Sum-of-norms clustering: theoretical guarantee and post-processing

by

Tao Jiang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 2020

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The work presented here was done in collaboration with my advisor Professor Stephen Vavasis and his student Chen Wen Zhai. I was a major contributor to all the results contained in the chapters.

Abstract

Sum-of-norms clustering is a method for assigning n points in $\mathbf{R}^d$ to K clusters, $1 \leq K \leq n$, using convex optimization. Recently, Panahi et al. [21] proved that sum-of-norms clustering is guaranteed to recover a mixture of Gaussians under the restriction that the number of samples is not too large. The first contribution of this thesis is to lift this restriction, i.e., show that sum-of-norms clustering can recover a mixture of Gaussians even as the number of samples tends to infinity. Our proof relies on an interesting characterization of clusters computed by sum-of-norms clustering that was developed inside a proof of the agglomeration conjecture by Chiquet et al. [8]. Because we believe this theorem has independent interest, we restate and reprove the Chiquet et al. [8] result herein.

Multiple algorithms have been proposed to solve the sum-of-norms clustering problem: subgradient descent by Hocking et al. [12], ADMM and ADA by Chi and Lange [6], stochastic incremental algorithm by Panahi et al. [21] and semismooth Newton-CG augmented Lagrangian method by Sun et al. [28]. All algorithms yield approximate solutions, even though an exact solution is demanded to determine the correct cluster assignment. The second contribution of this thesis is to close the gap between the output from existing algorithms and the exact solution to the optimization problem. We present a clustering test which identifies and certifies the correct clustering from an approximate solution yielded by any primal-dual algorithm. The test may not succeed if the approximation is inaccurate. However, we show the correct clustering is guaranteed to be found by a primal-dual path following algorithm after sufficiently many iterations, provided that the model parameter λ avoids a finite number of bad values. Numerical experiments are implemented to support our results.

Acknowledgements

I cannot express my gratitude enough to my advisor, Professor Stephen Vavasis, for his ceaselessly guidance, support and care. He introduced me to the amazing world of continuous optimization and helped me develop as a researcher. Being a brilliant mathematician himself, he has always been incredibly humble, acknowledged my work and celebrated my achievements. Working with Steve has been thoroughly enjoyable, and I am very grateful for having the privilege to have done so.

I would like to dedicate this thesis to my most loving parents for their unconditional love. They have always been supporting me to follow my passions, helping me become someone that I am proud of and granting me confidence in living my life to the fullest. These are the best gifts a child could ask for from her parents, and I am thankful to have them all.

I would like to thank all my exceptional friends for their love and company. All the laughter and tears, coffee and beers, movies and concerts, rockclimbing and hiking, dancing and fencing have made my journey at Waterloo extremely precious and memorable.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Clustering is perhaps the most central problem in unsupervised machine learning and has been studied for over 60 years [27]. The problem may be stated informally as follows. One is given n points, $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ lying in $\mathbf{R}^d$. One seeks to partition $\{1, \ldots, n\}$ into K sets $C_1, \ldots, C_K$ such that the $\boldsymbol{a}_i$'s for $i \in C_m$ are closer to each other than to the $\boldsymbol{a}_i$'s for $i \in C_{m'}$, $m' \neq m$.

Clustering is usually formulated as a non-convex optimization problem, which is combinatorially hard to solve and beset by nonoptimal local minimizers. Classical methods such as k-means and hierarchical clustering are prone to these issues. Meanwhile, issues of hardness and suboptimality of many nonconvex optimization problems are resolved by convex relaxation. At an affordable computational cost, convex relaxation yields a good solution to the original problem.

1.2 Sum-of-norms clustering

Pelckmans et al. [22], Hocking et al. [12] and Lindsten et al. [17] proposed the following convex formulation for the clustering problem:

$$\min_{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbf{R}^d} f'(\boldsymbol{x}) = \frac{1}{2} \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{a}_i\|^2 + \lambda \sum_{1 \leq i < j \leq n} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|. \tag{1.1}$$

This formulation is known in the literature as sum-of-norms clustering, convex clustering, or clusterpath clustering. Let $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*$ be the optimizer. (Note: (1.1) is strongly convex, hence the optimizer exists and is unique.) The cluster assignment is given by the $\boldsymbol{x}_i^*$'s: for i, i', if $\boldsymbol{x}_i^* = \boldsymbol{x}_{i'}^*$ then i, i' are assigned to the same cluster, else they are assigned to different clusters.

The first term of the objective function ensures $\boldsymbol{x}^*$ is a good approximation of the original data $\boldsymbol{a}$, while the second term penalizes the differences $\boldsymbol{x}_i^* - \boldsymbol{x}_{i'}^*$. As a result, the second term tends to make $\boldsymbol{x}_i^*$ equal to each other for many i. Furthermore, the tuning parameter λ controls the number of clusters indirectly. It is apparent that for $\lambda = 0$, each $\boldsymbol{a}_i$ is assigned to a different cluster of its own (unless $\boldsymbol{a}_i = \boldsymbol{a}_{i'}$ exactly), whereas for λ sufficiently large, the second summation drives all the $\boldsymbol{x}_i$'s to be equal (and hence there is one big cluster consisting of all n data points).

Throughout this thesis, we assume that all norms are Euclidean, although (1.1) has also been considered for other norms. In addition, some authors insert nonnegative weights in front of the terms in the above summations. Most of our results, however, require all weights identically 1, but we revisit the question of general weights in Sections 3.3 and 4.4.

1.3 Recovery of a mixture of Gaussians

Panahi et al. [21] developed several recovery theorems as well as a first-order optimization method for solving (1.1). Other authors, e.g., Sun et al. [28] have since extended these results. One of Panahi et al.'s results pertains to a mixture of spherical Gaussians, which is the a generative model for producing the data $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. Panahi et al. [21] proved that for the appropriate choice of λ, sum-of-norms clustering formulation (1.1) will exactly recover a mixture of Gaussians provided that the pairwise distance between Gaussian means are lower bounded. The lower bound is a function depending on the number of samples, the number of Gaussians and distribution parameters such as standard deviations.

One issue with this bound is that as the number of samples n tends to infinity, the bound seems to indicate that distinguishing the clusters becomes increasingly difficult (i.e., the Gaussian means have to be more distantly separated as $n \to \infty$).

The reason for this aspect of their bound is that their proof technique requires a gap of positive width (i.e., a region of $\mathbf{R}^d$ containing no sample points) between $\{\boldsymbol{a}_i : i \in C_m\}$ and $\{\boldsymbol{a}_i : i \in C_{m'}\}$ whenever $m \neq m'$. Clearly, such a gap cannot exist in the mixture-of-Gaussians distribution as the number of samples tends to infinity.

The first ambition of this thesis is to prove that (1.1) can recover a mixture of Gaussians even as $n \to \infty$. This is the content of Theorem 2 in Section 4.2 and 4.3 below. Naturally, under this hypothesis we cannot hope to correctly label all samples since, as $n \to \infty$, some of the samples associated with one mean will be placed arbitrarily close to another mean. Therefore, we are content in showing that (1.1) can correctly cluster the points lying within some fixed number of standard-deviations for each mean.

Our proof technique requires a cluster characterization theorem for sum-of-norms clustering derived by Chiquet et al. [8]. This result is not stated by these authors as a theorem, but instead appears as a sequence of steps inside a larger proof in a "supplementary material" appendix to their paper. Because we believe that this theorem is of independent interest, we restate it below and for the sake of completeness provide the proof (which is the same as the proof appearing in Chiquet et al.'s supplementary material). This material appears in Chapter 3. We conclude the recovery of a mixture of Gaussians with some experimental results in Section 4.5.

1.4 Identifying clusters from computation

To identify the correct clusters from an approximate solution, authors in practice propose the following approximate test with an artificial tolerance, $\epsilon > 0$. If the approximate solution $\boldsymbol{x}$ satisfies $\|\boldsymbol{x}_i - \boldsymbol{x}_{i'}\| \leq \epsilon$, i, i' are assigned to the same cluster. Otherwise, i, i' are assigned to different clusters. Hence, the value of artificial tolerance is critical. Unfortunately, to the best of our knowledge, neither the value of the tolerance nor the approximate test itself has been rigorously justified. The test is not robust. Since the relation $\|\boldsymbol{x}_i - \boldsymbol{x}_{i'}\| \leq \epsilon$ is not transitive, it is not clear how the test would cluster points i, j, k if $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \leq \epsilon, \|\boldsymbol{x}_j - \boldsymbol{x}_k\| \leq \epsilon$, and $\|\boldsymbol{x}_i - \boldsymbol{x}_k\| > \epsilon$. The test may not be accurate. The clusters obtained by the approximate test could deviate from the clusters corresponding to the optimizer of (1.1). The inaccuracy may lead to the failure of known properties of sum-of-norms clustering such as the recovery of a mixture of Gaussians as illustrated in Chapter 4 and the agglomeration property as illustrated in Section 3.2. It has been established in Section 1.3 that for an appropriate choice of λ, (1.1) exactly recovers a mixture of Gaussians. However, it is unknown if the recovery result is preserved when the approximate test is applied. Hocking et al. [12] conjectured that sum-of-norms clustering is agglomerative in the sense that as λ increases, clusters may fuse but never break apart. The conjecture was proven by Chiquet, Gutierrez and Rigaill [8] with some techniques which may not be applicable when the approximate test is implemented. Thus the agglomeration property may no longer hold.

The second result of this thesis is to present our clustering test and to justify it rigorously. The clustering test takes a primal and dual feasible solution for the second-order cone formulation of sum-of-norms clustering and attempts to determine all clusters. The test may report 'success' or 'failure'. If the test reports 'success', all clusters are correctly identified and a certificate is produced. The test and the proof of correctness are stated in Section 5.2. The proof heavily relies on the clustering characterization theorem in Chapter 3. The test requires the knowledge of a primal and dual feasible solution for the second-order cone formulation of sum-of-norms clustering, which can be constructed from the output of any primal-dual algorithm. The second-order cone formulation and some primal-dual algorithms are stated in Section 5.1. If a primal-dual path following algorithm is used, the test is guaranteed to report 'success' after a finite number of iterations except the test may never report 'success' when the λ value is at which clusters fuse to form a larger cluster. These results are shown in Section 5.4. The proof of the theoretical guarantee is a result of the properties of the central path, which are stated in Section 5.3. In Section 5.5, we present a few computational experiments to verify our test in practice.

Chapter 2

Literature Review

2.1 Clustering

Clustering was historically inspired by problems in anthropology and psychology [5]. It has become one of the most important techniques in data analysis. Over the past few decades, clustering has been adopted to solve various problems from a broad range of research areas such as computational biology, social science and image processing.

Given n data points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in $\mathbb{R}^d$ and a distance measure, clustering aims to partition $\{1, \ldots, n\}$ into K sets $C_1, \ldots, C_K$ such that points in the same set are closer to each other than those that are not.

k-means clustering is the best-known paradigm for clustering. Given the number of clusters k, k-means clustering is formulated as the following combinatorial optimization problem:

$$\min_{\{C_1,\ldots,C_k\}} \sum_{i=1}^{k} \sum_{j \in C_i} \|\boldsymbol{a}_j - \boldsymbol{\mu}_i\|^2 ,$$

where $\boldsymbol{\mu}_i = \arg\min_{\boldsymbol{\mu}} \sum_{j \in C_i} \|\boldsymbol{a}_j - \boldsymbol{\mu}\|^2$. The formulation is equivalent to a constrained version of nonnegative matrix factorization [9]. The problem is NP-hard to solve, and it is NP-hard to approximate to within some accuracy [27]. The best known method for

k-means clustering is Lloyd's algorithm as presented below:

Algorithm 1: Lloyd's algorithm

Initially partition $\{1, \ldots, n\}$ into k random subsets $C_1, \ldots, C_k$;
Alternate the following two operations:
For $m = 1, \ldots, k$, compute $\boldsymbol{\mu}_m := \frac{1}{|C_m|} \sum_{i \in C_m} \boldsymbol{a}_i$;
For $m = 1, \ldots, k$, define $C_m^{\text{NEW}} := \{i : \|\boldsymbol{a}_i - \boldsymbol{\mu}_m\| = \min_{m'} \|\boldsymbol{a}_i - \boldsymbol{\mu}_{m'}\|\}$.

At each iteration, Lloyd's algorithm does not increase the k-means objective value [27]. At termination, Lloyd's algorithm outputs a local minimizer, which may be suboptimal. Moreover, there are only nontrivial bounds on the suboptimality of the output, and the rate of convergence is yet to be established [27].

Soft k-means is the expectation-maximization (EM) approach for k-means clustering [27]. Similar to Lloyd's algorithm, EM approach determines clustering and updates cluster centroids at each iteration. However, clustering and centroids are computed based on probabilities. The EM algorithm consists of two steps. The expectation step computes the probability over a latent variable. The maximization step finds the maximizer of the expected log-likelihood, where the expectation is calculated according to the probability computed in the expectation step. For the detailed comparison between EM algorithm and Llyod algorithm, we refer the reader to [27].

Another straightforward clustering model is linkage-based clustering. The agglomerative algorithm starts with the trivial clustering of n singleton points. It proceeds to merge the nearest clusters of the current clustering [27]. The same procedure repeats until the stopping criterion is met. Two common stopping criteria are a fixed number of clusters and a distance upper bound [27].

The linkage-based clustering can be formulated as the following optimization problem:

$$\min_{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbf{R}^d} \quad \frac{1}{2} \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{a}_i\|^2 \tag{2.1a}$$

$$\text{s.t} \quad \sum_{i<j} 1_{\boldsymbol{x}_i \neq \boldsymbol{x}_j} \leq t \ , \quad \forall 1 \leq i < j \leq n \ . \tag{2.1b}$$

Similar to the k-means clustering model, this formulation is also combinatorially hard to solve and beset by nonoptimal local minimizers. Hocking et al. [12] proposed the following

convex relaxation of (2.1):

$$\min_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\in\mathbf{R}^d} \quad \frac{1}{2}\sum_{i=1}^{n}\|\boldsymbol{x}_i-\boldsymbol{a}_i\|^2 \tag{2.2a}$$

$$\text{s.t} \quad \sum_{i<j}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|\le t\ , \quad \forall 1\le i<j\le n\ . \tag{2.2b}$$

According to Hocking et al. [12], sum-of-norms clustering formulation (1.1) is also the Lagrangian relaxation of (2.2), as restated below

$$\min_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\in\mathbf{R}^d} f'(\boldsymbol{x})=\frac{1}{2}\sum_{i=1}^{n}\|\boldsymbol{x}_i-\boldsymbol{a}_i\|^2+\lambda\sum_{1\le i<j\le n}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|\,.$$

Pelckmans et al. [22] and Lindsten et al. [17] also proposed the same convex formulation (1.1) of the clustering problem independently.

2.2 Algorithm for sum-of-norms clustering

Many algorithms, both primal-only and primal-dual methods, have been proposed to solve (1.1). Two typical primal-only algorithms are the subgradient descent by Hocking et al. [12] and the stochastic incremental algorithm by Panahi et al. [21]. At each iteration of the subgradient descent, the algorithm computes the subgradient and the step size, which follows either a simple decreasing scheme or a line search strategy. The stochastic incremental algorithm is identical to an incremental proximal method in Bertsekas [2]. The algorithm scales well with the number of data points n, even though its rate of convergence is fairly weak according to Sun et al. [28].

Interior point methods are probably the most established primal-dual methods for convex optimization. They are widely used by optimization solvers such as CVX. In 2011, Lindsten et al. [17] applied CVX to solve the sum-of-norms clustering. Regrettably, Hocking et al. [12] remarked that CVX does not perform well on a large data set.

ADMM and AMA are two straightforward primal-dual methods for sum-of-norms clustering. They were first adopted to solve (1.1) by Chi and Lange [7]. In their paper, the unconstrained problem (1.1) is reformulated as the following constrained problem:

$$\min_{\boldsymbol{x},\boldsymbol{y}} \quad \frac{1}{2}\sum_{i=1}^{n}\|\boldsymbol{x}_i-\boldsymbol{a}_i\|^2+\lambda\sum_{1\le i<j\le n}\|\boldsymbol{y}_{ij}\| \tag{2.3a}$$

$$\text{s.t} \quad \boldsymbol{x}_i-\boldsymbol{x}_j-\boldsymbol{y}_{ij}=\boldsymbol{0}\ , \quad \forall 1\le i<j\le n\ . \tag{2.3b}$$

The first term is strongly convex, and the second term is non-smooth and convex. Hence, splitting is a natural strategy. Chi and Lange considered both ADMM, whose updates on the first block of variables are derived from minimizing the augmented Lagrangian function, and AMA, whose updates on the first block of variables are derived from minimizing the ordinary Lagrangian function. In the same paper, they also proposed an accelerated variant of AMA, which has been proven to be more efficient than ADMM. Nevertheless, both ADMM and AMA are prone to scalability issues.

Another well-known primal-dual algorithm is the semismooth Newton-CG augmented Lagrangian method (SSNAL) by Sun, Toh and Yuan [28]. The algorithm consists of a two-level nested loop. The outer loop of SSNAL is an augmented Lagrangian method (ALM) with an increasing step size. The inner loop a semismooth Newton-CG method to derive the primal update of ALM. To warm-start SSNAL, Sun et al. [28] implemented an inexact ADMM to generate an initial solution. SSNAL has been proven to be efficient. In practice, it also demonstrates high efficiency in various numerical experiments on simulated data sets such as half moons, unbalanced Gaussians and MINST.

2.3 Review of recovery

Recently, there have been various attempts to provide recovery guarantees for sum-of-norms clustering with uniform weights (1.1). Zhu et al. [31] showed that if a data set is generated by two well-separated cubes, then sum-of-norms clustering recovers the two clusters perfectly. The separation condition is rather strict: the distance between two cubes must be larger than a threshold dependent on the number of data points and the sizes of two cubes. Tan and Witten [29] studied the statistical properties of sum-of-norms clustering. Recently, Panahi et al. [21] developed several recovery results that certify recovery by sum-of-norms under rather mild conditions. Panahi et al. [21] also specialized their results for data sets such as a mixture of Gaussians and planted partitions. Sun, Toh and Yuan [28] extended these results to general weights: under some easy assumptions, perfect recovery is guaranteed for sum-of-norms clustering with general weights.

A related result by Radchenko and Mukherjee [24] analyzed the special case of a mixture of Gaussians with $K = 2$, $d = 1$ under slightly different hypotheses. Also, Mixon et al. [19] showed that Peng and Wei's semidefinite relaxation of clustering [23] can recover a mixture of Gaussians as $n \to \infty$, but this result requires nontrivial postprocessing of the semidefinite solution to recover the clusters.

2.4 Review of clustering identification

To resolve the issue of inaccuracy as mentioned in Section 1.4, Hocking et al. [12] developed a two-step method based on the approximate test as described in Section 1.4. The first step is detecting potential fusions using the approximate test. The artificial tolerance is chosen to be some fraction of the minimum distance between two data points, $\min_{1\leq i<i'\leq n} \|\boldsymbol{a}_i - \boldsymbol{a}_{i'}\|$. The second step is verifying potential fusions by checking if the detected fusions improve the objective value. Friedman et al. [10] presented a similar approach to detect fusions for a fused-lasso problem with coordinate descent algorithms. The algorithm includes a descent cycle, a fusion cycle and a smoothing cycle. The descent cycle employs coordinate descent to solve a fused-lasso problem. When the coordinate descent gets stuck, the algorithm enters the fusion cycle. The fusion cycle merges any adjacent pairs if the fusion of the pair decreases the objective value. However, it only examines the potential fusions of pairs, but it does not consider the fusions of three points or more. When the fusion cycle fails to merge any adjacent pairs, there may still exist a fusion of three points or more that improves the objective value. To resolve the issue, Friedman et al. [10] introduced a smoothing cycle. The smoothing cycle varies some parameters in the fused lasso problem, which allows fusions of more than two in the long run. Both methods by Friedman et al. [10] and Hocking et al. [12] guarantee a correct solution. Unfortunately, they are both very slow as they investigate all possible fusion events.

Chapter 3

Cluster characterization

3.1 Cluster characterization theorem

The following theorem is due to Chiquet et al. [8] appearing as a sequence of steps in a proof of the agglomeration conjecture. Refer to the next section for a discussion of the agglomeration conjecture. We restate the theorem here because it is needed for our analysis and because we believe it is of independent interest.

