SqueezeFit Linear Program: Fast and Robust Label-aware Dimensionality Reduction

A Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Mathematical Sciences in the Graduate School of The Ohio State University

By

Tien-hsin Lu, B.S.

Graduate Program in Mathematical Science

The Ohio State University

2020

Thesis Committee:

Dr. Dustin G. Mixon, Advisor

Dr. Dongbin Xiu

Abstract

We introduce the SqueezeFit linear program as a fast and robust dimensionality reduction method. This program is inspired by both the SqueezeFit semi-definite program [10] and scGeneFit [3], which is a linear program version of SqueezeFit that has been used to classify single cell RNA-sequence data with a given structured partition. The original SqueezeFit semi-definite program has a strong theoretical background but it exhibits slow runtimes with large data sets. In contrast, scGeneFit performs efficiently and robustly with scRNA-seq data given either flat or hierarchical label partitions, but it does not have much theoretical justification for its performance. The SqueezeFit linear program fills this computational and theoretical gap. After providing new theoretical guarantees, we illustrate the performance of the SqueezeFit linear program on real-world gene expression data.

Dedication

This is dedicated to all of my loved ones.

Acknowledgments

I would like to thank my advisor, Dr. Dustin G. Mixon, for introducing me to the field of mathematical data science and for all of the guidance he has given me throughout the past year. I am grateful for all of the invaluable insight and advice he has provided for this project. I would also like to thank Dr. Dongbin Xiu for serving as a part of my committee.

Also, thank you to Dr. Mixon for connecting me with Dr. Soledad Villar and Dr. Bianca Dumitrascu. A special thank you goes to Dr.Villar and Dumitrascu for providing scientific data sets, which made the completion of this project possible.

I would like to acknowledge the Ohio State University, the Graduate School, and the Department of Mathematics for providing such an amazing opportunity for me to learn and grow as a researcher in the field of mathematical sciences.

I would also like to give a shout out to my dear boyfriend, Richard Crews, for keeping me sane during this most stressful time. I would also like to acknowledge my friends and colleagues in this graduate program for their encouragements during the past two years. I appreciate them for motivating me to complete my work and stay on a schedule.

Finally, I would also like to thank my family for their love and support throughout my entire life.

Vita

1996 .... Born, Beijing, China

2017 .... B.S. Mathematics, University of California, Irvine

2018-present .... Graduate Teaching Associate, Department of Mathematics
The Ohio State University.

Publications

Fields of Study

Major Field: Mathematical Science

Contents

List of Tables

Table **Page**

List of Figures

Chapter 1: Introduction

In machine learning, data with too many features leads to high dimensionality and possible over-fitting. Furthermore, high-dimensional data can be extremely difficult to visualize and analyze. Overall, dimensionality reduction is a necessity. Various dimensionality reduction methods, including Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA), help scientists and industries compress data and reduce computation time. However, there are certain problems that can not easily be resolved by the existing methods, which has motivated researchers to develop alternate dimensionality reduction methods.

1.1 Motivating Problem: Marker Gene Selection

Single cell RNA-sequencing (RNA-seq) provides a wealth of information. Studies and technologies have provided valuable insights regarding the characterization of single cells among complex cell populations. Existing methodologies mostly rely on markers – fluorescence that tie the nearby genes that we are interested in that will be quantified [3]. Given a lot of cells of different types and the gene expression of each cell over multiple genes, we are eager to know an accurate and efficient way to find the handful of genes (i.e., the markers) that best enables classification by cell type.

The existing technologies, however, either fail to consider the hierarchical relationship between various types of cell and the interrelationship in the expression patterns across different genes [4], or are rather redundant under the circumstances that is unnecessary to obtain all cell types and markers to maintain the hierarchy [3]. Thus, the key challenge is to fill the methodological gap, with given single cell RNA-seq data and the hierarchy of cell type labels, by a robust and efficient selection approach which finds a few markers that allow classification [3].

1.2 Issues with Modern Classification: Adversarial Examples

Our problem is to identify what signal in the data should be used by the classifier, and modern classification technologies, i.e., neural networks, have similarly demonstrated the need for help with signal classification. Studies have shown that various models are vulnerable against adversarial attacks [5]. That is, when a small perturbation is applied to the correctly classified example, these models tend to misclassify the perturbed examples with high confidence [5]. See Figure 1.1 for a demonstration of an adversarial example.

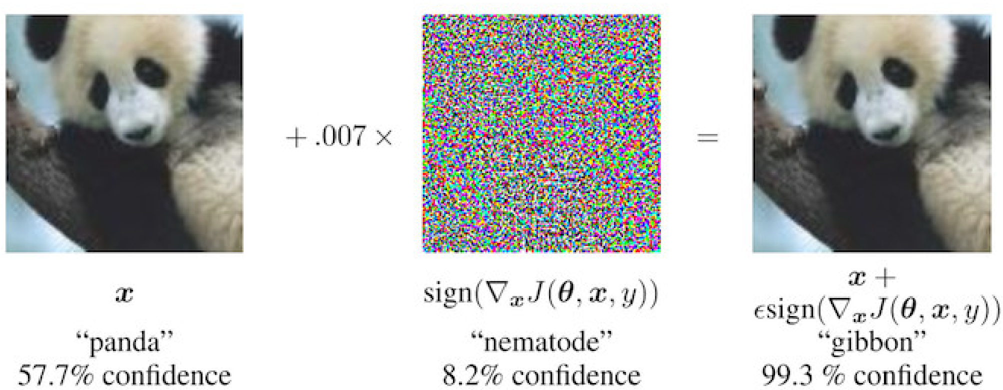

Figure 1.1: From [5]: GoogLeNet[14] changes its classification of the image with surprisingly high confidence after adding a slight perturbation to the image. Left: The model classifies the image as "panda" with 57.7% confidence; Middle: adding a small perturbation to the previous image; Right: the model classifies the slightly perturbed image as "gibbon" with 99.3% confidence.

One of the main reasons of the shared vulnerability against adversarial examples among various machine learning models, according to [5], is that the perturbations are aligned with the weight vectors of these models.

The adversarial examples shows our lack of understanding of the manner of neural networks. The challenge is to convey to the machine what portions of the signal are important. SqueezeFit (which will be introduced in the next section), however, is motivated by this and tries to overcome the shared weakness among machine learning methods, and communicate the signal precisely.

1.3 SqueezeFit

SqueezeFit is a label-aware dimensionality reduction method. Given data with various labels in a high-dimensional vector space, SqueezeFit finds a subspace of lower dimension on which the data can be projected while maintaining a desired distance between different points with different labels [10]. SqueezeFit, inspired by large margin nearest neighbor classification, relaxes the problem into a semi-definite program that recovers the projection. See Figure 1.2 for a visual demonstration of SqueezeFit comparing to other dimensionality reduction methods.

Following [10], consider a data set $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i\in\mathcal{I}}$ in $\mathbb{R}^d \times [k]$, where $\mathbf{x}_i$ is a sample in $\mathbb{R}^d$ with some index i, y_i is the corresponding label of $\mathbf{x}_i$ for all $i \in \mathcal{I}$, and $[k] = \{1, 2, ..., k\}$ represents the set of labels. SqueezeFit seeks the lowest dimensional subspace where samples with distinct labels are separated. Below, we show how this problem can be viewed as an optimization problem.

Let Π be an orthogonal projection onto some lower dimensional subspace. Let $\mathcal{Z}(\mathcal{D}) := \{\mathbf{x}_i - \mathbf{x}_j : i, j \in \mathcal{I}, y_i \neq y_j\}$ be the set of vector differences of samples with various labels,

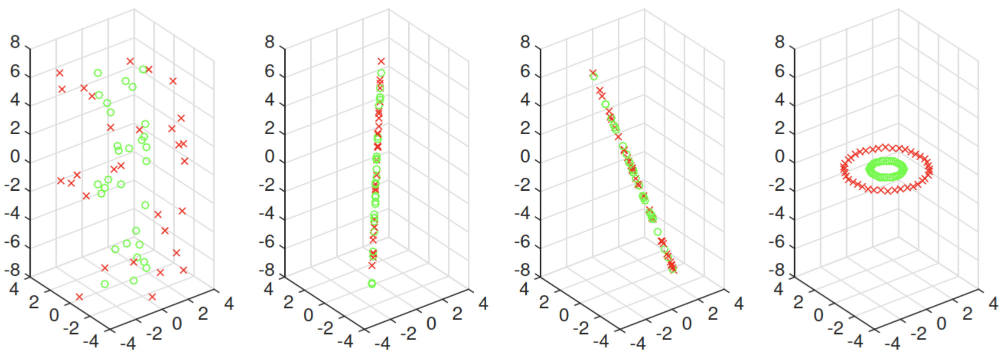

Figure 1.2: From [10]: (**Left**) Plot of data in $\mathbb{R}^3$; (**middle left**) Result of Principle Component Analysis (PCA). (**middle right**) Result of linear discriminant analysis (LDA). (**right**) SqueezeFit, different than PCA and LDA, projects onto a low-dimension subspace where desired distance between labels is maintained. SqueezeFit yields a good approximation and recovers the planted projection factor.

and let $\Delta > 0$ be a desired minimum separation between samples with different labels after the projection. We obtain the following optimization problem:

$$\textit{minimize} \quad \operatorname{rank}\Pi \quad \textit{subject to} \quad ||\Pi\mathbf{z}|| \geq \Delta \quad \forall \mathbf{z} \in \mathcal{Z}(\mathcal{D}), \quad \Pi^T = \Pi, \quad \Pi^2 = \Pi. \tag{1.1}$$

Following the convex relaxation of the previous equation from [10], **SqueezeFit Semi-definite Program** is obtained:

$$\textit{minimize} \quad \operatorname{tr}(M) \quad \textit{subject to} \quad \mathbf{z}^T M \mathbf{z} \geq \Delta^2 \quad \forall \mathbf{z} \in \mathcal{Z}(\mathcal{D}), \quad 0 \preccurlyeq M \preccurlyeq I. \qquad (\mathsf{sqz}(\mathcal{D}, \Delta))$$

SqueezeFit is not only amenable to theoretical analysis [10], but importantly, it can also be applied in scientific fields. In particular, the algorithm of SqueezeFit performed surprisingly well when dealing with the marker gene selection problem [3] mentioned in section 1.1. The adapted version of SqueezeFit to the marker gene problem, scGeneFit, is shown to be able to accurately discriminate cell types for a particular tissue type with known structured partition of cell type labels [3]. scGeneFit reduces $\mathsf{sqz}(\mathcal{D}, \Delta)$, the semi-definite program (SDP), to a linear program (LP) by taking advantage of an additional constraint given by the specific selection of marker genes. Following Theorem 7 from [10], SqueezeFit, in the context of the so-called projection factor recovery problem, is theoretically guaranteed to succeed with high probability within a certain range of signal-to-noise ratio (SNR).

