Mathematical Programming
https://doi.org/10.1007/s10107-018-1333-x

FULL LENGTH PAPER

Series A

When do birds of a feather flock together? *k*-Means, proximity, and conic programming

Xiaodong Li[1] · Yang Li[2] · Shuyang Ling[3] · Thomas Strohmer[2] · Ke Wei[4]

Received: 16 October 2017 / Accepted: 19 September 2018

Abstract
Given a set of data, one central goal is to group them into clusters based on some notion of similarity between the individual objects. One of the most popular and widely-used approaches is *k*-means despite the computational hardness to find its global minimum. We study and compare the properties of different convex relaxations by relating them to corresponding proximity conditions, an idea originally introduced by Kumar and Kannan. Using conic duality theory, we present an improved proximity condition under which the Peng–Wei relaxation of *k*-means recovers the underlying clusters exactly. Our proximity condition improves upon Kumar and Kannan and is comparable to that of Awashti and Sheffet, where proximity conditions are established for projective *k*-means. In addition, we provide a necessary proximity condition for the exactness of the Peng–Wei relaxation. For the special case of equal cluster sizes, we establish a different and completely localized proximity condition under which the Amini–Levina relaxation yields exact clustering, thereby having addressed an open problem by Awasthi and Sheffet in the balanced case. Our framework is not only deterministic and model-free but also comes with a clear geometric meaning which allows for further analysis and generalization. Moreover, it can be conveniently applied to analyzing various data generative models such as the stochastic ball models and Gaussian mixture models. With this method, we improve the current minimum separation bound for the stochastic ball models and achieve the state-of-the-art results of learning Gaussian mixture models.

Keywords Convex relaxation · *k*-Means · Clustering · Gaussian mixture model

Y. Li, S. Ling, T. Strohmer, and K. Wei acknowledge support from the NSF via Grants DMS 1620455 and DMS 1737943.

Extended author information available on the last page of the article

Published online: 29 September 2018

1 Introduction

k-Means clustering is one of the most well-known and widely-used clustering methods in unsupervised learning. Given N data points in $\mathbb{R}^m$, the goal is to partition them into k clusters by minimizing the total squared distance between each data point and the corresponding cluster center. It is a problem related to Voronoi tessellations [10]. However, k-means is combinatorial in nature since it is essentially equivalent to an integer programming problem [22]. Thus, minimizing the k-means objective function turns out to be an NP-hard problem, even if there are only two clusters [2] or if the data points are on a 2D plane [19].

Despite its hardness, numerous efforts have been made to develop effective and efficient heuristic algorithms to handle the k-means problem in practice. A famous example is Lloyd's algorithm [17] which was originally introduced for vector quantization and then became popular in data clustering due to its high efficiency and simplicity of implementation. One of the earliest convergence analyses of Lloyd's algorithm was given by Selim and Ismail [22]: Under certain conditions, the algorithm converges to a stationary point within a finite number of iterations but may fail to converge to a local minimum. A smoothed analysis given by Arthur, Manthey and Roglin [4] shows that the smoothed/expected number of iterations is bounded polynomially by N, k and m while the worst-case running time can be $2^{\Omega(N)}$ even for the case when data points are on a plane [24].

We are particularly interested in the semidefinite programming (SDP) relaxation for k-means by Peng and Wei [21], who observed that the k-means objective function can be written as the inner product between a projection matrix and a distance matrix constructed from the data, and the combinatorial constraints of the projection matrix can be convexified. Thus, whenever the Peng–Wei relaxation produces an output corresponding to a partition of the data set, the k-means problem is solved in polynomial time [27]. The details of the Peng–Wei relaxation will be explained in 2.

Theoretical properties of the Peng–Wei relaxation have also been studied under specific stochastic models in the literature. *Minimum separation conditions* were established in [5,13] to guarantee exact clustering for the stochastic ball models with balanced clusters (i.e., each cluster has the same number of points), while a similar study was conducted in [20] for the Gaussian mixture model.

Despite these efforts, the Peng–Wei relaxation is not yet thoroughly understood. Several fundamental questions of vital importance remain unexplored or require better answers, such as

- How do the number of clusters and the data dimension affect the performance of the Peng–Wei relaxation?
- How does the performance of the Peng–Wei relaxation depend on the balancedness of the cluster sizes and covariance structures within each cluster?
- Can the global minimum separation condition be localized?
- Under the special case of equal cluster sizes, does the tighter Amini–Levina relaxation [3] improve the Peng–Wei relaxation? If so, in which sense?

The studies in [5,13,20] reveal certain information about the Peng–Wei relaxation based on the assumption of sufficient minimum center separation: guaranteed exact

recovery in the case of the stochastic ball model [5,13] and learning of centers for the Gaussian mixture model [20]. The price to obtain such information, the requirement imposed upon the minimum center separation, is the homogeneity of the criteria forced on all different clusters. In other words, each pair of clusters, regardless of their shapes and cardinalities, must have their centers separated by a uniform distance determined by the *entire data set*. As a consequence of this "global" condition, the effect of an isolated but huge cluster ripples throughout the entire data set by raising the minimum center separation. Thus, a more "localized" condition, i.e., a condition on the center separation for each pair of clusters that relies largely on local information, is much desired. Such a more localized condition might pave the way to address the aforementioned fundamental questions regarding the Peng–Wei relaxation.

To that end, in this paper we introduce a proximity condition enabling us to relate the pairwise center distances to more localized quantities. Interestingly, it turns out that our proximity condition improves the one in [15] and is comparable to that in [6], the state-of-the-art proximity conditions in the literature of SVD-based projective *k*-means. Furthermore, under the Amini–Levina relaxation for clusters of equal cardinality, the associated proximity condition becomes even "fully localized", as it *only* involves information about pairs of clusters.

1.1 Organization of our paper

Our paper is organized as follows. In the remainder of this introductory section we present our aforementioned proximity condition, discuss its implication for various stochastic cluster models and briefly compare our results to the state of the art. In Sect. 2, we discuss *k*-means and its convex relaxation introduced by Peng and Wei. In Sect. 3, we show that the Peng–Wei relaxation yields the solution of the *k*-means objective as long as our proximity condition (1.1) is satisfied. A different proximity condition for the exactness of Amini–Levina relaxation is discussed in the same section. In 4, we consider the application of our framework to the stochastic ball model and the Gaussian mixture model. Numerical simulations that illustrate our theoretical findings are presented in Sect. 5. All proofs can be found in Sects. 6–8.

1.2 Proximity conditions under deterministic models

The idea of proximity conditions originates from the work [15] by Kumar and Kannan who use a proximity condition to characterize the performance of Lloyd's algorithm with an initialization given by an SVD-based projection under deterministic models. The result is later improved by Awasthi and Sheffet [6], who perform a finer analysis and redesign the proximity condition for the same algorithm. To the best of our knowledge, no such type of proximity conditions has been established for the Peng–Wei relaxation so far, and we will fill this gap in this paper.

Conceptually speaking, our proximity condition can be interpreted as follows:

For each pair of clusters, every point is closer to the center of its own cluster, while the bisector hyperplane of the centers keeps all points in the two clusters at a certain distance determined by global information of the data set.

Roughly speaking, the proximity condition characterizes for each pair of clusters how much closer each point is to the within-cluster center than the cross-cluster center. This is conceptually much more localized than minimum separation, which compares all pairwise center distances to a uniform quantity.

Let us introduce some necessary notation before we proceed to the exact statement of our proximity condition. Given a set of N data points $\Gamma = \{\boldsymbol{x}_l\}_{l=1}^N$ with k mutually disjoint clusters $\Gamma = \sqcup_{a=1}^k \Gamma_a$, we can re-index $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ according to the clusters: $\Gamma_a = \{\boldsymbol{x}_{a,i}\}_{1 \le i \le n_a}$ for all $1 \le a \le k$. Denote by $n_a = |\Gamma_a|$ the number of elements in Γ_a.

Denote the data matrix of the ath cluster by

$$\boldsymbol{X}_a^\top = \begin{bmatrix} \boldsymbol{x}_{a,1} & \ldots & \boldsymbol{x}_{a,n_a} \end{bmatrix} \in \mathbb{R}^{m \times n_a}.$$

Furthermore, define

$$\boldsymbol{c}_a = \frac{1}{n_a} \sum_{i=1}^{n_a} \boldsymbol{x}_{a,i}, \quad \boldsymbol{w}_{a,b} = \frac{\boldsymbol{c}_b - \boldsymbol{c}_a}{\|\boldsymbol{c}_b - \boldsymbol{c}_a\|}, \quad \text{and} \quad \overline{\boldsymbol{X}}_a = \boldsymbol{X}_a - \mathbf{1}_{n_a} \boldsymbol{c}_a^\top.$$

In other words, $\boldsymbol{c}_a$ is the sample mean (cluster center) of the ath cluster, $\boldsymbol{w}_{a,b}$ is the unit vector pointing from $\boldsymbol{c}_a$ to $\boldsymbol{c}_b$, and $\overline{\boldsymbol{X}}_a$ is the centered data matrix of the ath cluster. Now we are ready to give a mathematical characterization of the proximity condition.

Condition 1 (Proximity condition) *The partition $\Gamma = \sqcup_{a=1}^k \Gamma_a$ satisfies the proximity condition if for any $a \neq b$, there holds*

$$\min_{1 \le i \le n_a} \left\langle \boldsymbol{x}_{a,i} - \frac{\boldsymbol{c}_a + \boldsymbol{c}_b}{2}, \boldsymbol{w}_{b,a} \right\rangle > \frac{1}{2} \sqrt{\left(\sum_{l=1}^k \|\overline{\boldsymbol{X}}_l\|^2 \right) \left(\frac{1}{n_a} + \frac{1}{n_b} \right)}. \tag{1.1}$$

Here, $\|\overline{\boldsymbol{X}}_l\|$ is the operator norm of the matrix $\overline{\boldsymbol{X}}_l$.