Theorem 1. *Let $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*$ denote the optimizer of (1.1). For notational ease, let $\boldsymbol{x}^*$ denote the concatenation of these vectors into a single vector in $\mathbf{R}^{nd}$. Suppose that C is a nonempty subset of $\{1, \ldots, n\}$.*

(a) Necessary condition: If for some $\hat{\boldsymbol{x}} \in \mathbf{R}^d$, $\boldsymbol{x}_i^* = \hat{\boldsymbol{x}}$ for $i \in C$ and $\boldsymbol{x}_i^* \neq \hat{\boldsymbol{x}}$ for $i \notin C$ (i.e., C is one of clusters exactly determined by (1.1)), then there exist $\boldsymbol{z}_{ij}^*$ for $i, j \in C$, $i \neq j$, which solve

$$\begin{aligned} \boldsymbol{a}_i - \frac{1}{|C|} \sum_{l \in C} \boldsymbol{a}_l = \lambda \sum_{j \in C - \{i\}} \boldsymbol{z}_{ij}^* \quad & \forall i \in C, \\ \left\| \boldsymbol{z}_{ij}^* \right\| \leq 1 \quad & \forall i, j \in C, i \neq j, \\ \boldsymbol{z}_{ij}^* = -\boldsymbol{z}_{ji}^* \quad & \forall i, j \in C, i \neq j. \end{aligned} \tag{3.1}$$

(b) Sufficient condition: Suppose there exists a solution $\boldsymbol{z}_{ij}^*$ for $j \in C - \{i\}$, $i \in C$ to the conditions (3.1). Then there exists an $\hat{\boldsymbol{x}} \in \mathbf{R}^d$ such that the minimizer $\boldsymbol{x}^*$ of (1.1) satisfies $\boldsymbol{x}_i^* = \hat{\boldsymbol{x}}$ for $i \in C$.

Note: This theorem is an almost exact characterization of clusters that are determined by formulation (1.1). The only gap between the necessary and sufficient conditions is that the necessary condition requires that C be exactly all the points in a cluster, whereas the sufficient condition is sufficient for C to be a subset of the points in a cluster. The sufficient condition is notable because it does not require any hypothesis about the other $n - |C|$ points occurring in the input.

Proof. (Chiquet et al.) Proof for Necessity (a)
As $\boldsymbol{x}^*$ is the minimizer of the problem (1.1), and this objective function $f'(\boldsymbol{x})$ is convex, it follows that $\mathbf{0} \in \partial f'(\boldsymbol{x}^*)$, where $\partial f'(\boldsymbol{x}^*)$ denotes the subdifferential, that is, the set of subgradients of f' at $\boldsymbol{x}^*$. (See, e.g., [11] for background on convex analysis). Written explicitly in terms of the derivative of the squared-norm and subdifferential of the norm, this means that $\boldsymbol{x}^*$ satisfies the following condition:

$$\boldsymbol{x}_i^* - \boldsymbol{a}_i + \lambda \sum_{j \neq i} \boldsymbol{w}_{ij}^* = \mathbf{0} \qquad \forall i = 1, \ldots, n, \tag{3.2}$$

where $\boldsymbol{w}_{ij}^*$, $i = 1, \ldots, n$, $j = 1, \ldots, n$, $i \neq j$, are subgradients of the Euclidean norm function satisfying

$$\boldsymbol{w}_{ij}^* = \begin{cases} \frac{\boldsymbol{x}_i^* - \boldsymbol{x}_j^*}{\left\|\boldsymbol{x}_i^* - \boldsymbol{x}_j^*\right\|}, & \text{for } \boldsymbol{x}_i^* \neq \boldsymbol{x}_j^*, \\ \text{arbitrary point in } B(\mathbf{0}, 1), & \text{for } \boldsymbol{x}_i^* = \boldsymbol{x}_j^*, \end{cases}$$

with the requirement that $\boldsymbol{w}_{ij}^* = -\boldsymbol{w}_{ji}^*$ in the second case. Here, $B(\boldsymbol{c}, r)$ is notation for the closed Euclidean ball centered at $\boldsymbol{c}$ of radius r. Since $\boldsymbol{x}_i^* = \hat{\boldsymbol{x}}$ for $i \in C$, $\boldsymbol{x}_i^* \neq \hat{\boldsymbol{x}}$ for $i \notin C$, the KKT condition for $i \in C$ is rewritten as

$$\hat{\boldsymbol{x}} - \boldsymbol{a}_i + \lambda \sum_{j \notin C} \frac{\hat{\boldsymbol{x}} - \boldsymbol{x}_j^*}{\left\|\hat{\boldsymbol{x}} - \boldsymbol{x}_j^*\right\|} + \lambda \sum_{j \in C - \{i\}} \boldsymbol{w}_{ij}^* = \mathbf{0}, \tag{3.3}$$

Define $\boldsymbol{z}_{ij}^* = \boldsymbol{w}_{ij}^*$ for $i, j \in C$, $i \neq j$. Then

$$\left\|\boldsymbol{z}_{ij}^*\right\| \leq 1, \boldsymbol{z}_{ij}^* = -\boldsymbol{z}_{ji}^*, \forall i, j \in C, i \neq j.$$

Substitute $\boldsymbol{w}_{ij}^* = \boldsymbol{z}_{ij}^*$ into the equation (3.3) to obtain

$$\hat{\boldsymbol{x}} - \boldsymbol{a}_i + \lambda \sum_{j \notin C} \frac{\hat{\boldsymbol{x}} - \boldsymbol{x}_j^*}{\left\|\hat{\boldsymbol{x}} - \boldsymbol{x}_j^*\right\|} + \lambda \sum_{j \in C - \{i\}} \boldsymbol{z}_{ij}^* = \mathbf{0}, \tag{3.4}$$

Sum the preceding equation over $i \in C$, noticing that the last term cancels out, leaving

$$|C|\hat{\boldsymbol{x}} - \sum_{i \in C} \boldsymbol{a}_i + \lambda |C| \sum_{j \notin C} \frac{\hat{\boldsymbol{x}} - \boldsymbol{x}_j^*}{\left\|\hat{\boldsymbol{x}} - \boldsymbol{x}_j^*\right\|} = \mathbf{0},$$

which is rearranged to (renaming i to l):

$$\lambda \sum_{j \notin C} \frac{\hat{\boldsymbol{x}} - \boldsymbol{x}_j^*}{\left\|\hat{\boldsymbol{x}} - \boldsymbol{x}_j^*\right\|} = -\hat{\boldsymbol{x}} + \frac{1}{|C|} \sum_{l \in C} \boldsymbol{a}_l. \tag{3.5}$$

Subtract (3.5) from (3.4), simplify and rearrange to obtain

$$\boldsymbol{a}_i - \frac{1}{|C|} \sum_{l \in C} \boldsymbol{a}_l = \lambda \sum_{j \in C - \{i\}} \boldsymbol{z}_{ij}^* \qquad \forall i \in C, \tag{3.6}$$

as desired.

Proof for Sufficiency (b)

We will show that at the solution of (1.1), all the $\boldsymbol{x}_i^*$'s for $i \in C$ have a common value under the hypothesis that $\boldsymbol{z}_{ij}^*$ is a solution to the equation (3.1) for $i, j \in C$, $i \neq j$.

First, define the following intermediate problem. Let $\tilde{\boldsymbol{a}}$ denote the centroid of $\boldsymbol{a}_i$ for $i \in C$:

$$\tilde{\boldsymbol{a}} = \frac{1}{|C|} \sum_{l \in C} \boldsymbol{a}_l.$$

Consider the weighted problem sum-of-norms clustering problem with unknowns as follows: one unknown $\boldsymbol{x} \in \mathbf{R}^d$ is associated with C, and one unknown $\boldsymbol{x}_j$ is associated with each $j \notin C$ (for a total of $n - |C| + 1$ unknown vectors):

$$\min_{\boldsymbol{x}; \boldsymbol{x}_j} \frac{|C|}{2} \cdot \|\boldsymbol{x} - \tilde{\boldsymbol{a}}\|^2 + \frac{1}{2} \sum_{j \notin C} \|\boldsymbol{x}_j - \boldsymbol{a}_j\|^2 + \lambda |C| \sum_{j \notin C} \|\boldsymbol{x} - \boldsymbol{x}_j\| + \lambda \sum_{\substack{i, j \notin C \\ i < j}} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|. \tag{3.7}$$

This problem, being strongly convex, has a unique optimizer; denote the optimizing vectors $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{x}}_j$ for $j \notin C$.

The optimality conditions for (3.7) are:

$$|C|(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{a}}) + \lambda |C| \sum_{j \notin C} \boldsymbol{g}_j = \mathbf{0}, \tag{3.8}$$

$$\tilde{\boldsymbol{x}}_i - \boldsymbol{a}_i - \lambda |C| \boldsymbol{g}_i + \lambda \sum_{j \notin C \cup \{i\}} \boldsymbol{y}_{ij} = \mathbf{0} \qquad \forall i \notin C, \tag{3.9}$$

with subgradients defined as follows:

$$\boldsymbol{g}_j = \begin{cases} \frac{\tilde{\boldsymbol{x}}-\tilde{\boldsymbol{x}}_j}{\|\tilde{\boldsymbol{x}}-\tilde{\boldsymbol{x}}_j\|}, & \text{for } \tilde{\boldsymbol{x}}_j \neq \tilde{\boldsymbol{x}}, \\ \text{arbitrary in } B(\boldsymbol{0},1), & \text{for } \tilde{\boldsymbol{x}}_j = \tilde{\boldsymbol{x}}, \end{cases} \qquad \forall j \notin C,$$

and

$$\boldsymbol{y}_{ij} = \begin{cases} \frac{\tilde{\boldsymbol{x}}_i-\tilde{\boldsymbol{x}}_j}{\|\tilde{\boldsymbol{x}}_i-\tilde{\boldsymbol{x}}_j\|}, & \text{for } \tilde{\boldsymbol{x}}_i \neq \tilde{\boldsymbol{x}}_j, \\ \text{arbitrary in } B(\boldsymbol{0},1), & \text{for } \tilde{\boldsymbol{x}}_i = \tilde{\boldsymbol{x}}_j, \end{cases} \qquad \forall i,j \notin C, i \neq j,$$

with the proviso that in the second case, $\boldsymbol{y}_{ij} = -\boldsymbol{y}_{ji}$.

We claim that the solution for (1.1) given by defining $\boldsymbol{x}_i^* = \tilde{\boldsymbol{x}}$ for $i \in C$ while keeping the $\boldsymbol{x}_j^* = \tilde{\boldsymbol{x}}_j$ for $j \notin C$, where $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{x}}_j$ are the optimizers for (3.7) as in the last few paragraphs, is optimal for (1.1), which proves the main result. To show that this solution is optimal for (1.1), we need to provide subgradients to establish the necessary condition. Define $\boldsymbol{w}_{ij}$ to be the subgradients of $\boldsymbol{x}_i \mapsto \left\|\boldsymbol{x}_i - \tilde{\boldsymbol{x}}_j^*\right\|$ evaluated at $\tilde{\boldsymbol{x}}_i^*$ as follows:

$$\begin{aligned} \boldsymbol{w}_{ij} &= \boldsymbol{g}_j && \text{for } i \in C, j \notin C, \\ \boldsymbol{w}_{ij} &= \boldsymbol{y}_{ij} && \text{for } i,j \notin C,\ i \neq j, \\ \boldsymbol{w}_{ij} &= \boldsymbol{z}_{ij}^* && \text{for } i,j \in C,\ i \neq j, \end{aligned}$$

Before confirming that the necessary condition is satisfied, we first need to confirm that these are all valid subgradients. In the case that $i \in C$, $j \notin C$, we have constructed $\boldsymbol{g}_j$ to be a valid subgradient of $\boldsymbol{x} \mapsto \|\boldsymbol{x} - \tilde{\boldsymbol{x}}_j\|$ evaluated at $\tilde{\boldsymbol{x}}$, and we have taken $\boldsymbol{x}_i^* = \tilde{\boldsymbol{x}}$, $\boldsymbol{x}_j^* = \tilde{\boldsymbol{x}}_j$.

In the case that $i,j \notin C$, we have construct $\boldsymbol{y}_{ij}$ to be a valid subgradient of $\boldsymbol{x} \mapsto \|\boldsymbol{x} - \tilde{\boldsymbol{x}}_j\|$ evaluated at $\tilde{\boldsymbol{x}}_i$, and we have taken $\boldsymbol{x}_i^* = \tilde{\boldsymbol{x}}_i$, $\boldsymbol{x}_j^* = \tilde{\boldsymbol{x}}_j$.

In the case that $i,j \in C$, by construction $\boldsymbol{x}_i^* = \boldsymbol{x}_j^* = \tilde{\boldsymbol{x}}$, so any vector in $B(\boldsymbol{0},1)$ is a valid subgradient of $\boldsymbol{x} \mapsto \|\boldsymbol{x} - \tilde{\boldsymbol{x}}_j\|$ evaluated $\tilde{\boldsymbol{x}}_i$. Note that since $\boldsymbol{z}_{ij}^* \in B(\boldsymbol{0},1)$, then $\boldsymbol{w}_{ij}$ defined above also lies in $B(\boldsymbol{0},1)$.

Now we check the necessary conditions for optimality in (1.1). First, consider an $i \in C$:

$$\begin{aligned}
\tilde{\boldsymbol{x}}_i^* - \boldsymbol{a}_i + \lambda \sum_{j \neq i} \boldsymbol{w}_{ij} &= \tilde{\boldsymbol{x}} - \boldsymbol{a}_i + \lambda \sum_{j \in C - \{i\}} \boldsymbol{w}_{ij} + \lambda \sum_{j \notin C} \boldsymbol{w}_{ij} \\
&= \tilde{\boldsymbol{x}} - \boldsymbol{a}_i + \lambda \sum_{j \in C - \{i\}} \boldsymbol{z}_{ij}^* + \lambda \sum_{j \notin C} \boldsymbol{g}_j \\
&= \tilde{\boldsymbol{x}} - \boldsymbol{a}_i + \boldsymbol{a}_i - \frac{1}{|C|} \sum_{l \in C} \boldsymbol{a}_l + \lambda \sum_{j \notin C} \boldsymbol{g}_j \qquad \text{(by (3.1))} \\
&= \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{a}} + \lambda \sum_{j \notin C} \boldsymbol{g}_j \\
&= \boldsymbol{0} \qquad \text{(by (3.8)).}
\end{aligned}$$

Then we check for $i \notin C$:

$$\begin{aligned}
\tilde{\boldsymbol{x}}_i^* - \boldsymbol{a}_i + \lambda \sum_{j \neq i} \boldsymbol{w}_{ij} &= \tilde{\boldsymbol{x}}_i - \boldsymbol{a}_i + \lambda \sum_{j \in C} \boldsymbol{w}_{ij} + \lambda \sum_{j \notin C \cup \{i\}} \boldsymbol{w}_{ij} \\
&= \tilde{\boldsymbol{x}}_i - \boldsymbol{a}_i + \lambda \sum_{j \in C} (-\boldsymbol{g}_i) + \lambda \sum_{j \notin C \cup \{i\}} \boldsymbol{y}_{ij} \\
&= \tilde{\boldsymbol{x}}_i - \boldsymbol{a}_i - \lambda |C| \boldsymbol{g}_i + \lambda \sum_{j \notin C \cup \{i\}} \boldsymbol{y}_{ij} \\
&= \boldsymbol{0} \qquad \text{(by (3.9)).}
\end{aligned}$$

□

3.2 Agglomeration Conjecture

Recall that when $\lambda = 0$, each $\boldsymbol{a}_i$ is in its own cluster in the solution to (1.1) (provided the $\boldsymbol{a}_i$'s are distinct), whereas for sufficiently large λ, all the points are in one cluster. Hocking et al. [12] conjectured that sum-of-norms clustering with equal weights has the following agglomeration property: as λ increases, clusters merge with each other but never break up. This means that the solutions to (1.1) as λ ranges over $[0, \infty)$ induce a tree of hierarchical clusters on the data.

This conjecture was proved by Chiquet et al. using Theorem 1. Consider a $\bar{\lambda} \geq \lambda$ and

its corresponding sum-of-norms cluster model:

$$\min_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n} \frac{1}{2}\sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{a}_i\|^2 + \bar{\lambda}\sum_{i<j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|\,. \tag{3.10}$$

Corollary 1.1. *(Chiquet et al.) If there is a C such that minimizer $\boldsymbol{x}^*$ of (1.1) satisfies $\boldsymbol{x}_i^* = \hat{\boldsymbol{x}}$ for $i \in C$, $\boldsymbol{x}_i^* \neq \hat{\boldsymbol{x}}$ for $i \notin C$ for some $\hat{\boldsymbol{x}} \in \mathbf{R}^d$, then there exists an $\hat{\boldsymbol{x}}' \in \mathbf{R}^d$ such that the minimizer of (3.10), $\bar{\boldsymbol{x}}^*$, satisfies $\bar{\boldsymbol{x}}_i^* = \hat{\boldsymbol{x}}'$ for $i \in C$.*

The corollary follows from Theorem 1. If C is a cluster in the solution of (1.1), then by the necessary condition, there exist subgraidents $\boldsymbol{z}_{ij}^*$ satisfying (3.1) for λ. If we scale each of these subgradients by $\lambda/\bar{\lambda}$, we now obtain a solution to (3.1) for with λ replaced by $\bar{\lambda}$, and the theorem states that this is sufficient for the points in C to be in the same cluster in the solution to (3.10).

Let *fusion values* denote the values of λ at which clusters fuse to form a larger cluster. According to agglomeration theorem 1.1, there are at most n fusion values.

It should be noted that Hocking et al. construct an example of unequally-weighted sum-of-norms clustering in which the agglomeration property fails. It is still mostly an open question to characterize for which norms and for which families of unequal weights the agglomeration property holds. Refer to Chi and Steinerberger [6] for some recent progress.

3.3 Extension to other weights

Several authors, e.g., Sun et al. [28] have introduced weights into either the first or second or both summations in (1.1). One purpose for introducing weights is to be able to eliminate many of the terms in the second summation (i.e., use a weight of 0 on those terms) in order to reduce the number of terms in the objective function to $o(n^2)$ for the purpose of efficient computation. For example, Sun et al. use exponentially decaying weights as in (4.11) below that are zeroed out for $\boldsymbol{a}_i$'s sufficiently far apart. The Chiquet et al. characterization theorem, however, does not extend to fully general weights. (The obstacle is that the left-hand side of (3.5) does not cancel out the third term on the left-hand side of (3.4) for general weights.) The most general class of weights for which the theorem applies is *multiplicative weights*, which are as follows. Each data point $\boldsymbol{a}_i$ for $i = 1, \ldots, n$ is associated with a positive weight r_i. Then both terms in (1.1) are weighted as follows:

$$\min_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n \in \mathbf{R}^d} \frac{1}{2}\sum_{i=1}^{n} r_i\|\boldsymbol{x}_i - \boldsymbol{a}_i\|^2 + \lambda\sum_{i<j} r_i r_j\|\boldsymbol{x}_i - \boldsymbol{x}_j\|. \tag{3.11}$$

Therefore, our recovery theorem also extends to multiplicative weights, which is the subject of the rest of this section. A small computational experiment reported in Section 4.5 suggests recovery of a mixture of Gaussians may also be possible with exponentially decaying weights.

We can draw the same conclusions as Theorem 1 when (3.1) in the necessary and sufficient conditions is replaced with the following system of equalities and a norm inequality:

$$\begin{aligned} \boldsymbol{a}_i - \sum_{l \in C} \frac{r_l}{\sum_{l' \in C} r_{l'}} \boldsymbol{a}_l &= \lambda \sum_{j \in C - \{i\}} r_j \boldsymbol{z}_{ij}^* && \forall i \in C, \\ \left\| \boldsymbol{z}_{ij}^* \right\| &\leq 1 && \forall i, j \in C, i \neq j, \\ \boldsymbol{z}_{ij}^* &= -\boldsymbol{z}_{ji}^* && \forall i, j \in C, i \neq j. \end{aligned} \tag{3.12}$$

The proof of this generalization is analogous to the proof of Theorem 1, which we omit.

An analogous agglomeration conjecture for this setting was shown by Chiquet et al. [8], i.e. the path of solutions to (3.11) as λ ranges over $[0, \infty)$ contains no splits for multiplicative weights.