According to [10], it takes 50 minutes for a standard Macbook Air 2013 to run SqueezeFit when the data set contains 800 data points with dimension 100. The SDP is indeed slow on large data sets, but it has a strong theoretical backing. In contrast, scGeneFit, a linear program version of SqueezeFit, is much faster and shows its robustness and efficiency in the context of the marker selection problem. Yet, the LP version has no developed theoretical foundation.

1.4 Our Approach

This thesis aims to fill the gap between SqueezeFit and scGeneFit by providing a theoretical treatment for an LP version of SqueezeFit. In addition to providing a detailed theoretical analysis, we will also illustrate several numerical experiments. As we will see, the LP version can decrease the running time while maintaining the performance of the original SDP, and has a known application (marker gene selection).

In the next chapter, we will derive the SqueezeFit Linear Program and the dual program, and discuss weak and strong duality using general convex theory. We will discuss SqueezeFit LP in the context of the projection factor recovery problem in chapter 3. In chapter 4, we will illustrate the results of several numerical implementations and experiments. A summary of our results and some insights of a few research questions for future work will be discussed in the last chapter.

Chapter 2: Derivation of SqueezeFit Linear Program

2.1 Review of General Convex Theory

SqueezeFit is convex optimization program. Before we discuss $\mathsf{sqzLP}(\mathcal{D}, \Delta)$, let us review some general convex theory. See [2] for more details. We will use the following terminologies, notations and definitions mentioned in this section throughout the development of the LP version of SqueezeFit.

Definition 1. (Convex function) A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if it satisfies,

$$f_i(\alpha \mathbf{x} + \beta \mathbf{y}) \leq \alpha f_i(\mathbf{x}) + \beta f_i(\mathbf{y}), \tag{2.1}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and for all $\alpha, \beta \in \mathbb{R}_{\geq 0}$ with $\alpha + \beta = 1$.

Definition 2. A **convex optimization problem** has the form

$$\textit{minimize} \quad f(\mathbf{x}) \quad \textit{subject to} \quad g_i(\mathbf{x}) \leq b_i, \quad i = 1, \ldots m, \tag{2.2}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the optimization variable, $b_1, \ldots b_m \in \mathbb{R}$ are the limits for the constraints, and the objective function $f : \mathbb{R}^n \to \mathbb{R}$, and the constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}\ \forall i = 1, \ldots m$ are convex.

Linear programs and semidefinite programs are both convex optimization problems. The general form of linear programs is the following:

$$\textit{minimize} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad \mathbf{a}_i^T\mathbf{x} \le b_i, \quad i = 1, 2, \ldots, m, \tag{LP}$$

where $\mathbf{c}, \mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$ are called the problem parameters describing the objective and constraint functions.

Definition 3. A set K is a **cone** if for every $\mathbf{x} \in K$ and $\alpha \ge 0$, we have $\alpha x \in K$, it is a **convex cone** if for every $\mathbf{x}, \mathbf{y} \in K$ and $\alpha_1, \alpha_2 \ge 0$, we have $\alpha_1\mathbf{x} + \alpha_2\mathbf{y} \in K$.

Definition 4. Let K be a set, then the **dual cone** of K is defined as $K^\star := \{\mathbf{y} : \mathbf{x}^T\mathbf{y} \ge 0 \quad \forall \mathbf{x} \in K\}$.

A **Cone Program** has the form of

$$\textit{minimize} \quad \mathbf{c}^T\mathbf{x} \quad \textit{subject to} \quad A\mathbf{x} - \mathbf{b} \in L, \quad \mathbf{x} \in K, \tag{2.3}$$

where L and K are cones. The corresponding **Dual Cone Program** is

$$\textit{maximize} \quad \mathbf{b}^T\mathbf{y} \quad \textit{subject to} \quad \mathbf{c} - A^T\mathbf{y} \in K^\star, \quad \mathbf{y} \in L^\star, \tag{2.4}$$

where $K^\star$ and $L^\star$ are the dual cones of K and L respectively.

The optimal value of the dual problem provides the best lower bound for the primal problem that can be obtained from a convex combination of the primal constraints. The following definitions describe the relationship between these two values:

Definition 5. (**Weak Duality**) Let $d^\star$ be the optimal value of the dual problem, and let $p^\star$ be the optimal value of the primal problem. If $d^\star \le p^\star$, we say that **weak duality** holds. The difference between $p^\star$ and $d^\star$, i.e., $p^\star - d^\star$ is called the **optimal duality gap** of the original problem.

When the duality gap is zero, we obtain strong duality:

Definition 6. (**Strong Duality**) If $d^\star = p^\star$, we say that **strong duality** holds.

Strong duality holds means that the best lower bound can be achieved by the dual problem, hence, we say that the duality is tight.

Definition 7. If a dual feasible (μ, v) is found, that means a lower bound on the optimal value of the primal problem is found. Thus, we say that a dual feasible point (μ, v) provides a **certificate** for that lower bound.

Definition 8. Let $x^\star$ be a primal optimal, and let $(\mu^\star, v^\star)$ be a dual optimal point when strong duality holds. Then the following condition

$$\mu_i^\star > 0 \Rightarrow f_i(x^\star) = 0 \quad \text{for each } i \tag{2.5}$$

or,

$$f_i(x^\star) < 0 \Rightarrow \mu_i^\star = 0 \quad \text{for each } i \tag{2.6}$$

is known as **complementary slackness**.

2.2 Motivation

SqueezeFit seeks the lowest dimensional subspace where samples with distinct labels are separated. In Chapter 1, we mentioned that SqueezeFit can be written as an optimization problem. Consider $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i\in\mathcal{I}}$ in $\mathbb{R}^d \times [k]$, where $\mathbf{x}_i$ is a sample in $\mathbb{R}^d$ with some index i, y_i is the corresponding label of $\mathbf{x}_i$ for all $i \in \mathcal{I}$, and $[k] = \{1, 2, ..., k\}$ is the set of labels. Let Π be an orthogonal projection onto some lower dimensional subspace, we have [10]

$$\textit{minimize} \quad \operatorname{rank}\Pi \quad \textit{subject to} \quad ||\Pi\mathbf{z}|| \geq \Delta \quad \forall \mathbf{z} \in \mathcal{Z}(\mathcal{D}), \quad \Pi^T = \Pi, \quad \Pi^2 = \Pi, \tag{2.7}$$

where $\mathcal{Z}(\mathcal{D}) := \{\mathbf{x}_i - \mathbf{x}_j : i, j \in \mathcal{I}, y_i \neq y_j\}$.

We aimed to resolve the difficulty of running the SDP SqueezeFit on large data sets, so we considered adding a new constraint to the program $\mathsf{sqz}(\mathcal{D}, \Delta)$ and make it a linear program (LP), which, in general, is faster than a SDP. The new constraint allows us to get both the speed and relevance to scGeneFit, which is a LP version of $\mathsf{sqz}(\mathcal{D}, \Delta)$ and is a convex relaxation of the previous equation. The constraint makes the optimization variable M to be a diagonal matrix due to the special objective of selecting marker genes [3]. We take inspirations from scGeneFit, and restrict M to be a diagonal matrix. We obtain

$$\textit{minimize} \quad \operatorname{tr}(M) \quad \textit{subject to} \quad \mathbf{z}^T M \mathbf{z} \geq \Delta^2 \quad \forall \mathbf{z} \in \mathcal{Z}(\mathcal{D}), \quad 0 \preccurlyeq M \preccurlyeq I, \quad M \textit{ diagonal.} \tag{2.8}$$

2.3 SqueezeFit Linear Program

To obtain the LP version of SqueezeFit, we want to adapt our (2.8) to a general LP form (see equation (LP)). Since we enforce M to be a diagonal matrix, we can "vectorize" M by its diagonal, and denote it as $\mathbf{m}$, i.e., $\mathbf{m}$ is a vector contains the diagonal entries of M. Hence, we obtain $\operatorname{tr}(M) = \sum_i \mathbf{m}_i$. We can simplify the above equation (2.8) to

$$\textit{minimize} \quad \sum_i \mathbf{m}(i) \quad \textit{subject to} \quad \sum_i \mathbf{z}(i)^2 \mathbf{m}(i) \geq \Delta^2 \quad \forall \mathbf{z} \in \mathcal{Z}(\mathcal{D}), \quad 0 \leq \mathbf{m}(i) \leq 1 \quad \forall i. \tag{$\mathsf{sqzLP}(\mathcal{D}, \Delta)$}$$

2.4 Deriving the Dual Cone Program of $\mathsf{sqzLP}(\mathcal{D}, \Delta)$

To derive the dual program of $\mathsf{sqzLP}(\mathcal{D}, \Delta)$, we follow (2.3) and (2.4) from Definition 4 in Section 2.1 by taking $\mathbf{x} = \mathbf{m}$, $\mathbf{c} = \mathbb{1}_d$, i.e., the all ones column vector of dimension d. Let $\mathbf{z}^{\circ 2}$ denote the element-wise square of vector $\mathbf{z}$. Let $K = \mathbb{R}^d_{\geq 0}$, and $L = \mathbb{R}^{|\mathcal{Z}|+d}_{\geq 0}$. Consider

$$A = \left[\begin{array}{c} \overline{Z} \\ \hline -I \end{array}\right], \quad \mathbf{b} = \left[\begin{array}{c} \Delta^2 \mathbb{1}_{|\mathcal{Z}|} \\ \hline -\mathbb{1}_d \end{array}\right],$$

where $\overline{Z} \in \mathbb{R}^{|\mathcal{Z}| \times d}$ has the rows $(\mathbf{z}^{\circ 2})^T$ for each $\mathbf{z} \in \mathcal{Z}(\mathcal{D})$ and I is the $d \times d$ identity matrix.

Plugging into the dual cone program format in (2.3), we obtain

$$\textit{maximize} \quad \left[\begin{array}{c} \Delta^2 \mathbb{1}_{|\mathcal{Z}|} \\ \hline -\mathbb{1}_d \end{array}\right]^T \mathbf{y} \quad \textit{subject to} \quad \mathbb{1}_d - \left[\, \overline{Z} \mid -I \,\right] \mathbf{y} \succcurlyeq 0, \quad \mathbf{y} \succcurlyeq 0. \tag{2.9}$$

Performing the inner product and combining the terms, we obtain a dual cone program for $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ when $\mathcal{Z}(\mathcal{D})$ is infinite:

$$\begin{aligned} \textit{sup} \quad & \Delta^2 \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) - \sum_{j \in [d]} \mathbf{v}_j \\ \textit{subject to} \quad & \mathbf{1} - \sum_{\mathbf{z} \in Z(D)} \mu(\mathbf{z}) \mathbf{z}^{\circ 2} + \mathbf{v} \geq \mathbf{0}, \quad \mathbf{v} \geq \mathbf{0}, \quad \mu \geq 0, \end{aligned} \tag{$\mathsf{dual}(\mathcal{D}, \Delta)$}$$

where $\mu : \mathcal{Z}(\mathcal{D}) \to \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^d$ are the decision variables.

