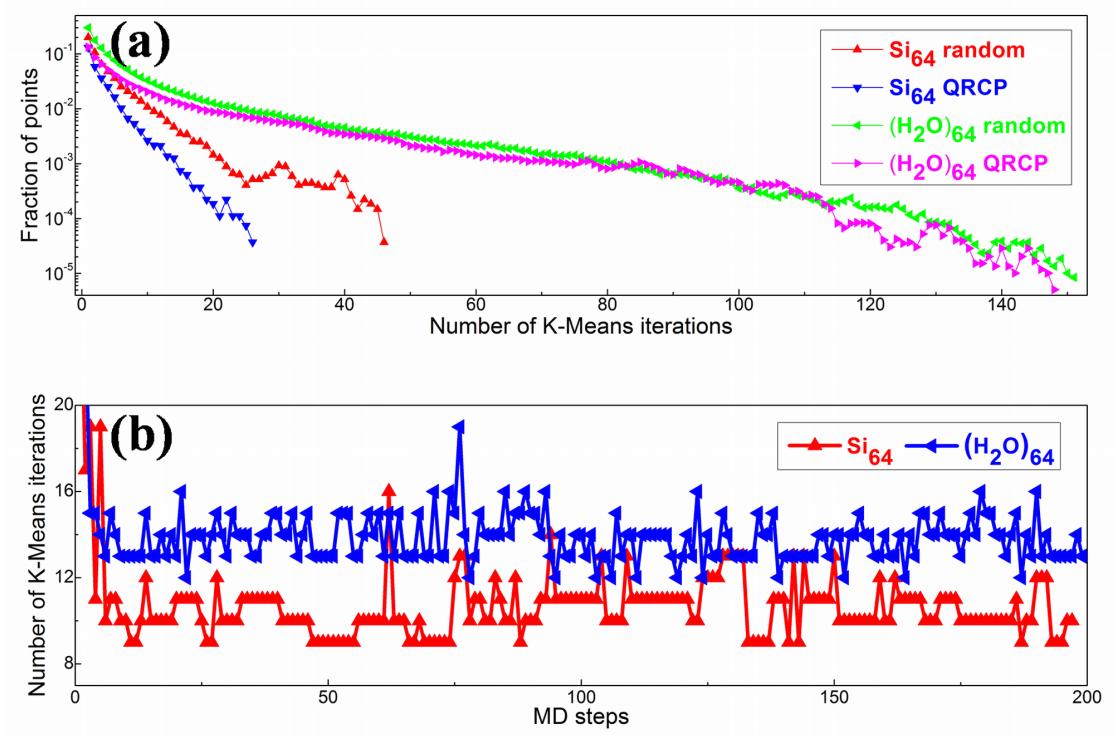


Figure 5.4: Comparison of the ISDF-CVT method by using either random or QRCP initialization for hybrid DFT AIMD simulations on bulk silicon system Si_{64} and liquid water system $(\text{H}_2\text{O})_{64}$, including (a) the fraction of points what switch cluster in each K-Means iteration and (b) the number of K-Means iterations during each MD step.

CVT method yields a potential energy trajectory that is much smoother compared to that obtained from QRCP. This is because the interpolation points obtained from CVT are driven by the electron density, which varies smoothly along the MD trajectory. Such properties do not hold for the QRCP method. This means that the CVT method can be more effective when a smooth potential energy surface is desirable, such as in the case of geometry optimization. The absolute error of the potential energy from the CVT method is coincidentally smaller than that from QRCP, but again we are not aware of any reason for this behavior to hold in general.