Chapter 4

Recovery of mixture of Gaussians [1]

4.1 Mixture of Gaussians

In this chapter, we assume the data $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ is generated by a mixture of Gaussians, which is the same generative model adopted in [21]. The parameters of the generative model are K means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K \in \mathbf{R}^d$, K variances $\sigma_1^2, \ldots, \sigma_k^2$, and K probabilities $w_1, \ldots, w_k$, all positive and summing to 1. One draws n i.i.d. samples as follows. First, an index $m \in \{1, \ldots, K\}$ is selected at random according to probabilities $w_1, \ldots, w_K$. Next, a point $\boldsymbol{a}$ is chosen according to the spherical Gaussian distribution $N(\boldsymbol{\mu}_m, \sigma_m^2 I)$. Note that the covariance matrix could also be arbitrary, and our result could be extended to an arbitrary covariance matrix. Our assumption is aligned with the assumption in Panahi et al. [21].

Panahi et al. [21] proved that for the appropriate choice of λ, sum-of-norms clustering formulation (1.1) will exactly recover a mixture of Gaussians (i.e., each point will be labeled with m if it was selected from $N(\boldsymbol{\mu}_m, \sigma_m^2 I)$) provided that for all m, m', $1 \leq m < m' \leq K$,

$$\|\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\| \geq \frac{CK\sigma_{\max}}{w_{\min}} \text{polylog}(n), \tag{4.1}$$

where $C = \frac{1}{n}\sum_i \boldsymbol{a}_i$, $\sigma_{max} = \max_m \sigma_m$ and $w_{min} = \min_m w_m$. As $n \to \infty$, $\boldsymbol{\mu}_m$'s have to be more distantly separated. Hence, distinguishing the clusters becomes increasingly difficult.

[1] This chapter is based on *Recovery of a mixture of Gaussians by sum-of-norms clustering* by Jiang, Vavasis and Zhai [15]

4.2 Main recovery theorem

In this section, we present our main result about recovery of mixture of Gaussians. As noted in the introduction, a theorem stating that every point is labeled correctly is not possible in the setting of $n \to \infty$, so we settle for a theorem stating that points within a constant number of standard deviations from the means are correctly labeled.

Theorem 2. *Let the vertices $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbf{R}^d$ be generated from a mixture of K Gaussian distributions with parameters $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$, $\sigma_1^2, \ldots, \sigma_K^2$, and $w_1, \ldots, w_K$. Let $\theta > 0$ be given, and let*

$$V_m = \{i : \|\boldsymbol{a}_i - \boldsymbol{\mu}_m\| \leq \theta\sigma_m\}, \quad m = 1, \ldots, K.$$

Let $\epsilon > 0$ be arbitrary. Then for any $m = 1, \ldots, K$, with probability exponentially close to 1 (and depending on ϵ; see (4.5)) as $n \to \infty$, for the solution $\boldsymbol{x}^*$ to (1.1), the points indexed by V_m are in the same cluster provided

$$\lambda \geq \frac{2\theta\sigma_m}{(F(\theta, d)w_m - \epsilon)n}. \tag{4.2}$$

Here, $F(\theta, d)$ denotes the cumulative density function of the chi distribution with d degrees of freedom (which tends to 1 rapidly as θ increases). Furthermore, the cluster associated with V_m is distinct from the cluster associated with $V_{m'}$, $1 \leq m < m' \leq K$ with probability exponentially close to 1 as $n \to \infty$ (see (4.6)), provided that

$$\lambda < \frac{\|\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\|}{2(n-1)}. \tag{4.3}$$

In order to state a simpler bound, we can fix some values. For example, let us take $\theta = 2d^{1/2}$ and let $c_d = F(2d^{1/2}, d)$. The Chernoff bound implies that $c_d \to 1$ exponentially fast in d. Let $w_{\min} = \min_{m=1,\ldots,K} w_m$ and $\sigma_{\max} = \max_{m=1,\ldots,K} \sigma_m$. Finally, let us take $\epsilon = c_d w_{\min}/2$. Then the above theorem states there is a λ such that with probability tending to 1 exponentially fast in n, the points in V_m, for any $m = 1, \ldots, K$ are each in the same cluster, and these clusters are distinct, provided that

$$\min_{1 \leq m < m' \leq K} \|\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\| > \frac{16\sqrt{d}\sigma_{\max}}{c_d w_{\min}}. \tag{4.4}$$

Compared to the Panahi et al. bound (4.1), we have removed the dependence of the right-hand side on n as well as the factor of K. (The dependence of the Panahi et al. bound on d is not made explicit so we cannot compare the two bounds' dependence on d. Note that there is still an implicit dependence on K in (4.4) since necessarily $w_{\min} \leq 1/K$.)

4.3 Proof of the main theorem

Proof. Let $\epsilon > 0$ be fixed. Fix an $m \in \{1, \ldots, K\}$. First, we show that all the points indexed by V_m are in the same cluster. The usual technique for proving a recovery result is to find subgradients to satisfy the sufficient condition, which in this case is Theorem 1 taking C in the theorem to be V_m. Observe that conditions (3.1) involve equalities and norm inequalities. A standard technique in the literature (see, e.g., Candès and Recht [4]) is to find the least-squares solution to the equalities and then prove that it satisfies the inequalities. This is the technique we adopt herein. The conditions (3.1) are in sufficiently simple form that we can write down the least-squares solution in closed form; it turns out to be:

$$\boldsymbol{z}_{ij}^* = \frac{1}{\lambda |V_m|}(\boldsymbol{a}_i - \boldsymbol{a}_j) \quad \forall i, j \in V_m, i \neq j.$$

It follows by construction (and is easy to check) that this formula satisfies the equalities in (3.1), so the remaining task is to show that the norm bound $\left\|\boldsymbol{z}_{ij}^*\right\| \leq 1$ is satisfied. By definition of V_m, $\|\boldsymbol{a}_i - \boldsymbol{a}_j\| \leq 2\theta\sigma_m$. The probability that an arbitrary sample $\boldsymbol{a}_i$ is associated with mean $\boldsymbol{\mu}_m$ is w_m. Furthermore, with probability $F(\theta, d)$, this sample satisfies $\|\boldsymbol{a}_i - \boldsymbol{\mu}_m\| \leq \theta\sigma_m$, i.e., $i \in V_m$. Since the second choice in the mixture of Gaussians is conditionally independent from the first, the overall probability that $i \in V_m$ is $F(\theta, d)w_m$. Therefore, $E[|V_m|] = F(\theta, d)w_m n$. It follows that the probability that $|V_m| \geq (F(\theta, d)w_m - \epsilon)n$ is exponentially close to 1 as $n \to \infty$ for a fixed $\epsilon > 0$. Specifically,

$$\text{Prob}\left[|V_m| \geq (F(\theta, d)w_m - \epsilon)n\right] \geq 1 - \exp(-2\epsilon^2 n), \tag{4.5}$$

by Hoeffding's inequality for the binomial distribution [13]. Thus, provided

$$\lambda \geq 2\theta\sigma_m / ((F(\theta, d)w_m - \epsilon)n),$$

we have constructed a solution to (3.1) with probability exponentially close to 1 as $n \to \infty$.

For the second part of the theorem, suppose $1 \leq m < m' \leq K$. For each sample $\boldsymbol{a}_i$ associated with $\boldsymbol{\mu}_m$ satisfying $\|\boldsymbol{a}_i - \boldsymbol{\mu}_m\| \leq \theta\sigma_m$ (i.e., lying in V_m), the probability is $1/2$ that

$$(\boldsymbol{a}_i - \boldsymbol{\mu}_m)^T(\boldsymbol{\mu}_{m'} - \boldsymbol{\mu}_m) \leq 0$$

by the fact that the spherical Gaussian distribution has mirror-image symmetry about any hyperplane through its mean. Therefore, with probability exponentially close to 1 as $n \to \infty$, we can assume that at least one $i \in V_m$ satisfies the above inequality. In particular,

$$\text{Prob}\left[\exists i \in V_m \text{ s.t. } (\boldsymbol{a}_i - \boldsymbol{\mu}_m)^T(\boldsymbol{\mu}_{m'} - \boldsymbol{\mu}_m) \leq 0\right] \geq 1 - 2^{-|V_m|}, \tag{4.6}$$

(Note that, as noted above, $|V_m|$ grows linearly with n with probability exponentially close to 1 as $n \to \infty$.) Similarly, with probability exponentially close to 1, at least one sample $i' \in V_{m'}$ satisfies

$$(\boldsymbol{a}_{i'} - \boldsymbol{\mu}_{m'})^T(\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}) \leq 0.$$

Then

$$\begin{aligned}
\|\boldsymbol{a}_i - \boldsymbol{a}_{i'}\|^2 &= \|\boldsymbol{a}_i - \boldsymbol{\mu}_m - \boldsymbol{a}_{i'} + \boldsymbol{\mu}_{m'} + \boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\|^2 \\
&= \|\boldsymbol{a}_i - \boldsymbol{\mu}_m - \boldsymbol{a}_i + \boldsymbol{\mu}_{m'}\|^2 + 2(\boldsymbol{a}_i - \boldsymbol{\mu}_m)^T(\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}) \\
&\quad - 2(\boldsymbol{a}_{i'} - \boldsymbol{\mu}_{m'})^T(\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}) + \|\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\|^2 \\
&\geq \|\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\|^2, \qquad (4.7)
\end{aligned}$$

where, in the final line, we used the two inequalities derived earlier in this paragraph.

Consider the first-order optimality conditions for equation (1.1), which are given by (3.2). Apply the triangle inequality to the summation in (3.2) to obtain,

$$\|\boldsymbol{x}_i^* - \boldsymbol{a}_i\| \leq \lambda(n-1), \text{ and} \qquad (4.8)$$
$$\|\boldsymbol{x}_{i'}^* - \boldsymbol{a}_{i'}\| \leq \lambda(n-1). \qquad (4.9)$$

Therefore,

$$\begin{aligned}
\|\boldsymbol{x}_i^* - \boldsymbol{x}_{i'}^*\| &= \|\boldsymbol{a}_i - \boldsymbol{a}_{i'} + \boldsymbol{x}_i^* - \boldsymbol{a}_i - \boldsymbol{x}_{i'}^* + \boldsymbol{a}_{i'}\| \\
&\geq \|\boldsymbol{a}_i - \boldsymbol{a}_{i'}\| - \|\boldsymbol{x}_i^* - \boldsymbol{a}_i\| - \|\boldsymbol{x}_{i'}^* - \boldsymbol{a}_{i'}\| \qquad \text{(by the triangle inequality)} \\
&\geq \|\boldsymbol{\mu}_{m'} - \boldsymbol{\mu}_m\| - 2\lambda(n-1) \qquad \text{(by (4.7), (4.8), and (4.9)).}
\end{aligned}$$

Therefore, we conclude that $\boldsymbol{x}_i^* \neq \boldsymbol{x}_{i'}^*$, i.e., that V_m and $V_{m'}$ are not in the same cluster, provided that the right-hand side of the preceding inequality is positive, i.e.,

$$\lambda < \frac{\|\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\|}{2(n-1)}.$$

This concludes the proof of the second statement. □

4.4 Extension to multiplicative weights

With the new theorem of cluster characterization, we can derive the conditions about recovery of mixture of Gaussians in the case of multiplicative weights noted in (3.11), as an

extension to Theorem 2. This requires a further modeling assumption on the distribution of the weights. As before, assume each data item $\boldsymbol{a}_i$, $i = 1, \ldots, n$ is chosen from a mixture of K Gaussians. Assume that the weight r_i associated with data item $\boldsymbol{a}_i$ is chosen independently at random according to $r_i \sim \Omega_m$. Here, $m \in \{1, \ldots, K\}$ denotes the specific Gaussian associated with $\boldsymbol{a}_i$. The distributions $\Omega_1, \ldots, \Omega_K$ are all assumed to be supported in a single bounded interval $[0, R]$. Denote the mean of Ω_m as $\bar{r}_m$, $m = 1, \ldots, K$. Assume these means are all positive: $0 < \bar{r}_m \leq R$.

The main result is that for any $m = 1, \ldots, K$, with probability exponentially close to 1 (and depending on ϵ) as $n \to \infty$, for the solution $\boldsymbol{x}^*$ computed by (3.11), the points in V_m are in the same cluster provided that

$$\lambda \geq \frac{2\theta\sigma_m}{(F(\theta, d)w_m - \epsilon)n\bar{r}_m},$$

and the cluster associated with V_m is distinct from the cluster associated with $V_{m'}$, $1 \leq m < m' \leq K$, provided that

$$\lambda < \frac{\|\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\|}{2(n-1)(\bar{r} - \epsilon)},$$

where $\bar{r}$ is the overall mean of the r_i's, that is, $\bar{r} = w_1\bar{r}_1 + \cdots + w_K\bar{r}_K$.

Similar techniques from the proof of Theorem 2 are used to prove the recovery of the multiplicative-weight problem. First, we can construct a solution to (3.12) as follows

$$\boldsymbol{z}_{ij}^* = \frac{1}{\lambda r'_m}(\boldsymbol{a}_i - \boldsymbol{a}_j) \quad \forall i, j \in V_m, i \neq j,$$

where $r'_m = \sum_{l \in V_m} r_l$ and $\bar{\boldsymbol{a}}_m = \sum_{l \in V_m} \frac{r_l}{r'_m}\boldsymbol{a}_l$. Our task is to prove that the norm bound $\left\|\boldsymbol{z}_{ij}^*\right\| \leq 1$ holds. By definition of V_m, $\|\boldsymbol{a}_i - \boldsymbol{a}_j\| \leq 2\theta\sigma_m$. As before, the probability that $|V_m| \geq (F(\theta, d)w_m - \epsilon_1)n$ is exponentially close to 1 as $n \to \infty$ for a fixed $\epsilon_1 > 0$. Furthermore, the probability that $r'_m \geq (\bar{r}_m - \epsilon_2)|V_m|$ is exponentially close to 1 by Hoeffding's inequality [13] as $n \to \infty$ for fixed ϵ_2. Thus, provided

$$\lambda \geq \frac{2\theta\sigma_m}{(F(\theta, d)w_m\bar{r}_m - \epsilon)n},$$

we have constructed a solution to (3.12) with probability exponentially close to 1, which implies that all points in V_m are in the same cluster.

Turn now to the analysis of the upper bound on λ. The first-order optimality conditions of (3.11) imply the following inequalities by applying the triangle inequality to the

summation of subgradients

$$\|\boldsymbol{x}_i^* - \boldsymbol{a}_i\| \leq \lambda \sum_{j \neq i} r_j, \quad \forall i \tag{4.10}$$

By the same argument in the proof of Theorem 2, there exist at least one $i \in V_m, i' \in V_{m'}$ satisfying the following inequality with probability exponentially close to 1

$$\|\boldsymbol{x}_i^* - \boldsymbol{x}_{i'}^*\| \geq \|\boldsymbol{\mu}_{m'} - \boldsymbol{\mu}_m\| - \lambda \sum_{j \neq i} r_j - \lambda \sum_{j \neq i'} r_j \quad \text{(by (4.7), (4.10))}.$$

Therefore, we conclude that $\boldsymbol{x}_i^* \neq \boldsymbol{x}_{i'}^*$, i.e., that V_m and $V_{m'}$ are not in the same cluster, provided that for all $i \in V_m, i' \in V_{m'}$

$$\lambda < \frac{\|\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\|}{\sum_{j \neq i} r_j + \sum_{j \neq i'} r_j}.$$

Applying Hoeffding's bound again, we can claim that for any $\epsilon > 0$, with probability tending to 1 exponentially fast with n, this inequality will hold provided that

$$\lambda < \frac{\|\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\|}{2(n-1)(\bar{r} - \epsilon)}.$$

4.5 Computational experiments

In this section, we perform experiments in which a solver for sum-of-norms clustering is applied to a set of points drawn from a mixture of Gaussians. Four experiments are performed to address four questions: (1) How flexibly can λ be chosen? (2) How does the recovery depend on d, the space dimension? (3) How does the recovery degrade as σ (the standard deviation of the Gaussians) increases? and (4) Does the result hold for general weights?

Even though we do not attempt to test sum-of-norms clustering on other data sets in this section, we are investigating its performance on a half-moon data set in Section 5.5. For other data sets, we also refer the reader to [7, 12]. Moreover, the reader may refer to [12] for comparison of sum-of-norms to other clustering algorithms.

In all cases, the code used is our own Julia [3] implemention of the Chi-Lange [7] ADMM solver. Each iteration of this solver requires $O(n^2 d)$ operations since the objective function contains $O(n^2)$ terms, each involving vectors of length d. We observed that the number

of iterations to reach a fixed tolerance scales linearly with n. This means that the overall running time scales cubically with n. Our convergence tolerance ϵ_{tol} was taken to be 10^{-6} in all cases. This tolerance corresponds to the quantities ϵ^{pri} and ϵ^{rel} in the supplemental material of [7]. These parameters correspond to the absolute and relative precisions, which control the primal and dual precisions. The algorithm terminates when the primal and dual residuals are bounded by the precisions respectively. With this tolerance, the runs described below took approximately 27 hours total on an Intel Xeon processor single-threaded.

After termination, clusters were recovered from the approximately converged solution $\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n$ as follows. An i is selected arbitrarily from $\{1, \ldots, n\}$. Then all vectors j such that $\|\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}_j\| \leq \sqrt{\epsilon_{\text{tol}}}$ are assigned to a cluster. These j's (including i itself) are then deleted from the list of nodes, and the process is repeated until all nodes are used up. Call these recovered clusters $R_1, \ldots, R_{K'}$. The question of how to best retrieve clusters from an approximate solution of (1.1) is nontrivial and is left as a topic for future research.

Then V_m, $m = 1, \ldots, K$, are mapped to one of these recovered clusters, i.e., a mapping $\ell : \{1, \ldots, K\} \to \{1, \ldots, K'\}$ is computed such that $R_{\ell(m)}$ contains the most number of elements of V_m. In other words,

$$\ell(m) := \operatorname*{argmax}_{m'=1,\ldots,K'} \#(V_m \cap R_{m'}),$$

for each $m = 1, \ldots, K$, with ties broken arbitrarily. (Here, $\#(\cdot)$ denotes set-cardinality.) This mapping $\ell(\cdot)$ is not necessarily injective.

Then three scores are computed:

$$s_1 = \frac{1}{\#(V_1 \cup \cdots \cup V_m)} \sum_{m=1}^{K} \#(V_m \cap R_{\ell(m)}),$$

which is the fraction of entries in $V_1 \cup \cdots \cup V_m$ correctly clustered,

$$s_2 = \frac{1}{n} \sum_{m=1}^{K} \#\{i \in \{1, \ldots, n\} : \boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{\mu}_m, \sigma_m^2 I) \text{ and } i \in R_{\ell(m)}\},$$

the fraction of entries of all n data points correctly clustered, and

$$s_3 = \frac{\#\ell(\{1, \ldots, K\})}{K},$$

the number of distinct recovered clusters divided by the true number. Note that as λ increases, one would expect s_1 and s_2 to increase while s_3 decreases, since clusters expand as λ increases.

The first experiment is meant to determine whether choices λ outside the range specified by Theorem 2 can still recover clusters. For this experiment we chose $n = 1000$, $d = K = 6$, $w_i = 1/6$ and $\boldsymbol{\mu}_i = \boldsymbol{e}_i$ (ith column of the identity matrix) for $i = 1, \ldots, 6$, $\sigma = 0.0094$, and $\theta = 2.0$. This choice of σ is made so that the upper and lower bounds on λ in Theorem 2 are nearly equal to a single value $\lambda^* = 7.0 \cdot 10^{-4}$. Then we tested recovery for $\lambda = \kappa\lambda^*$ with $\kappa = 1/4, 1/2, 1, 2, 4$, as shown in Table 4.1.

The data in Table 4.1 indicates that the recovery is perfect between $\lambda^*/2$ and $2\lambda^*$. As the theorem predicts, as λ increases, a greater number of V_m's is recovered, but a smaller number of V_m's are distinct. This table suggests that a strengthening may exist of our main theorem in which both inequalities are less restrictive, but not by orders of magnitude.

Table 4.1: Recovery for varying λ. Value λ^* is the essentially unique value satisfying the two inequalities of Theorem 2.