The proximity condition has a very intuitive geometric interpretation, see also Fig. 1. Suppose the partition of data points satisfies the proximity condition. Then each pair of clusters Γ_a and Γ_b can be separated by a plane through the bisector of their sample means $\boldsymbol{c}_a$ and $\boldsymbol{c}_b$. Moreover, the distance between every point in those two clusters and the bisector must be greater than the right hand side of (1.1). This geometric interpretation can be further illustrated by rewriting (1.1): Denote by $h_{a,b} = \|\boldsymbol{c}_a - \boldsymbol{c}_b\|$ the distance between the two centers $\boldsymbol{c}_a$ and $\boldsymbol{c}_b$. Moreover, define

$$\tau_{a,b} = \max\{\max(\boldsymbol{u}_{a,b}), \max(\boldsymbol{u}_{b,a})\} \quad \text{where} \quad \boldsymbol{u}_{a,b} = \overline{\boldsymbol{X}}_a \boldsymbol{w}_{a,b} \text{ for } 1 \le a, b \le k.$$

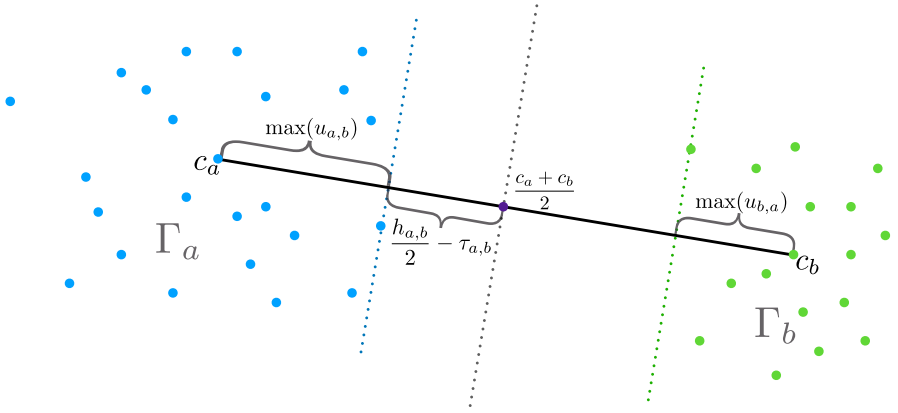

Fig. 1 Proximity condition: If the partition of data points satisfies the proximity condition, then each pair of clusters Γ_a and Γ_b can be separated by a plane through the bisector of their sample means $\boldsymbol{c}_a$ and $\boldsymbol{c}_b$, and the distance between each individual point in those two clusters and the bisector is greater than the right hand side of (1.1)

Clearly, $\tau_{a,b}$ is the maximum signed projection distance over all the data points in the clusters Γ_a and Γ_b. As illustrated in Fig. 1, one can easily check that the left hand side of proximity condition (1.1) is in fact equal to $\frac{1}{2}h_{a,b} - \tau_{a,b}$ which is the shortest distance between the midpoint $\frac{\boldsymbol{c}_a+\boldsymbol{c}_b}{2}$ and the projections of all the data points in Γ_a and Γ_b on the line connecting $\boldsymbol{c}_a$ and $\boldsymbol{c}_b$. This observation gives us the following proposition.

Proposition 1 *The proximity condition* (1.1) *is equivalent to*

$$h_{a,b} > 2\tau_{a,b} + \sqrt{\sum_{l=1}^{k} \|\overline{\boldsymbol{X}}_l\|^2 \left(\frac{1}{n_a} + \frac{1}{n_b}\right)}, \quad \forall a \neq b. \tag{1.2}$$

Besides showing that the proximity condition (1.1) guarantees the exactness of Peng–Wei relaxation, we also obtain a necessary proximity condition. If a deterministic mixture fails to fulfill the necessary condition, exact recovery by the Peng–Wei relaxation is provably impossible.

Awasthi and Sheffet's has raised an open question in [6]: can the pairwise separation condition be fully localized, i.e., depend only on information of the corresponding pair of clusters? We apply the Amini and Levina's relaxation [3], originally intended to address the weak assortativity issue in community detection among networks, to convexify the k-means problem in the case of balanced clusters. Surprisingly, we end up with a completely localized proximity condition for the exactness of the convex relaxation, thus solving Awasthi and Sheffet's open problem for the balanced case.

Furthermore, beyond the scope of the Peng–Wei relaxation of k-means, the proximity condition itself provides an algorithm that can accept answers to the NP-hard k-means problem (although it is not able to reject an answer). For a given solution to k-means, one can simply check whether the proximity condition holds, and if it

does hold, then the solution is provably the unique global minimum. The time cost is proportional to $\mathcal{O}(kN + m^2 N)$. Assuming the number of clusters k and the dimension of data m are fixed, the time complexity is linear in the total number of points N, which improves the quasilinear-time algorithm proposed in [13] in terms of the time complexity.

1.3 Comparison to existing proximity conditions in the literature

As mentioned before, in the literature of projective k-means, proximity conditions have been proposed in [15] and later improved in [6]. In this section we compare our proximity conditions with these existing results.

Denote $\overline{\boldsymbol{W}} = [\overline{\boldsymbol{X}}_1^\top, \ldots, \overline{\boldsymbol{X}}_k^\top]^\top$. By our notation, the original Kumar-Kannan proximity condition [15] is equivalent to

$$h_{a,b} > 2\tau_{a,b} + Ck\left(\frac{1}{\sqrt{n_a}} + \frac{1}{\sqrt{n_b}}\right)\|\overline{\boldsymbol{W}}\|, \quad \forall a \neq b,$$

for some large absolute constant $C > 0$. The fact that $\max_{1\le l\le k}\|\overline{\boldsymbol{X}}_l\| \le \|\overline{\boldsymbol{W}}\|$ implies $\sqrt{\sum_{l=1}^k \|\overline{\boldsymbol{X}}_l\|^2} \le \sqrt{k}\|\overline{\boldsymbol{W}}\|$. Therefore, our proximity condition (1.2) is strictly weaker than the Kumar-Kannan condition by at least a factor of $\sqrt{k}$.

The comparison between (1.1) and the Awasthi-Sheffet conditions in [6] is less straightforward. Theorem 4 therein states that consistent clustering is guaranteed by projective k-means plus Lloyd's algorithm as long as

$$h_{a,b} > \max\left\{2\tau_{a,b} + C\left(\frac{1}{\sqrt{n_a}} + \frac{1}{\sqrt{n_b}}\right)\|\overline{\boldsymbol{W}}\|,\ C\sqrt{k}\left(\frac{1}{\sqrt{n_a}} + \frac{1}{\sqrt{n_b}}\right)\|\overline{\boldsymbol{W}}\|\right\} \quad \forall a \neq b. \tag{1.3}$$

Compared to our proximity condition (1.1), the second term on the right-hand side of 1.3 could be more stringent given the fact $\sqrt{\sum_{l=1}^k \|\overline{\boldsymbol{X}}_l\|^2} \le \sqrt{k}\|\overline{\boldsymbol{W}}\|$, whereas the first term is less stringent than ours since

$$\|\overline{\boldsymbol{W}}\|^2 = \|\overline{\boldsymbol{W}}^\top\overline{\boldsymbol{W}}\| = \left\|\sum_{a=1}^k \overline{\boldsymbol{X}}_a^\top\overline{\boldsymbol{X}}_a\right\| \le \sum_{a=1}^k \left\|\overline{\boldsymbol{X}}_a^\top\overline{\boldsymbol{X}}_a\right\| = \sum_{a=1}^k \left\|\overline{\boldsymbol{X}}_a\right\|^2.$$

Therefore, it is fair to say our proximity condition is comparable to the Awasthi-Sheffet condition.

1.4 Implications under stochastic models

We should emphasize that in order to prove our main results, we benefit a lot from the existing primal-dual analyses in [5,13]. The major difference between our analysis and [5,13] is that we aim at deriving proximity conditions under deterministic models rather than establishing minimum separation results under stochastic models.

However, we are still curious about what minimum separation conditions our proximity condition can yield when applied to both the stochastic ball model and the Gaussian mixture model. Before presenting conditions given by our proximity condition, we first review the state-of-the-art results on both models.

Existing work on the Peng–Wei relaxation: The stochastic ball model can be viewed as a special case of mixture models where the distributions of sample data points are compactly supported on k disjoint unit balls in $\mathbb{R}^m$. The clusters are balanced and the covariance structure is fairly rigid since all the distributions are assumed to be identical and isotropic.

Let Δ be the minimal separation between the cluster centers. In [5], it is proven that the Peng–Wei relaxation achieves exact recovery provided $\Delta > 2\sqrt{2}(1+1/\sqrt{m})$, where the lower bound of Δ is independent of the number of clusters k. Another bound of Δ is given in [13] stating that exact recovery is guaranteed if $\Delta > 2 + k^2/m$ which is near-optimal in the $m \gg k^2$ regime.

The Gaussian mixture model (GMM) as a stochastic model is more flexible. This model is characterized by its density function which is a weighted sum of the density functions of Gaussian or subgaussian distributions. In [20], assuming the Gaussian distributions are identical and isotropic, Mixon, Villar and Ward prove that the Peng–Wei relaxation learns the Gaussian centers for balanced clusters when the center separations are required to be above $k\sigma$, where $\sigma \boldsymbol{I}$ is the common covariance of all Gaussian distributions.

Existing work on other algorithms: Clustering Gaussian mixture models has received extensive attention in machine learning and statistics communities. Besides [20], a lot of progress has been made in developing efficient algorithms for this task. Among them are a family of algorithms here referred to as the projective k-means [1,6,9,14,15,18,25]. In general, the projective k-means works in two steps: first project all the data points onto a lower dimensional space usually based on singular value decomposition (SVD), and then classify each point by heuristic methods such as single linkage clustering in [1] or Lloyd's algorithm in [6].

Vempala and Wang [25] show that if each pairwise center separation is larger than a quantity determined by the number of clusters k, the dimension m and the variances of the clusters, the projective algorithm can classify a mixture of k isotropic Gaussians with high probability. Achlioptas and McSherry [1] show that SVD-based projection followed by single-linkage clustering is able to classify all the sampled data points accurately if the center separation of each pair of clusters is greater than the operator norm of the covariance matrix and the weights of the two clusters plus a term which depends on the concentration properties of the distributions in the mixture. The algorithm studied by Kannan and Kumar in [15]—the work that first devises the idea of *proximity condition*—also begins with an SVD-based projection and proceeds by Lloyd's algorithm which is initialized by an unspecified near-optimal solution to the k-means problem. As stated before, its technical results are improved by Awatshi and Sheffet in [6]. Recently, Lu and Zhou [18] provide a more detailed estimation of misclassification rate for each iteration of Lloyd's algorithm with initialization given by spectral methods [14].

Table 1 Comparison of results on GMM: the separation bound for [25] only applies to mixtures of isotropic Gaussian distributions and the bound for [20] is used to guarantee learning cluster centers instead of recovering the labels of data points

Authors	Separation bounds	Algorithms	Exact	Year
Vempala and Wang [25]	$\mathcal{O}(k^{1/4}\log^{1/4}(m))$	Projective k-means	Yes	2004
Achlioptas and McSherry [1]	$\mathcal{O}(k + k^{1/2}\log^{1/2} N)$	Projective k-means	Yes	2005
Kumar and Kannan [15]	$\mathcal{O}(k(\text{polylog}(N)))$	Projective k-means	Yes	2010
Awasthi and Sheffet [6]	$\mathcal{O}(k^{1/2}(\text{polylog}(N)))$	Projective k-means	Yes	2012
Lu and Zhou [18]	$\mathcal{O}(k^{3/2})$	Projective k-means	No	2016
Mixon et al. [20]	$\mathcal{O}(k)$	SDP k-means	No	2017
Our work	$\mathcal{O}(k^{1/2} + \log^{1/2}(kN))$	SDP k-means	Yes	–

Our results: We can easily apply the proximity condition to the stochastic ball model and the Gaussian mixture model. The corresponding recovery guarantees are competitive with or improve upon other state-of-the-art results.