Let's first verify weak duality:

Theorem 1. (Weak Duality) If $\mathbf{x}$ is primal feasible and $\mathbf{y}$ is dual feasible, and let $\mathbf{b}$ and $\mathbf{c}$ be what we defined above, then $\langle \mathbf{b}, \mathbf{y} \rangle \leq \langle \mathbf{c}, \mathbf{x} \rangle$, i.e., weak duality holds.

Proof. Let $\mathbf{x}, \mathbf{c}, \mathbf{y}, \mathbf{b}, L$ and K be what we defined above. Since $\mathbf{x}$ is primal feasible and $\mathbf{y}$ is dual feasible, $\mathbf{y} \in L^\star$, $A\mathbf{x} - \mathbf{b} \in L$. Then $\langle \mathbf{y}, A\mathbf{x} - \mathbf{b} \rangle \geq 0$ implies

$$\langle \mathbf{y}, A\mathbf{x} \rangle \geq \langle \mathbf{y}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{y} \rangle. \tag{2.10}$$

Then we have

$$\langle \mathbf{b}, \mathbf{y} \rangle \leq \langle \mathbf{y}, A\mathbf{x} \rangle = \langle A^T \mathbf{y}, \mathbf{x} \rangle \leq \langle \mathbf{c}, \mathbf{x} \rangle, \tag{2.11}$$

and hence weak duality. $\square$

2.5 Dual Certificate for SqueezeFit LP

Next, we want to show that $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ admits a dual certificate. Hence, the following definition from [10] is necessary for our next theorem:

Definition 9. (Following Definition 4 of [10]) We call the vectors with smallest length in $\mathcal{Z}(\mathcal{D})$, if exists, the **contact vectors** of $\mathcal{D}$, and we say the data set $\mathcal{D} = \{(x_i, y_i)\}_{i\in\mathcal{I}}$ is **Δ-fixed** if there exist $\mathbf{m} \in \arg\mathsf{sqzLP}(\mathcal{D}, \Delta)$ such that $\mathrm{diag}(\mathbf{m}^{1/2})x_i = x_i$ for all $i \in \mathcal{I}$.

Following [10], for Δ-fixed data, strong duality can be characterized with the following theorem:

Theorem 2. Let $\mathcal{D} = \{(x_i, y_i)\}_{i\in\mathcal{I}}$ be Δ-fixed. Then $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ admits a dual certificate $(\mu, \mathbf{v})$ if and only if the contact vectors (see Definition 9) of $\mathcal{D}$ span $\mathrm{span}\{x_i\}_{i\in\mathcal{I}}$.

Proof. (This proof follows the proof of Theorem 15 of [10].)

($\Leftarrow$) Pick any finite subset of the set of contact vectors of $\mathcal{D}$, denoted as $\mathcal{Z}_0$. Suppose $\mathcal{Z}_0$ spans $\mathrm{span}\{x_i\}_{i\in\mathcal{I}}$. Let $\xi := \sum_{\mathbf{z}\in\mathcal{Z}_0} \mathbf{z}^{\circ 2}$. Let $\eta = \min_{i\in\mathcal{T}} \sum_{z\in\mathcal{Z}_0} \mathbf{z}(i)^2$, where $\mathbf{z}(i)$ here denotes the i-th element of $\mathbf{z}$. Let

$$\mu(\mathbf{z}) = \begin{cases} \frac{1}{\eta} & \text{if } \mathbf{z} \in \mathcal{Z}_0 \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

and pick $\mathbf{v}$ such that

$$\mathbf{v}(i) = \begin{cases} \sum_i \mu(\mathbf{z})\mathbf{z}(i)^2 - 1 & \text{if } i \in \mathcal{T} \\ 0 & \text{otherwise} \end{cases} \tag{2.13}$$

One can easily show that (μ, v) is feasible in the dual problem with objective value $\sum_i \mathbf{m}(i)$. By Lemma 10 in [10], (μ, v) is therefore a dual certificate secured by weak duality.

($\Rightarrow$) Consider $\mathbf{m}_{opt} = (\mathbb{1}_{|\mathcal{T}|}, \mathbb{0}_{d-|\mathcal{T}|})^T \in \mathbb{R}^d$, where $\mathbf{m}_{opt}(i) = 1$ when $\mathbf{z}_i \neq \mathbf{0}$. We claim that if (μ, v) is feasible in 1.1, then so is $(\mu, \mathbf{v} \circ \mathbf{m}_{opt})$. The feasibility follows from

$$\mathbb{1} - \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) \mathbf{z}^{\circ 2} + \mathbf{v} \circ \mathbf{m}_{opt} = \mathbb{1} - \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) \mathbf{z}^{\circ 2} + \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) \mathbf{z}^{\circ 2} - \mathbf{m}_{opt} \succcurlyeq 0$$

and the objective value is larger since $\sum_i (\mathbf{v} \circ \mathbf{m}_{opt})(i) \leq \sum_i \mathbf{v}(i)$. Therefore, without loss of generality, $(\mu, \mathbf{v})$ has $\mathrm{im}(\mathrm{diag}(\mathbf{v})) \subseteq \mathrm{im}(\mathrm{diag}(\mathbf{v} \circ \mathbf{m}_{opt}))$.

Consider $\mathbf{q} := \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) \frac{\Delta^2}{\|\mathbf{z}\|^2} \mathbf{z}^{\circ 2} - \mathbf{v}$. Then $\sum_i \mathbf{q}(i) = |\mathcal{T}|$, which equals to the dual value of $(\mu, \mathbf{v})$. The orthogonal projection onto $\mathrm{span}\{\mathbf{x}_i\}_{i \in \mathcal{T}}$, denoted as Π, satisfies $\mathrm{im}(\mathrm{diag}(\mathbf{v})) \subseteq \mathrm{im}(\mathrm{diag}(\mathbf{v} \circ \mathbf{m}_{opt})) \subseteq \mathrm{im}(\Pi)$. Let

$$\begin{aligned} \mathbf{a} &:= \mathbb{1}_{\mathcal{T}} - \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z})(1 - \frac{\Delta^2}{\|\mathbf{z}\|^2}) \mathbf{z}^{\circ 2} - \mathbf{q} \\ &= \mathbb{1}_{\mathcal{T}} - \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) \mathbf{z}^{\circ 2} + \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) \frac{\Delta^2}{\|\mathbf{z}\|^2} \mathbf{z}^{\circ 2} - \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) \frac{\Delta^2}{\|\mathbf{z}\|^2} \mathbf{z}^{\circ 2} + \mathbf{v} \\ &= \mathbb{1}_{\mathcal{T}} - \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) \mathbf{z}^{\circ 2} + \mathbf{v} \geq \mathbf{0} \end{aligned}$$

By Lemma 10(i) in [10], Π is in the optimal set of $\mathsf{sqzLP}(\mathcal{D}, \Delta)$, and so $\mathrm{tr}(\Pi) = \sum_i \mathbf{q}(i)$. Therefore, the dual feasibility of $(\mu, \mathbf{v})$ implies a stronger inequality

$$0 \leq \sum_i \mathbf{a}(i) = - \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z})(\|\mathbf{z}\|^2 - \Delta^2) \leq 0. \tag{2.14}$$

Hence, we can conclude:

(a) We must have that $\mathbf{a} = \mathbf{0}$ since $\mathbf{a} \geq \mathbf{0}$ and $\sum_i \mathbf{a}(i) = 0$.

(b) $\sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z})(\|\mathbf{z}\|^2 - \Delta^2) = 0$ implies that $\mu(\mathbf{z}) \neq 0$ only when $\mathbf{z}$ is a contact vector of $\mathcal{D}$, and $\mu(\mathbf{z}) = 0$ otherwise, by Lemma 12 of [10].

Consider the span of contact vectors of $\mathcal{D}$, denoted as T, by conclusion (a), we have

$$\mathbb{1}_{\mathcal{T}} - \sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) \mathbf{z}^{\circ 2} + \mathbf{v} = 0,$$

$$\sum_{\mathbf{z} \in \mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) \mathbf{z}^{\circ 2} = \mathbb{1}_{\mathcal{T}} + \mathbf{v}.$$

By conclusion (b), we obtain:

$$\text{span}\{x_i\}_{i\in\mathcal{I}} = \text{im}(\Pi) \subseteq T \subseteq \text{span}\{x_i\}_{i\in\mathcal{I}}$$

Hence, $T = \text{span}\{x_i\}_{i\in\mathcal{I}}$ as desired. □

Let $\mathcal{Z}_0$ be the set of contact vectors of $\mathcal{D} = \{(x_i, y_i)\}_{i\in\mathcal{I}}$. If $\mathcal{D}$ is Δ-fixed, then $m_{opt} = (\mathbb{1}_{|\mathcal{T}|}, \mathbb{0}_{d-|\mathcal{T}|}) \in \mathbb{R}^d$, where $m_{opt}(i) = 1$ when $\mathbf{z}_i \neq \mathbf{0}$. Then, $\langle m_{opt}, m_{opt} - \sum_{\mathbf{z}} \mu(\mathbf{z})\mathbf{z}^{\circ 2} + \mathbf{v}\rangle = \langle m_{opt}, \sum_{\mathbf{z}} \mu(\mathbf{z})\mathbf{z}^{\circ 2} - \mathbf{v}\rangle = |\mathcal{T}| = r$.