λ	s_1 (% of V_m recovered)	s_2 (total % recovered)	s_3 (% distinct clusters)
$\lambda^*/4$	38/304	39/1000	6/6
$\lambda^*/2$	304/304	1000/1000	6/6
λ^*	304/304	1000/1000	6/6
$2\lambda^*$	304/304	1000/1000	6/6
$4\lambda^*$	304/304	1000/1000	1/6

In the second experiment, we varied d and K. Note that as d and K get larger for fixed n, we move away from the asymptotic range in which Theorem 2 applies since the size of each cluster shrinks. On the other hand, as n is fixed while d and k get larger, we are closer to the range of parameters for which the Panahi et al. result applies. For these tests, we fixed $n = 1000$, looped over $d = K = 4, 16, 64$ and $\theta = \sqrt{d}$ (so that $\theta = 2, 4, 8$). Note that this variation of θ with respect to d is chosen so that $F(\theta, d)$ is about the same value (between .5 and .6) for all three trials. As in the previous experiment, we chose $w_i = 1/K$ and $\boldsymbol{\mu}_i = \boldsymbol{e}_i$ (ith column of the identity matrix) for $i = 1, \ldots, K$. Finally, we chose σ so that the upper and lower bounds in Theorem 2 were equal, and we chose λ to be this unique value of λ. (Note that σ shrinks like $d^{3/2}$ for this variation of parameters.)

We found that in all three cases, all 1000 points were clustered correctly into K distinct clusters (so no table is presented). This robust behavior is not predicted by our theorem, since the arguments in the theorem are weak if n/K is small. See further comments on this matter in Chapter 6.

The next experiment considers the effect of increasing σ. For this experiment we fixed $d = 1$, $K = 2$, $n = 1000$, $\mu_1 = 0$, $\mu_2 = 1$, $w_1 = w_2 = 1/2$, $\theta = 1$. Let $\lambda_{\max}$ be the value appearing on the right-hand side of (4.3). In all trials, we fixed $\lambda = \lambda_{\max}$, which does not depend on σ. We chose σ^* to be the value of σ that makes the right-hand sides of (4.2) and (4.3) equal. Then we increased σ by factors of $\sqrt{2}$ to observe the effect on recovery. The results appear in Table 4.2. Note that the method continues to be robust for values of σ modestly outside the range that we have established, but then the behavior quickly decays. It is likely that we could have gotten better performance by carefully tuning λ.

Table 4.2: Recovery for varying σ. Here, σ^* is the unique value that makes the right-hand sides of (4.2) and (4.3) equal.

σ	s_1 (% of V_m recovered)	s_2 (total % recovered)	s_3 (% distinct clusters)
σ^*	700/700	996/1000	2/2
$2^{1/2}\sigma^*$	700/700	950/1000	2/2
$2\sigma^*$	700/700	742/1000	2/2
$2^{3/2}\sigma^*$	108/700	108/1000	2/2
$4\sigma^*$	46/700	46/1000	2/2

The last experiment is a study of exponentially decaying weights, which is a case in which our theory does not apply. Similar to Yuan et al. [30], we used the following weighting:

$$\min \frac{1}{2}\sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{a}_i\|^2 + \lambda \sum_{i<j} \exp(-\phi \|\boldsymbol{a}_i - \boldsymbol{a}_j\|^2) \|\boldsymbol{x}_i - \boldsymbol{x}_j\|, \tag{4.11}$$

where $\phi > 0$ is a tuning parameter. Note that for ϕ close to 0, this formulation recovers equal weights, whereas as $\phi \to \infty$, the weights in the second term tend to 0 and hence each $\boldsymbol{a}_i$ will end up in its own cluster. In the case of [30], the exponentially decaying weights are truncated to 0 for points sufficiently far apart in order to improve computational efficiency (by removing most of the terms from the second summation of (1.1)). However, since our study here does not concern efficiency, we did not truncate any terms. We chose $n = 1000$, $d = K = 6$, $\sigma = .0094$, λ as the unique value that satisfies (4.2) and (4.3) if ϕ were zero (equal weights), $\theta = 2$, The results in Table 4.3 show that for exponentially decaying weights, the correct clusters are recovered provided that ϕ is not too large, i.e., the weights do not fall to 0 too quickly.

Table 4.3: Recovery for varying ϕ.

ϕ	s_1 (% of V_m recovered)	s_2 (total % recovered)	s_3 (% distinct clusters)
500	304/304	999/1000	6/6
1000	304/304	901/1000	6/6
1500	92/304	144/1000	6/6
2000	14/304	14/1000	6/6

Chapter 5

Clustering test and guarantee [1]

5.1 Feasibility and complementary slackness

In this section, we consider a second-order cone (SOCP) formulation of (1.1). Both feasibility and complementary slackness are stated. A second-order cone program can be directly solved by a feasible interior-point method. For infeasible algorithms such as the ADMM proposed by Chi and Lange [7], we construct a feasible solution for the SOCP from the outputs of such algorithms.

5.1.1 Second-order cone formulation

We first present the equivalent SOCP formulation to (1.1), which will be derived in this section. The SOCP formulation can be written in both standard dual and standard primal forms. The standard dual form is as follows.

$$\min_{\boldsymbol{x},s,t} \quad f(\boldsymbol{x},s,t) = \sum_{i=1}^{n} s_i + \lambda \sum_{1\le i<j\le n} t_{ij} \tag{5.1a}$$

$$\text{s.t} \quad t_{ij} \ge \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \ , \quad \forall 1 \le i < j \le n \ , \tag{5.1b}$$

$$s_i \ge \left\| \begin{pmatrix} \boldsymbol{x}_i - \boldsymbol{a}_i \\ s_i - 1 \end{pmatrix} \right\| \ , \quad \forall i = 1, \dots, n \ . \tag{5.1c}$$

[1] This chapter is based on *On identifying clusters from sum-of-norms clustering computation* by Jiang and Vavasis [14].

The standard primal form is as follows.

$$\min_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},s,u,t} \quad f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},s,u,t) = \sum_{i=1}^{n} s_i + \lambda \sum_{1\le i<j\le n} t_{ij} \tag{5.2a}$$

$$\text{s.t} \quad \boldsymbol{x}_i - \boldsymbol{x}_j - \boldsymbol{y}_{ij} = \boldsymbol{0}\ , \quad \forall 1 \le i < j \le n\ , \tag{5.2b}$$

$$\boldsymbol{x}_i - \boldsymbol{z}_i = \boldsymbol{a}_i\ , \quad \forall i = 1,\dots,n\ , \tag{5.2c}$$

$$s_i - u_i = 1\ , \quad \forall i = 1,\dots,n\ , \tag{5.2d}$$

$$t_{ij} \ge \|\boldsymbol{y}_{ij}\|\ , \quad \forall 1 \le i < j \le n\ , \tag{5.2e}$$

$$s_i \ge \left\| \begin{pmatrix} \boldsymbol{z}_i \\ u_i \end{pmatrix} \right\|\ , \quad \forall i = 1,\dots,n\ . \tag{5.2f}$$

The SOCP formulation of the dual problem is as follows.

$$\max_{\boldsymbol{\delta},\boldsymbol{\beta},\gamma} \quad h(\boldsymbol{\delta},\boldsymbol{\beta},\gamma) = \sum_{i=1}^{n} \boldsymbol{a}_i^T \boldsymbol{\beta}_i + \sum_{i=1}^{n} \gamma_i \tag{5.3a}$$

$$\text{s.t} \quad -\sum_{j=1}^{i-1} \boldsymbol{\delta}_{ji} + \sum_{j=i+1}^{n} \boldsymbol{\delta}_{ij} + \boldsymbol{\beta}_i = \boldsymbol{0}\ , \quad \forall i = 1,\dots,n\ , \tag{5.3b}$$

$$\lambda \ge \|\boldsymbol{\delta}_{ij}\|\ , \quad \forall 1 \le i < j \le n\ , \tag{5.3c}$$

$$1 - \gamma_i \ge \left\| \begin{pmatrix} \boldsymbol{\beta}_i \\ \gamma_i \end{pmatrix} \right\|\ , \quad \forall i = 1,\dots,n\ . \tag{5.3d}$$

Although the dual form (5.1) is more compact than the primal form (5.2), we will be usng (5.2) for the rest of this thesis because we make explicit reference to the additional variables in (5.2). Both primal and dual problems are feasible, and Slater condition holds for both of them. Consider the following primal and dual feasible solution:

$$\boldsymbol{x}_i = \boldsymbol{a}_i, \boldsymbol{z}_i = \boldsymbol{0}, s_i = 1, u_i = 0,\ \forall i = 1,\dots,n; \quad \boldsymbol{y}_{ij} = \boldsymbol{a}_i - \boldsymbol{a}_j, t_{ij} = \|\boldsymbol{a}_i - \boldsymbol{a}_j\| + 1,\ \forall 1 \le i < j \le n;$$

$$\boldsymbol{\delta}_{ij} = \boldsymbol{0},\ \forall 1 \le i < j \le n; \quad \boldsymbol{\beta}_i = \boldsymbol{0}, \gamma_i = 0,\ \forall i = 1,\dots,n;$$

which is a also primal and dual Slater point. Hence, strong duality holds since the problem is formulated as convex optimization.

We derive the SOCP (5.2) as follows. Introduce auxiliary variables $\boldsymbol{y}_{ij}$ and $\boldsymbol{z}_i$ and constraints (5.2b) and (5.2c). Introduce variable t_i and constraint (5.2e). Introduce variables s_i and u_i satisfying (5.2d) and

$$s_i \ge \frac{\|\boldsymbol{x}_i - \boldsymbol{a}_i\|}{2} + \frac{1}{2}, \quad \forall i = 1,\dots,n.$$

Multiply the constraint in the previous line by 2, and add $s_i^2 - 2s_i$ to both sides. Simplify the inequality, and substitute u_i to obtain constraint (5.2f). The objective function has the following upper bound using auxiliary variables and new constraints:

$$\begin{aligned} f'(\boldsymbol{x}) &= \frac{1}{2}\sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{a}_i\|^2 + \lambda \sum_{1\le i<j\le n} \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \\ &= \frac{1}{2}\sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{a}_i\|^2 + \lambda \sum_{1\le i<j\le n} \|\boldsymbol{y}_{ij}\| \\ &\le \sum_{i=1}^{n} s_i - \frac{n}{2} + \lambda \sum_{1\le i<j\le n} t_{ij} \\ &= f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, s, u, t) - \frac{n}{2}. \end{aligned} \tag{5.4}$$

Notice that (1.1) is a minimization problem. For every feasible solution $\boldsymbol{x}$ to (1.1), we can construct a feasible solution $(\boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}', s', u', t')$ to (5.2) such that $\boldsymbol{x} = \boldsymbol{x}'$ and the upper bound (5.4) is achieved. Hence, we can replace the objective function $f'(\boldsymbol{x})$ with the linear function $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, s, u, t)$ as shown in (5.2a). The original problem (1.1) and the SOCP (5.2) are indeed equivalent. Since we omit the constant term $\frac{n}{2}$ in the objective function of the SOCP, the objective values of (1.1) at $\boldsymbol{x}$ and (5.2) at the corresponding solution $(\boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}', s', u', t')$ differ by a constant $\frac{n}{2}$.

With a standard procedure to derive SOCP dual, we obtain the dual formulation (5.3).

For the clustering test, we require primal-dual feasibility for the above primal and dual SOCP. Such a primal and dual feasible solution can be obtained by applying a feasible primal-dual interior-point method to the problem above. Each iterate of the algorithm is feasible, so is the output. Nevertheless, the output may not be feasible for our SOCP when a general primal-dual algorithm is used to solve for (1.1). Luckily, given that the output is close to the feasible set, we are able to find a small perturbation on the output to attain feasibility. The rest of the section elaborates on the perturbation and validates feasibility for the perturbed solution.

Now let us consider a general primal-dual algorithm which solves (1.1) and yields an output that is either in or close to the feasible set. The dual problem of (1.1) is as follows.

$$\max_{\boldsymbol{\delta}} \quad h'(\boldsymbol{\delta}) = -\frac{1}{2}\sum_{i=1}^{n} \left\| \sum_{j=1}^{i-1} \boldsymbol{\delta}_{ji} - \sum_{j=i+1}^{n} \boldsymbol{\delta}_{ij} \right\|^2 - \sum_{1\le i<j\le n} \langle \boldsymbol{\delta}_{ij}, \boldsymbol{a}_i - \boldsymbol{a}_j \rangle \tag{5.5a}$$

$$\text{s.t} \quad \|\boldsymbol{\delta}_{ij}\| \le \lambda\,, \quad \forall 1 \le i < j \le n\,. \tag{5.5b}$$

Notice the formulation is equivalent to the SOCP dual (5.3). We obtained (5.5) by eliminating $\boldsymbol{\beta}$ and γ in (5.3). The objective function $h'(\boldsymbol{\delta})$ has the following lower bound:

$$\begin{aligned} h'(\boldsymbol{\delta}) &= -\frac{1}{2}\sum_{i=1}^{n}\|\sum_{j=i+1}^{n}\boldsymbol{\delta}_{ij} - \sum_{j=1}^{i-1}\boldsymbol{\delta}_{ji}\|^2 - \sum_{1\le i<j\le n}\langle \boldsymbol{\delta}_{ij}, \boldsymbol{a}_i - \boldsymbol{a}_j\rangle \\ &\ge \sum_{i=1}^{n}\gamma_i - \frac{n}{2} + \sum_{i=1}^{n}\langle \boldsymbol{a}_i, \boldsymbol{\beta}_i\rangle \\ &= h(\boldsymbol{\delta}, \boldsymbol{\beta}, \gamma) - \frac{n}{2}. \end{aligned} \tag{5.6}$$

For every feasible solution $\boldsymbol{\delta}$ to (5.5), we can construct a feasible solution $(\boldsymbol{\delta}', \boldsymbol{\beta}', \gamma')$ to (5.3) such that $\boldsymbol{\delta} = \boldsymbol{\delta}'$ and the lower bound (5.6) is achieved. The objective values of (5.5) at $\boldsymbol{\delta}$ and (5.3) at the corresponding solution $(\boldsymbol{\delta}', \boldsymbol{\beta}', \gamma')$ differ by a constant $\frac{n}{2}$.

Let $(\boldsymbol{x}, \boldsymbol{\delta})$ denote the output yielded by the primal-dual algorithm. To construct a feasible solution from $(\boldsymbol{x}, \boldsymbol{\delta})$, we first update $\boldsymbol{\delta}$ as follows

$$\boldsymbol{\delta}_{ij} \leftarrow \begin{cases} \frac{\lambda \boldsymbol{\delta}_{ij}}{\|\boldsymbol{\delta}_{ij}\|}, & \text{if } \|\boldsymbol{\delta}_{ij}\| > \lambda, \\ \boldsymbol{\delta}_{ij}, & \text{otherwise.} \end{cases}$$

The updated $\boldsymbol{\delta}_{ij}$ has norm no more than λ. Notice that the perturbation is small provided that the dual solution was already close to the feasible set. Next, define the following variables:

$$\begin{aligned} \boldsymbol{y}_{ij} &= \boldsymbol{x}_i - \boldsymbol{x}_j, && \forall 1 \le i < j \le n, \\ \boldsymbol{z}_i &= \boldsymbol{x}_i - \boldsymbol{a}_i, && \forall i = 1, \ldots, n, \\ s_i &= \frac{1}{2}(1 + \|\boldsymbol{z}_i\|^2), && \forall i = 1, \ldots, n, \\ u_i &= \frac{1}{2}(-1 + \|\boldsymbol{z}_i\|^2), && \forall i = 1, \ldots, n, \\ t_{ij} &= \|\boldsymbol{y}_{ij}\|, && \forall 1 \le i < j \le n, \\ \boldsymbol{\beta}_i &= \sum_{j=1}^{i-1}\boldsymbol{\delta}_{ji} - \sum_{j=i+1}^{n}\boldsymbol{\delta}_{ij}, && \forall i = 1, \ldots, n, \\ \gamma_i &= \frac{1}{2}(1 - \|\boldsymbol{\beta}_i\|^2), && \forall i = 1, \ldots, n. \end{aligned} \tag{5.7}$$

It can be easily verified that these newly defined variables, $\boldsymbol{x}$ from the primal-dual algorithm and the updated $\boldsymbol{\delta}$ form a primal and dual feasible solution for the SOCP. Notice that the inequality (5.4) achieves equality at $\boldsymbol{x}$ and $\{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, s, u, t\}$, and the inequality (5.6) achieves

equality at the updated $\boldsymbol{\delta}$ and $\{\boldsymbol{\delta}, \boldsymbol{\beta}, \gamma\}$. Hence, the original objective function value at $(\boldsymbol{x}, \boldsymbol{\delta})$ and the SOCP objective function value at the updated solution differ by a constant $\frac{n}{2}$.

5.1.2 Complementary slackness

Let $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, s, u, t, \boldsymbol{\delta}, \boldsymbol{\beta}, \gamma)$ be a primal and dual feasible solution for the SOCP formulation of sum-of-norms clustering. Let us define $\boldsymbol{\epsilon}^{ij} = \begin{pmatrix} \epsilon_1^{ij} \\ \boldsymbol{\epsilon}_2^{ij} \end{pmatrix}$ for all $1 \leq i < j \leq n$ and $\boldsymbol{\sigma}^i = \begin{pmatrix} \sigma_1^i \\ \boldsymbol{\sigma}_2^i \\ \sigma_3^i \end{pmatrix}$ for all $i = 1, \ldots, n$ as follows:

$$t_{ij}\lambda + \boldsymbol{y}_{ij}^T \boldsymbol{\delta}_{ij} = \epsilon_1^{ij}, \quad \forall 1 \leq i < j \leq n, \tag{5.8}$$

$$t_{ij}\boldsymbol{\delta}_{ij} + \lambda \boldsymbol{y}_{ij} = \boldsymbol{\epsilon}_2^{ij}, \quad \forall 1 \leq i < j \leq n, \tag{5.9}$$

$$s_i(1 - \gamma_i) + \boldsymbol{z}_i^T \boldsymbol{\beta}_i + u_i \gamma_i = \sigma_1^i, \quad \forall i = 1, \ldots, n, \tag{5.10}$$

$$s_i \boldsymbol{\beta}_i + (1 - \gamma_i)\boldsymbol{z}_i = \boldsymbol{\sigma}_2^i, \quad \forall i = 1, \ldots, n, \tag{5.11}$$

$$s_i \gamma_i + (1 - \gamma_i) u_i = \sigma_3^i, \quad \forall i = 1, \ldots, n. \tag{5.12}$$

At the optimizer, there hold $\boldsymbol{\epsilon} = \mathbf{0}, \boldsymbol{\sigma} = \mathbf{0}$ by KKT conditions. The system of equalities above becomes the complementary slackness condition. At an approximate solution, the right-hand sides $\boldsymbol{\epsilon}, \boldsymbol{\sigma}$ are non-zero. If $\boldsymbol{\epsilon}^{ij} = \begin{pmatrix} \mu' \\ \mathbf{0} \end{pmatrix}, \boldsymbol{\sigma}^i = \begin{pmatrix} \mu' \\ \mathbf{0} \\ 0 \end{pmatrix}$ for all $i = 1, \ldots, n$ and for all $1 \leq i < j \leq n$, we refer the corresponding solution as a μ'-centered solution, and μ' is the central-path parameter. Otherwise, an upper bound on general right-hand sides $\boldsymbol{\epsilon}, \boldsymbol{\sigma}$

can be derived from the duality gap:

$$
\begin{aligned}
& f'(\boldsymbol{x}) - h'(\boldsymbol{\delta}) \\
=& f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, s, u, t) - h(\boldsymbol{\delta}, \boldsymbol{\beta}, \gamma) \quad \text{(By (5.4) and (5.6))} \\
=& \sum_{i=1}^{n} s_i - \sum_{i=1}^{n} \gamma_i + \sum_{i=1}^{n} \langle \boldsymbol{x}_i - \boldsymbol{a}_i, \boldsymbol{\beta}_i \rangle + \lambda \sum_{1 \le i < j \le n} t_{ij} - \sum_{i=1}^{n} \langle \boldsymbol{x}_i, \boldsymbol{\beta}_i \rangle \\
& \text{(By adding and subtracting } \sum_{i=1}^{n} \langle \boldsymbol{x}_i, \boldsymbol{\beta}_i \rangle) \\
=& \sum_{i=1}^{n} s_i - \sum_{i=1}^{n} \gamma_i + \sum_{i=1}^{n} \langle \boldsymbol{x}_i - \boldsymbol{a}_i, \boldsymbol{\beta}_i \rangle + \lambda \sum_{1 \le i < j \le n} t_{ij} - \sum_{i=1}^{n} \langle \boldsymbol{x}_i, \sum_{j=1}^{i-1} \boldsymbol{\delta}_{ji} - \sum_{j=i+1}^{n} \boldsymbol{\delta}_{ij} \rangle \quad \text{(By (5.3b))} \\
=& \sum_{i=1}^{n} s_i - \sum_{i=1}^{n} \gamma_i + \sum_{i=1}^{n} \langle \boldsymbol{x}_i - \boldsymbol{a}_i, \boldsymbol{\beta}_i \rangle + \lambda \sum_{1 \le i < j \le n} t_{ij} - \sum_{1 \le i < j \le n} \langle \boldsymbol{x}_j - \boldsymbol{x}_i, \boldsymbol{\delta}_{ij} \rangle \\
& \text{(By expanding the summation)} \\
=& \sum_{i=1}^{n} (s_i - \gamma_i + \langle \boldsymbol{x}_i - \boldsymbol{a}_i, \boldsymbol{\beta}_i \rangle) + \sum_{1 \le i < j \le n} (\lambda t_{ij} + \langle \boldsymbol{y}_{ij}, \boldsymbol{\delta}_{ij} \rangle) \quad \text{(By (5.2b))} \\
=& \sum_{i=1}^{n} (s_i (1 - \gamma_i) + \langle \boldsymbol{z}_i, \boldsymbol{\beta}_i \rangle + u_i \gamma_i) + \sum_{1 \le i < j \le n} (\lambda t_{ij} + \langle \boldsymbol{y}_{ij}, \boldsymbol{\delta}_{ij} \rangle) \quad \text{(By (5.2c), (5.2d))} \\
=& \sum_{i=1}^{n} \sigma_1^i + \sum_{1 \le i < j \le n} \epsilon_1^{ij}
\end{aligned}
$$