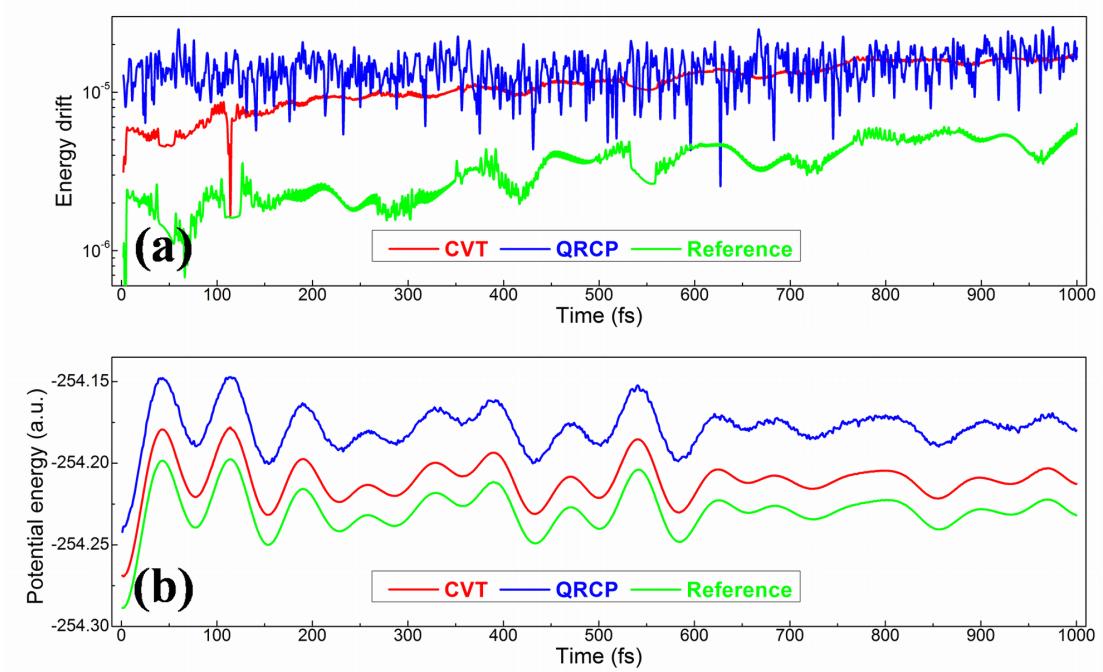


Figure 5.5: Comparison of hybrid HSE06 DFT AIMD simulations by using the ISDF-CVT and ISDF-QRCP methods as well as exact nested two-level SCF iteration procedure as the reference on the bulk silicon Si_{64} , including (a) relatively energy drift and (b) potential energy during MD steps.

We also apply the ACE-ISDF (CVT) and ACE-ISDF (QRCP) methods for hybrid DFT AIMD simulations on liquid water system $(\text{H}_2\text{O})_{64}$ under the NVT ensemble to sample the radial distribution function in Figure 5.6. We find that the results from all three methods agree very well, and our result is in quantitative agreement with previous hybrid functional calculations [24], which uses a different exchange-correlation functional (PBE0) and Van der Waals functional (TS-vdW) [75].