- For the stochastic ball model, we show that $\Delta > 2 + \mathcal{O}(\sqrt{k/m})$ is sufficient to guarantee the exact recovery of the Peng–Wei relaxation, which improves the separation condition $\Delta > 2 + k^2/m$ in [13] when k is large. Moreover, our result applies to a broader class of stochastic ball models where each cluster can have a different number of points and may even satisfy a different probability distribution as long as the support of density function is contained within a unit ball.
- For the Gaussian mixture model, we summarize our result for the Peng–Wei relaxation and other state-of-the-art results for both the Peng–Wei relaxation and projective k-means in Table 1. It has been shown in [20] that the centers of a Gaussian mixture can be accurately estimated by Peng–Wei relaxation provided the minimal separation is $\mathcal{O}(k)$. In contrast, our proximity provides a different minimal separation condition $\mathcal{O}(k^{1/2} + \log^{1/2}(kN))$, which is smaller than $\mathcal{O}(k)$ if k is large and N not too large. Our separation condition is better than [15] and comparable to [6] for projective k-means. Though our bound loses a $k^{1/4}$ factor vis-à-vis the one in [25] for the special case of spherical Gaussian mixtures, we can handle more general Gaussian mixtures where the density functions do not have to be spherical or identical.

1.5 Notation

Let $\mathbf{1}_{\Gamma_a}$ be the indicator vector of $\Gamma_a \subseteq \Gamma$. $\mathbf{1}_n$ is an $n \times 1$ vector with all entries equal to 1. Given any two real matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in $\mathbb{R}^{m\times n}$, we define the inner product as $\langle \boldsymbol{U}, \boldsymbol{V}\rangle = \text{Tr}(\boldsymbol{U}\boldsymbol{V}^\top) = \sum_{i=1}^m \sum_{j=1}^n U_{ij}V_{ij}$. For a vector $\boldsymbol{v}$, $\max(\boldsymbol{v})$ is equal to the largest entry of $\boldsymbol{v}$. We denote $\boldsymbol{Z} \geq 0$ if $\boldsymbol{Z}$ is a nonnegative matrix, i.e., each entry is nonnegative; $\boldsymbol{Z} \succeq 0$ if $\boldsymbol{Z}$ is a symmetric positive semi-definite matrix. Besides, we also use the notation listed below throughout the paper.

m Dimension of data
k Number of clusters

Γ	Set of N data points in $\mathbb{R}^m$
Γ_a	The ath cluster
N	Total number of data points
n_a	Number of points in the ath cluster
$\mathcal{S}^N$	Set of $N \times N$ symmetric matrices
$\mathcal{S}^N_+$	Set of $N \times N$ positive semi-definite matrices
$\mathbb{R}^{N\times N}_+$	Set of $N \times N$ nonnegative matrices
$\boldsymbol{W}$	Data matrix of all N data points
$\boldsymbol{X}_a$	Data matrix of the ath cluster
$\overline{\boldsymbol{X}}_a$	Centered data matrix of the ath cluster
$\boldsymbol{D}$	Squared distance matrix
$\boldsymbol{X}$	Ground-truth solution to the SDP relaxation of k-means
$\boldsymbol{Y}^{(a,b)}$	Submatrix of any $N \times N$ matrix $\boldsymbol{Y}$ given by $\{y_{s,t}\}_{s\in\Gamma_a, t\in\Gamma_b}$
$\boldsymbol{x}_{a,i}$	The ith data point in the ath cluster
$\boldsymbol{\mu}_a$	Population mean of the ath cluster in a generative model
$\boldsymbol{c}_a$	Sample mean of the ath cluster
$\boldsymbol{w}_{a,b}$	Unit vector pointing from $\boldsymbol{c}_a$ to $\boldsymbol{c}_b$
$\boldsymbol{u}_{a,b}$	Signed projection distance given by $\boldsymbol{u}_{a,b} = \overline{\boldsymbol{X}}_a \boldsymbol{w}_{a,b}$
$h_{a,b}$	Distance between $\boldsymbol{c}_a$ and $\boldsymbol{c}_b$
$\tau_{a,b}$	Maximum signed projection distance determined by $\boldsymbol{u}_{a,b}$ and $\boldsymbol{u}_{b,a}$

2 k-Means and the Peng–Wei relaxation

In this section, we briefly review the formulation of k-means and its SDP relaxation introduced by Peng and Wei [21]. Let $\Gamma = \{\boldsymbol{x}_l\}_{l=1}^N$ be a set of N data points in $\mathbb{R}^m$. k-means attempts to divide Γ into k disjoint clusters by seeking a solution to the following minimization problem:

$$\min_{\{\Gamma_a\}_{a=1}^k} \min_{\{\boldsymbol{\gamma}_a\}_{a=1}^k} \sum_{a=1}^k \sum_{l\in\Gamma_a} \left\|\boldsymbol{x}_l - \boldsymbol{\gamma}_a\right\|^2,$$

where $\{\Gamma_a\}_{a=1}^k$ form a partition of Γ (i.e., $\sqcup_{a=1}^k \Gamma_a = \Gamma$ and $\Gamma_a \sqcap \Gamma_b = \emptyset$ if $a \neq b$). For any given partition $\{\Gamma_a\}_{a=1}^k$, choosing $\boldsymbol{\gamma}_a$ as the centroid $\boldsymbol{\gamma}_a = \boldsymbol{c}_a = \frac{1}{|\Gamma_a|}\sum_{j\in\Gamma_a} \boldsymbol{x}_j$ $(a = 1, \ldots, k)$ minimizes the objective function. Therefore, the k-means problem is equivalent to:

$$\min_{\{\Gamma_a\}_{a=1}^k} \sum_{a=1}^k \sum_{l\in\Gamma_a} \|\boldsymbol{x}_l - \boldsymbol{c}_a\|^2, \tag{2.1}$$

Given an arbitrary partition $\{\Gamma_a\}_{a=1}^k$ of Γ, let $\mathbf{1}_{\Gamma_a}$ $(a = 1, \ldots, k)$ be the indicator function of the ath cluster. That is,

$$\mathbf{1}_{\Gamma_a}(l) = \begin{cases} 1 & \text{if } l \in \Gamma_a, \\ 0 & \text{otherwise.} \end{cases}$$

A simple calculation can reveal that

$$\frac{1}{|\Gamma_a|}\sum_{l\in\Gamma_a,s\in\Gamma_a}\|\boldsymbol{x}_l-\boldsymbol{x}_s\|^2=2\sum_{l\in\Gamma_a}\|\boldsymbol{x}_l-\boldsymbol{\gamma}_a\|^2$$

and hence,

$$\begin{aligned}\sum_{a=1}^{k}\sum_{l\in\Gamma_a}\left\|\boldsymbol{x}_l-\boldsymbol{\mu}_a\right\|^2&=\frac{1}{2}\sum_{a=1}^{k}\frac{1}{|\Gamma_a|}\sum_{l\in\Gamma_a,s\in\Gamma_a}\|\boldsymbol{x}_l-\boldsymbol{x}_s\|^2\\&=\frac{1}{2}\sum_{a=1}^{k}\frac{1}{|\Gamma_a|}\langle\mathbf{1}_{\Gamma_a}\mathbf{1}_{\Gamma_a}^{\top},\boldsymbol{D}\rangle,\end{aligned}$$

where $\boldsymbol{D}\in\mathbb{R}^{N\times N}$ is the distance matrix with the (l,s)th entry being given by $\boldsymbol{D}_{l,s}=\|\boldsymbol{x}_l-\boldsymbol{x}_s\|^2$. Therefore, we can rewrite the k-means problem as

$$\begin{aligned}&\min\quad\langle\boldsymbol{Z},\boldsymbol{D}\rangle\\&\text{s.t.}\quad\boldsymbol{Z}=\sum_{a=1}^{k}\frac{1}{|\Gamma_a|}\mathbf{1}_{\Gamma_a}\mathbf{1}_{\Gamma_a}^{\top}\text{ with }\sqcup_{a=1}^{k}\Gamma_a=\Gamma\text{ and }\Gamma_a\sqcap\Gamma_b=\emptyset\text{ for }a\neq b.\end{aligned}\tag{2.2}$$

It is self-evident that (2.2) is a non-convex problem due to the combinatorial nature of the feasible set. Indeed, (2.2) is an NP-hard problem [2]. Despite this, it can be easily verified that $\boldsymbol{Z}=\sum_{a=1}^{k}\frac{1}{|\Gamma_a|}\mathbf{1}_{\Gamma_a}\mathbf{1}_{\Gamma_a}^{\top}$ satisfies the following four properties:

$$\boldsymbol{Z}\succeq 0,\quad\boldsymbol{Z}\geq 0,\quad\boldsymbol{Z}\mathbf{1}_N=\mathbf{1}_N,\quad\mathrm{Tr}(\boldsymbol{Z})=k.$$

Replacing the constraint in (2.2) by the above four properties leads to the SDP relaxation of k-means introduced by Peng and Wei in [21],

$$\begin{aligned}&\min\quad\langle\boldsymbol{Z},\boldsymbol{D}\rangle\\&\text{s.t.}\quad\boldsymbol{Z}\succeq 0,\quad\boldsymbol{Z}\geq 0,\quad\boldsymbol{Z}\mathbf{1}_N=\mathbf{1}_N,\quad\mathrm{Tr}(\boldsymbol{Z})=k,\end{aligned}\tag{2.3}$$

which will be the focus of this paper.

The Peng–Wei relaxation is a convex problem and can be solved in polynomial time using the interior-point method [27]. We denote by $\boldsymbol{X}$ the optimal solution to the Peng–Wei relaxation. Clearly, every feasible point of (2.2) is also feasible for (2.3); so once the optimal solution to (2.3) has the form $\boldsymbol{X}=\sum_{a=1}^{k}\frac{1}{|\Gamma_a|}\mathbf{1}_{\Gamma_a}\mathbf{1}_{\Gamma_a}^{\top}$, it must be an optimal solution to the k-means problem. Therefore, the question of central importance is:

When is the solution to (2.3) *of the form* $\boldsymbol{X}=\sum_{a=1}^{k}\frac{1}{|\Gamma_a|}\mathbf{1}_{\Gamma_a}\mathbf{1}_{\Gamma_a}^{\top}$?