Each term in the both summations is non-negative as shown below:

$$\sigma_1^i = s_i(1 - \gamma_i) + \langle \boldsymbol{z}_i, \boldsymbol{\beta}_i \rangle + u_i \gamma_i = \frac{1}{2}(\|\boldsymbol{z}_i\|^2 + 2\langle \boldsymbol{z}_i, \boldsymbol{\beta}_i \rangle + \|\boldsymbol{\beta}_i\|^2) \ge 0, \quad \forall i = 1, \ldots, n,$$

$$\epsilon_1^{ij} = \lambda t_{ij} + \langle \boldsymbol{y}_{ij}, \boldsymbol{\delta}_{ij} \rangle \ge \lambda t_{ij} - \|\boldsymbol{y}_{ij}\| \, \|\boldsymbol{\delta}_{ij}\| \ge \lambda t_{ij} - \lambda \|\boldsymbol{y}_{ij}\| = 0, \quad \forall 1 \le i < j \le n.$$

Define $\mu := f'(\boldsymbol{x}) - h'(\boldsymbol{\delta})$ to be the duality gap at the feasible solution. Notice that our choice of μ is the not usual central-path parameter μ used in the primal-dual interior-point method. As mentioned earlier, we adopt the notation μ' to denote the the central-path parameter, and the relationship between μ and μ' is that $\mu = (n + 0.5(n+1)n)\mu'$. Since n is a known constant, $O(\mu)$ and $O(\mu')$ can be used interchangeably.

Combined with the non-negativity condition, $\sigma_1^i, \epsilon_1^{ij}$ satisfy $\sigma_1^i \le \mu$ for all $i = 1, \ldots, n$ and $\epsilon_1^{ij} \le \mu$ for all $1 \le i < j \le n$. At termination, the duality gap μ at the feasible solution is small, which implies the right-hand sides $\sigma_1^i, \epsilon_1^{ij}$ are also well bounded.

We now have $\sigma_1^i, \epsilon_1^{ij}$ upper bounded in terms of μ, and the remainder of the section is to establish upper bounds on $\left\|\boldsymbol{\epsilon}_2^{ij}\right\|$ and $\left\|\begin{pmatrix}\boldsymbol{\sigma}_2^i \\ \sigma_3^i\end{pmatrix}\right\|$. In fact, in (5.13) and (5.14) below, we show that both are upper bounded by $O(\sqrt{\mu})$. Consider a general setting of second-order cone programming.

Lemma 3. *Let $\boldsymbol{p} \in K_1 \bigotimes \cdots \bigotimes K_n, \boldsymbol{q} \in K_1^* \bigotimes \cdots \bigotimes K_n^*$ denote a primal and dual feasible solution for a second-order cone program where $K_1, \ldots, K_n$ are second-order cones and $K_1^*, \ldots, K_n^*$ are the corresponding dual cones. Let $\boldsymbol{x}, \boldsymbol{z}$ be subvectors of $\boldsymbol{p}, \boldsymbol{q}$ respectively such that $\boldsymbol{x} \in K_i$ and $\boldsymbol{z} \in K_i^*$. Let $\boldsymbol{x} = \begin{pmatrix} x_0 \\ \bar{\boldsymbol{x}} \end{pmatrix}, \boldsymbol{z} = \begin{pmatrix} z_0 \\ \bar{\boldsymbol{z}} \end{pmatrix}$. If $\boldsymbol{x}^T \boldsymbol{z} \leq \mu$, then $\|z_0 \bar{\boldsymbol{x}} + x_0 \bar{\boldsymbol{z}}\| \leq \sqrt{2 x_0 z_0 \mu}$.*

Proof. If $x_0 = 0$, then $\|\bar{\boldsymbol{x}}\| \leq x_0 = 0$ by feasibility assumption. Hence, $\bar{\boldsymbol{x}} = \boldsymbol{0}$, which implies $\|z_0 \bar{\boldsymbol{x}} + x_0 \bar{\boldsymbol{z}}\| = 0 \leq \sqrt{2 x_0 z_0 \mu}$. Similarly, if $z_0 = 0$, then $\boldsymbol{z}$ satisfies $\|z_0 \bar{\boldsymbol{x}} + x_0 \bar{\boldsymbol{z}}\| = 0 \leq \sqrt{2 x_0 z_0 \mu}$ by the same argument.

Otherwise, $x_0 > 0, z_0 > 0$, and we derive the following inequalities

$$\begin{aligned}
\boldsymbol{x}^T \boldsymbol{z} = x_0 z_0 + \bar{\boldsymbol{x}}^T \bar{\boldsymbol{z}} &\leq \mu \\
\Rightarrow 1 + \frac{\bar{\boldsymbol{x}}^T}{x_0} \frac{\bar{\boldsymbol{z}}}{z_0} &\leq \frac{\mu}{x_0 z_0} \quad (\text{Since } x_0 > 0, z_0 > 0) \\
\Rightarrow \left\| \frac{\bar{\boldsymbol{x}}}{x_0} + \frac{\bar{\boldsymbol{z}}}{z_0} \right\|^2 = \left\| \frac{\bar{\boldsymbol{x}}}{x_0} \right\|^2 + \left\| \frac{\bar{\boldsymbol{z}}}{z_0} \right\|^2 + 2 \frac{\bar{\boldsymbol{x}}^T}{x_0} \frac{\bar{\boldsymbol{z}}}{z_0} &\leq 2 - 2 + \frac{2\mu}{x_0 z_0} \quad (\text{Since } x_0 \geq \|\bar{\boldsymbol{x}}\|, z_0 \geq \|\bar{\boldsymbol{z}}\|) \\
\Rightarrow \left\| \frac{\bar{\boldsymbol{x}}}{x_0} + \frac{\bar{\boldsymbol{z}}}{z_0} \right\| &\leq \sqrt{\frac{2\mu}{x_0 z_0}} \\
\Rightarrow \|z_0 \bar{\boldsymbol{x}} + x_0 \bar{\boldsymbol{z}}\| &\leq \sqrt{2 x_0 z_0 \mu}.
\end{aligned}$$

□

Let $\bar{\boldsymbol{a}} := \frac{1}{n} \sum_{i=1}^n \boldsymbol{a}_i$ denote the centroid of all data points. Let $\boldsymbol{x}_1' := \boldsymbol{x}_2' := \ldots := \boldsymbol{x}_n' := \bar{\boldsymbol{a}}$. Then the primal objective value of the original sum-of-norms formulation at $\boldsymbol{x}'$ is

$$f'(\boldsymbol{x}') = \frac{1}{2} \sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 .$$

Let $\boldsymbol{\delta}_{ij}' = \boldsymbol{0}$ for all $1 \leq i < j \leq n$. Then $\boldsymbol{\delta}'$ is a feasible solution to the dual problem of the original formulation and the dual objective value at $\boldsymbol{\delta}'$ is

$$h'(\boldsymbol{\delta}') = 0.$$

Let f^* and h^* denote the primal and dual optimal values of the SOCP respectively, which must satisfy the following inequality by strong duality:

$$\frac{n}{2} = h'(\boldsymbol{\delta}') + \frac{n}{2} \leq h^* = f^* \leq f'(\boldsymbol{x}') + \frac{n}{2} = \frac{1}{2}\sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + \frac{n}{2}.$$

At the feasible solution $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, s, u, t, \boldsymbol{\delta}, \boldsymbol{\beta}, \gamma)$, the objective value is at a distance of at most μ away from the optimal value, which implies

$$\sum_{i=1}^{n} s_i + \lambda \sum_{1\leq i<j\leq n} t_{ij} \leq f^* + \mu \leq \frac{1}{2}\sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + \frac{n}{2} + \mu,$$

which is rearranged to

$$\sum_{i=1}^{n} \left(s_i - \frac{1}{2}\right) + \lambda \sum_{1\leq i<j\leq n} t_{ij} \leq \frac{1}{2}\sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + \mu.$$

Moreover, by feasibility, $s_i \geq \frac{1}{2}$ holds for all $i = 1, \ldots, n$ and $t_{ij} \geq 0$ holds for all $1 \leq i < j \leq n$. Hence,

$$s_i \leq \frac{1}{2}\sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + \frac{1}{2} + \mu, \quad t_{ij} \leq \frac{1}{\lambda}\left(\frac{1}{2}\sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + \mu\right).$$

As $t_{ij}\lambda + \boldsymbol{y}_{ij}^T\boldsymbol{\delta}_{ij} = \epsilon_1^{ij} \leq \mu$, $\left\|\boldsymbol{\epsilon}_2^{ij}\right\|$ has the following upper bound by Lemma 3

$$\left\|\boldsymbol{\epsilon}_2^{ij}\right\| = \|t_{ij}\boldsymbol{\delta}_{ij} + \lambda \boldsymbol{y}_{ij}\| \leq \sqrt{2t_{ij}\lambda\mu} \leq \sqrt{\sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 \mu + 2\mu^2}. \tag{5.13}$$

Similarly, at the feasible solution, the dual objective value is at a distance of at most μ away from the optimal value, which implies

$$\sum_{i=1}^{n} \boldsymbol{a}_i^T\boldsymbol{\beta}_i + \sum_{i=1}^{n} \gamma_i \geq h^* - \mu \geq \frac{n}{2} - \mu,$$

which is rearranged to

$$\sum_{i=1}^{n} \left(\frac{1}{2} - \gamma_i\right) \leq \sum_{i=1}^{n} \boldsymbol{a}_i^T\boldsymbol{\beta}_i + \mu.$$

By feasibility, $\frac{1}{2} - \gamma_i \geq 0$, which implies

$$1 - \gamma_i \leq \frac{1}{2} + \sum_{i=1}^{n} \boldsymbol{a}_i^T \boldsymbol{\beta}_i + \mu.$$

Since $\lambda \geq \|\boldsymbol{\delta}_{ij}\|$, $\|\boldsymbol{\beta}_i\|$ satisfies

$$\|\boldsymbol{\beta}_i\| = \left\| \sum_{j=1}^{i-1} \boldsymbol{\delta}_{ji} - \sum_{j=i+1}^{n} \boldsymbol{\delta}_{ij} \right\| \leq (n-1)\lambda.$$

By Cauchy-Schwartz inequality,

$$\boldsymbol{a}_i^T \boldsymbol{\beta}_i \leq \|\boldsymbol{a}_i\| \cdot \|\boldsymbol{\beta}_i\| \leq (n-1)\lambda \|\boldsymbol{a}_i\|.$$

Therefore, $1 - \gamma_i$ satisfies

$$1 - \gamma_i \leq \frac{1}{2} + \sum_{l=1}^{n} (n-1)\lambda \|\boldsymbol{a}_l\| + \mu.$$

Since $s_i(1-\gamma_i) + \boldsymbol{z}_i^T \boldsymbol{\beta}_i + u_i \gamma_i = \sigma_1^i$, $\left\| \begin{pmatrix} \boldsymbol{\sigma}_2^i \\ \sigma_3^i \end{pmatrix} \right\|$ has the following upper bound by Lemma 3

$$\begin{aligned} \left\| \begin{pmatrix} \boldsymbol{\sigma}_2^i \\ \sigma_3^i \end{pmatrix} \right\| &= \left\| \begin{pmatrix} s_i \boldsymbol{\beta}_i + (1-\gamma_i)\boldsymbol{z}_i \\ s_i \gamma_i + (1-\gamma_i) u_i \end{pmatrix} \right\| \leq \sqrt{2 s_i (1-\gamma_i) \mu} \\ &\leq \sqrt{2 \cdot \left(\frac{1}{2} \sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + \frac{1}{2} + \mu \right) \cdot \left(\frac{1}{2} + \sum_{l=1}^{n} (n-1)\lambda \|\boldsymbol{a}_l\| + \mu \right) \cdot \mu} \qquad (5.14) \\ &\leq \sqrt{\left(\sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + 1 + 2\mu \right) \cdot \left(\frac{1}{2} + \sum_{l=1}^{n} (n-1)\lambda \|\boldsymbol{a}_l\| + \mu \right) \mu}. \end{aligned}$$

5.2 Clustering test

Given a primal and dual feasible solution $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, s, u, t, \boldsymbol{\delta}, \boldsymbol{\beta}, \gamma)$ with a duality gap μ, we find candidate clusters as follows. First, select an index i from $\{1, \ldots, n\}$ arbitrarily. Construct a ball of radius $\mu^{0.75}$ about $\boldsymbol{x}_i$. Create a candidate cluster with all indices k such that $\boldsymbol{x}_k$ is located in the ball about $\boldsymbol{x}_i$ (i.e. $\{k : \|\boldsymbol{x}_i - \boldsymbol{x}_k\| \leq \mu^{3/4}\}$). Now find an index j

that is not in any candidate cluster and construct a ball about $\boldsymbol{x}_j$. Repeat until all data points are used up.

If the output of the primal-dual algorithm is not feasible for the SOCP, we construct a feasible solution as described in the previous section. With the feasible solution, we define

$$\boldsymbol{g}_{ij} := \begin{cases} -\boldsymbol{\delta}_{ij}, & \text{if } i < j, \\ \boldsymbol{\delta}_{ji}, & \text{if } j < i. \end{cases}$$

For any candidate cluster C, compute $\boldsymbol{q}_{ij} := \boldsymbol{g}_{ij} + \frac{1}{m} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j - \boldsymbol{\omega}_i + \boldsymbol{\omega}_j) + \frac{1}{m} \sum_{k \notin C} (\boldsymbol{g}_{ik} - \boldsymbol{g}_{jk})$ for all $i, j \in C, i \neq j$, denoted as *Chiquet-Gutierrez-Rigail (CGR) subgradients*. Check if the following two conditions hold:

***CGR subgradient condition*:** All CGR subgradients $\boldsymbol{q}_{ij}$ satisfy the CGR inequality $\|\boldsymbol{q}_{ij}\| \leq \lambda$.

***Separation condition*:** All candidate clusters are separated at distance of at least 2τ, where $\tau = \sqrt{2\mu}$.

If both conditions hold for all candidate clusters, then the test terminates and reports 'success'. Each candidate cluster is a real cluster given by the optimal solution, thus all clusters are correctly identified. The $\boldsymbol{q}_{ij}$'s serve as certificates. If either condition fails for any candidate cluster, the test reports 'failure'. One has to run more iterations of the algorithm to decrease the duality gap μ. Repeat the process until the test reports 'success'. Note that this test is algorithm-independent, but it does require the algorithm to be of primal-dual type.

The CGR subgradients condition certifies that each cluster we identify is indeed a cluster or part of a larger cluster by Theorem 1 in Section 3.1. This is presented in Section 5.2.1. The separation condition certifies that there is no super-cluster with more than one cluster we identify by the following theorem:

Theorem 4. *Define $\tau > 0$ such that the true optimizer and the approximate solution are at distance of at most τ away (i.e. $\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \tau$). If there exist $i, j \in C$ such that $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| > 2\tau$, then C is not a cluster or part of a larger cluster.*

Therefore, we determine all clusters correctly when the test succeeds.

5.2.1 CGR subgradients and clustering corollary

Let $C \subseteq [n]$ denote a subset of points. Let $m := |C|$ denote the cardinality of C.

Lemma 5. *For all $i, j \in C, i \neq j$, define $\boldsymbol{q}_{ij} := \boldsymbol{g}_{ij} + \frac{1}{m} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j - \boldsymbol{\omega}_i + \boldsymbol{\omega}_j) + \frac{1}{m} \sum_{k \notin C} (\boldsymbol{g}_{ik} - \boldsymbol{g}_{jk})$. Then $\boldsymbol{q}_{ij}$ satisfies*

$$\boldsymbol{a}_i - \bar{\boldsymbol{a}} = \sum_{j \in C \setminus \{i\}} \boldsymbol{q}_{ij}, \quad \forall i \in C \tag{5.15}$$

$$\boldsymbol{q}_{ij} = -\boldsymbol{q}_{ji}, \quad \forall i, j \in C, i \neq j, \tag{5.16}$$

where $\bar{\boldsymbol{a}} = \frac{1}{m} \sum_{i \in C} \boldsymbol{a}_i$.