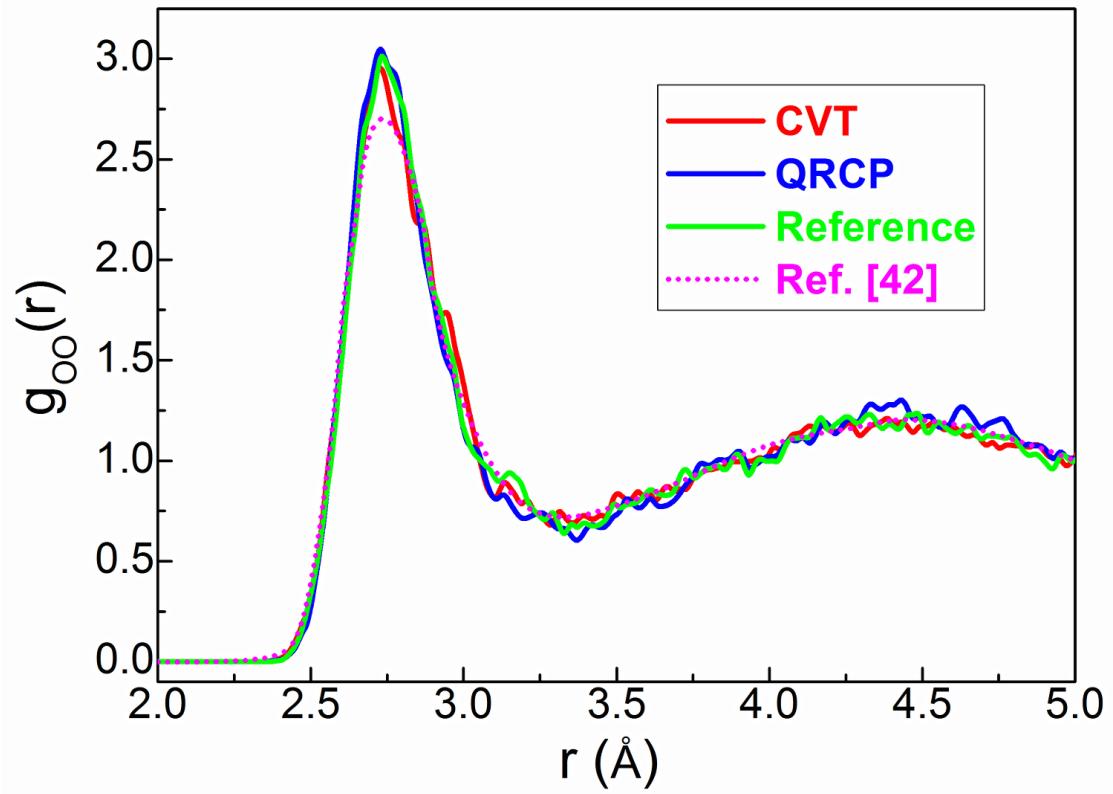


Figure 5.6: The oxygen-oxygen radial distribution functions $g_{OO}(r)$ of liquid water system (H_2O)₆₄ at $T = 295$ K obtained from hybrid HSE06 + DFT-D2 AIMD simulations with the ISDF-CVT and ISDF-QRCP methods, exact nested two-level SCF iteration procedure (as the reference) as well as previous hybrid PBE0 + TS-vdW calculation [24].

5.6 Conclusion

In this work, we demonstrate that the interpolative separable density fitting decomposition (ISDF) can be efficiently performed through a separated treatment of interpolation points and interpolation vectors. We find that the centroidal Voronoi tessellation method (CVT) provides an effective choice of interpolation points using only the electron density as the input information. The resulting interpolation points are by design inhomogeneous in the real space, concentrated at regions where the electron density is significant, and are well separated from each other. These are all key ingredients for obtaining a low

rank decomposition that is accurate and a well conditioned set of interpolation vectors. We demonstrate that the CVT-based ISDF decomposition can be an effective strategy for reducing the cost hybrid functional calculations for large systems. The CVT-based method achieves similar accuracy when compared with that obtained from QRCP, with significantly improved efficiency. Since the solution of the CVT method depends continuously with respect to the electron density, we also find that the CVT method produces a smoother potential energy surface than that by the QRCP method in the context of ab initio molecular dynamics simulation. Our analysis indicates that it might be possible to further improve the quality of the interpolation points by taking into account the gradient information in the weight vector. We also expect that the CVT-based strategy can also be useful in other contexts where the ISDF decomposition is applicable, such as ground state calculations with rung-5 exchange-correlation functionals, and excited state calculations. These will be explored in the future work.

5.7 Acknowledgments

This work was partly supported by the National Science Foundation under grant No. DMS-1652330, the DOE under grant No. DE-SC0017867, the DOE CAMERA project (L. L.), and by the DOE Scientific Discovery through Advanced Computing (SciDAC) program (K. D., W. H. and L. L.). The authors thank the National Energy Research Scientific Computing (NERSC) center and the Berkeley Research Computing (BRC) program at the University of California, Berkeley for making computational resources available. We thank Anil Damle and Robert Saye for useful discussions.