Once strong duality holds, a dual certificate of a given optimizer may be desired. Hence, we developed the following lemma:

Lemma 3. (Complementary Slackness) Suppose $(\mu, \mathbf{v})$ is a dual certificate of $\mathsf{sqzLP}(\mathcal{D}, \Delta)$. For any optimal $\mathbf{m}$ of $\mathsf{sqzLP}(\mathcal{D}, \Delta)$, the optimal set of $\mathsf{dual}(\mathcal{D}, \Delta)$ is given by the set of points $(\mu, \mathbf{v})$ that are feasible in the dual program $\mathsf{dual}(\mathcal{D}, \Delta)$ and also satisfy the following:

$$\text{supp}(\mu) \subseteq \{\mathbf{z} \in \mathcal{Z}(\mathcal{D}) : \mathbf{m}^T\mathbf{z} = \Delta^2\}, \tag{2.15}$$

$$v_i = 0, \forall i \notin \mathcal{T}. \tag{2.16}$$

Proof. Suppose $(\mu, \mathbf{v})$ is feasible in $\mathsf{dual}(\mathcal{D}, \Delta)$. Hence $\mathbb{1} - \sum_{\mathbf{z}\in\mathcal{Z}(\mathcal{D})} \mu(\mathbf{z})\mathbf{z}^{\circ 2} + \mathbf{v} \geq \mathbf{0}$. Since $m_i \geq 0, \forall i$, we have

$$\langle \mathbf{m}, \mathbb{1} - \sum_{\mathbf{z}\in\mathcal{Z}(\mathcal{D})} \mu(\mathbf{z})\mathbf{z}^{\circ 2} + \mathbf{v}\rangle \geq 0 \tag{2.17}$$

After separating the inner product and doing some rearrangement, we get

$$\langle \mathbf{m}, \mathbb{1}\rangle \geq \langle \mathbf{m}, \sum_{\mathbf{z}\in\mathcal{Z}(\mathcal{D})} \mu(\mathbf{z})\mathbf{z}^{\circ 2}\rangle - \langle \mathbf{m}, \mathbf{v}\rangle \tag{2.18}$$

We can rewrite (2.18) as

$$\sum_i \mathbf{m}(i) \geq \sum_{\mathbf{z}\in\mathcal{Z}(\mathcal{D})} \mu(\mathbf{z})\mathbf{m}^T\mathbf{z}^{\circ 2} - \mathbf{m}^T\mathbf{v} \geq \Delta^2 \sum_{\mathbf{z}\in\mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) - \mathbf{m}^T\mathbf{v} \geq \Delta^2 \sum_{\mathbf{z}\in\mathcal{Z}(\mathcal{D})} \mu(\mathbf{z}) - \sum_i \mathbf{v}(i). \tag{2.19}$$

Since $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ admits a dual certificate, the optimal set of $\mathsf{dual}(\mathcal{D}, \Delta)$ is the set of feasible points $(\mu, \mathbf{v})$ in $\mathsf{dual}(\mathcal{D}, \Delta)$ in which all of the inequalities from (2.19) achieve equality.

Since $\mathbf{z}$ is supported on $\mathcal{T}$, $\Delta^2 = \sum_{i \in \mathcal{T}} \mathbf{z}(i)^2 = \|\mathbf{z}\|^2$. Hence $\mathbf{z}$ is a contact vector. Therefore, $\mu(\mathbf{z}) \neq 0$ only if $\mathbf{z}$ is a contact vector. Hence, the equality of the second inequality can be written as (2.15). The equality of the third inequality implies (2.16), whereas the equality of the first inequality only occurs when $\mathbf{m} = \mathbf{0}$. □

By complementary slackness, the primal value of $\mathbf{m}$ equals to $(\mu, \mathbf{v})$, i.e., the dual value. Hence, Weak duality suggests $\mathbf{m}$ is feasible in $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ [10].

Now that we have discussed the duality theory of the LP version of SqueezeFit, in the next chapter, we will study the behavior of $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ when under a certain model of data.

Chapter 3: Projection Factor Recovery with the SqueezeFit Linear Program

3.1 Review of Probability Concepts

Before we discuss the Projection Factor Recovery model, let us review some general probability concepts. See [15], [11] and [12] for more details.

Definition 10. (**Markov's Inequality**) Let $X \geq 0$ be random variable, then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}, \quad \textit{for any } a > 0. \tag{3.1}$$

Definition 11. (**Chebyshev's Inequality**) Given any random variable X, then for any positive b,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq b) \leq \frac{Var(X)}{b^2}. \tag{3.2}$$

Definition 12. (**Chernoff Bound**) Consider a random variable X, let ψ be the *log-moment generating function* of X, and let $\psi^\star$ be its *Legendre dual*:

$$\psi(\lambda) := \log(\mathbb{E}[e^{\lambda(X-\mathbb{E}X)}]), \quad \psi^\star(t) = \sup_{\lambda \geq 0}\{\lambda t - \psi(\lambda)\}. \tag{3.3}$$

Then we have $\mathbb{P}(X - \mathbb{E}X \geq t) \leq e^{-\psi^\star(t)}$.

With the Chernoff Bound, we can obtain Gaussian tail bounds and Chi-squared tail bounds:

Lemma 4. (**Gaussian Tails**) Let $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{P}[X - \mathbb{E}(X) \geq t] \leq e^{-t^2/2\sigma^2} \tag{3.4}$$

Furthermore, if we let $\mu = 0, \sigma = 1$, i.e., $X \sim \mathcal{N}(0,1)$, we have [11]

$$\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/2} \quad \textit{for every} \quad t > 0 \tag{3.5}$$

Lemma 5. (**Chi-Squared Tails**) Let X has Chi-squared distribution with n degrees of freedom, then, for every $t > 0$, we have

$$\mathbb{P}[|X - n| \geq t] \leq 2\exp(-\frac{1}{7}\min(\frac{t^2}{n}, t)). \tag{3.6}$$

Definition 13. (**Cauchy-Schwarz Inequality**) For any two random variables X and Y, we have

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}. \tag{3.7}$$

where equality holds if and only if $X = \alpha Y$ for some constant $\alpha \in \mathbb{R}$.

Theorem 6. (**Central Limit Theorem**) Suppose $\{X_i\}_{i=1}^{\infty}$ is a sequence of iid random variables with zero mean and unit variance. Then $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ converges in distribution to the standard Gaussian distribution.

Theorem 7. (**Bernstein's Inequality**) Suppose $\{X_i\}_{i=1}^{\infty}$ is a sequence of random variables with zero mean and variance σ^2, and $|X_i| \leq b$. Then for every $t > 0$, we have

$$\mathbb{P}\{\sum_{i=1}^{n} \geq t\} \leq 2\exp(-\frac{1}{3}\cdot\min(\frac{t^2}{n\sigma^2}, \frac{t}{b})). \tag{3.8}$$

3.2 Projection Factor Recovery

We plant a subspace onto which the data will be projected, and we want to identify the conditions under which $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ recovers this subspace. We make use of the following definition from [10] to handle the projection factor recovery for $\mathsf{sqzLP}(\mathcal{D}, \Delta)$:

Definition 14. Consider $\mathcal{D}_0 = \{(\mathbf{x}_i, y_i)\}_{i\in[a]}$ in $\mathbb{R}^d \times [k]$, select any $\sigma > 0$. For each $i \in [a]$, draw $\{g_{ij}\}_{j\in[b]}$ independently from $\mathcal{N}(0, \sigma^2)$, and consider the perturbed data set $\mathcal{D} = \{(\mathbf{x}_i + g_{ij}, y_i)\}_{i\in[a],j\in[b]}$, follow definition 5 of [10],we say $\mathcal{D}$ is drawn from the **projection factor model**, and we denote it as $\mathcal{D} \sim \text{PFM}(\mathcal{D}_0, \sigma^2, b)$.

Definition 15. We define E as the edge set of the graph $G = (\mathcal{I}, E)$, where $E \subseteq \{(i, j) : i, j \in \mathcal{I}\}$ and $\mathcal{Z}_0 = \{\mathbf{x}_i - \mathbf{x}_j : (i, j) \in E\}$.

For simplicity, we assume that $G = (\mathcal{I}, E)$ is necessarily *k-partite*, where k is the number of labels in the data set $\mathcal{D}$, i.e., $k = |\{y_i, i \in \mathcal{I}\}|$, and every vertex v of $G = (\mathcal{I}, E)$ has degree $deg(v) \leq 1$. Following [10], now we want to show that the dual certificate for the perturbed data set $\mathcal{D}$ is a predictable perturbation of the dual certificate for the original data $\mathcal{D}_0$ by the next theorem.

Theorem 8. (Main result) Let $\mathcal{D}_0 = \{(\mathbf{x}_i, y_i)\}_{i\in[a]}$ in $\mathbb{R}^d \times [k]$. Consider the projection factor model $\mathcal{D} \sim \text{PFM}(\mathcal{D}_0, \sigma^2, b)$. Let E be the edge set of the graph $G = (\mathcal{I}, E)$. Then SqueezeFit LP has arg $\mathsf{sqzLP}(\mathcal{D}, \Delta)$= $\{\mathbb{1}_T\}$ with probability at least $1 - 2|T^c| \cdot exp(-\frac{|E|}{7})$, provided

$$\sigma^2 \lesssim (\tfrac{1}{|E|} \cdot \min_{i\in\mathcal{T}} \textstyle\sum_{z\in\mathcal{Z}_0} z(i)^2)^{1/2},$$

where $\mathcal{Z}_0$ is the set of contact vectors of $\mathcal{D}_0$.

Proof. Consider the set of contact vectors $\mathcal{Z}_0$ from Definition 9. Clearly, we have

$$\mathcal{Z}_0 \subseteq \mathcal{Z} \coloneqq \{(\mathbf{x}_i - \mathbf{x}_j), i, j \in \mathcal{I}, y_i \neq y_j\}. \tag{3.9}$$

Since noises are added on $\mathcal{T}^c$ and we want $\mathbf{m}_{opt} = \mathbb{1}_{\mathcal{T}}$, we want to select a dual certificate for a perturbed Δ-fixed data. Let

$$\eta = \min_{i\in\mathcal{T}} \sum_{z\in\mathcal{Z}_0} \mathbf{z}(i)^2, \tag{3.10}$$

and let

$$\mu(\mathbf{z}) = \begin{cases} \frac{1}{\eta} & \text{if } \mathbf{z} \in \mathcal{Z}_0 \\ 0 & \text{otherwise} \end{cases}, \tag{3.11}$$

we want

$$1 \geq \sum_{\mathbf{z}\in\mathcal{Z}_0} \mu(\mathbf{z})\mathbf{z}(j)^2, \forall j \in \mathcal{T}^c, \text{ where } \sum_{\mathbf{z}\in\mathcal{Z}_0} \mu(\mathbf{z})\mathbf{z}(j)^2 = \frac{\sum_{\mathbf{z}\in\mathcal{Z}_0} \mathbf{z}(j)^2}{\eta}, \tag{3.12}$$

and by rearranging the terms, we obtain

$$\sum_{\mathbf{z}\in\mathcal{Z}_0} \mathbf{z}(j)^2 \leq \eta = \min_{i\in\mathcal{T}} \sum_{\mathbf{z}\in\mathcal{Z}_0} \mathbf{z}(i)^2, \tag{3.13}$$

and hence, we want

$$\max_{j\in\mathcal{T}^c} \sum_{\mathbf{z}\in\mathcal{Z}_0} \mathbf{z}(j)^2 \leq \min_{i\in\mathcal{T}} \sum_{\mathbf{z}\in\mathcal{Z}_0} \mathbf{z}(i)^2. \tag{3.14}$$

Let $S = \{i \in \mathcal{I} : deg(i) = 1\}$, and consider $\{\mathbf{x}_{ij}\}_{i\in\mathcal{I}, j\in\mathcal{T}^c}$ is independently and identically $\mathcal{N}(0, \sigma^2)$ distributed. Then for $(\mathbf{x}_i(j) - \mathbf{x}_{i'}(j))^2 \sim \mathcal{N}(0, 2\sigma^2)$ over E, we have

$$\begin{aligned}
&\max_{j\in\mathcal{T}^c} \sum_{\mathbf{z}\in\mathcal{Z}_0} \mathbf{z}(j)^2 = \max_{j\in\mathcal{T}^c} 2 \sum_{\{i,i'\}\in E} (\mathbf{x}_i(j) - \mathbf{x}_{i'}(j))^2 \\
&= \max_{j\in\mathcal{T}^c} 2 \sum_{\{i,i'\}\in E} (2\sigma^2\mathcal{N}(0,1))^2 \\
&= \max_{j\in\mathcal{T}^c} 8\sigma^4 \sum_{\{i,i'\}\in E} (\mathcal{N}(0,1))^2 \\
&= 8\sigma^4 \times (\text{max of } |\mathcal{T}^c| \quad i.i.d \quad \chi^2(|E|) \text{ random variables})
\end{aligned}$$