Proof. Substitute the primal constraint (5.2d) into the perturbed complementary slackness (5.12) to obtain the following equality of γ_i and s_i

$$1 - \gamma_i = s_i - \sigma_3^i, \quad \forall i = 1, \ldots, n.$$

Substitute the equality above into (5.11) and divide both sides by s_i to obtain the following equation of $\boldsymbol{\beta}_i$ in terms of $\boldsymbol{z}_i$

$$\boldsymbol{\beta}_i = -\boldsymbol{z}_i + \boldsymbol{\omega}_i, \quad \forall i = 1, \ldots, n.$$

Notice that the operation is valid because $s_i \geq \frac{1}{2}$ by the primal constraint (5.2d) and (5.2f). Substitute the primal constraint (5.2c) and the equality above into the dual constraint (5.3b) yielding

$$-\sum_{j=1}^{i-1} \boldsymbol{\delta}_{ji} + \sum_{j=i+1}^{n} \boldsymbol{\delta}_{ij} - \boldsymbol{x}_i + \boldsymbol{a}_i + \boldsymbol{\omega}_i = \boldsymbol{0}, \quad \forall i = 1, \ldots, n.$$

With the definition of $\boldsymbol{g}_{ij}$, the equality above is rewritten as

$$-\boldsymbol{x}_i + \boldsymbol{a}_i + \boldsymbol{\omega}_i - \sum_{j \neq i} \boldsymbol{g}_{ij} = \boldsymbol{0}, \quad \forall i = 1, \ldots, n. \tag{5.17}$$

By (5.17), we have the following equality holds for all $i \in C$

$$-\boldsymbol{x}_i + \boldsymbol{a}_i + \boldsymbol{\omega}_i - \sum_{j \in C \setminus \{i\}} \boldsymbol{g}_{ij} - \sum_{k \notin C} \boldsymbol{g}_{ik} = \boldsymbol{0}. \tag{5.18}$$

Sum (5.18) over all $i \in C$ and divide the new equality by m to obtain

$$-\frac{1}{m} \sum_{i \in C} \boldsymbol{x}_i + \bar{\boldsymbol{a}} + \frac{1}{m} \sum_{i \in C} \boldsymbol{\omega}_i - \frac{1}{m} \sum_{i \in C} \sum_{k \notin C} \boldsymbol{g}_{ik} = \boldsymbol{0}. \tag{5.19}$$

Change the index in (5.19) from i to j. Subtract (5.19) from (5.18) to obtain

$$-\boldsymbol{x}_i+\frac{1}{m}\sum_{j\in C}\boldsymbol{x}_j+\boldsymbol{a}_i-\bar{\boldsymbol{a}}+\boldsymbol{\omega}_i-\frac{1}{m}\sum_{j\in C}\boldsymbol{\omega}_j-\sum_{j\in C\setminus\{i\}}\boldsymbol{g}_{ij}+\frac{1}{m}\sum_{j\in C}\sum_{k\notin C}(\boldsymbol{g}_{jk}-\boldsymbol{g}_{ik})=\boldsymbol{0},\quad \forall i\in C,$$

which is rearranged to

$$\begin{aligned}\boldsymbol{a}_i-\bar{\boldsymbol{a}}&=\boldsymbol{x}_i-\frac{1}{m}\sum_{j\in C}\boldsymbol{x}_j-\boldsymbol{\omega}_i+\frac{1}{m}\sum_{j\in C}\boldsymbol{\omega}_j+\sum_{j\in C\setminus\{i\}}\boldsymbol{g}_{ij}+\frac{1}{m}\sum_{j\in C}\sum_{k\notin C}(\boldsymbol{g}_{ik}-\boldsymbol{g}_{jk})\\&=\sum_{j\in C\setminus\{i\}}\left[\frac{1}{m}(\boldsymbol{x}_i-\boldsymbol{x}_j-\boldsymbol{\omega}_i+\boldsymbol{\omega}_j)+\boldsymbol{g}_{ij}+\frac{1}{m}\sum_{k\notin C}(\boldsymbol{g}_{ik}-\boldsymbol{g}_{jk})\right]\\&=\sum_{j\in C\setminus\{i\}}\boldsymbol{q}_{ij}\quad\text{(By definition)},\quad\forall i\in C.\end{aligned}$$

Moreover, by the definition of $\boldsymbol{q}_{ij}$, we observe the following property for all $i,j\in C, i\neq j$

$$\begin{aligned}\boldsymbol{q}_{ij}&=\boldsymbol{g}_{ij}+\frac{1}{m}\cdot(\boldsymbol{x}_i-\boldsymbol{x}_j-\boldsymbol{\omega}_i+\boldsymbol{\omega}_j)+\frac{1}{m}\sum_{k\notin C}(\boldsymbol{g}_{ik}-\boldsymbol{g}_{jk})\\&=-\boldsymbol{g}_{ji}-\frac{1}{m}\cdot(\boldsymbol{x}_j-\boldsymbol{x}_i-\boldsymbol{\omega}_j+\boldsymbol{\omega}_i)-\frac{1}{m}\sum_{k\notin C}(\boldsymbol{g}_{jk}-\boldsymbol{g}_{ik})\\&=-\boldsymbol{q}_{ji}\end{aligned}$$

□

Corollary 5.1. *If $\|\boldsymbol{q}_{ij}\|\leq\lambda$ holds for all $i\neq j, i,j\in C$ where C is a candidate cluster, then C is a cluster or part of a larger cluster.*

The proof of the corollary follows trivially by Theorem 1 and Corollary 1.1.

5.2.2 Duality gap and distinct clustering corollary

As derived earlier in Section 3.2, the duality gap at the feasible solution is:

$$f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},s,u,t)-h(\boldsymbol{\delta},\boldsymbol{\beta},\gamma)=\sum_{i=1}^{n}s_i+\lambda\sum_{1\leq i<j\leq n}t_{ij}-\sum_{i=1}^{n}\boldsymbol{a}_i^T\boldsymbol{\beta}_i-\sum_{i=1}^{n}\gamma_i=\sum_{i=1}^{n}\sigma_1^i+\sum_{1\leq i<j\leq n}\epsilon_1^{ij}=:\mu. \tag{5.20}$$

By the property of strong convexity of f', we have

$$\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}^*\|^2 \leq f'(\boldsymbol{x}) - f'(\boldsymbol{x}^*) = f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},s,u,t) - f(\boldsymbol{x}^*,\boldsymbol{y}^*,\boldsymbol{z}^*,s^*,u^*,t^*),$$

which is further bounded as follows by weak duality

$$\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}^*\|^2 \leq f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},s,u,t) - h(\boldsymbol{\delta},\boldsymbol{\beta},\gamma). \tag{5.21}$$

Then, for any primal-dual algorithm, the distance between the approximate solution and the optimizer is given by

$$\tau = \|\boldsymbol{x}-\boldsymbol{x}^*\| \leq \sqrt{2\mu} \tag{5.22}$$

Corollary 5.2. *Let C denote a candidate cluster. If for all $i \in C, j \notin C$ it holds that $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| > 2\tau$ where $\tau = \sqrt{2\mu}$ for any primal-dual algorithm, then there does not exist a super-cluster which strictly contains C.*

The proof follows directly from Theorem 4.

5.3 Properties of the central path

In this section, we explore the properties of the central path for a primal-dual path following algorithm. These properties play a fundamental role in the proof of our main theorem in Section 6. In the main theorem, we state that if a primal-dual path following algorithm is used, our clustering test will eventually succeed after a finite number of iterations when λ is not at any fusion value. The proof of the ultimate success relies on the linear convergence to the optimal primal-dual pair, which will be shown to be satisfied in the remainder of this section.

Unfortunately, there are very few theorems about the central path of second-order-cone programming in literature. The only applicable result that we are aware of is due to Nesterov and Tuncel [20]. Their work showed that with a primal-dual interior point method, the primal μ'-centered iterates converge to the primal analytic center superlinearly under two assumptions. One of the assumptions is that there is a unique dual optimizer. This assumption does not hold for SON clustering in general. There are often infinitely many dual optimal solutions.

In spite of the lack of convergence analysis for SOCP, there are established theorems from semidefinite programming (SDP). SDP specializes to SOCP. With some standard

techniques, we can easily rewrite our SOCP problem as SDP and apply the primal-dual path following algorithm to solve the new SDP. The following theorem states that the μ'-centered iterates converge to the analytic center superlinearly.

Theorem 6 (Luo et al. [18]). *Assume the semidefinite program has a strictly complementary solution and the iterates of the algorithm converge tangentially to the central path. Let $(X(\mu'), Z(\mu'))$ denote a μ'-centered primal-dual pair. Let (X^a, Z^a) denote the analytic centers of the primal and dual optimal sets. Let $\mu' \in (0,1)$ be the central path parameter. There holds*

$$\|X(\mu') - X^a\| = O(\mu'), \quad \|Z(\mu') - Z^a\| = O(\mu').$$

Tangential convergence to the central path is a terminology adopted by Luo et al. [18]. It was first defined in [16] by Kojima et al. as follows. Let (X^r, Z^r) denote the solution generated by some algorithm at the r^{th} iteration. Let (X^a, Z^a) denote an optimal solution. Then the sequence $\{(X_r, Z_r)\}$ converges to the optimizer (X^a, Z^a) tangentially to the central path if

$$\lim_{r\to\infty} \left\| \sqrt{X_r} Z_r \sqrt{X_r} - \mu'_r I \right\|_F / \mu'_r = 0,$$

where $\mu'_r = tr(X_r Z_r)$. Assume a primal-dual path following algorithm satisfying the assumptions of Luo et al., and it is applied to solve the SOCP as SDP. To employ Theorem 6, we show that our SOCP has a strictly complementary optimizer. Notice that this statement is not necessarily true for all values of λ. One has to assume λ is not at any fusion value. When λ is exactly at a fusion value λ^*, strict complementarity may fail. The failure is not surprising since any arbitrarily small negative perturbation $\lambda^* + \epsilon$ yields a different clustering. In other words, complete cluster identification for these fusion values is ill-posed. Thus it is unreasonable to expect an algorithm that satisfies a guarantee for such a problem. There are at most n fusion values as a result of Theorem 1.1.

It is worth remarking that Theorem 6 does not directly apply to a primal-dual SOCP interior-point method. SOCP is a special case of SDP, yet the central path of SOCP is not just a simple projection of the SDP central path. The reason is that the log-barrier function for SOCP is not a specialization of the log-barrier function for SDP. Let $\boldsymbol{x}$ denote a primal feasible solution for an SOCP where $\boldsymbol{x} = \begin{pmatrix} x_0 \\ \bar{\boldsymbol{x}} \end{pmatrix} \in \mathbb{R}^{d+1}$ satisfies $x_0 \geq \|\bar{\boldsymbol{x}}\|$. Then the log-barrier function inherited from SDP reformulation would be

$$\phi_{SDP}(\boldsymbol{x}) = -\ln(x_0^2 - \|\bar{\boldsymbol{x}}\|^2) - (d-1)\ln x_0,$$

while the log-barrier function inherited from the original SOCP would be

$$\phi_{SOCP}(\boldsymbol{x}) = -\ln(x_0^2 - \|\bar{\boldsymbol{x}}\|^2).$$

The removal of the second term actually accelerates the convergence.

We suspect that an SOCP interior-point method should also satisfy a bound analogous to Theorem 6, but we are not aware of a proof in the literature.

5.3.1 Strict complementarity

By specializing the definition of strict complementarity in SDP to SOCP [1], a primal and dual optimal solution satisfies strict complementarity if and only if

$$t_{ij} + \lambda > \|\boldsymbol{y}_{ij} + \boldsymbol{\delta}_{ij}\|, \quad \forall 1 \le i < j \le n, \tag{5.23}$$

$$s_i + 1 - \gamma_i > \left\| \begin{pmatrix} \boldsymbol{z}_i + \boldsymbol{\beta}_i \\ u_i + \gamma_i \end{pmatrix} \right\|, \quad \forall i = 1, ..., n \tag{5.24}$$

The following theorem is a sufficient condition for strict complementarity of (5.2) and (5.3).

Theorem 7. *If $\lambda > 0$ is a parameter value at which fusion does not occur, then there exists a strictly complementary primal-dual optimal solution to SOCP (5.2) and (5.3) at λ.*

To prove Theorem 7, we consider a new optimization problem and construct such a strictly complementary primal-dual optimal solution from the new problem. Let λ_1, λ_2 be the two successive fusion values such that $\lambda \in (\lambda_1, \lambda_2)$. Note that it is possible for $\lambda_1 = 0$ or $\lambda_2 = \infty$. Let $(\boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}', s', u', t', \boldsymbol{\delta}', \boldsymbol{\beta}', \gamma')$ denote a primal and dual optimal solution at λ_1. Let $C_1, C_2, ..., C_K$ denote the clusters identified by the optimal solution above. When $\lambda_1 = 0$, there are n clusters, and each cluster is a singleton set. When λ_1 is the largest fusion value, there is only one cluster containing all n points. Define $\bar{\boldsymbol{a}}_k := \frac{1}{|C_k|} \sum_{i \in C_k} \boldsymbol{a}_i$. Consider the following optimization problem:

$$\min_{\boldsymbol{p}_1, ..., \boldsymbol{p}_K \in \mathbf{R}^d} \frac{1}{2} \sum_{k=1}^{K} |C_k| \, \|\boldsymbol{p}_k - \bar{\boldsymbol{a}}_k\|^2 + \lambda \sum_{1 \le k < k' \le K} |C_k| \cdot |C_{k'}| \, \|\boldsymbol{p}_k - \boldsymbol{p}_{k'}\| \,. \tag{5.25}$$

Let $\boldsymbol{p}$ denote the optimal solution of (5.25).

Lemma 8. *Vector $\boldsymbol{p}$ satisfies $\boldsymbol{p}_k \neq \boldsymbol{p}_{k'}$ for all $k, k' \in [K], k \neq k'$.*

Proof. For the purpose of contradiction, we may assume there exist $\hat{k} \neq \hat{k}'$ such that $\boldsymbol{p}_{\hat{k}} = \boldsymbol{p}_{\hat{k}'}$. Let $\boldsymbol{x}_i^* = \boldsymbol{p}_k$ for $i \in C_k, k \in [K]$. By the first-order optimality condition of (5.25) at $\boldsymbol{p}$, there exist $\boldsymbol{g}_{kk'} \in \partial\|\boldsymbol{p}_k - \boldsymbol{p}_{k'}\|$ for all $k \neq k'$ such that

$$\mathbf{0} = \boldsymbol{p}_k - \bar{\boldsymbol{a}}_k + \lambda \sum_{k' \neq k} |C_{K'}| \cdot \boldsymbol{g}_{kk'} = \boldsymbol{p}_k - \boldsymbol{a}_i + \boldsymbol{a}_i - \bar{\boldsymbol{a}}_k + \lambda \sum_{k' \neq k} |C_{K'}| \cdot \boldsymbol{g}_{kk'}. \tag{5.26}$$

Define $\boldsymbol{g}'$ with $\boldsymbol{\delta}'$ as before

$$\boldsymbol{g}'_{ij} := \begin{cases} -\boldsymbol{\delta}'_{ij}, & \text{if } i < j, \\ \boldsymbol{\delta}'_{ji}, & \text{if } j < i. \end{cases}$$

By the feasibility and complementary slackness in Section 5.1, the dual solution satisfies

$$\boldsymbol{a}_i - \bar{\boldsymbol{a}}_k = \sum_{j \in C_k - \{i\}} \boldsymbol{g}'_{ij}, \quad \forall i \in C_k, k \in [K], \qquad \text{and} \quad \left\|\boldsymbol{g}'_{ij}\right\| = \left\|\boldsymbol{\delta}'_{ij}\right\| \leq \lambda_1, \quad \forall i \neq j. \tag{5.27}$$

Substitute (5.27) to (5.26) to obtain

$$\begin{aligned} 0 &= \boldsymbol{p}_k - \boldsymbol{a}_i + \sum_{j \in C_k - \{i\}} \boldsymbol{g}'_{ij} + \lambda \sum_{k' \neq k} |C_{K'}| \cdot \boldsymbol{g}_{kk'} \\ &= \boldsymbol{x}_i^* - \boldsymbol{a}_i + \sum_{j \in C_k - \{i\}} \boldsymbol{g}'_{ij} + \lambda \sum_{k' \neq k} |C_{K'}| \cdot \boldsymbol{g}_{kk'}, \end{aligned}$$

satisfying (3.2) at i. As $i \in C_k, k \in [K]$ are chosen arbitrarily, the equality (3.2) holds for all i hence $\boldsymbol{x}^*$ is an optimal solution to (1.1). Since $\boldsymbol{p}_{\hat{k}} = \boldsymbol{p}_{\hat{k}'}$, we have $\boldsymbol{x}_i^* = \boldsymbol{x}_j^*$ for all $i, j \in C_{\hat{k}} \cup C_{\hat{k}'}$. By the agglomerative properties of the clusterpath, cluster $C_{\hat{k}}, C_{\hat{k}'}$ merge at some $\lambda' \in (\lambda_1, \lambda]$, which contradicts our choice of λ_2. That concludes our proof. □

By Lemma 8, the objective function is differentiable at $\boldsymbol{p}$. Hence, there holds

$$|C_k|(\boldsymbol{p}_k - \bar{\boldsymbol{a}}_k) + \lambda |C_k| \sum_{k' \neq k} |C_{k'}| \cdot \frac{\boldsymbol{p}_k - \boldsymbol{p}_{k'}}{\|\boldsymbol{p}_k - \boldsymbol{p}_{k'}\|} = \mathbf{0}, \quad \forall k \in [K]. \tag{5.28}$$

Define the following primal-dual solution:

$$
\begin{aligned}
\boldsymbol{x}_i^* &= \boldsymbol{p}_k, \quad \forall i \in C_k, k \in [K] \\
\boldsymbol{y}_{ij}^* &= \boldsymbol{x}_i^* - \boldsymbol{x}_j^*, \quad \forall 1 \le i < j \le n \\
\boldsymbol{z}_i^* &= \boldsymbol{x}_i^* - \boldsymbol{a}_i, \quad \forall i = 1, \dots, n, \\
s_i^* &= \frac{1}{2}(1 + \|\boldsymbol{z}_i^*\|^2), \quad \forall i = 1, \dots, n \\
u_i^* &= \frac{1}{2}(-1 + \|\boldsymbol{z}_i^*\|^2), \quad \forall i = 1, \dots, n \\
t_{ij}^* &= \|\boldsymbol{y}_{ij}^*\|, \quad \forall 1 \le i < j \le n \\
\boldsymbol{\delta}_{ij}^* &= \begin{cases} \boldsymbol{\delta}'_{ij}, & \text{if } i < j \text{ and } i, j \in C_k \\ \lambda \frac{\boldsymbol{x}_j^* - \boldsymbol{x}_i^*}{\|\boldsymbol{x}_j^* - \boldsymbol{x}_i^*\|}, & \text{otherwise} \end{cases} \qquad \forall 1 \le i < j \le n \\
\boldsymbol{\beta}_i^* &= -\boldsymbol{z}_i^*, \quad \forall i = 1, \dots, n \\
\gamma_i^* &= \frac{1}{2}(1 - \|\boldsymbol{\beta}_i^*\|^2), \quad \forall i = 1, \dots, n
\end{aligned}
\tag{5.29}
$$

Lemma 9. *The solution defined by* (5.29) *is optimal for SOCP* (5.2) *and* (5.3) *at* λ.

Proof. By construction, the primal constraints (5.2b), (5.2c), (5.2d), (5.2e), (5.2f), the dual constraints (5.3c), (5.3d), and the complementary slackness conditions (5.8), (5.9), (5.10), (5.11) and (5.12) with $\boldsymbol{\epsilon} = \mathbf{0}, \boldsymbol{\sigma} = \mathbf{0}$ are automatically satisfied. It remains to check if the solution satisfies (5.3b).

Verification for (5.3b)**:** For any $i \in C_k$ with some $k \in [K]$, (5.3b) is rewritten as

follows due to (5.27) and (5.28)

$$
\begin{aligned}
&-\sum_{j=1}^{i-1}\boldsymbol{\delta}_{ji}^{*}+\sum_{j=i+1}^{n}\boldsymbol{\delta}_{ij}^{*}+\boldsymbol{\beta}_{i}^{*}\\
=&-\sum_{j<i,j\in C_k-\{i\}}\boldsymbol{\delta}_{ji}^{\prime}+\sum_{j>i,j\in C_k-\{i\}}\boldsymbol{\delta}_{ij}^{\prime}+\lambda\sum_{k\neq k^{\prime}}|C_{k^{\prime}}|\frac{\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}}{\|\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}\|}+\boldsymbol{a}_{i}-\boldsymbol{x}_{i}^{*}\\
=&-\sum_{j\in C_k-\{i\}}\boldsymbol{g}_{ij}^{\prime}+\lambda\sum_{k\neq k^{\prime}}|C_{k^{\prime}}|\frac{\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}}{\|\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}\|}+\boldsymbol{a}_{i}-\boldsymbol{x}_{i}^{*}\\
=&\bar{\boldsymbol{a}}_{k}-\boldsymbol{a}_{i}+\lambda\sum_{k\neq k^{\prime}}|C_{k^{\prime}}|\frac{\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}}{\|\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}\|}+\boldsymbol{a}_{i}-\boldsymbol{p}_{i}\\
=&\bar{\boldsymbol{a}}_{k}+\lambda\sum_{k\neq k^{\prime}}|C_{k^{\prime}}|\frac{\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}}{\|\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}\|}-\boldsymbol{p}_{i}\\
=&\mathbf{0}.
\end{aligned}
$$

By KKT conditions, the solution defined above forms an optimal primal-dual pair.