CHAPTER 6
CONCLUSION

BIBLIOGRAPHY

- [1] E A. Erosheva. Bayesian estimation of the grade of membership model. *Bayesian Stat.*, 7, 01 2003.
- [2] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. *Journal of machine learning research*, 9(Sep):1981–2014, 2008.
- [3] A. Anandkumar, D. Hsu, and S. Kakade. A method of moments for mixture models and hidden markov models. In *COLT*, 2012.
- [4] Anima Anandkumar, Dean P. Foster, Daniel Hsu, Sham Kakade, and Yi-Kai Liu. A spectral algorithm for latent Dirichlet allocation. In *NIPS*, 2012.
- [5] Animashree Anandkumar, Daniel Hsu, and Sham M Kakade. A method of moments for mixture models and hidden markov models. In *Conference on Learning Theory*, pages 33–1, 2012.
- [6] F. Aquilante, T. B. Pedersen, and R. Lindh. Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals. *J. Chem. Phys.*, 126:194106, 2007.
- [7] S. Arora, R. Ge, and A. Moitra. Learning topic models – going beyond SVD. In *FOCS*, 2012.
- [8] Sanjeev Arora, Rong Ge, Yonatan Halpern, David Mimno, Ankur Moitra, David Sontag, Yichen Wu, and Michael Zhu. A practical algorithm for topic modeling with provable guarantees. In *ICML*, 2013.
- [9] David Arthur and Sergei Vassilvitskii. How slow is the k-means method? In *Proceedings of the twenty-second annual symposium on Computational geometry*, pages 144–153. ACM, 2006.

- [10] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.
- [11] Franz Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*, 23(3):345–405, 1991.
- [12] A. S. Banerjee, L. Lin, W. Hu, C. Yang, and J. E. Pask. Chebyshev polynomial filtered subspace iteration in the discontinuous galerkin method for large-scale electronic structure calculations. *J. Chem. Phys.*, 145:154101, 2016.
- [13] Trapit Bansal, Chiranjib Bhattacharyya, and Ravindran Kannan. A provable SVD-based algorithm for learning topics in dominant admixture corpus. In *NIPS*, 2014.
- [14] Axel D Becke. A multicenter numerical integration scheme for polyatomic molecules. *The Journal of chemical physics*, 88(4):2547–2553, 1988.
- [15] N. H. F. Beebe and J. Linderberg. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. Quantum Chem.*, 12:683, 1977.
- [16] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. *JMLR*, 2003.
- [17] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.
- [18] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.

- [19] D. R. Bowler and T. Miyazaki. O(n) methods in electronic structure calculations. *Rep. Prog. Phys.*, 75:036503, 2012.
- [20] T. F. Chan and P. C. Hansen. Some applications of the rank revealing QR factorization. *SIAM J. Sci. Statist. Comput.*, 13:727–741, 1992.
- [21] Jonathan Chang, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David M. Blei. Reading tea leaves: How humans interpret topic models. In *NIPS*, 2009.
- [22] William Dawson and François Gygi. Performance and accuracy of recursive subspace bisection for hybrid dft calculations in inhomogeneous systems. *J. Chem. Theory Comput.*, 11(10):4655–4663, 2013.
- [23] Weicong Ding, Prakash Ishwar, and Venkatesh Saligrama. Most large topic models are approximately separable. In *ITA, 2015*, pages 199–203. IEEE, 2015.
- [24] R. A. DiStasio, B. Santra, Z. Li, X. Wu, and R. Car. The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. *J. Chem. Phys.*, 141:084502, 2014.
- [25] Qiang Du, Max Gunzburger, Lili Ju, and Xiaoqiang Wang. Centroidal voronoi tessellation algorithms for image compression, segmentation, and multichannel restoration. *Journal of Mathematical Imaging and Vision*, 24(2):177–194, 2006.
- [26] J Willard Gibbs. *Elementary principles in statistical mechanics*. Courier Corporation, 2014.
- [27] Nicolas Gillis and Stephen A Vavasis. Fast and robust recursive algorithms for separable nonnegative matrix factorization. *IEEE transactions on pattern analysis and machine intelligence*, 36(4):698–714, 2014.