3 Exact recovery guarantees

3.1 Exact clustering and proximity conditions

In a nutshell our following main theorem states that the proximity condition (1.1) implies the exactness of the Peng–Wei relaxation (2.3):

Theorem 2 (Main theorem) *Suppose the partition* $\{\Gamma_a\}_{a=1}^k$ *obeys the proximity condition* (1.1). *Then the minimizer of the Peng–Wei relaxation* (2.3) *is unique and given by* $\boldsymbol{X} = \sum_{a=1}^k \frac{1}{|\Gamma_a|} \mathbf{1}_{\Gamma_a} \mathbf{1}_{\Gamma_a}^\top$.

Since the global minimum of (2.3) is always smaller than that of (2.1), Theorem 2 implies that the proximity condition provides a simple algorithm that is able to accept answers to the k-means problem.

Corollary 1 (Algorithm accepting answers to k-means) *If a partition* $\Gamma = \sqcup_{a=1}^k \Gamma_a$ *satisfies the proximity condition (1), then it is the unique global minimum to the k-means objective function.*

Note that each data point $\boldsymbol{x}_{a,i}$ appears $k-1$ times on the left hand side of (1), and it takes $\mathcal{O}(m^2 n_a)$ amount of time to compute each matrix operator norm using the Golub-Reisch SVD algorithm [11]. Thus, the time cost to examine the proximity condition is proportional to $\mathcal{O}(kN + m^2 N)$.

To the best of our knowledge, k-means problem has not been shown in NP or not. The proximity condition does not change this fact. We want to emphasize that the polynomial time examination of the proximity condition (1) does not imply that an answer to the k-means problem can be verified in polynomial time since it does not accept all correct answers. A different approach that leverages the dual certificate associated with the Peng–Wei relaxation to test under certain conditions the optimality of a candidate k-means solution can be found in [13]. The algorithm proposed in [13] tests the optimality of a candidate solution in quasilinear time. Hence, our method improves the time complexity by a logarithmic factor.

While the main theorem provides a sufficient condition for the Peng–Wei relaxation to exactly recover a given partition, the following theorem gives a necessary condition.

Theorem 3 (Necessary condition) *Suppose* $\boldsymbol{X} = \sum_{a=1}^k \frac{1}{|\Gamma_a|} \mathbf{1}_{\Gamma_a} \mathbf{1}_{\Gamma_a}^\top$ *is a global minimum of* (2.3). *Then the partition* $\{\Gamma_a\}_{a=1}^k$ *must satisfy*

$$h_{a,b} \geq \tau_{a,b} + \sqrt{\tau_{a,b}^2 + \max_t \|\overline{\boldsymbol{X}}_t\|^2 \left(\frac{1}{n_a} + \frac{1}{n_b}\right)}, \quad \forall a \neq b. \tag{3.1}$$

Notice that as long as $\boldsymbol{X}$ is a solution to (2.3), $\{\Gamma_a\}_{a=1}^k$ must be a global minimum to the k-means. In other words, it is harder for a deterministic mixture to be exactly recovered by the Peng–Wei relaxation than being the global minimum to the k-means. It remains unclear whether this necessary condition (Theorem 3) is only necessary for the Peng–Wei relaxation or is necessary for the k-means itself as well.

3.2 Balanced case: Amini–Levina relaxation and proximity condition

One special case of interest is the balanced case where each cluster has the same number of points, i.e. $|\Gamma_1| = \cdots = |\Gamma_k| = n$. We have seen in Section 2 that the k-means problem can be rewritten as (2.2):

$$\begin{aligned} &\min \quad \langle \boldsymbol{Z}, \boldsymbol{D} \rangle \\ &\text{s.t.} \quad \boldsymbol{Z} = \sum_{a=1}^{k} \frac{1}{|\Gamma_a|} \mathbf{1}_{\Gamma_a} \mathbf{1}_{\Gamma_a}^{\top} \text{ with } \sqcup_{a=1}^{k} \Gamma_a = \Gamma \text{ and } \Gamma_a \sqcap \Gamma_b = \emptyset \text{ for } a \neq b. \end{aligned} \tag{3.2}$$

With the balanced assumption, i.e., the cardinalities of all clusters being the same, it is easy to verify that $\boldsymbol{Z} = \sum_{a=1}^{k} \frac{1}{n} \mathbf{1}_{\Gamma_a} \mathbf{1}_{\Gamma_a}^{\top}$ obeys the following four constraints:

$$\boldsymbol{Z} \succeq 0, \quad \boldsymbol{Z} \geq 0, \quad \boldsymbol{Z}\mathbf{1}_N = \mathbf{1}_N, \quad \operatorname{diag}(\boldsymbol{Z}) = \frac{1}{n}\mathbf{1}_N.$$

This leads to the Amini–Levina relaxation of k-means, which was first introduced in [3] for community detection under balanced case in order to address the weak assortativity issue:

$$\begin{aligned} &\min \quad \langle \boldsymbol{Z}, \boldsymbol{D} \rangle \\ &\text{s.t.} \quad \boldsymbol{Z} \succeq 0, \quad \boldsymbol{Z} \geq 0, \quad \boldsymbol{Z}\mathbf{1}_N = \mathbf{1}_N, \quad \operatorname{diag}(\boldsymbol{Z}) = \frac{1}{n}\mathbf{1}_N. \end{aligned} \tag{3.3}$$

As with the analyses on the Peng–Wei relaxation, once the optimal solution to (3.3) takes the form $\boldsymbol{X} = \sum_{a=1}^{k} \frac{1}{n} \mathbf{1}_{\Gamma_a} \mathbf{1}_{\Gamma_a}^{\top}$, the Amini–Levina relaxation gives an optimal solution to the k-means problem with balanced assumption. Once again, we ask the same question for Peng and Wei's relaxation: *When is the solution to* (3.3) *of the form* $\boldsymbol{X} = \sum_{a=1}^{k} \frac{1}{n} \mathbf{1}_{\Gamma_a} \mathbf{1}_{\Gamma_a}^{\top}$?

Unsurprisingly, the answer is another proximity condition specially tailored for Amini and Levina's relaxation.

Condition 4 (Proximity condition for balanced clusters) *A partition* $\Gamma = \sqcup_{a=1}^{k} \Gamma_a$ *with* $|\Gamma_1| = \cdots = |\Gamma_k| = n$ *satisfies the proximity condition for balanced clusters if for any* $a \neq b$, *there holds*

$$\min_{1 \leq i \leq n_a} \left\langle \boldsymbol{x}_{a,i} - \frac{\boldsymbol{c}_a + \boldsymbol{c}_b}{2}, \boldsymbol{w}_{b,a} \right\rangle > \sqrt{\frac{k}{4n} \left(\|\overline{\boldsymbol{X}}_a\|^2 + \|\overline{\boldsymbol{X}}_b\|^2 \right)}. \tag{3.4}$$

Similar to the general case, the proximity condition for balanced clusters also has an equivalent formulation:

$$h_{a,b} > 2\tau_{a,b} + \sqrt{\frac{k}{n} \left(\|\overline{\boldsymbol{X}}_a\|^2 + \|\overline{\boldsymbol{X}}_b\|^2 \right)}. \tag{3.5}$$

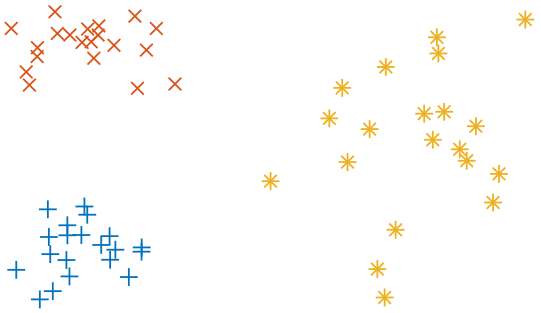

Fig. 2 An example of three clusters in the plane. Each contains 20 points. The proximity for the general case (1.1) fails for this instance. However, the proximity condition for balanced clusters (3.4) is satisfied and hence ensures the partition is optimal to the k-means problem with balanced assumption

Theorem 5 (Exact recovery for balanced clusters) *Suppose the partition* $\{\Gamma_a\}_{a=1}^k$ *with* $|\Gamma_1| = \cdots = |\Gamma_k| = n$ *obeys the proximity condition for balanced clusters* (3.4)*. Then the minimizer of the Amini–Levina relaxation* (3.3) *is unique and given by* $\boldsymbol{X} = \sum_{a=1}^k \frac{1}{n} \mathbf{1}_{\Gamma_a} \mathbf{1}_{\Gamma_a}^\top$*. Therefore, the partition* $\{\Gamma_a\}_{a=1}^k$ *can be recovered exactly by the Amini–Levina relaxation.*

Compared with the proximity condition for Peng and Wei's relaxation (1.1), the proximity condition for Amini and Levina's relaxation distinguishes itself by decoupling the clusters in the sense that each of the $k(k-1)$ inequalities in (3.4) only depends on the two clusters involved in the inequality. In the case of balanced clusters, this immediately solves the open question posed by Awasthi and Sheffet [6], which asks if such a proximity condition exists.

The completely localized proximity condition is particularly meaningful when there are a few abnormal clusters whose covariance matrices are huge in matrix operator norm, but at the same time being away from all the other clusters. In this case, the proximity condition for Amini and Levina's relaxation has far better chance than that for Peng and Wei's relaxation to detect a reasonable partition of the data set. Figure 2 provides such an example.

Analogously, we can also prove a necessary condition for the Amini–Levina relaxation, which can be compared with Theorem 3 for the general case.

Theorem 6 (Necessary condition for balanced clusters) *Suppose* $\boldsymbol{X} = \sum_{a=1}^k \frac{1}{|\Gamma_a|} \mathbf{1}_{\Gamma_a} \mathbf{1}_{\Gamma_a}^\top$ *is a global minimum of* (3.3)*. Then the partition* $\{\Gamma_a\}_{a=1}^k$ *must satisfy*

$$h_{a,b} \geq \tau_{a,b} + \sqrt{\tau_{a,b}^2 + \frac{1}{n}\left(\|\overline{\boldsymbol{X}}_a\|^2 + \|\overline{\boldsymbol{X}}_b\|^2\right)}, \quad \forall a \neq b. \tag{3.6}$$

4 Results under random models

Next we apply the proximity condition (1.1) to data sets generated from the generalized stochastic ball model and the Gaussian mixture model, respectively. We first give a formal definition for each model and then present the minimal separation condition which is sufficient to guarantee the exact recovery of underlying clusters by the Peng–

Wei relaxation. The minimal separation conditions are established by verifying the proximity condition (1) for those two random models. For proofs, see Sects. 8.2 and 8.3.

4.1 Stochastic ball model

The definition of generalized stochastic ball model is given as follows where we only assume the support of the density function is contained in the unit ball of $\mathbb{R}^m$ for all clusters.