□

Lemma 10. *The solution defined by* (5.29) *is strictly complementary.*

Proof. The strict complementarity is equivalent to (5.23) and (5.24), which can be easily checked as shown below

Verification for (5.23)**:** Let $1\leq i<j\leq n$. If $\boldsymbol{y}_{ij}^{*}=\mathbf{0}$, then there exists some $k\in[K]$ such that $i,j\in C_k$. By definition, $t_{ij}^{*}=0$ and $\boldsymbol{\delta}_{ij}^{*}=\boldsymbol{\delta}_{ij}$. Notice that $\boldsymbol{\delta}_{ij}$ is the optimal dual solution to (1.1) at λ_1, then it satisfies $\|\boldsymbol{\delta}_{ij}\|\leq\lambda_1<\lambda$ by the definition of λ. Hence,

$$
t_{ij}^{*}+\lambda=\lambda>\|\boldsymbol{\delta}_{ij}\|=\|\boldsymbol{\delta}_{ij}^{*}\|=\|\boldsymbol{y}_{ij}^{*}+\boldsymbol{\delta}_{ij}^{*}\|.
$$

If $\boldsymbol{y}_{ij}^{*}\neq\mathbf{0}$, then there exist $k,k^{\prime}\in[K]$ such that $i\in C_k, j\in C_{k^{\prime}}$ and $k\neq k^{\prime}$. By definition, $t_{ij}^{*}=\|\boldsymbol{y}_{ij}^{*}\|=\|\boldsymbol{p}_{k}-\boldsymbol{p}_{k^{\prime}}\|$ and $\boldsymbol{\delta}_{ij}^{*}=\lambda\frac{\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}}{\|\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}\|}$. Hence,

$$
t_{ij}^{*}+\lambda=\|\boldsymbol{p}_{k}-\boldsymbol{p}_{k^{\prime}}\|+\lambda>\left|\|\boldsymbol{p}_{k}-\boldsymbol{p}_{k^{\prime}}\|-\lambda\right|=\left\|\boldsymbol{p}_{k}-\boldsymbol{p}_{k^{\prime}}+\lambda\frac{\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}}{\|\boldsymbol{p}_{k^{\prime}}-\boldsymbol{p}_{k}\|}\right\|=\left\|\boldsymbol{y}_{ij}^{*}+\boldsymbol{\delta}_{ij}^{*}\right\|.
$$

Verification for (5.24): Let $i \in [n]$. By construction,

$$s_i^* + 1 - \gamma_i^* = \|\boldsymbol{z}_i^*\|^2 + 1 > 0 = \left\| \begin{pmatrix} \boldsymbol{z}_i^* - \boldsymbol{z}_i^* \\ -\frac{1}{2}(1 - \|\boldsymbol{z}_i\|^2) + \frac{1}{2}(1 - \|\boldsymbol{z}_i\|^2) \end{pmatrix} \right\| = \left\| \begin{pmatrix} \boldsymbol{z}_i^* + \boldsymbol{\beta}_i^* \\ u_i^* + \gamma_i^* \end{pmatrix} \right\|$$

Since the indices are chosen arbitrarily, the solution defined above is strictly complementary. □

With three lemmas presented in this section, there exists a strictly complementary optimal solution (as defined by (5.29)) to SOCP (5.2) and (5.3).

5.4 Test Guarantee

In Section 4, we validated our test theoretically in the sense that if the test succeeds, it is guaranteed that the correct clusters are found. In this section, we show that the test succeeds after a finite number of iterations of a certain interior point method, provided that λ is not at any fusion value. Specifically, we prove that the two conditions in our test are guaranteed to hold for a primal-dual path following algorithm satisfying the assumptions of Luo et al. [18] when the duality gap μ is sufficiently small.

Theorem 11. *If λ is not a fusion value, then there exists $\mu_0 > 0$ such that both CGR subgradients and separation conditions in the test are satisfied for any duality gap $\mu \leq \mu_0$ for a primal-dual path following algorithm satisfying the assumptions of Luo et al. [18].*

Let $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, s, u, t, \boldsymbol{\delta}, \boldsymbol{\beta}, \gamma)$ denote a primal and dual feasible solution. Let $C_1, C_2, ..., C_K$ denote the clusters obtained at optimum. Let $\mu' \in (0, 1)$ denote the central path parameter and let μ denote the duality gap at the feasible solution. By Theorem 6, there hold

$$\|\boldsymbol{x}(\mu') - \boldsymbol{x}^a\| = O(\mu'), \quad \|\boldsymbol{\delta}(\mu') - \boldsymbol{\delta}^a\| = O(\mu')$$

where $\boldsymbol{x}(\mu'), \boldsymbol{\delta}(\mu')$ are μ'-centered solution and $\boldsymbol{x}^a, \boldsymbol{\delta}^a$ are the analytic centers of the primal and dual optimal sets respectively. Moreover, since the iterates converge tangentially to the central path, we may assume the size of the central path neighborhood to be as follows

$$\|\boldsymbol{x} - \boldsymbol{x}(\mu')\| = O(\mu'), \quad \|\boldsymbol{\delta} - \boldsymbol{\delta}(\mu')\| = O(\mu').$$

Luo et al. [18] validated the assumption above for their interior point algorithm, which is a generalization of the Mizuno-Todd-Ye predictor-corrector method for linear programming. Combine the two sets of equations above and employ the triangle inequality to obtain

$$\|\boldsymbol{x} - \boldsymbol{x}^a\| = O(\mu'), \quad \|\boldsymbol{\delta} - \boldsymbol{\delta}^a\| = O(\mu').$$

As the duality gap μ is of a linear order of the central path parameter μ', the equalities above are rewritten as

$$\|\boldsymbol{x} - \boldsymbol{x}^a\| = O(\mu), \quad \|\boldsymbol{\delta} - \boldsymbol{\delta}^a\| = O(\mu).$$

Define $p, p' \geq 0$ such that $\|\boldsymbol{x}_i - \boldsymbol{x}_i^a\| \leq p\mu$ for all i and $\left\|\boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{ij}^a\right\| \leq p'\mu$ for all distinct pairs (i, j). Then, for all distinct pairs (i, j) in any cluster C_k, there holds $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \leq 2p\mu$. Moreover, define $q > 0$ such that all $\boldsymbol{x}_i^a$'s in different clusters are at least q apart, which implies that $\boldsymbol{x}_i$'s in different clusters are separated by a distance of at least $q - 2p\mu$. We may assume the duality gap satisfies $\mu < \frac{q}{2p}$. Notice that this assumption is guaranteed to be true after a finite number of iterations.

Let $C := C_k$ for some $k \in [K]$. By Lemma 3, there hold $\left\|\boldsymbol{\epsilon}_2^{ij}\right\| \leq \sqrt{\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 \mu + 2\mu^2}$ for all $i < j$ and

$$\left\| \begin{pmatrix} \boldsymbol{\sigma}_2^i \\ \sigma_3^i \end{pmatrix} \right\| \leq \sqrt{\left(\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + 1 + 2\mu \right) \cdot \left(\frac{1}{2} + \sum_{l=1}^n (n-1)\lambda \|\boldsymbol{a}_l\| + \mu \right) \mu}$$

for all i.

5.4.1 Bound the CGR subgradient

Bound $\|\boldsymbol{\delta}_{ij}\|$

Lemma 12. *For all $i, j \in C, i \neq j$, the following inequality holds*

$$\|\boldsymbol{\delta}_{ij}\| \leq \lambda - r + p'\mu$$

where $r := \min_{l \neq l', l, l' \in C_k, k \in [K]} (\lambda - \|\boldsymbol{\delta}_{ll'}^a\|) > 0$.

Proof. Let $i, j \in C$ and $i \neq j$. By the definition of analytic center and strict complementarity,

$$\|\boldsymbol{\delta}_{ll'}^a\| < \lambda,$$

holds for all $l \neq l', l, l' \in C_k, k \in [K]$. Hence, $r > 0$ by definition. Moreover, r also satisfies

$$\left\|\boldsymbol{\delta}_{ij}^a\right\| \leq \lambda - r, \quad \forall i, j \in C, i \neq j.$$

Since $\left\|\boldsymbol{\delta}_{ij} - \boldsymbol{\delta}_{ij}^a\right\| \leq p'\mu$, we obtain

$$\|\boldsymbol{\delta}_{ij}\| \leq \lambda - r + p'\mu, \quad \forall i, j \in C, i \neq j.$$

□

Bound $\|\boldsymbol{g}_{ik} - \boldsymbol{g}_{jk}\|$

Lemma 13. *For all $i, j \in C$ and $k \notin C$, the following inequality holds*

$$\|\boldsymbol{g}_{ik} - \boldsymbol{g}_{jk}\| \leq \frac{4\lambda p\mu}{q - 2p\mu} + \frac{2\sqrt{\sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 \mu + 2\mu^2}}{q - 2p\mu} + \frac{\mu}{q - 2p\mu}$$

Proof. Let $i, j \in C$ and $k \notin C$. Without loss of generality, we may assume $i < j < k$. Hence, $\boldsymbol{g}_{ik} = -\boldsymbol{\delta}_{ik}, \boldsymbol{g}_{jk} = -\boldsymbol{\delta}_{jk}$. By (5.9), we derive

$$t_{ik}\boldsymbol{\delta}_{ik} - t_{jk}\boldsymbol{\delta}_{jk} = -\lambda \boldsymbol{y}_{ik} + \lambda \boldsymbol{y}_{jk} + \boldsymbol{\epsilon}_2^{ik} - \boldsymbol{\epsilon}_2^{jk} = -\lambda(\boldsymbol{x}_i - \boldsymbol{x}_j) + \boldsymbol{\epsilon}_2^{ik} - \boldsymbol{\epsilon}_2^{jk}. \tag{5.30}$$

Adding the term $(t_{jk} - t_{ik})\boldsymbol{\delta}_{jk}$ to both sides of the equality to obtain

$$t_{ik}(\boldsymbol{\delta}_{ik} - \boldsymbol{\delta}_{jk}) = (t_{jk} - t_{ik})\boldsymbol{\delta}_{jk} - \lambda(\boldsymbol{x}_i - \boldsymbol{x}_j) + \boldsymbol{\epsilon}_2^{ik} - \boldsymbol{\epsilon}_2^{jk}.$$

Notice that $t_{ik} \geq \|\boldsymbol{y}_{ik}\| = \|\boldsymbol{x}_i - \boldsymbol{x}_k\| \geq q - 2p\mu > 0$ by the primal constraint (5.2e) and our assumption on the duality gap. Divide the equality above by t_{ik} to obtain

$$\boldsymbol{\delta}_{ik} - \boldsymbol{\delta}_{jk} = \frac{t_{jk} - t_{ik}}{t_{ik}}\boldsymbol{\delta}_{jk} - \frac{\lambda(\boldsymbol{x}_i - \boldsymbol{x}_j)}{t_{ik}} + \frac{\boldsymbol{\epsilon}_2^{ik} - \boldsymbol{\epsilon}_2^{jk}}{t_{ik}}.$$

Substitute the definition of $\boldsymbol{g}$ into the equality above to obtain

$$\boldsymbol{g}_{ik} - \boldsymbol{g}_{jk} = \frac{t_{ik} - t_{jk}}{t_{ik}}\boldsymbol{\delta}_{jk} + \frac{\lambda(\boldsymbol{x}_i - \boldsymbol{x}_j)}{t_{ik}} - \frac{\boldsymbol{\epsilon}_2^{ik} - \boldsymbol{\epsilon}_2^{jk}}{t_{ik}}. \tag{5.31}$$

By the perturbed complementary slackness (5.8), the primal constraint (5.2e) and the Cauchy-Schwarz inequality, we derive the following inequality

$$\epsilon_1^{ik} = t_{ik}\lambda + \boldsymbol{y}_{ik}^T\boldsymbol{\delta}_{ik} \geq t_{ik}\lambda - \|\boldsymbol{y}_{ik}\| \cdot \|\boldsymbol{\delta}_{ik}\| \geq t_{ik}\lambda - \|\boldsymbol{y}_{ik}\| \cdot \lambda,$$

which yields an upper bound on t_{ik}

$$t_{ik} \leq \|\boldsymbol{y}_{ik}\| + \frac{\epsilon_1^{ik}}{\lambda}.$$

Combined with the primal constraint (5.2e) at t_{jk} and the triangle inequality, we obtain the following

$$t_{ik} - t_{jk} \leq \|\boldsymbol{y}_{ik}\| + \frac{\epsilon_1^{ik}}{\lambda} - \|\boldsymbol{y}_{jk}\| \leq \|\boldsymbol{y}_{ik} - \boldsymbol{y}_{jk}\| + \frac{\epsilon_1^{ik}}{\lambda} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\| + \frac{\epsilon_1^{ik}}{\lambda}. \tag{5.32}$$

The same inequality holds for $t_{jk} - t_{ik}$ due to the symmetry of (5.32). By (5.31), (5.32) and triangle inequality, the norm bound of $\boldsymbol{g}_{ik} - \boldsymbol{g}_{jk}$ is as follows

$$\begin{aligned}
\|\boldsymbol{g}_{ik} - \boldsymbol{g}_{jk}\| &\leq \frac{|t_{ik} - t_{jk}| \cdot \|\boldsymbol{\delta}_{jk}\|}{t_{ik}} + \frac{\lambda \|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{t_{ik}} + \frac{\left\|\boldsymbol{\epsilon}_2^{ik}\right\| + \left\|\boldsymbol{\epsilon}_2^{jk}\right\|}{t_{ik}} \quad \text{(By triangle inequality)} \\
&\leq \frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\| + \frac{\epsilon_1^{ik}}{\lambda}}{t_{ik}} \|\boldsymbol{\delta}_{jk}\| + \frac{\lambda \|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{t_{ik}} + \frac{\left\|\boldsymbol{\epsilon}_2^{ik}\right\| + \left\|\boldsymbol{\epsilon}_2^{jk}\right\|}{t_{ik}} \quad \text{(By (5.32))} \\
&\leq \frac{2\lambda \|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{t_{ik}} + \frac{\left\|\boldsymbol{\epsilon}_2^{ik}\right\| + \left\|\boldsymbol{\epsilon}_2^{jk}\right\|}{t_{ik}} + \frac{\epsilon_1^{ik}}{t_{ik}} \quad \text{(By (5.2e) and (5.30)).}
\end{aligned}$$

Since $i, j \in C$ and $k \notin C$, there hold $t_{ik} \geq \|\boldsymbol{y}_{ik}\| = \|\boldsymbol{x}_i - \boldsymbol{x}_k\| \geq q - 2p\mu$ and $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \leq 2p\mu$. Moreover, there also hold $\epsilon_1^{ik} \leq \mu$, $\left\|\boldsymbol{\epsilon}_2^{ik}\right\| \leq \sqrt{\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 \mu + 2\mu^2}$ and $\left\|\boldsymbol{\epsilon}_2^{jk}\right\| \leq \sqrt{\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 \mu + 2\mu^2}$. Hence, $\|\boldsymbol{g}_{ik} - \boldsymbol{g}_{jk}\|$ is further upper bounded as follows

$$\|\boldsymbol{g}_{ik} - \boldsymbol{g}_{jk}\| \leq \frac{4\lambda p\mu}{q - 2p\mu} + \frac{2\sqrt{\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 \mu + 2\mu^2}}{q - 2p\mu} + \frac{\mu}{q - 2p\mu} \tag{5.33}$$

□

Bound $\|\boldsymbol{\omega}_i\|$

Lemma 14. *For all $i \in C$, it holds*

$$\|\boldsymbol{\omega}_i\| \leq 2\sqrt{2\left(\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + 1 + 2\mu\right) \cdot \left(\frac{1}{2} + \sum_{l=1}^n (n-1)\lambda \|\boldsymbol{a}_l\| + \mu\right)\mu}.$$

Proof. Let $i \in C$. By definition,

$$\boldsymbol{\omega}_i = \frac{\sigma_3^i}{s_i}\boldsymbol{z}_i + \frac{1}{s_i}\boldsymbol{\sigma}_2^i.$$

By the primal constraint (5.2f), we have

$$\|\boldsymbol{z}_i\| \leq \sqrt{2s_i - 1} \leq \sqrt{2s_i}, \quad s_i \geq \frac{1}{2},$$

which implies

$$\frac{\|\boldsymbol{z}_i\|}{s_i} \leq \sqrt{\frac{2}{s_i}} \leq \sqrt{4} = 2, \quad \frac{1}{s_i} \leq 2.$$

Coupled with triangle inequality, these two inequalities yield

$$\|\boldsymbol{\omega}_i\| \leq \frac{\|\boldsymbol{z}_i\|}{s_i}\sigma_3^i + \frac{1}{s_i}\left\|\boldsymbol{\sigma}_2^i\right\| \leq 2\sigma_3^i + 2\left\|\boldsymbol{\sigma}_2^i\right\|.$$

Moreover, since $\left\|\begin{pmatrix}\boldsymbol{\sigma}_2^i \\ \sigma_3^i\end{pmatrix}\right\| \leq \sqrt{\left(\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + 1 + 2\mu\right) \cdot \left(\frac{1}{2} + \sum_{l=1}^n (n-1)\lambda \|\boldsymbol{a}_l\| + \mu\right)\mu}$ holds for any $i \in [n]$ by Lemma 3 and the duality gap,

$$(\sigma_3^i)^2 + \left\|\boldsymbol{\sigma}_2^i\right\|^2 \leq \left(\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + 1 + 2\mu\right) \cdot \left(\frac{1}{2} + \sum_{l=1}^n (n-1)\lambda \|\boldsymbol{a}_l\| + \mu\right)\mu.$$

which implies the following inequality since $(a+b)^2 \leq 2a^2 + 2b^2$

$$(\sigma_3^i + \left\|\boldsymbol{\sigma}_2^i\right\|)^2 \leq 2 \cdot \left(\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + 1 + 2\mu\right) \cdot \left(\frac{1}{2} + \sum_{l=1}^n (n-1)\lambda \|\boldsymbol{a}_l\| + \mu\right)\mu.$$

Therefore, the following holds as $i \in C$ is chosen arbitrarily:

$$\|\boldsymbol{\omega}_i\| \leq 2\sigma_3^i + 2\left\|\boldsymbol{\sigma}_2^i\right\| \leq 2\sqrt{2 \cdot \left(\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + 1 + 2\mu\right) \cdot \left(\frac{1}{2} + \sum_{l=1}^n (n-1)\lambda \|\boldsymbol{a}_l\| + \mu\right)\mu}.$$

□

Bound CGR subgradients

Lemma 15. *For all $i, j \in C$ and $i < j$, there holds*

$$\begin{aligned}\|\boldsymbol{q}_{ij}\| \leq & \lambda - r + p'\mu + \frac{1}{m} \cdot \left(2p\mu + 4\sqrt{2 \cdot \left(\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + 1 + 2\mu\right) \cdot \left(\frac{1}{2} + \sum_{l=1}^n (n-1)\lambda \|\boldsymbol{a}_l\| + \mu\right)\mu}\right) \\ & + \frac{n-m}{m}\left(\frac{4\lambda p\mu}{q - 2p\mu} + \frac{2\sqrt{\sum_{l=1}^n \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 \mu + 2\mu^2}}{q - 2p\mu} + \frac{\mu}{q - 2p\mu}\right).\end{aligned}$$

Proof. Let $i, j \in C$ and $i < j$. By triangle inequality,

$$\|\boldsymbol{q}_{ij}\| \leq \|\boldsymbol{\delta}_{ij}\| + \frac{1}{m} \cdot (\|\boldsymbol{x}_i - \boldsymbol{x}_j\| + \|\boldsymbol{\omega}_i\| + \|\boldsymbol{\omega}_j\|) + \frac{1}{m} \sum_{k \notin C} \|\boldsymbol{g}_{ik} - \boldsymbol{g}_{jk}\| .$$

With the assumptions on the distance between points,

$$\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \leq 2p\mu.$$

By Lemma 12, 13, Lemma 14 and the inequality above, we obtain

$$\begin{aligned}\|\boldsymbol{q}_{ij}\| \leq & \lambda - r + p'\mu + \frac{1}{m} \cdot \left(2p\mu + 4\sqrt{2 \cdot \left(\sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 + 1 + 2\mu \right) \cdot \left(\frac{1}{2} + \sum_{l=1}^{n} (n-1)\lambda \|\boldsymbol{a}_l\| + \mu \right) \mu} \right) \\ & + \frac{n-m}{m} \left(\frac{4\lambda p\mu}{q - 2p\mu} + \frac{2\sqrt{\sum_{l=1}^{n} \|\bar{\boldsymbol{a}} - \boldsymbol{a}_l\|^2 \mu + 2\mu^2}}{q - 2p\mu} + \frac{\mu}{q - 2p\mu} \right) .\end{aligned} \tag{5.34}$$

□

5.4.2 Proof of Theorem 11

Proof. We rewrite (5.34) with $O(\cdot)$ notation to obtain the following inequality

$$\|\boldsymbol{q}_{ij}\| \leq \lambda - r + O(\sqrt{\mu}), \quad \forall i, j \in C_k, i \neq j, k \in [K],$$

since $C = C_k$ is an arbitrarily cluster. As $r > 0$ by Lemma 12, there exists $\mu_1 > 0$ such that for all $\mu \leq \mu_1$, $\|\boldsymbol{q}_{ij}\| \leq \lambda$ holds for all $i, j \in C_k, i \neq j, k \in [K]$. Here concludes the proof of the CGR subgradient condition.

Since $q > 0$, there exists $\mu_2 > 0$ such that $2\sqrt{2\mu_2} < q - 2p\mu_2$. Hence, for all $\mu \leq \mu_2$, all clusters are separated at distance of at least $2\sqrt{2\mu_2}$. Here concludes the proof of the second condition.