- [28] S. Goedecker. Linear scaling electronic structure methods. *Rev. Mod. Phys.*, 71:1085–1123, 1999.
- [29] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins Univ. Press, Baltimore, fourth edition, 2013.
- [30] S. Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27(15):1787–1799, 2006.
- [31] M. Guidon, J. Hutter, and J. Vandevondele. Auxiliary density matrix methods for Hartree-Fock exchange calculations. *J. Chem. Theory Comput.*, 6:2348–2364, 2010.
- [32] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.
- [33] C. Hartwigsen, S. Goedecker, and J. Hutter. Relativistic separable dual-space gaussian pseudopotentials from H to Rn. *Phys. Rev. B*, 58:3641, 1998.
- [34] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Erratum: "hybrid functionals based on a screened coulomb potential" [J. Chem. Phys. 118, 8207 (2003)]. *J. Chem. Phys.*, 124:219906, 2006.
- [35] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695, 1985.
- [36] Patrik O. Hoyer. Non-negative matrix factorization with sparseness constraints. *JMLR*, 2004.
- [37] Daniel Hsu, Sham M Kakade, and Tong Zhang. A spectral algorithm for learning

- hidden markov models. *Journal of Computer and System Sciences*, 78(5):1460–1480, 2012.
- [38] W. Hu, L. Lin, and C. Yang. DGDFT: A massively parallel method for large scale density functional theory calculations. *J. Chem. Phys.*, 143(12):124110, 2015.
- [39] W. Hu, L. Lin, and C. Yang. Edge reconstruction in armchair phosphorene nanoribbons revealed by discontinuous galerkin density functional theory. *Phys. Chem. Chem. Phys.*, 17(47):31397–31404, 2015.
- [40] Wei Hu, Lin Lin, and Chao Yang. Interpolative separable density fitting decomposition for accelerating hybrid density functional calculations with applications to defects in silicon. *J. Chem. Theory Comput.*, accepted, 2017.
- [41] Wei Hu, Lin Lin, and Chao Yang. Projected commutator diis method for accelerating hybrid functional electronic structure calculations. *J. Chem. Theory Comput.*, accepted, 2017.
- [42] Kejun Huang, Xiao Fu, and Nikolaos D. Sidiropoulos. Anchor-free correlated topic modeling: Identifiability and algorithm. In *NIPS*, 2016.
- [43] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul. Introduction to variational methods for graphical models. *Machine Learning*, pages 183–233, 1999.
- [44] H. Koch, A. Sánchez de Merás, and T. B. Pedersen. Reduced scaling in electronic structure calculations using cholesky decompositions. *J. Chem. Phys.*, 118:9481–9484, 2003.
- [45] Alex Kulesza, N Raj Rao, and Satinder Singh. Low-rank spectral learning. In *Artificial Intelligence and Statistics*, pages 522–530, 2014.

- [46] Thomas K Landauer, Peter W Foltz, and Darrell Laham. An introduction to latent semantic analysis. *Discourse processes*, 25(2-3):259–284, 1998.
- [47] Moontae Lee, David Bindel, and David Mimno. Robust spectral inference for joint stochastic matrix factorization. In *NIPS*, 2015.
- [48] Moontae Lee, David Bindel, and David Mimno. From correlation to hierarchy: Practical topic modeling via spectral inference. In *12th INFORMS Workshop on Data Mining and Decision Analytics*, 2017.
- [49] Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In *NIPS*, 2014.
- [50] Dawen Liang, Jaan Altosaar, Laurent Charlin, and David M Blei. Factorization meets the item embedding: Regularizing matrix factorization with item co-occurrence. In *Proceedings of the 10th ACM conference on recommender systems*, pages 59–66. ACM, 2016.
- [51] L. Lin, J. Lu, L. Ying, and W. E. Adaptive local basis set for kohn-sham density functional theory in a discontinuous galerkin framework i: Total energy calculation. *J. Comput. Phys.*, 231(4):2140–2154, 2012.
- [52] L. Lin, Z. Xu, and L. Ying. Adaptively compressed polarizability operator for accelerating large scale ab initio phonon calculations. *Multiscale Model. Simul.*, 15:29–55, 2017.
- [53] Lin Lin. Adaptively compressed exchange operator. *J. Chem. Theory Comput.*, 12(5):2242–2249, 2016.
- [54] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