Let $e = |E|$ and consider $\{u_j\}_{j\in\mathcal{T}^c} \sim \chi^2(e)$, then for some s, we have

$$\mathbb{P}\{\max_{j\in\mathcal{T}^c} u_j > s\} = \mathbb{P}\{\exists j \in \mathcal{T}^c \ \ s.t. \ \ u_j > s\} \leq \sum_{j\in\mathcal{T}^c} \mathbb{P}\{u_j > s\} = |\mathcal{T}^c| \times \mathbb{P}\{u_j > s\}. \tag{3.15}$$

Since for some $x \sim \chi^2(n)$, and for some t, by Lemma 5, we have

$$\mathbb{P}\{(x - n) \geq t\} \leq \mathbb{P}\{|x - n| \geq t\} \leq 2 \cdot exp(-\frac{1}{7} \cdot \min(\frac{t^2}{n}, t)). \tag{3.16}$$

Put $s \coloneqq e + t$, then

$$\mathbb{P}\{\max_{j\in\mathcal{T}^c} u(j) > e + t\} \leq |\mathcal{T}^c| \cdot \mathbb{P}\{u_j > e + t\} \leq |\mathcal{T}^c| \cdot 2 \cdot exp(-\frac{1}{7} \cdot \min(\frac{t^2}{e}, t)). \tag{3.17}$$

Consider $t = e$, we have

$$\mathbb{P}\{\max_{j \in \mathcal{T}^c} u(j) > 2e\} \leq |\mathcal{T}^c| \cdot 2 \cdot exp(-\frac{e}{7}). \tag{3.18}$$

Hence,

$$\begin{aligned}
&\max_{j \in \mathcal{T}^c} \sum_{\mathbf{z} \in \mathcal{Z}_0} \mathbf{z}(j)^2 \\
&= 8\sigma^4 \times (\text{max of } |\mathcal{T}^c| \quad i.i.d \quad \chi^2(|E|) \text{ random variables}) \\
&\leq 8\sigma^4 \times O(|E|) \text{ with probability greater than } 1 - |\mathcal{T}^c| \cdot 2 \cdot exp(-\frac{e}{7})
\end{aligned}$$

Therefore, arg $\mathsf{sqzLP}(\mathcal{D}, \Delta)$= $\{\mathbb{1}_T\}$ with probability at least $1 - |\mathcal{T}^c| \cdot 2 \cdot exp(-\frac{e}{7})$, if

$$\sigma^2 \lesssim (\tfrac{1}{|E|} \cdot \min_{i \in \mathcal{T}} \textstyle\sum_{z \in \mathcal{Z}_0} \mathbf{z}(i)^2)^{1/2}$$

as desired. □

There are a few models that are related to our projection factor recovery model, and we will discuss them in Appendix A. In the next chapter, we will discuss a few implementations and the results of our numerical experiments of the model using planted and real data sets.

Chapter 4: Numerical Experiments and Results

In this section, we will discuss several different implementations of SqueezeFit Linear Program, and our numerical experiments. We used CVX [7][6] on MATLAB [9] for the implementations.

4.1 Implementation Variants

4.1.1 Direct Implementation

We implement the convex optimization problem $\mathsf{sqz}(\mathcal{D}, \Delta)$ and $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ directly. Figure 4.1 shows the plots of result of $\mathsf{sqz}(\mathcal{D}, \Delta)$ and $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ for a planted projection factor recovery in $\mathbb{R}^3$. We generated similar projection factor recovery problems with different levels of perturbation. We ran 100 trials on different values of σ and calculated the successful rates of either method, i.e., the quotient between the number of successes (error less than 10^{-4}) and the number of trials. We plotted the successful rates for the two methods, shown in Figure 4.2. Our numerical experiments showed that $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ can perform as good as $\mathsf{sqz}(\mathcal{D}, \Delta)$ in projection factor recovery problems, and be able to recover the planted projection factors more accurately with higher noise level.

We are not only interested in the performance of SqueezeFit LP comparing to the original SDP, but also the running time when dealing with data with higher dimensions. We generated random known models with different and higher dimensions using the idea of

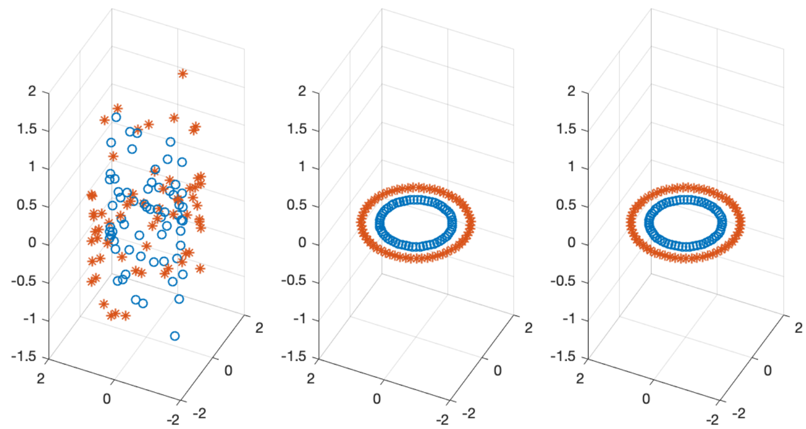

Figure 4.1: (**Left**)Points in $\mathbb{R}^3$ in two classes drawn from a random known model with an unknown projection factor. (**Middle**) SqueezeFit yields a good approximation and recovers the projection factor. Similar to Figure 1.2. (**Right**) The Linear Program version of SqueezeFit yields the same result as the original SqueezeFit, the Semidefinite Program.

Dimension	LP(seconds)	SDP(seconds)
4	0.27785	2.7187
8	0.21925	2.8649
16	0.2763	5.8024
32	0.66274	40.866
64	0.70162	607.6813
128	1.9643	3039.5166

Table 4.1: Empirical running time for $\mathsf{sqz}(\mathcal{D}, \Delta)$ and $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ dealing with 1000 data points of dimension 4, 8, 16, 32, 64 and 128. The linear program shows great advantage against the semidefinite program regarding time complexities.

hyper-sphere, and timed SqueezeFit LP and SDP. We used a 2017 standard MacBook Pro with 3.1 GHz Dual-Core Intel Core i5 processor. Table 4.1 shows the results of the empirical running time to run $\mathsf{sqz}(\mathcal{D}, \Delta)$ and $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ with various data dimensions. We observed

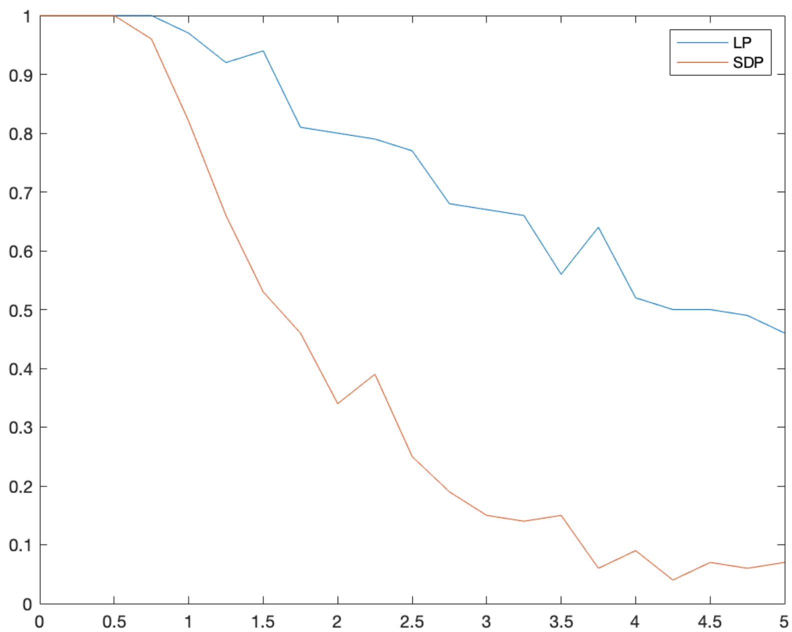

Figure 4.2: Projection Recovery of a random known model with different levels of perturbation (see Definition 14) and with $\Delta = 0.2$, $d = 3$ and $n = 63$. The **horizontal axis** represents the values of σ, and the **vertical axis** represents the successful rates of the two method.

that it took $\mathsf{sqz}(\mathcal{D}, \Delta)$ much longer - nearly 800 times and nearly 1500 times longer than $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ - to analyze data with dimension 64 and 128 correspondingly.

4.1.2 Implementation Improvements

We observed that even with small d, the program is slow when finding the set $\mathcal{Z}(\mathcal{D})$ with direct implementation. Lemma 11 of [10] suggests that the Δ constraints are typically not tight. We follow [10] and relax $\mathcal{Z}(\mathcal{D})$: Fix s to be any integer greater than 1, and denote the indices of the s nearest neighbors to x_i as $S(i,p)$, where $p \in [k]$ is its label, and let

$$\mathcal{Z}_s(\mathcal{D}) := \bigcup_{i=[n]} \bigcup_{p\in[k], p\neq y_i} \{x_i - x_j : j \in S(i,p)\}. \tag{4.1}$$

Replace $\mathcal{Z}(\mathcal{D})$ in $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ with $\mathcal{Z}_s(\mathcal{D})$. We obtain

$$\textit{minimize} \quad \sum_i \mathbf{m}(i) \quad \textit{subject to} \quad \sum_i \mathbf{z}(i)^2 \mathbf{m}(i) \geq \Delta^2 \quad \forall \mathbf{z} \in \mathcal{Z}_s(\mathcal{D}), \quad 0 \leq \mathbf{m}(i) \leq 1 \quad \forall i \tag{$\mathsf{sqzLP}^s(\mathcal{D}, \Delta)$}$$

Instead of having $O(n^2)$ constraints, the above relaxation reduces it to $O(sn)$ constraints. As a result, it requires less time to calculate the set of constraining vectors and reduces the total running time of the linear program (See Table 4.2). Furthermore, illustrated by Figure 4.3, our numerical experiments showed that the relaxed program yields close approximations to the original program optimizers.

n	Original	Relaxed
100	1.0195	0.24453
200	1.6064	0.20526
400	6.5114	0.22791
800	29.0561	0.26727
1600	151.0611	0.53171

Table 4.2: Empirical running time for $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ and $\mathsf{sqzLP}^s(\mathcal{D}, \Delta)$ dealing with n data points of dimension 4 and $s = 1$, where $n = 100, 200, 400, 800, 1600$. The relaxation in $\mathsf{sqzLP}^s(\mathcal{D}, \Delta)$ reduces the total running time of the linear program by a great amount especially when $n \geq 200$.