Definition 1 (*Generalized stochastic ball model*) Let $\{\boldsymbol{\mu}_a\}_{a=1}^k$ be a set of k deterministic vectors in $\mathbb{R}^m$. For each $1 \le a \le k$, $\mathcal{D}_a$ is a distribution supported on the unit ball of $\mathbb{R}^m$ with a covariance matrix $\boldsymbol{\Sigma}_a$ and $\{\boldsymbol{r}_{a,i}\}_{i=1}^{n_a}$ are i.i.d. zero-mean random vectors drawn from the distribution $\mathcal{D}_a$. The ath cluster is formed by $\{\boldsymbol{x}_{a,i}\}_{i=1}^{n_a}$, where $\boldsymbol{x}_{a,i} = \boldsymbol{\mu}_a + \boldsymbol{r}_{a,i}$ for $1 \le i \le n_a$.

Corollary 2 *Denote* $\sigma_{\max}^2 = \max_{1\le a\le k} \|\boldsymbol{\Sigma}_a\|$, $N = \sum_{a=1}^k n_a$, $w_{\min} = \frac{1}{N}\min_{1\le a\le k} n_a$, *and* $\Delta = \min_{a\ne b} \|\boldsymbol{\mu}_a - \boldsymbol{\mu}_b\|$. *For the generalized stochastic ball model, we draw* n_a *points from the* a*th ball for each* $1 \le a \le k$. *The Peng–Wei relaxation achieves exact recovery with probability at least* $1 - N^{-\gamma}$ *if* $N \ge \frac{4}{w_{\min}} \log(4kmN^{\gamma})$ *and*

$$\Delta \ge 2 + \sqrt{\frac{2}{w_{\min}}}\sigma_{\max} + 7\sqrt{\frac{t}{w_{\min}}}, \tag{4.1}$$

where $t = \sqrt{\frac{4\log(4kmN^{\gamma})}{Nw_{\min}}}$ *and* $\gamma > 0$. *In particular, if* $n_a = n$ *for all* a, $w_{\min} = \frac{1}{k}$ *and each* $\mathcal{D}_a$ *is a uniform distribution over the unit ball of* $\mathbb{R}^m$, *then* (4.1) *can be simplified to*

$$\Delta \ge 2 + \sqrt{\frac{2k}{m+2}} + 7\sqrt{tk}$$

by noting that $\sigma_{\max}^2 = \|\boldsymbol{\Sigma}_a\| = \frac{1}{m+2}$.

Remark 1 As the number of data points N goes to infinity provided k and $w_{\min}$ are fixed, the value of $t = \sqrt{\frac{4\log(4kmN^{\gamma})}{Nw_{\min}}}$ vanishes. So asymptotically the minimal separation condition reduces to $\Delta > 2 + \sqrt{\frac{2k}{m+2}}$ when $n_a = n$ and $\boldsymbol{\Sigma}_a = \frac{1}{m+2}\boldsymbol{I}_m$. Note that we only assume that the distribution is supported on the unit ball, so rotation-invariant distributions which are assumed in [12,13] are also included. Compared with the result in [12,13] where $\Delta > 2 + \frac{k^2}{m}$ is required, we have achieved a better bound when k is large.

We can also apply the necessary lower bound (Theorem 3) to the generalized stochastic ball model. To illustrate this, let us study a special case where the following Corollary holds.

Corollary 3 *For the generalized ball model, if for all* $1 \le a \le k$ *we have* $n_a = n$, *then with high probability, the Peng–Wei relaxation fails to achieve exact recovery provided that* N *is large enough and*

$$\Delta < 1 + \sqrt{1 + 2\sigma_{\max}^2}.$$

If for any a, $\mathcal{D}_a$ is the uniform distribution over the unit ball, the bound becomes

$$\Delta < 1 + \sqrt{1 + \frac{2}{m+2}}.$$

4.2 Gaussian mixture model

The definition of Gaussian mixture model is given below, followed by the minimal separation condition for the exactness of the Peng–Wei relaxation.

Definition 2 (*Gaussian mixture model*) Consider a mixture of k Gaussian distributions $\mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$ in $\mathbb{R}^m$ with a set of weights $\{w_a\}_{a=1}^k$ obeying $w_a \geq 0$ and $\sum_{a=1}^k w_a = 1$. The probability density function of this mixture model is

$$p(\boldsymbol{x}) = \sum_{a=1}^{k} w_a \, p_{\mathcal{N}}(\boldsymbol{x}; \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a), \quad \boldsymbol{x} \in \mathbb{R}^m,$$

where $p_{\mathcal{N}}(\boldsymbol{x}; \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$ is the probability density function of the Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$.

Corollary 4 *Denote $\sigma_{\max}^2 = \max_{1\leq a\leq k}\{\|\boldsymbol{\Sigma}_a\|\}$, $w_{\min} = \min_{1\leq a\leq k}\{w_a\}$ and $\Delta = \min_{a\neq b}\|\boldsymbol{\mu}_a - \boldsymbol{\mu}_b\|$. For the Gaussian mixture model, the Peng–Wei relaxation achieves exact recovery with probability at least $1 - 6N^{-1}$ if*

$$\Delta \geq \sigma_{\max}\left(\frac{2}{\sqrt{w_{\min}}} + 4\sqrt{2}\log^{1/2}(kN^2) + q(N; m, k, w_{\min})\right),$$

where $q(N; m, k, w_{\min}) = o(1)$ if $N \gg m^2k^2\log(k)/w_{\min}$. In particular, if $n_a = n$ and $\boldsymbol{\Sigma}_a = \boldsymbol{I}_m$ for all $1 \leq a \leq k$, then the above condition reduces to

$$\Delta \geq 2\sqrt{k} + 4\sqrt{2}\log^{1/2}(kN^2) + q(N; m, k, 1/k),$$

and $q(N; m, k, 1/k) = o(1)$ if $N \gg m^2k^3\log(k)$.

5 Numerical experiments

Consider applying the Peng–Wei relaxation to the generalized stochastic ball model. When the total number of the data points N becomes large enough, the parameter t vanishes and the sufficient lower bound predicted by Corollary 2 as in (4.1) becomes

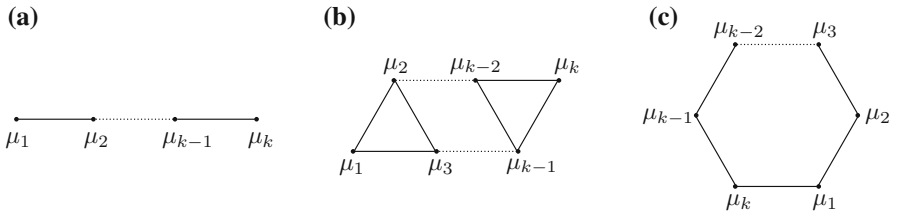

Fig. 3 Illustration of three instructive centroidal geometries. The minimal separation Δ is the distance between two adjacent centers. Our bound refers to (5.1) with parameters calculated for the given distribution. The state-of-the-art bound (5.2) is the bound proved by [5,13]

$$\Delta \geq 2 + \sigma_{\max}\sqrt{\frac{2}{w_{\min}}}. \tag{5.1}$$

The state-of-the-art bound for the stochastic ball model proved in [5,13] is

$$\Delta > \min\left\{2\sqrt{2}\left(1+\frac{1}{\sqrt{m}}\right), 2+\frac{k^2}{m}\right\}. \tag{5.2}$$

The exact phase transition bound, above which exact recovery can be achieved by the Peng–Wei relaxation of k-means, is smaller than both of the above sufficient lower bounds. As one would expect, the actual lower bound is hard to find in practice. The major difficulty occurs when the number of clusters k is greater than 2. In this case, when creating an instance of the stochastic ball model with prescribed minimal separation distance Δ, there are infinitely many possible ways to place the centers and this cannot be resolved by translation, rotation, and scaling. To address this, we investigate the worst case where centers are packed as compactly as possible while points in each cluster are chosen in the most scattered way. We have a better chance finding a more accurate lower bound under this arrangement.

Three instructive centroidal geometries, the geometries formed by the locations of the centers, are considered, and we call them circle-shaped geometry, line-shaped geometry, and hive-shaped geometry respectively. Centers are packed compactly under these shapes, especially the hive-shaped geometry. We can rescale the three geometries to change the minimal separation distance Δ. An illustration of these geometries formed by the locations of the centers is shown in Fig. 3.

We let the number of data points in each cluster be $n_a = 100$. Hence, the total number of points $N = 100k$. As a result, $w_{\min} = 1/k$. These n_a points are equispaced points on the unit circle centered at $\boldsymbol{\mu}_a$. The data points are chosen in this way since it maximizes the variance. Because the data is isotropic and the variance is equal to 1, we have $\sigma_{\max} = 1/\sqrt{m} = 1/\sqrt{2}$.

For k and m chosen above, we can see that our bound is an improvement to the state-of-the-art result. Overall, it is still a meaningful addition to the state-of-the-art result. Nevertheless, it is not yet tight. Figure 4 shows that the actual lower bound is almost independent of the parameter k, while our theory still relies on the assumption that $\Delta \geq 2 + \mathcal{O}(\sqrt{k/m})$.

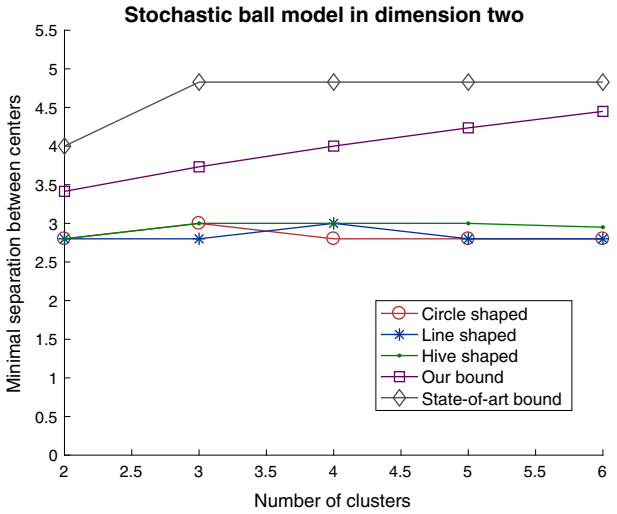

Fig. 4 Numerical experiment on the stochastic ball model with dimension 2 and number of clusters varying from 2 to 6. The sufficient lower bound here is the bound proved in Corollary 2. The Peng–Wei relaxation (SDP) is solved by SDPNAL+v0.5 (beta) [28,29]

Another parameter that may affect the bound is the dimension m. To reveal dependence of the bound on the dimension, we fix the number of clusters k to be 2 and let the dimension m vary between 2 and 10. The center separation Δ is chosen among 100 equispaced number between 2 and 4. The number of points in each cluster n_a is equal to $25 \times 2^{m-1}$, so there are $N = 50 \times 2^{m-1}$ in total. The distribution $\mathcal{D}_a$ for each ball is the uniform distribution on the unit sphere centered at $\boldsymbol{\mu}_a$. For any fixed pair of m and Δ, we generate 20 instances of the stochastic ball model.