- [55] J. Lu and K. Thicke. Cubic scaling algorithms for RPA correlation using interpolative separable density fitting. *J. Comput. Phys.*, 351:187 – 202, 2017.
- [56] J. Lu and L. Ying. Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost. *J. Comput. Phys.*, 302:329–335, 2015.
- [57] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proc. of the Fifth Berkeley Symp. On Math. Stat. and Prob.*, volume 1, pages 281–297, 1967.
- [58] S. Manzer, P. R. Horn, N. Mardirossian, and M. Head-Gordon. Fast, accurate evaluation of exact exchange: The occ-RI-K algorithm. *J. Chem. Phys.*, 143(2):024113, 2015.
- [59] R. Martin. *Electronic Structure – Basic Theory and Practical Methods*. Cambridge Univ. Pr., West Nyack, NY, 2004.
- [60] NN Medvedev. The algorithm for three-dimensional voronoi polyhedra. *Journal of computational physics*, 67(1):223–229, 1986.
- [61] R. B. Murphy, M. D. Beachy, R. A. Friesner, and M. N. Ringnalda. Pseudospectral localized möller–plesset methods: Theory and calculation of conformational energies. *J. Chem. Phys.*, 103:1481, 1995.
- [62] Radford M Neal. *Probabilistic inference using Markov chain Monte Carlo methods*, 1993.
- [63] Radford M Neal et al. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.

- [64] Frank Neese, Frank Wennmohs, Andreas Hansen, and Ute Becker. Efficient, approximate and parallel hartree–fock and hybrid dft calculations. a “chain-of-spheres” algorithm for the hartree–fock exchange. *Chem. Phys.*, 356:98–109, 2009.
- [65] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511, 1984.
- [66] Robert L Ogniewicz and Olaf Kübler. Hierarchic voronoi skeletons. *Pattern recognition*, 28(3):343–359, 1995.
- [67] R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill. Tensor hypercontraction. II. Least-squares renormalization. *J. Chem. Phys.*, 137:224106, 2012.
- [68] R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill. Discrete variable representation in electronic structure theory: Quadrature grids for least-squares tensor hypercontraction. *J. Chem. Phys.*, 138:194107, 2013.
- [69] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. GloVe: Global vectors for word representation. In *EMNLP*, 2014.
- [70] Jonathan K Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.
- [71] X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler. Resolution-of-identity approach to Hartree-Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New J. Phys.*, 14:053020, 2012.
- [72] G. Reynolds, T. J. Martinez, and E. A. Carter. Local weak pairs spectral and pseudospectral singles and doubles configuration interaction. *J. Chem. Phys.*, 105:6455, 1996.

- [73] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.
- [74] A. Szabo and N.S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. McGraw-Hill, New York, 1989.
- [75] A. Tkatchenko and M. Scheffler. Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.*, 102:073005, 2009.
- [76] M. Wainwright and M. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, pages 1–305, 2008.
- [77] F. Weigend. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.*, 4:4285–4291, 2002.
- [78] G. Zhang, L. Lin, W. Hu, C. Yang, and J. E. Pask. Adaptive local basis set for Kohn–Sham density functional theory in a discontinuous Galerkin framework ii: Force, vibration, and molecular dynamics calculations. *J. Comput. Phys.*, 335:426–443, 2017.