In summary, relaxing the $\mathcal{Z}(\mathcal{D})$ constraints by replacing $\mathcal{Z}(\mathcal{D})$ with $\mathcal{Z}_s(\mathcal{D})$ is considerably faster when $n \geq 200$ while remains the good performance of the original $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ program. We will use the algorithm of $\mathsf{sqzLP}^s(\mathcal{D}, \Delta)$ for further numerical experiments in the next two sections due to its convenient features.

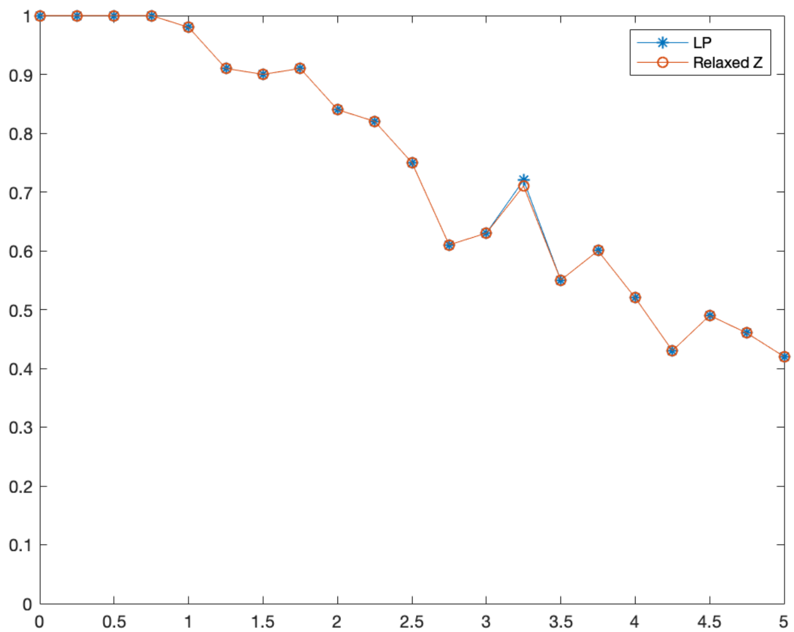

Figure 4.3: Projection Recovery of a random known model with different levels of perturbation (see Definition 14) and with $\Delta = 0.2$, $d = 3$ and $n = 63$. The **horizontal axis** represents the values of σ, and the **vertical axis** represents the successful rates. The successful rates of $\mathsf{sqzLP}(\mathcal{D}, \Delta)$ and $\mathsf{sqzLP}^s(\mathcal{D}, \Delta)$ are shown to be almost identical according to our numerical results.

4.2 Relation to Theoretical Results

Before we apply more data sets to the program, let's compare the numerical results in the previous section to the theoretical results in chapter 3. Again, we used a random known model with different levels of perturbation and with $\Delta = 0.2$. The result of Theorem 8 assures for this specific known model, with $\sigma^2 \leq 0.390802687774539$, the probability of successfully finding the accurate optimizer is at least 0.999753180391827. We observed in Table 4.3 that the results of our numerical experiments confirms our previous illustrated theories.

σ	σ^2	Probability
0	0	1
0.125	0.015625	1
0.25	0.0625	1
0.375	0.14062	1
0.5	0.25	1
0.625	0.39062	1
0.75	0.5625	0.98
0.875	0.76562	1
1	1	0.97

Table 4.3: Projection Factor Recovery of a random known model with different levels of perturbation and with $\Delta = 0.2$ and $n = 63$.

4.3 Application

SqueezeFit Linear Program is motivated and inspired by the Marker Gene Selection Problem mentioned in Chapter 1. Since we have confirmed that the numerical experiments using simulated data agree with the previously developed theories, we now want to study the performance of SqueezeFit LP when applying it to scRNA-seq data with given structured partition. Here, we followed [3] and illustrates the result of SqueezeFit LP of two data sets with different given partition structure of cell types: Cord blood mononuclear cell with Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq), and mouse cortex single-cells.

4.3.1 Cord Blood Mononuclear cell (CBMC)

We applied SqueezeFit LP to a data set obtained by a CBMC study - CITE-seq [13], which is a method that “combines highly multiplexed protein marker detection with unbiased transcriptome profiling for thousands of single cells” [13], and allows “simultaneous detection of single cell transcriptomes and protein markers” [13]. The data set contains more than

8000 CBMC single-cell profiles and has a dimension of 500 representing the "top highly variable genes from both human and mouse cell lines" [3]. With the given flat partition structure of the 13 distinct cell types population, we applied $\mathsf{sqzLP}^s(\mathcal{D}, \Delta)$ to this data set with a sampling rate of 10%, and limited the size of $\mathcal{Z}_s(\mathcal{D})$ to be at most 5000. $\mathsf{sqzLP}^s(\mathcal{D}, \Delta)$ is expected to identify the marker genes to project the data on while maintaining desired separation of distinct cell type population to achieve dimensionality reduction. It took the LP 20.162791 seconds to find the solution. Figure 4.4 illustrates our results.

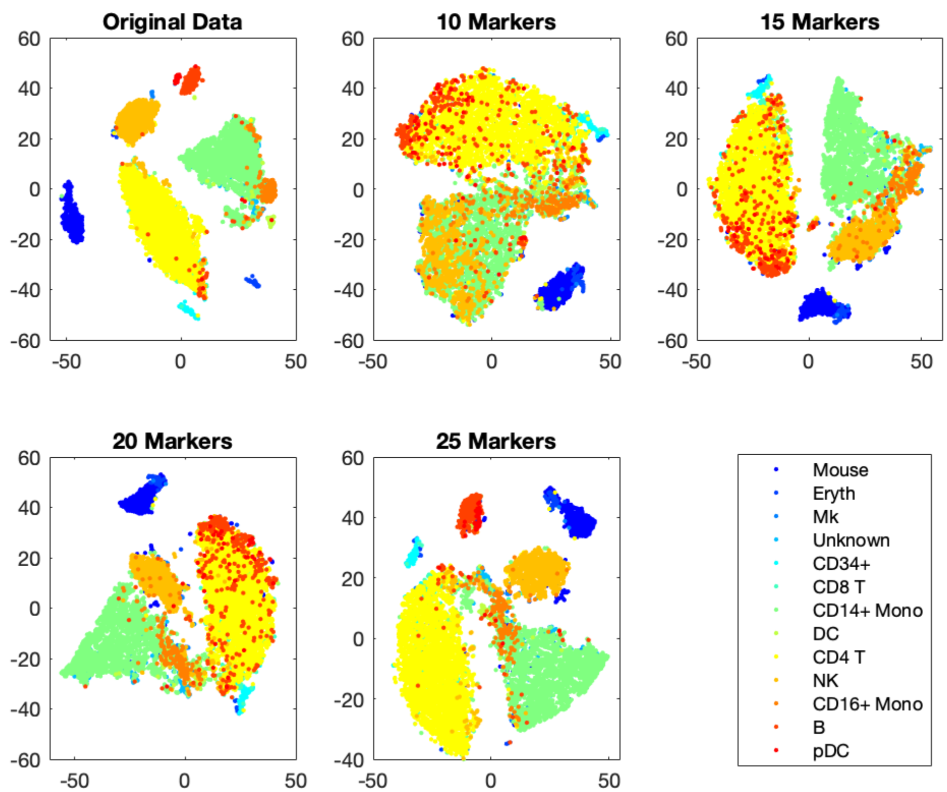

Figure 4.4: T-distributed stochastic neighbor embedding (t-SNE) visualization of results of SqueezeFit LP with a sample rate of 10% and a maximum of 5000 constraints. Figure shows projection of the single-cell expression profiles of CBMC [13] on 10, 15, 20, 25 marker genes. It seems that 25 marker genes are sufficient to distinguish 13 different population.

4.3.2 Single-cell analysis on mouse cortical cell

We applied SqueezeFit LP to another scRNA-seq data set obtained from [16]. This data set represents cell types in the mouse cortex and hippocampus and the study provides a hierarchy of those cell types [3] [16]. The data set contains 3005 data points of dimension 4000. We applied $\mathsf{sqzLP}^s(\mathcal{D}, \Delta)$ with a sampling rate of 5% and a maximum of 600 constraints. Figure 4.5 illustrates our result of SqueezeFit LP using the first layer of labels. It takes the LP 230.921854 seconds to find the solution. Figure 4.6 shows the performance of $\mathsf{sqzLP}^s(\mathcal{D}, \Delta)$ on the second layer of the hierarchical labels. It takes 255.813360 seconds to find the marker genes. Similar to the result of [3], we found that SqueezeFit LP performed fairly well in the astrocytes-ependymal, pyramidal CA1, and oligodendrocytes subpopulations, but poorly when distinguishing cell subtypes in the microgolia subpopulation, which might be directly caused by the existence of rare cell subtypes [3].

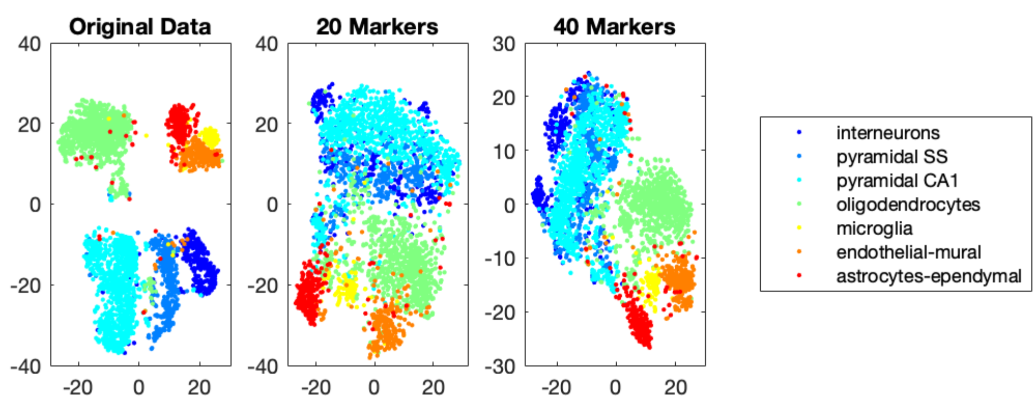

Figure 4.5: t-SNE visualization of the performance of SqueezeFit LP applying on a set of scRNA-seq data identifying mouse cortical cell-type population using the first layer labels [16] with a sampling rate of 5% and a maximum of 600 constraints. **Left** provides a t-SNE visualization of the original data using the first layer of labels. The rest of the figure shows projection of the single-cell expression on 20 and 40 markers.