From Fig. 5, it is evident that neither our bound nor the state-of-the-art bound is tight. The blue line, which represents the bound $\Delta \geq 2 + \frac{2}{m}$, fits our empircal result the best. Based on the observation of dependence between the empirical lower bound and the parameters k and m as in Figs. 4 and 5, we formulate a conjecture as stated below.

Conjecture 7 *For a mixture generated by the generalized stochastic ball model, the Peng–Wei relaxation achieves exact recovery with high probability if*

$$\Delta \geq 2 + \mathcal{O}\left(\frac{1}{m}\right), \tag{5.3}$$

provided that the total number of points N is large enough.

After the completion of this manuscript, a semidefinite relaxation based on graph cuts has been proposed in [16] to overcome the performance limits of Peng–Wei relaxation, which provides a new alternative way to learn the stochastic ball models.

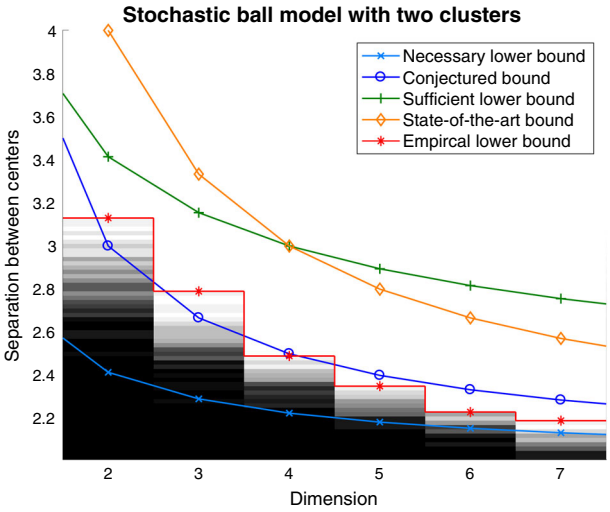

Fig. 5 Numerical experiment on the stochastic ball model with 2 clusters and dimension varying from 2 to 7. For given dimension and separation, the lighter the color is, the higher the probability of success is. The sufficient lower bound here is the bound given by Corollary 2, while the necessary lower bound is obtained by applying Theorem 3 directly to the stochastic ball model, which is $1 + \sqrt{1 + 2/m}$ in this case. Being constrained by computational resources, we are not able to sample more points in higher dimension since the time cost is prohibitive. This infers that the right half of the empirical lower bound is potentially smaller than the exact phase transition bound, which is what we are trying to approximate in this experiment. The Peng–Wei relaxation (SDP) is executed via SDPNAL+v0.5 (beta) [28,29]

6 Proofs for Section 3.1

We will prove the main theorem and related results under the proximity condition given in Proposition 1. The proof for the equivalence of the two proximity conditions is presented at the end of this section. The key ingredient in the proof of the main theorem is to construct a dual variable to certify the optimality of the desired solution $\boldsymbol{X} = \sum_{a=1}^{k} \frac{1}{|\Gamma_a|} \mathbf{1}_{\Gamma_a} \mathbf{1}_{\Gamma_a}^{\top}$ based on the conic duality theorem in convex optimization [7].

6.1 Conic duality

We first rewrite (2.3) as a cone program in standard form which naturally leads to its dual formulation. Noting that $\boldsymbol{Z}$ is a symmetric variable, the Peng–Wei relaxation of k-means (2.3) is equivalent to the following optimization problem:

$$\begin{aligned} \min \quad & \langle \boldsymbol{Z}, \boldsymbol{D} \rangle \\ \text{s.t.} \quad & \boldsymbol{Z} \succeq \mathbf{0}, \quad \boldsymbol{Z} \geq \mathbf{0}, \quad \frac{1}{2}(\boldsymbol{Z} + \boldsymbol{Z}^{\top})\mathbf{1}_N = \mathbf{1}_N, \quad \mathrm{Tr}(\boldsymbol{Z}) = k. \end{aligned} \tag{6.1}$$

Let $\mathcal{K} = \mathcal{S}_+^N \cap \mathbb{R}_+^{N \times N}$, the intersection of two self-dual cones: the positive semi-definite cone $\mathcal{S}_+^N$ and the nonnegative cone $\mathbb{R}_+^{N \times N}$. By definition, it is a pointed[1] and

[1] $\mathcal{K}$ is pointed if for $\mathbf{Z} \in \mathcal{K}$ and $-\mathbf{Z} \in \mathcal{K}$, $\mathbf{Z}$ must be $\mathbf{0}$, see Chapter 2 in [7].

closed convex cone with a nonempty interior. Moreover, its dual cone[2] is given by $\mathcal{K}^* = \mathcal{S}_+^N + \mathbb{R}_+^{N\times N} = \{\boldsymbol{B} + \boldsymbol{Q} : \boldsymbol{B} \geq \boldsymbol{0},\ \boldsymbol{Q} \succeq \boldsymbol{0}\}$. Let $\mathcal{A}$ be a linear map $\mathcal{A}$ from $\mathcal{S}^N$ to $\mathbb{R}^{N+1}$ defined as follows:

$$\mathcal{A}(\boldsymbol{Z}): \quad \boldsymbol{Z} \to \begin{bmatrix} \langle \boldsymbol{Z}, \boldsymbol{I}_N \rangle \\ \frac{1}{2}(\boldsymbol{Z} + \boldsymbol{Z}^\top)\boldsymbol{1}_N) \end{bmatrix}.$$

We can express (6.1) in the form of a standard cone program,

$$\min \quad \langle \boldsymbol{Z}, \boldsymbol{D} \rangle, \quad \text{s.t.} \quad \mathcal{A}(\boldsymbol{Z}) = \begin{bmatrix} k \\ \boldsymbol{1}_N \end{bmatrix}, \quad \boldsymbol{Z} \in \mathcal{K}. \tag{6.2}$$

Thus, using the standard derivation in Lagrangian duality theory [8], the dual problem of (6.1) can be easily obtained and given by

$$\max \quad -kz - \langle \boldsymbol{\alpha}, \boldsymbol{1}_N \rangle, \quad \text{s.t.} \quad \boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}) \in \mathcal{K}^*, \tag{6.3}$$

where $\boldsymbol{\lambda} = \begin{bmatrix} z \\ \boldsymbol{\alpha} \end{bmatrix} \in \mathbb{R}^{N+1}$ is the dual variable with respect to the affine constraints and

$$\mathcal{A}^*(\boldsymbol{\lambda}) := \frac{1}{2}(\boldsymbol{\alpha}\boldsymbol{1}_N^\top + \boldsymbol{1}_N\boldsymbol{\alpha}^\top) + z\boldsymbol{I}_N \tag{6.4}$$

is the adjoint operator of $\mathcal{A}$ under the canonical inner product over $\mathbb{R}^{N\times N}$.

6.2 Optimality condition

This subsection presents a necessary and sufficient condition for $\boldsymbol{X} = \sum_{a=1}^{k} \frac{1}{|\Gamma_a|}\boldsymbol{1}_{\Gamma_a}\boldsymbol{1}_{\Gamma_a}^\top$ to be the global minimum of the Peng–Wei relaxation. The result is summarized in Proposition 2, which follows from the complementary slackness in the conic duality theory. Moreover, a stronger sufficient condition has been established for the uniqueness of $\boldsymbol{X}$ in Proposition 3.

Theorem 8 (Conic Duality Theorem, Theorem 2.4.1 in [7]) *There hold:*

1. *If the primal problem is strictly feasible and bounded below, then the dual program is solvable*[3] *and the optimal values of the primal/dual problems are equal to each other;*
2. *If the dual problem is strictly feasible and bounded above, then the primal program is solvable and the optimal values of the primal/dual problems are equal to each other;*
3. *Assume either the primal problem or the dual problem is bounded and strictly feasible. Then* $(\boldsymbol{Z}, \boldsymbol{\lambda})$ *is a pair of primal/dual optimum if and only if either the duality gap is zero or the complementary slackness holds.*

[2] The dual cone of $\mathcal{K}$ is defined as $\{\boldsymbol{W} : \langle \boldsymbol{W}, \boldsymbol{Z} \rangle \geq 0, \forall \boldsymbol{Z} \in \mathcal{K}\}$; in particular, there holds $(\mathcal{K}^*)^* = \mathcal{K}$.

[3] The primal problem or dual problem is solvable if it is feasible, bounded and the optimal value is attained.

The following lemma, tailored to (6.1) and (6.3), simply follows from the strict feasibility of (6.1) or (6.3) and Theorem 8.

Lemma 1 *Both primal/dual problems* (6.1) *and* (6.3) *are strictly feasible and bounded below/above. Therefore, they are are solvable (so the optimal values are attained). Moreover,* $(\boldsymbol{X}, \boldsymbol{\lambda})$ *is a pair of primal/dual optima if and only if the complementary slackness holds:* $\langle \boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}), \boldsymbol{X} \rangle = 0$ *where* $\boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}) \in \mathcal{K}^*$.

Proof Consider $\widetilde{\boldsymbol{Z}} = \frac{1-\lambda}{N} \mathbf{1}_N \mathbf{1}_N^\top + \lambda \boldsymbol{I}_N$, where $\lambda = \frac{k-1}{N-1} > 0$ for $k \geq 2$. Note that $\widetilde{\boldsymbol{Z}} \succeq \lambda \boldsymbol{I}_N \succ \mathbf{0}$ and $\widetilde{\boldsymbol{Z}} \geq \frac{1-\lambda}{N} \mathbf{1}_N \mathbf{1}_N^\top > \mathbf{0}$. So $\widetilde{\boldsymbol{Z}}$ is in the interior of $\mathcal{K}$. It is also easy to verify that $\widetilde{\boldsymbol{Z}}$ satisfies the other two equality constraints. This shows (6.1) is strictly feasible. In addition, we can see that the objective function in (6.1) is also nonnegative since both $\boldsymbol{Z}$ and $\boldsymbol{D}$ are entrywise nonnegative. In conclusion, the primal problem is strictly feasible and bounded below by 0.

Note that $\boldsymbol{J}_{N\times N} = \mathbf{1}_N \mathbf{1}_N^\top$ is a strictly positive symmetric matrix. For the dual problem (6.3), we can take $\boldsymbol{\alpha} = \mathbf{0}$ and let z be a sufficiently large positive number such that

$$\boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}) = \underbrace{\boldsymbol{J}_{N\times N}}_{\text{a positive matrix}} + \underbrace{(\boldsymbol{D} + z\boldsymbol{I}_N - \boldsymbol{J}_{N\times N})}_{\text{a positive definite matrix}}$$

is in the interior of $\mathcal{K}^*$. Hence, the dual program is also strictly feasible. Its optimal value is bounded above because it is always smaller than the optimal value of the primal problem.