Let $\mu_0 = \min\{\mu_1, \mu_2\}$, then both CGR subgradient and separation conditions are satisfied for any $\mu \leq \mu_0$. □

5.5 Computational experiments

In this section, we conduct experiments in which a Chi-Lange ADMM solver [7] and our clustering test for sum-of-norms clustering are applied to a simulated dataset of two normally distributed half moons. We intend to answer the following questions: (1) How does the performance of our test depend on λ? and (2) How does the recovery of two half moons depend on λ?

Our algorithm is implemented in Julia [3]. It terminates if the clustering test succeeds, or if the maximum number of iterations is reached. In the algorithm, the code tests for clustering every t iterations of the ADMM solver. The value of t is taken to be 8 in our experiment. At the end of every t iterations, the solver yields a primal solution and a dual solution, from which our algorithm constructs a primal and dual feasible pair for the SOCP formulation by (5.7). With the feasible solution, the algorithm then creates candidate clusters, computes duality gap and constructs CGR subgradients. The code checks for the CGR subgradient condition and separation condition. If both conditions hold, the clustering test reports 'success'. Otherwise, the code runs t more iterations of the ADMM solver and repeats the clustering test. The detailed algorithm is outlined as follows. Each iteration of the ADMM solver is of complexity $O(n^2 d)$.

Algorithm 2: Find clusters

$C \leftarrow \{1, \ldots, n\}$;
$k \leftarrow 1$;
while $C \neq \emptyset$ **do**
  Choose $i \in C$ arbitrarily;
  Create a cluster $R_k \leftarrow \{j : \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \leq \mu^{3/4}\}$ (including i itself);
  Delete all these points in R_k from C;
  $k \leftarrow k + 1$;
Return candidate clusters $\{R_1, R_2, \ldots, R_{K'}\}$;

Algorithm 3: An ADMM algorithm with our clustering test

Result: Clustering assignment
Initialize $(\boldsymbol{x}, \boldsymbol{\delta})$;
while *clustering test fails or maximum number of iterations is not reached* **do**
  for $l = 1, 2, \ldots, t$ **do**
    ADMM updates by Chi and Lange [7];
  end
  Construct a feasible solution for SOCP by (5.29) from the current ADMM iterate;
  Compute the duality gap μ;
  Run Algorithm 2 to find clusters $\{R_1, R_2, \ldots, R_{K'}\}$;
  Compute CGR subgradients from dual variables for $\{R_1, R_2, \ldots, R_{K'}\}$;
  Check the CGR subgradient condition; Check that no two clusters are distance $\leq 2\sqrt{2\mu}$ of each other;
  Mark the clustering test 'success' if both conditions pass and mark it 'failure' otherwise;
end
Return recovered clusters $\{R_1, R_2, \ldots, R_{K'}\}$.

To assess the performance of recovery, we employ the Rand index by Rand [25]. Rand index is a measure which specifically evaluates the performance of clustering. It compares two clusterings $\{R_1, \ldots, R_{K'}\}$ and $\{V_1, \ldots, V_K\}$ in a pairwise manner. If a pair of data points are placed in the same cluster in both clusterings, or if a pair of data points are placed in different clusters in both clusterings, then this pair is called a similar assignment and it contributes to the measure of similarity between two clusterings. We define the

following two sets of similar assignments on all distinct pairs of instances:

$$S := \{(i,j) : 1 \leq i < j \leq n \text{ such that there exist } m, m' \text{ satisfying } i, j \in V_m \cap R_{m'}\},$$
$$D := \{(i,j) : 1 \leq i < j \leq n \text{ such that } i \in V_{m_1}, j \in V_{m_2}, m_1 \neq m_2, \text{ and } i \in R_{m'_1}, j \in R_{m'_2}, m'_1 \neq m'_2\}.$$

Then Rand index is defined as the fraction of all distinct pairs which are similar assignments:

$$R = \frac{|S| + |D|}{\binom{n}{2}},$$

where $|\cdot|$ denotes the cardinality function. The value of R ranges from 0 to 1. When $R = 0$, two clusterings are completely dissimilar. When $R = 1$, two clusterings are identical. A higher Rand index indicates a higher level of similarity. A random assignment to clusters in the case of equally sized clusters, $K = 2$ yields expected Rand index of 0.5.

The experiment is conducted on a simulated dataset of two normally distributed half moons with 500 instances. The angle of two half moons follows a Gaussian distribution with a mean of 0 and a standard deviation of $\frac{\pi}{6}$. A random noise which follows a two-dimensional Gaussian distribution with a mean of 0 and a standard deviation of 0.05 displaces the instances from the moons. Fifty linearly spaced values of λ are taken from the range $[10^{-8}, 0.00496]$. The range is determined empirically. Furthermore, the maximum number of iterations is chosen to be 50,000. It took approximately 15 hours total on an Intel Xeon processor single-threaded to complete the experiment.

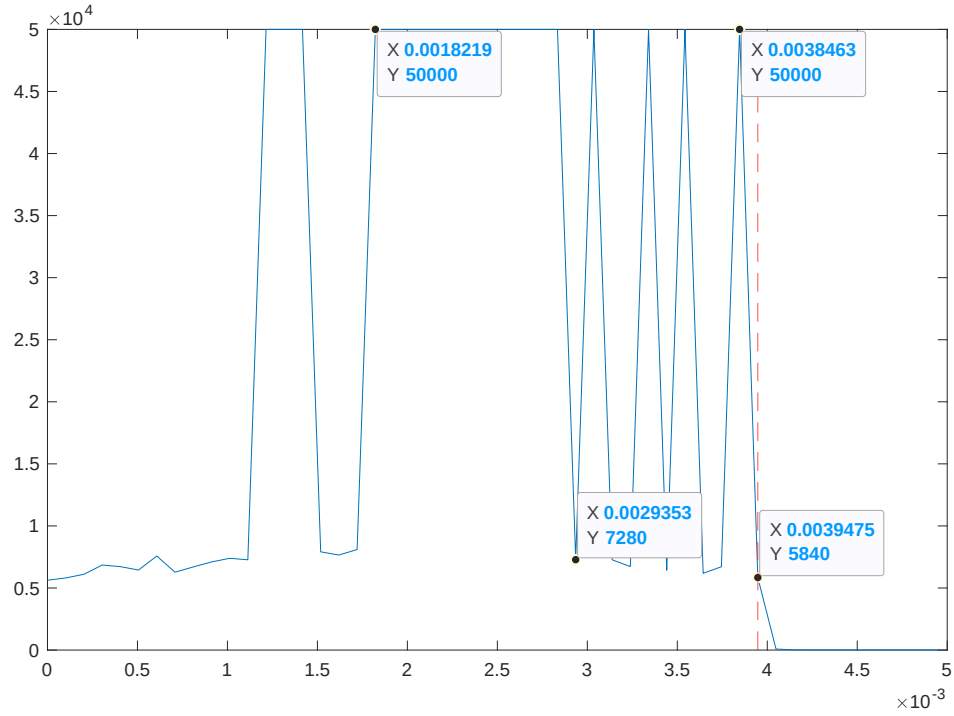

Figure 5.1: Iteration counts versus λ

Our first objective is to evaluate the performance of our clustering test. At 32 out of 50 values of λ, the clustering test succeeds before the maximum number of iterations is reached. When λ is in the range between $\lambda = 0.0018219$ and $\lambda = 0.0028341$, the algorithm repeatedly reaches the iteration threshold before the test succeeds as shown in Figure 5.1. The performance is interpretable with theories discussed earlier. The clustering test is not guaranteed to succeed when λ is at a fusion value, and the test performs poorly near a fusion value as shown in Figure 5.1. When $n = 500$, there are at most 500 fusion values. All fusion values are in the range between $\lambda = 0.00040$ and $\lambda = 0.00405$ as observed in the experiment. Hence, fusion occurs frequently, and massive fusion values are located densely in a small region. Thus, in our experiment, it is very likely that the λ we pick is near or at a fusion value, which leads to the poor performance of our clustering test.

We anticipate that the clustering test improves with fewer data points, and it is indeed the case. The same experiment is also implemented for 200 instances generated from two normally distributed half moons with the same parameters. At 89 out of 100 values of λ, our clustering test succeeds before the maximum number of iterations is reached.

The experiment also attempts to explore the relationship between λ value and the

recovery of half moons. To evaluates the recovery, we compute the Rand index with the recovered clustering and the generative clustering. The figure below shows Rand index against λ values. The value of Rand index increases monotonically and peaks at $\lambda = 0.00395$, where the clustering test succeeds and the Rand index achieves a value of 0.949.

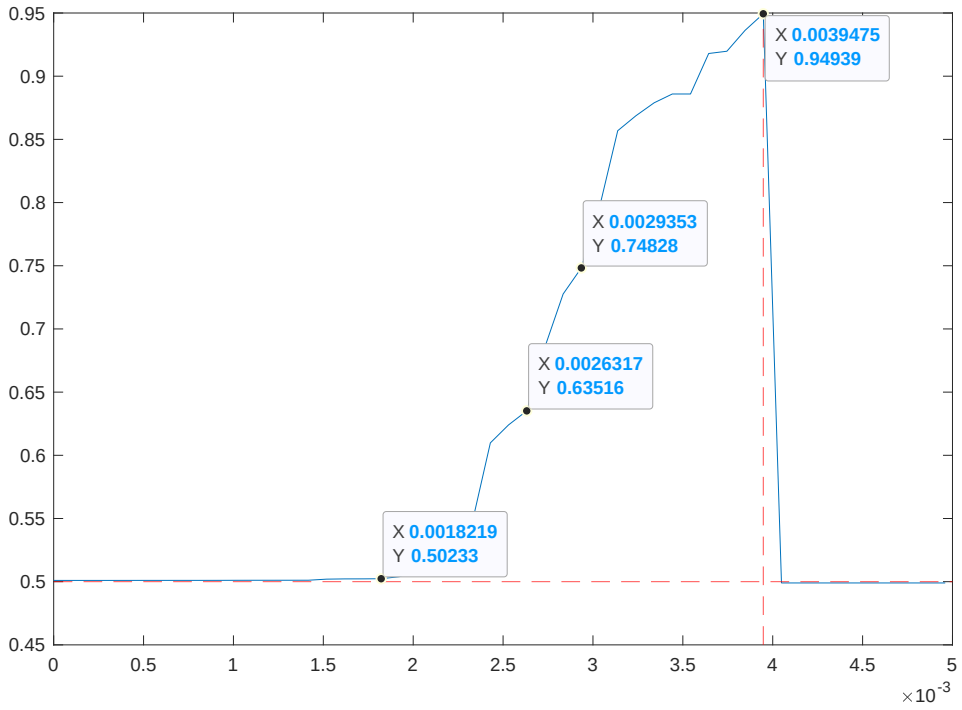

Figure 5.2: Rand index versus λ

To illustrate the clustering at $\lambda = 0.00395$, we also plot the two half moons and color the clusters. Red instances belong to one cluster, and blue instances belong to another cluster. Yellow instances are assigned to clusters of singleton points, and they are identified as noises.

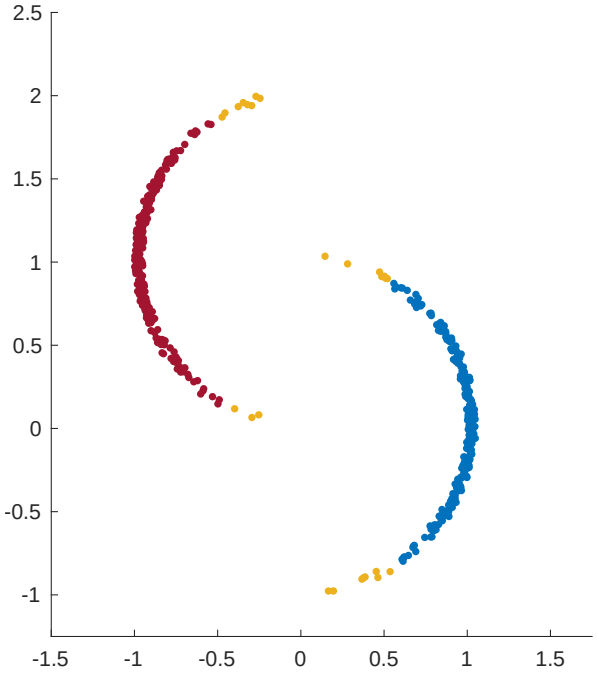

Figure 5.3: Labeled points with clustering at $\lambda = 0.00395$

Sum-of-norms clustering with equal weights performs poorly on standard half moons [7] and normally distributed half moons with a large standard deviation. To resolve the issue, many authors such as Sun et al. [28] apply exponentially decaying weights to the sum-of-norms clustering. The exponentially decaying weight of pair (i, j) is determined by the distance between original data $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$. The weight is set to zero if j is not among $i's$ k-nearest neighbors. Otherwise, the weight is computed as follows

$$w_{ij} = \exp(-\phi \left\| a_i - a_j \right\|^2)$$

where ϕ is a nonnegative parameter. Assigning weights in this manner implicitly imposes a prior hypothesis that the nearest-neighbor structure corresponds to true clustering, which is certainly the case for the standard half-moon data set. Chi and Lange [7] assess the effect of the number of nearest neighbors k and the parameter ϕ on SON clustering with numerical experiments on a half-moon dataset of 100 points. Setting $k = 10$ and $\phi = 0.5$ yields the best clustering. Choosing $k = 50$ and $\phi = 0$ results in a similar clustering pattern to our experiment: clusters only form until late then all points quickly coalesce to one cluster. At any value of λ, SON clustering could not identify two half moons with high accuracy. When $k = 10$ and $\phi = 0$, or $k = 50$ and $\phi = 0.5$, SON clustering correctly

identifies clusters for the easier points but fails to cluster points located at the lower tip of the right moon and the upper tip of the left moon.

Chapter 6

Discussion

The analysis of the mixture of Gaussians in Chapter 4 used only standard bounds and simple properties of the normal distribution, so it should be apparent to the reader that many extensions of this result (e.g., Gaussians with a more general covariance matrix, uniform distributions, many kinds of deterministic distributions) are possible. The key technique is Theorem 1, which essentially decouples the clusters from each other so that each can be analyzed in isolation. Such a theorem does not apply to most other clustering algorithms, or even to sum-of-norm clustering in the case of non-multiplicative weights, so obtaining similar results for other algorithms remains a challenge.

An interesting question concerns the ranges of parameters for which the Panahi et al. result (which requires an upper bound on n), or its extension due to Sun et al. applies versus our bound (which assumes $n \to \infty$). Our result, stated loosely, is that the probability of correct labeling of points a fixed number of standard deviations from the means goes to 1 exponentially fast in n, whereas the other result states that all points are correctly labeled with probability that goes to 1 exponentially fast in the ratio

$$\frac{\min_{1 \leq m < m' \leq K} \|\boldsymbol{\mu}_m - \boldsymbol{\mu}_{m'}\|}{\max_{1 \leq m \leq K} \sigma_m}.$$

Is it possible to stitch the two results together into a theorem that encompasses all values of n? One of our computational experiments suggests that this may be possible.

We also proposed a test to determine all clusters from an approximate solution yielded from any primal-dual type method. If the test reports 'success', then the clusters are correctly identified. Moreover, if a primal-dual path following method that maintains close proximity to the central path is used, the test is guaranteed to report 'success' after a finite

number of iterations at non-fusion values of λ, where strict complementarity holds. A few natural questions concerning strict complementarity and the test itself are (1) Is there a rigorous test that works when strict complementarity fails? (2) What is the complexity for our clustering test since it depends on the choice of λ values? (3) Is the test guaranteed to work for a general primal-dual algorithm? (4) Can one identify clusters correctly from a primal-only algorithm?

References

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95, 12 2001.

[2] Dimitri P. Bertsekas. Incremental proximal methods for large scale convex optimization. *Mathematical Programming*, 129(2):163, June 2011.

[3] J. Bezanson, A. Edelman, S. Karpinski, and V.B. Shah. Julia: A fresh approach to numerical computing. *SIAM Rev.*, 59(1):65–98, 2017.

[4] Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717, Apr 2009.

[5] Raymond B. Cattell. The description of personality: Principles and findings in a factor analysis. *The American Journal of Psychology*, 58(1):69–90, 1945.

[6] E. Chi and S. Steinerberger. Recovering trees with convex clustering. https://arxiv.org/abs/1806.11096, 2018.

[7] Eric C. Chi and Kenneth Lange. Splitting methods for convex clustering. *Journal of Computational and Graphical Statistics*, 24(4):994–1013, 2015. PMID: 27087770.

[8] J. Chiquet, P. Gutierrez, and G. Rigaill. Fast tree inference with weighted fusion penalties. *Journal of Computational and Graphical Statistics*, 26:205–216, 2017.

[9] Chris Ding, Xiaofeng He, and Horst D. Simon. *On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering*, pages 606–610. Society for Industrial and Applied Mathematics, 2005.

[10] Jerome Friedman, Trevor Hastie, Holger Höfling, and Robert Tibshirani. Pathwise coordinate optimization. *Ann. Appl. Stat.*, 1(2):302–332, 12 2007.

[11] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Springer, 2012.

[12] T. Hocking, A. Joulin, F. Bach, and J.-P. Vert. Clusterpath: An algorithm for clustering using convex fusion penalties. In *International Conference on Machine Learning*, 2011.

[13] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Stat. Assoc*, 58:13–30, 1963.

[14] T. Jiang and S. Vavasis. On identifying clusters from sum-of-norms clustering computation. https://arxiv.org/abs/2006.11355, 2020.

[15] T. Jiang, S. Vavasis, and C. W. Zhai. Recovery of a mixture of gaussians by sum-of-norms clustering. https://arxiv.org/abs/1902.07137, 2019.

[16] M. Kojima, M. Shida, and S. Shindoh. Local convergence of predictor—corrector infeasible-interior-point algorithms for sdps and sdlcps. *Mathematical Programming*, 80:129–160, 1998.

[17] F. Lindsten, H. Ohlsson, and L. Ljung. Clustering using sum-of-norms regularization: With application to particle filter output computation. In *IEEE Statistical Signal Processing Workshop (SSP)*, 2011.

[18] Zhi-Quan Luo, Jos F. Sturm, and Shuzhong Zhang. Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming. *SIAM Journal on Optimization*, 8(1):59–81, 1998.

[19] Dustin G Mixon, Soledad Villar, and Rachel Ward. Clustering subgaussian mixtures by semidefinite programming. *Information and Inference: A Journal of the IMA*, 6(4):389–415, 03 2017.

[20] Yu. Nesterov and L. Tuncel. Local superlinear convergence of polynomial-time interior-point methods for hyperbolicity cone optimization problems. *SIAM Journal on Optimization*, 26(1):139–170, 2016.

[21] A. Panahi, D. Dubhashi, F. Johansson, and C. Bhattacharyya. Clustering by sum of norms: Stochastic incremental algorithm, convergence and cluster recovery. *Journal of Machine Learning Research*, 70, 2017.

[22] K. Pelckmans, J. De Brabanter, J. A. K. Suykens, and B. De Moor. Convex cluster shrinkage. Available on-line at ftp://ftp.esat.kuleuven.ac.be/sista/kpelckma/ccs_pelckmans2005.pdf, 2005.

[23] Jiming Peng and Yu Wei. Approximating K-means-type clustering via semidefinite programming. *SIAM J. Optim.*, 18(1):186–205, 2007.

[24] Peter Radchenko and Gourab Mukherjee. Convex clustering via l1 fusion penalization. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1527–1546, 2017.

[25] William M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

[26] S. Shalev-Shwartz. Online learning and online convex optimization. *Foundations and trends in machine learning*, 4:107–194, 2011.

[27] S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge University Press, 2014.

[28] D. Sun, K.-C. Toh, and Y. Yuan. Convex clustering: model, theoretical guarantees and efficient algorithm. https://arxiv.org/abs/1810.02677, 2018.

[29] Kean Ming Tan and Daniela Witten. Statistical properties of convex clustering. *Electron. J. Statist.*, 9(2):2324–2347, 2015.

[30] Y. Yuan, D. Sun, and K.-C. Toh. An efficient semismooth Newton based algorithm for convex clustering. https://arxiv.org/abs/1802.07091, 2018.

[31] Changbo Zhu, Huan Xu, Chenlei Leng, and Shuicheng Yan. Convex optimization procedure for clustering: Theoretical revisit. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 1619–1627. Curran Associates, Inc., 2014.