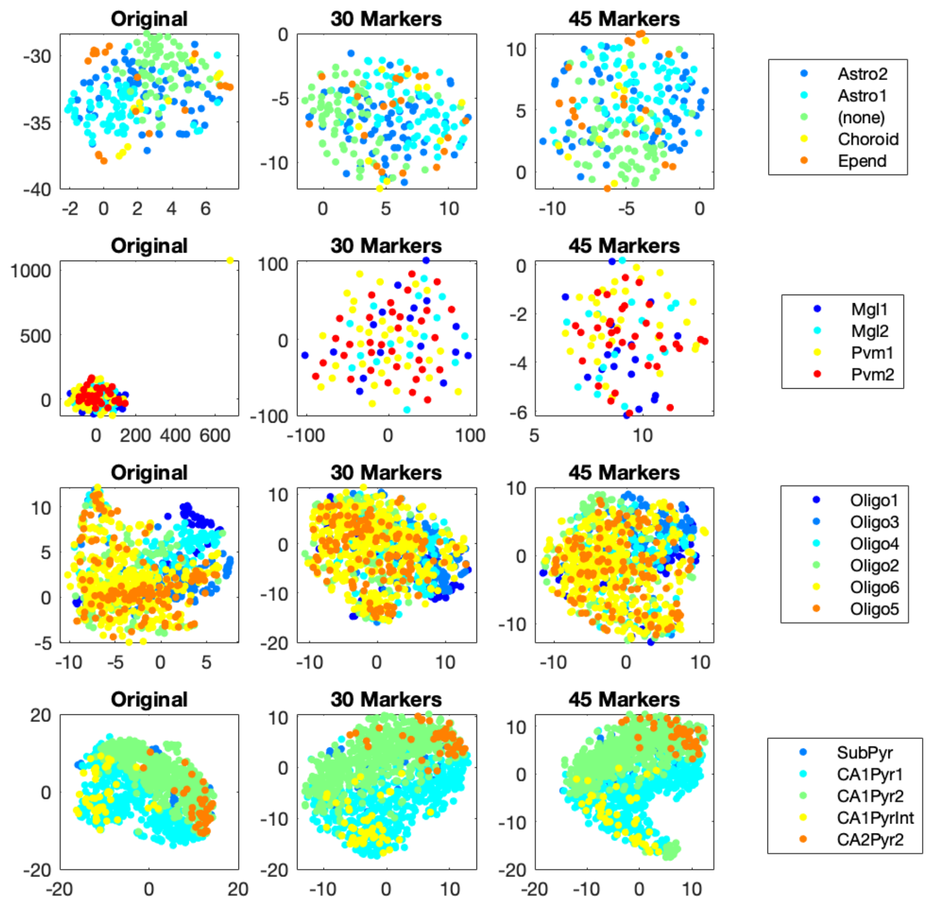

Figure 4.6: t-SNE visualization of the performance of SqueezeFit LP applying on the a set of scRNA-seq data identifying mouse cortical cell-type population using the second layer of labels [16] with a sampling rate of 5% and a maximum of 600 constraints. Figure shows the projection of the single-cell expression of various subpopulation on 30 and 40 markers. SqueezeFit LP performed fairly nicely in certain subpopulations including astrocytes-ependymal(**row 1**), pyramidal CA1(**row 4**), and oligodendrocytes(**row 3**), but poorly in the microgolia(**row 2**) subpopulation.

From the previous experiments on the two different scRNA-seq data sets, we conclude that SqueezeFit LP is able to find the marker genes so that the projection separates distinct cell types of scRNA-seq data sets with given structured partition - either flat (e.g. CBMC) or hierarchical (mouse cortex) [3]. Though the program seems more accurate and efficient on a flat partition, but is also able to handle and recover more complicated structures [3]. In the next section, we will summarize our theoretical and numerical results together, and discuss a few research questions for future work.

Chapter 5: Discussion

Inspired by the SqueezeFit semi-definite program [10] and scGeneFit [3], we proposed SqueezeFit Linear Program. We followed [10], and guaranteed the LP theoretically in the context of the projection factor recovery, mentioned in Chapter 3. We investigated the performance of SqueezeFit LP through numerical experiments in Chapter 4. We observed that SqueezeFit LP performed much faster than the original program when dealing with data with high dimension and improved the performance in projection factor recovery with planted data sets. We also applied two scRNA-seq data sets with different types of partition structure. It seems that SqueezeFit LP can handle flat and hierarchical relationships between distinct class labels without taking too much time. In summary, SqueezeFit LP fills the gap between the two predecessors: SqueezeFit LP overcomes the difficulty to run the original SqueezeFit SDP on large data, and provides a theoretical backing for scGeneFit. In this section, we will discuss a few questions and topics that came across throughout the development of SqueezeFit LP that can be interesting for future research work.

First, we noticed that SqueezeFit LP produces less adequate results in distinct subtypes under certain subpopulations when rare cell types emerge (see Figure 4.6 for details). We may be able to make a small adjustment to the program, for example, certain thresholds during the sampling process to overcome this issue. Also, the difficulty of placing the unlabeled cells, mentioned in [3], may be reduced by a relaxation of the categorical labeling to a manifold surface constraint.

We also realized that the diagonal constraint is specifically good for the marker gene selection problem due to its special objective. However, the diagonal constraint may not perform as well as the marker gene selection problem in other classification problems, for example, image classification. In the future, it will be interesting to investigate the performance of SqueezeFit in the context of a different classification problem by adding a slightly modified constraint.

Furthermore, the numerical experiments of SqueezeFit focused on k-nearest neighbor classifiers [10]. However, convolution neural networks are currently the best known algorithms for image classification [10]. For example, ImageNet [8], a deep convolutional neural network is able to classify millions of high-resolution images of various of classes with low error rates. As SqueezeFit aspires to overcome adversarial attacks (see in Chapter 1), enforcing convolution-friendly constraints in SqueezeFit can be another potential future project [10].

Bibliography

[1] Emmanuel Abbe, Afonso S. Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *CoRR*, abs/1405.3267, 2014.

[2] Stephen Boyd and Lieven Bandenberghe. *Convex Optimization*. International series of monographs on physics. Cambridge, 2004.

[3] Bianca Dumitrascu, Soledad Villar, Dustin G. Mixon, and Barbara E. Engelhardt. Optimal marker gene selection for cell type discrimination in single cell analyses. *bioRxiv*, 2019.

[4] Greg Finak, Andrew McDavid, Masanao Yajima, Jingyuan Deng, Vivian Gersuk, Alex K. Shalek, Chloe K. Slichter, Hannah W. Miller, M. Julianna McElrath, Martin Prlic, Peter S. Linsley, and Raphael Gottardo. Mast: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell rna-seq data. *Genome Biology*, 2015.

[5] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv*, 2014.

[6] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

[7] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, mar 2014.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25, 01 2012.

[9] MATLAB. *version 9.4.0 (R2018a)*. The MathWorks Inc., Natick, Massachusetts, 2018.

[10] Culver McWhirter, Dustin G. Mixon, and Soledad Villar. Squeezefit: Label-aware dimensionality reduction by semidefinite programming. *arXiv*, 2018.

[11] Dustin G. Mixon. Mathematics of data science. `https://dustingmixon.wordpress.com/2019/01/07/math-8610-mathematics-of-data-science/`, 2020.

[12] Hossein Pishro-Nik. *Introduction to Probability, Statistics, and Random Processes*. Kappa Research LLC, 2014.

[13] Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K. Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. Simultaneous epitope and transcriptome measurement in single cells. *NATURE METHODS*, 14(9):865 – 868, 2017.

[14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, and Vincent Vanhouckeand Andrew Rabinovich. Going deeper with convolutions. *arXiv*, 2014.

[15] Roman van Handel. Probability in high dimension. *Princeton*, 2016.

[16] Amit Zeisel, Ana B. Muñoz-Manchado, Simone Codeluppi, Peter Lönnerberg, Gioele La Manno, Anna Juréus, Sueli Marques, Hermany Munguba, Liqun He, Christer Betsholtz, Charlotte Rolny, Gonçalo Castelo-Branco, Jens Hjerling-Leffler, and Sten Linnarsson. Cell types in the mouse cortex and hippocampus revealed by single-cell rna-seq. *Science*, 347(6226):1138–1142, 2015.

Appendix A: Related Models

Here, we discuss two models that are related to our Projection Factor Recovery Model in Chapter 3: Planted Clique and Stochastic Block Model.

A.1 Planted Clique

The clique problem is one of the problems that seeks clusters given a collection of objects: Given a graph $G = (V, E)$, the clique problem aims to identify the largest **clique** in G, where clique is defined as a subset of all vertices V in which all elements are pairwise adjacent [11]. We are interested in finding the **clique number** $\omega(G)$ of the graph, which is defined as the size of the largest clique in G. However, the problem is hard to solve - there is no polynomial-time algorithm to determine the clique number of graphs in general [11]. In fact, the problem of determining whether there exists a clique of a certain size (positive integer) in a given graph is NP-complete [11].

One can consider an **independent set** in a graph $G = (V, E)$, which is defined as a collection of nonadjacent vertices in G, which is also a clique in the graph complement $\overline{G}$. Hence, we are also interested in finding $\alpha(G)$, the size of the largest independent set in G, known as the **independence number** [11]. Denote S to be an independent set of $G = (V, E)$, we can write the problem of finding the independence number as the following program:

$$\textit{maximize} \quad |S| \quad \textit{subject to} \quad S \subseteq V, \quad \binom{S}{2} \cap E = \emptyset. \tag{A.1}$$

After rewriting the previous program in terms of functions of the form $x : V \to \mathbb{R}$, and a tight relaxation (see details in [11]), we have the following semi-definite program:

$$\textit{maximize} \quad \langle J, X\rangle \quad \textit{subject to} \quad X_{ij} = 0 \quad \forall\{i,j\} \in E, \quad \operatorname{tr} X = 1, \quad X \succcurlyeq 0, \quad \operatorname{rank} X = 1. \tag{A.2}$$

We can obtain the Lovász number by relaxing the rank constraint, hence the Lovász number $\theta(G)$ is given by the optimal value of the following program:

$$\textit{maximize} \quad \langle J, X\rangle \quad \textit{subject to} \quad X_{ij} = 0 \quad \forall\{i,j\} \in E, \quad \operatorname{tr} X = 1, \quad X \succcurlyeq 0, \tag{Lovász SDP}$$

where $J = \mathbb{1}\mathbb{1}^T$, and $X = xx^T$. The dual program is given by

$$\textit{minimize} \quad \lambda \quad \textit{subject to} \quad \lambda I \succcurlyeq A, \quad A_{ij} = 1 \quad \forall\{i,j\} \in E, \tag{A.3}$$

where A is defined as the adjacency matrix of G [11].