Therefore, the application of Theorem 8 implies that $(\boldsymbol{X}, \boldsymbol{\lambda})$ is a pair of primal/dual optima if and only if the complementary slackness holds, i.e., $\langle \boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}), \boldsymbol{X} \rangle = 0$ where $\boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}) \in \mathcal{K}^*$ and $\boldsymbol{X} \in \mathcal{K}$. □

Remark 2 The complementary slackness is indeed equivalent to the zero duality gap since the optimal values of both problems are attained and there holds

$$\langle \boldsymbol{D}, \boldsymbol{X} \rangle = -\langle \mathcal{A}^*(\boldsymbol{\lambda}), \boldsymbol{X} \rangle = -\langle \boldsymbol{\lambda}, \mathcal{A}(\boldsymbol{X}) \rangle = -\left\langle \boldsymbol{\lambda}, \begin{bmatrix} k \\ \mathbf{1}_N \end{bmatrix} \right\rangle = -kz - \langle \boldsymbol{\alpha}, \mathbf{1}_N \rangle.$$

In the following lemma, we will derive a more explicit expression for complementary slackness which will be used in the analysis later. By definition of $\mathcal{K}^*$, the matrix $\boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda})$ must be in the form of

$$\boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}) = \boldsymbol{B} + \boldsymbol{Q}, \tag{6.5}$$

where $\boldsymbol{B} \geq \mathbf{0}$, $\boldsymbol{Q} \succeq 0$ and both of them are symmetric.

Lemma 2 *The complementary slackness* $\langle \boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}), \boldsymbol{X} \rangle = 0$ *is equivalent to*

$$\boldsymbol{B}^{(a,a)} = \mathbf{0} \text{ for all } 1 \leq a \leq k, \quad \text{and} \quad \boldsymbol{Q}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{Q} = \mathbf{0}, \tag{6.6}$$

where $\boldsymbol{B} \geq \mathbf{0}$ *and* $\boldsymbol{Q} \succeq \mathbf{0}$ *obeys* (6.5) *for some* $\boldsymbol{\lambda}$. *It follows immediately that* $\boldsymbol{Q}^{(a,b)} \mathbf{1}_{n_b} = \mathbf{0}$ *for* $1 \leq a, b \leq k$. *Moreover,* (6.6) *implies that the dual variable*

$\boldsymbol{\lambda} = \begin{bmatrix} z \\ \boldsymbol{\alpha} \end{bmatrix}$ *satisfies*

$$\boldsymbol{\alpha}_a = -\frac{2}{n_a}\boldsymbol{D}^{(a,a)}\mathbf{1}_{n_a} + \frac{1}{n_a^2}\langle \boldsymbol{D}^{(a,a)}, \boldsymbol{J}_{n_a \times n_a}\rangle \mathbf{1}_{n_a} - \frac{z}{n_a}\mathbf{1}_{n_a}, \tag{6.7}$$

where $\boldsymbol{\alpha}_a$ *is the ath block of* $\boldsymbol{\alpha}$ *given by* $\{\alpha_i\}_{i\in\Gamma_a}$.

Proof It suffices to prove (6.6) from $\langle \boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}), \boldsymbol{X}\rangle = 0$ since the other direction is trivial. Note that the complementary slackness is equivalent to $\langle \boldsymbol{B} + \boldsymbol{Q}, \boldsymbol{X}\rangle = 0$ for some $\boldsymbol{B} \succeq \mathbf{0}$ and $\boldsymbol{Q} \succeq \mathbf{0}$. Since $\boldsymbol{X} \succeq \mathbf{0}$ and $\boldsymbol{X} \succeq 0$, it follows that $\langle \boldsymbol{B}, \boldsymbol{X}\rangle = \langle \boldsymbol{Q}, \boldsymbol{X}\rangle = 0$. From $\langle \boldsymbol{B}, \boldsymbol{X}\rangle = 0$ and $\boldsymbol{B} \succeq \mathbf{0}$, we have

$$\langle \boldsymbol{B}^{(a,a)}, \boldsymbol{J}_{n_a \times n_a}\rangle = 0 \Longleftrightarrow \boldsymbol{B}^{(a,a)} = \mathbf{0}$$

where $\boldsymbol{X}^{(a,a)} = \boldsymbol{J}_{n_a \times n_a}$. Since both $\boldsymbol{X}$ and $\boldsymbol{Q}$ are positive semi-definite matrices, we have

$$0 = \langle \boldsymbol{X}, \boldsymbol{Q}\rangle = \mathrm{Tr}(\boldsymbol{X}\boldsymbol{Q}) = \|\boldsymbol{X}^{1/2}\boldsymbol{Q}^{1/2}\|_F^2,$$

which gives $\boldsymbol{Q}^{1/2}\boldsymbol{X}^{1/2} = \boldsymbol{X}^{1/2}\boldsymbol{Q}^{1/2} = \mathbf{0}$ and in turn implies $\boldsymbol{Q}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{Q} = \mathbf{0}$.

Now we proceed to derive (6.7). Following from $\boldsymbol{Q}^{(a,a)}\mathbf{1}_{n_a} = \mathbf{0}$ and $\boldsymbol{B}^{(a,a)} = \mathbf{0}$, we obtain

$$\begin{aligned} \boldsymbol{Q}^{(a,a)}\mathbf{1}_{n_a} &= \boldsymbol{D}^{(a,a)}\mathbf{1}_{n_a} + \frac{1}{2}(n_a\boldsymbol{\alpha}_a + \boldsymbol{\alpha}_a^\top \mathbf{1}_{n_a}\mathbf{1}_{n_a}) + z\mathbf{1}_{n_a} = \mathbf{0}, \\ \mathbf{1}_{n_a}^\top \boldsymbol{Q}^{(a,a)}\mathbf{1}_{n_a} &= \mathbf{1}_{n_a}^\top \boldsymbol{D}^{(a,a)}\mathbf{1}_{n_a} + n_a\boldsymbol{\alpha}_a^\top \mathbf{1}_{n_a} + n_a z = \mathbf{0}, \end{aligned}$$

where $\boldsymbol{Q} = \boldsymbol{D} + \frac{1}{2}(\boldsymbol{\alpha}\mathbf{1}_N^\top + \mathbf{1}_N\boldsymbol{\alpha}^\top) + z\boldsymbol{I}_N - \boldsymbol{B}$ follows from $\boldsymbol{B} + \boldsymbol{Q} = \boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda})$ and the definition of $\mathcal{A}^*$, see (6.5) and (6.4). From the second equation above, we get $\boldsymbol{\alpha}_a^\top \mathbf{1}_{n_a} = -\frac{1}{n_a}\mathbf{1}_{n_a}^\top \boldsymbol{D}^{(a,a)}\mathbf{1}_{n_a} - z$. Substituting it into the first one gives

$$\begin{aligned} \boldsymbol{\alpha}_a &= \frac{1}{n_a}\left(-2\boldsymbol{D}^{(a,a)}\mathbf{1}_{n_a} - \boldsymbol{\alpha}_a^\top \mathbf{1}_{n_a}\mathbf{1}_{n_a} - 2z\mathbf{1}_{n_a}\right) \\ &= \frac{1}{n_a}\left(-2\boldsymbol{D}^{(a,a)}\mathbf{1}_{n_a} + \frac{1}{n_a}\mathbf{1}_{n_a}\mathbf{1}_{n_a}^\top \boldsymbol{D}^{(a,a)}\mathbf{1}_{n_a} - z\mathbf{1}_{n_a}\right), \end{aligned}$$

which completes the proof. □

Because of (6.7), the effective dual variables are only z and $\boldsymbol{B}^{(a,b)}$ with $a \neq b$ since $\boldsymbol{\alpha}$ can be fully represented by a function of z if the complementary slackness holds, and plugging $\boldsymbol{\alpha}$ back into the expression of $\boldsymbol{Q}$ in (6.5) gives

$$\boldsymbol{Q} = z(\boldsymbol{I}_N - \boldsymbol{E}) + \boldsymbol{M} - \boldsymbol{B}, \tag{6.8}$$

where

$$\begin{aligned}\boldsymbol{E}^{(a,b)} &= \frac{1}{2}\left(\frac{1}{n_a}+\frac{1}{n_b}\right)\boldsymbol{J}_{n_a\times n_b},\\ \boldsymbol{M}^{(a,b)} &= \boldsymbol{D}^{(a,b)} - \left(\frac{1}{n_a}\boldsymbol{D}^{(a,a)}\boldsymbol{J}_{n_a\times n_b}+\frac{1}{n_b}\boldsymbol{J}_{n_a\times n_b}\boldsymbol{D}^{(b,b)}\right)\\ &\quad+\frac{1}{2}\left(\frac{1}{n_a^2}\langle \boldsymbol{D}^{(a,a)},\boldsymbol{J}_{n_a\times n_a}\rangle+\frac{1}{n_b^2}\langle \boldsymbol{D}^{(b,b)},\boldsymbol{J}_{n_b\times n_b}\rangle\right)\boldsymbol{J}_{n_a\times n_b}. \end{aligned}\tag{6.9}$$

In particular, if $a = b$,

$$\begin{aligned}\boldsymbol{E}^{(a,a)} &= \frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a},\\ \boldsymbol{M}^{(a,a)} &= \left(\boldsymbol{I}_{n_a}-\frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a}\right)\boldsymbol{D}^{(a,a)}\left(\boldsymbol{I}_{n_a}-\frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a}\right). \end{aligned}\tag{6.10}$$

On the other hand, if $\boldsymbol{B}\succeq 0$, $\boldsymbol{B}^{(a,a)}=\boldsymbol{0}$ for all $1\le a\le k$, and $\boldsymbol{Q}\succeq 0$ has the form of (6.8), then one can easily verify that $\boldsymbol{QX}=0$ since $\langle \boldsymbol{Q},\boldsymbol{X}\rangle=0$, and $\boldsymbol{B}+\boldsymbol{Q}=\boldsymbol{D}+\mathcal{A}^*(\boldsymbol{\lambda})$ for z in (6.8) and $\boldsymbol{\alpha}$ in (6.7). Therefore, Lemma 2 implies that $\boldsymbol{X}$ is a global minimizer of (6.1).

In summary, we have established a necessary and sufficient condition for $\boldsymbol{X}$ to be a global minimizer of the Peng–Wei relaxation of k-means.