Similar to our Projection Factor Recovery problem mentioned in Chapter 3 which aims for conditions and probabilities of recovering a planted projection factor, the **planted clique** model seeks graphs for which Lovász SDP exactly recovers a maximum clique of the graph complement [11]. The following lemma and theorem from [11] will answer the question (the proofs can also be found in [11]).

Lemma 9. (Lemma 4.1.6 in [11]) Suppose there exists an feasible point (λ, A) for (A.3) such that $A\mathbb{1}_S = |S|\mathbb{1}_S$ and $\lambda_2(A) \leq |S|$. Then $\frac{1}{|S|}\mathbb{1}_S\mathbb{1}_S^T$ is the unique optimizer of (Lovász SDP), and S is the unique maximum independent set of G.

Denote $G(n,p,k)$ to be a random graph model, which is derived from drawing $G_0 \sim G(n,p)$, the Erdös-Rényi random graph, and $S \sim \mathrm{Unif}\binom{V}{k}$, and then adding all edges in between S and G_0 to the desired graph model. One can expect the Lovász number of $\overline{G}$, since G is a slightly perturbed version of G_0, to also have the order of $\sqrt{n}$ [11]. Consider the case $p = \frac{1}{2}$, we have the following theorem:

Theorem 10. (Theorem 4.1.7 from [11]) There exists a universal constant $c > 0$ such that if $k \geq c\sqrt{n}$, then the planted clique in $G \sim G(n, \frac{1}{2}, k)$ is exactly recovered by Lovász's primal program for $\overline{G}$ with probability approaching 1 as $n \to \infty$.

The theorem tells us that when k is sufficiently larger than $\sqrt{n}$, the Lovász SDP can exactly recover the independent set S, and achieves the goal of the planted clique model. We observed the similarity between the projection factor recovery model for SqueezeFit LP and the planted clique model for the clique problem: when it appears to be difficult computationally to find a solution, one may treat the problem with a convex relaxation and seek the cases which the problem can be solved.

A.2 Stochastic Block Model

The Stochastic Block Model(SBM) is another model that exhibits cluster behaviors [1]. Consider a graph $G = (V, E) \sim SBM(n, p, q)$, where n is an even number representing the number vertices in the graph, and E is a random subset of $\binom{V}{2}$ with a specific distribution: Consider S and S^c as two communities of the graph. Draw $S \sim \text{Unif}\binom{V}{n/2}$. The events $\{\{i, j\} \in E\}$ are independent and have probability p if i and j are in the same community, and have probability q otherwise. The model aims to recover two communities of a given graph [1][11].

Consider a given graph $G = (V, E) \sim SBM(n, p, q)$, let $\mathbf{g} \in \{-1, 1\}^n$ be the solution to the problem, i.e., for any vertex i, we have $\mathbf{g}_i = 1$ if $i \in S$, and $\mathbf{g}_i = -1$ if $i \in S^c$. Consider the case when $p > q$, and let $p = \alpha \cdot \frac{\log n}{n}$ and $q = \beta \cdot \frac{\log n}{n}$, which is motivated by the Erdös-Rényi model [11]. Given the adjacency matrix A of the graph, the algorithm for recovering the two communities can be written as the following program [1]:

$$\textit{maximize} \quad \mathbf{x}^T A\mathbf{x} \quad \textit{subject to} \quad \mathbf{x}_i = \pm 1. \tag{A.4}$$

We can consider a convex relaxation to the previous problem and obtain the following semi-definite program [1]:

$$\textit{maximize} \quad \mathrm{tr}(AX) \quad \textit{subject to} \quad X_{ii} = 1 \quad X \succcurlyeq 0, \tag{A.5}$$

where $X = \mathbf{x}\mathbf{x}^T$ for a vector $x \in \mathbb{R}^n$ with entries ± 1, and for any i-th entry $\mathbf{x}_i = 1$ if the vertex i is in one community and -1 otherwise. The following theorem gives the conditions that guarantees exact recovery :

Theorem 11. ([1], [11]) If $(\alpha - \beta)^2 > 8(\alpha + \beta) + \frac{8}{3}(\alpha - \beta)$, the SDP (A.5) has a unique solution given by $\mathbf{g}\mathbf{g}^T$ with high probability.

The above theorem shows us the high probability to fully recover the communities of a given graph in polynomial time. The detailed proof of the previous theorem can be found in [1], and the following definition and lemma are the keys:

Definition 16. (See Definition 2 of [1]) Let G_+ be a subgraph of G which includes the edges that connect within one community, and D_+ the degree matrix of G_+. Let G_- and D_- be defined similarly. The **Stochastic Block Model Laplacian** is defined as

$$\mathcal{L} = D_+ - D_- - A, \tag{A.6}$$

where A is the adjacency matrix of the graph G.

Lemma 12. (See Lemma 6 in [1]) If $2\mathcal{L} + I_n - \mathbb{1}\mathbb{1}^T \succcurlyeq 0$ and $\lambda_2(2\mathcal{L} + I_n - \mathbb{1}\mathbb{1}^T) > 0$, then (A.5) has a unique solution given by the outer product of $\mathbf{g}$.

The proof of the above lemma can be found in [1]. We again observed the similarity between the Stochastic Block Model and our Projection Factor Recovery problem: we give a convex relaxation of a combinatorial problem, and apply convex and probabilistic theorems, which helps us find the condition and probability of a solving a problem exactly in polynomial time.

Appendix B: Proofs of Supporting Theorems

Here, we provide the proofs of a few supporting lemmas and theorems.

B.1 Proof of Supporting Lemmas

B.1.1 Proof of Lemma 4

Proof. [12] Consider $\psi(\lambda)$ and $\psi^{\star}(t)$ in the definition of Chernoff Bound (see Definition 12) and, let $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathbb{E}[e^{\lambda(X-\mathbb{E}(X))}] = e^{\lambda^2\sigma^2/2}$. Let

$$\psi(\lambda) = \frac{\lambda^2\sigma^2}{2}, \quad \psi^{\star}(t) = \frac{t^2}{2\sigma^2}. \tag{B.1}$$

We obtain the Gaussian tails: $\mathbb{P}[X - \mathbb{E}(X) \geq t] \leq e^{-t^2/2\sigma^2}$. □

B.1.2 Proof of Lemma 5

Proof. [11] Use a change of variable $y = (1-2s)x$ when computing the moment generating function of X, where X has Chi-squared distribution with n degrees of freedom, then

$$\mathbb{E}e^{sX} = (1-2s)^{-n/2}, \quad s \in (-\infty, \frac{1}{2}). \tag{B.2}$$

We can bound the right-hand tail with Chernoff bound, and obtain

$$\mathbb{P}\{X - n \geq t\} \leq e^{-st} \cdot \mathbb{E}e^{s(X-n)} = \exp(-(n+t)s - \frac{n}{2}\log(1-2s)). \tag{B.3}$$

Consider $s = \frac{t}{2(n+t)} \in (0, \frac{1}{2})$, which is the minimizer, we have

$$\mathbb{P}\{X - n \geq t\} \leq \exp(-\frac{t}{2} \cdot (1 - \frac{n}{t}) \log(1 + \frac{t}{n})) \leq \exp(-\frac{t}{2} \cdot \frac{2}{7} \min(\frac{t}{n}, 1)), \tag{B.4}$$

by assuming $t < n$, we have

$$\mathbb{P}\{X - n \geq t\} \leq \exp(-\frac{t}{7}). \tag{B.5}$$

Similarly, the left-hand tail can be bounded with Chernoff. We have

$$\mathbb{P}\{n - X \geq t\} \leq e^{-st} \cdot \mathbb{E}e^{s(n-X)} = \exp((n-t)s - \frac{n}{2} \log(1 + 2s)). \tag{B.6}$$

Assuming $t < n$, we consider the minimizer of the above bound $s = \frac{t}{2(n-t)}] > 0$. We have

$$\mathbb{P}\{n - X \geq t\} \leq \exp(\frac{t}{2} \cdot (1 + \frac{n}{t} \log(1 - \frac{t}{n})) \leq \exp(\frac{t}{2} \cdot -\frac{t^2}{4n}) = \exp(-\frac{t^2}{4n}). \tag{B.7}$$

Combining the previous bounds, we obtain

$$\mathbb{P}\{|X - n| \geq t\} \leq 2 \cdot \exp(-\frac{1}{7} \min(\frac{t^2}{n}, t)), \tag{B.8}$$

as desired. □

B.2 Proof of Supporting Theorems

B.2.1 Proof of Theorem 7

Proof. ([11]) Apply Chernoff bound and properties of exponential functions. We have

$$\mathbb{P}\{\sum_{i=1}^{n} X_i \geq t\} \leq \exp(-st)\mathbb{E}(s \sum_{i=1}^{n} X_i) = e^{-st} \prod_{i=1}^{n} \mathbb{E}(e^{sX_i}). \tag{B.9}$$

Since $\mathbb{E}e^{s|X|} \leq e^{sb} < \infty$, we can express the moment generating function in terms of the moments. Hence,

$$\mathbb{E}e^{sX_i} = \mathbb{E}\sum_{k=0}^{\infty}\frac{(sX)^k}{k!} = \sum_{k=0}^{\infty}\frac{s^k\mathbb{E}(X^k)}{k!}$$
$$\leq 1 + \sigma^2\sum_{k=2}^{\infty}\frac{s^k b^{k-2}}{k!}$$
$$= 1 + \frac{\sigma^2}{b^2}(e^{sb} - 1 - sb)$$
$$\leq \exp(\frac{\sigma^2}{b^2}(e^{sb} - 1 - sb)).$$

Hence,

$$\mathbb{P}\{\sum_{i=1}^{n} X_i \geq t\} \leq \exp(-st)\mathbb{E}(s\sum_{i=1}^{n} X_i) = e^{-st}\prod_{i=1}^{n}\mathbb{E}(e^{sX_i})$$
$$\leq \exp(-\frac{n\sigma^2}{b^2}\cdot g(\frac{bt}{n\sigma^2}))$$

, where $g(z) = (1+z)\log(1+z) - z \geq \frac{1}{3}\min(z^2, z)$ for $z \in (0,\infty)$. Hence,

$$\mathbb{P}\{\sum_{i=1}^{n} X_i \geq t\} \leq \exp(-\frac{n\sigma^2}{b^2}\cdot g(\frac{bt}{n\sigma^2})) \leq -\frac{1}{3}\min(\frac{\mathrm{t}^2}{\mathrm{n}\sigma^2}, \frac{\mathrm{t}}{\mathrm{b}}), \tag{B.10}$$

as desired, where the same result can be obtained through similar work when applying Chernoff bound to $\{-X_i\}_{i=1}^{n}$. Hence,

$$\mathbb{P}\{|\sum_{i=1}^{n} X_i| \geq t\} \leq -\frac{1}{3}\min(\frac{\mathrm{t}^2}{\mathrm{n}\sigma^2}, \frac{\mathrm{t}}{\mathrm{b}})$$

as desired. □