Proposition 2 (Optimality condition) *Any feasible pair of $\boldsymbol{Q}\succeq\boldsymbol{0}$ and $\boldsymbol{B}\succeq\boldsymbol{0}$ where $\boldsymbol{Q}$ has the form of (6.8) and $\boldsymbol{B}^{(a,a)}=\boldsymbol{0}$ for all $1\le a\le k$, certifies $\boldsymbol{X}$ to be a global minimum of (6.1). Conversely, if $\boldsymbol{X}$ is a global minimum of (6.1), then such a pair of $(\boldsymbol{Q},\boldsymbol{B})$ (or $(z,\boldsymbol{B})$) must exist.*

The optimality condition we have established is essentially equivalent to that of [12]. However, we use conic duality theory in [7] to show the strong duality holds, and both primal/dual solutions exist for Peng–Wei relaxation by constructing a Slater's constraint qualification. This lays the foundation to derive the necessary condition for the tightness of Peng–Wei relaxation, which is not fully addressed in [12].

In other words, the optimality condition in Proposition 2 is not strong enough to guarantee that $\boldsymbol{X}$ is a unique solution to (6.1). The following proposition provides a sufficient condition for the uniqueness of $\boldsymbol{X}$ by imposing a stricter condition on $\boldsymbol{B}$.

Proposition 3 (A sufficient condition for the uniqueness of global minimum) *Any feasible pair of $\boldsymbol{Q}\succeq\boldsymbol{0}$ and $\boldsymbol{B}\succeq\boldsymbol{0}$, where $\boldsymbol{Q}$ has the form of (6.8), $\boldsymbol{B}^{(a,a)}=\boldsymbol{0}$ for all $1\le a\le k$, and $\boldsymbol{B}^{(a,b)}>\boldsymbol{0}$ for all $a\ne b$, certifies $\boldsymbol{X}$ to be a unique global minimum of (6.1).*

Proof Proposition 2 implies $\boldsymbol{X}$ is a global minimum of (6.1). Let $\widetilde{\boldsymbol{X}}\in\mathbb{R}^{N\times N}$ be an arbitrary feasible solution satisfying $\widetilde{\boldsymbol{X}}\boldsymbol{1}_N=\boldsymbol{1}_N$, $\mathrm{Tr}(\widetilde{\boldsymbol{X}})=k$, $\widetilde{\boldsymbol{X}}\succeq 0$ and $\widetilde{\boldsymbol{X}}\ge 0$. We will prove $\boldsymbol{X}$ is a unique solution by showing that if $\widetilde{\boldsymbol{X}}\ne\boldsymbol{X}$, there holds

$$\langle \boldsymbol{D},\widetilde{\boldsymbol{X}}-\boldsymbol{X}\rangle>0.$$

We start with $\langle \boldsymbol{Q}, \widetilde{\boldsymbol{X}} - \boldsymbol{X} \rangle$. Since $\boldsymbol{Q} \succeq 0$, $\widetilde{\boldsymbol{X}} \succeq 0$, and $\langle \boldsymbol{Q}, \boldsymbol{X} \rangle = 0$, it follows that

$$\langle \boldsymbol{Q}, \widetilde{\boldsymbol{X}} - \boldsymbol{X} \rangle = \langle \boldsymbol{Q}, \widetilde{\boldsymbol{X}} \rangle \geq 0.$$

By the definition of $\boldsymbol{Q}$, and the fact $\widetilde{\boldsymbol{X}} \mathbf{1}_N = \boldsymbol{X} \mathbf{1}_N = \mathbf{1}_N$ and $\mathrm{Tr}(\widetilde{\boldsymbol{X}}) = \mathrm{Tr}(\boldsymbol{X}) = k$, there holds,

$$\langle \boldsymbol{Q}, \widetilde{\boldsymbol{X}} - \boldsymbol{X} \rangle = \langle \boldsymbol{D}, \widetilde{\boldsymbol{X}} - \boldsymbol{X} \rangle - \langle \boldsymbol{B}, \widetilde{\boldsymbol{X}} - \boldsymbol{X} \rangle \geq 0.$$

Since the supports of $\boldsymbol{B}$ and $\boldsymbol{X}$ are disjoint, one has $\langle \boldsymbol{B}, \boldsymbol{X} \rangle = 0$. Therefore, in order to show $\langle \boldsymbol{D}, \widetilde{\boldsymbol{X}} - \boldsymbol{X} \rangle > 0$, it suffices to prove that $\langle \boldsymbol{B}, \widetilde{\boldsymbol{X}} \rangle > 0$, which will be done by contradiction.

Suppose $\langle \boldsymbol{B}, \widetilde{\boldsymbol{X}} \rangle = \sum_{a \neq b} \langle \boldsymbol{B}^{(a,b)}, \widetilde{\boldsymbol{X}}^{(a,b)} \rangle = 0$. Then we have $\widetilde{\boldsymbol{X}}^{(a,b)} = 0$ which follows from $\boldsymbol{B}^{(a,b)} > 0$ for all $a \neq b$ and $\widetilde{\boldsymbol{X}} \geq 0$. Therefore, the support of $\widetilde{\boldsymbol{X}}$ must be the same as that of $\boldsymbol{X}$. Note that $\widetilde{\boldsymbol{X}}$ is a positive semi-definite matrix which satisfies $\widetilde{\boldsymbol{X}} \mathbf{1}_N = \mathbf{1}_N$ and $\mathrm{Tr}(\widetilde{\boldsymbol{X}}) = k$. So for any $1 \leq a \leq k$, $\widetilde{\boldsymbol{X}}^{(a,a)} \mathbf{1}_{n_a} = \mathbf{1}_{n_a}$. This means that 1 is an eigenvalue of $\widetilde{\boldsymbol{X}}$ with multiplicity at least k. Since all the eigenvalues of $\widetilde{\boldsymbol{X}}$ are nonnegative and their sum is equal to $\mathrm{Tr}(\widetilde{\boldsymbol{X}}) = k$, $\widetilde{\boldsymbol{X}}$ has only k nonzero eigenvalues and all of them are 1. Thus, each $\widetilde{\boldsymbol{X}}^{(a,a)}$ is a rank one matrix. It follow that $\widetilde{\boldsymbol{X}}^{(a,a)} = \frac{1}{n_a} \mathbf{1}_{n_a} \mathbf{1}_{n_a}^\top = \boldsymbol{X}^{(a,a)}$ since $\widetilde{\boldsymbol{X}}^{(a,a)} \mathbf{1}_{n_a} = \mathbf{1}_{n_a}$ and $\widetilde{\boldsymbol{X}}^{(a,a)}$ is symmetric. This contradicts the assumption $\widetilde{\boldsymbol{X}} \neq \boldsymbol{X}$. □

6.3 Sufficient condition for dual certificate

We will further reduce the sufficient condition in Proposition 3 to one that will be used in the construction of the dual certificate. As suggested by that proposition, we need to find a number $z \in \mathbb{R}$ and a symmetric matrix $\boldsymbol{B} \in \mathbb{R}_+^{N \times N}$ such that the following sufficient condition holds:

$$\boldsymbol{Q} \succeq \mathbf{0}, \quad \boldsymbol{B}^{(a,b)} > \mathbf{0}, \quad \boldsymbol{B}^{(a,a)} = \mathbf{0} \quad \forall a \neq b, \tag{6.11}$$

where $\boldsymbol{Q}$ is given in (6.8). As a result $\boldsymbol{Q}$, satisfies $\boldsymbol{Q}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{Q} = \mathbf{0}$ automatically.

In order to present our final sufficient optimality condition, we first introduce two linear subspaces. Note that $\boldsymbol{X}$ is clearly a projection matrix satisfying $\boldsymbol{X}^2 = \boldsymbol{X}$. Let T and $T^\perp$ be two linear subspaces in $\mathbb{R}^{N \times N}$ defined as

$$T = \{\boldsymbol{X}\boldsymbol{Y} + \boldsymbol{Y}\boldsymbol{X} - \boldsymbol{X}\boldsymbol{Y}\boldsymbol{X} : \boldsymbol{Y} \in \mathbb{R}^{N \times N}\},$$
$$T^\perp = \{(\boldsymbol{I}_N - \boldsymbol{X})\boldsymbol{Y}(\boldsymbol{I}_N - \boldsymbol{X}) : \boldsymbol{Y} \in \mathbb{R}^{N \times N}\}.$$

Denote by $\mathcal{P}_T : \mathbb{R}^{N \times N} \to T$ and $\mathcal{P}_{T^\perp} : \mathbb{R}^{N \times N} \to T^\perp$ the corresponding projection operators. We use subscripts to denote projections, for example letting $\mathcal{P}_T(\boldsymbol{B}) = \boldsymbol{B}_T$ and $\mathcal{P}_{T^\perp}(\boldsymbol{B}) = \boldsymbol{B}_{T^\perp}$. For any $\boldsymbol{Z} \in \mathbb{R}^{N \times N}$, it can be easily verified that the (a, b)th block of $\boldsymbol{Z}_T$ and $\boldsymbol{Z}_{T^\perp}$ are

$$\boldsymbol{Z}_{T}^{(a,b)} = \frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a}\boldsymbol{Z}^{(a,b)} + \frac{1}{n_b}\boldsymbol{Z}^{(a,b)}\boldsymbol{J}_{n_b\times n_b} - \frac{1}{n_a n_b}\boldsymbol{J}_{n_a\times n_a}\boldsymbol{Z}^{(a,b)}\boldsymbol{J}_{n_b\times n_b}, \quad (6.12)$$

$$\boldsymbol{Z}_{T^\perp}^{(a,b)} = \left(\boldsymbol{I}_{n_a} - \frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a}\right)\boldsymbol{Z}^{(a,b)}\left(\boldsymbol{I}_{n_b} - \frac{1}{n_b}\boldsymbol{J}_{n_b\times n_b}\right). \quad (6.13)$$

Proposition 4 *The optimality condition with uniqueness in* (6.11) *is equivalent to*

$$\begin{aligned}
&z\mathcal{P}_{T^\perp}(\boldsymbol{I}_N) + \boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp} \succeq \boldsymbol{0},\\
&\boldsymbol{M}_{T}^{(a,b)} - \boldsymbol{B}_{T}^{(a,b)} - \frac{z(n_a+n_b)}{2n_a n_b}\boldsymbol{J}_{n_a\times n_b} = \boldsymbol{0}, \quad \forall a \neq b,\\
&\boldsymbol{B}^{(a,b)} = (\boldsymbol{B}^{(b,a)})^\top, \quad \boldsymbol{B}^{(a,a)} = \boldsymbol{0}, \quad \boldsymbol{B}^{(a,b)} > \boldsymbol{0}, \quad \forall\, a \neq b.
\end{aligned} \quad (6.14)$$

Proof We first show that (6.11) implies (6.14), and then show the other direction.

(6.11)$\Longrightarrow$(6.14): Noting that $\boldsymbol{E} \in T$, $\mathcal{P}(\boldsymbol{I}_N) = \boldsymbol{I}_N - \boldsymbol{X}$ and $\boldsymbol{Q}$ has the form of (6.8), the projection of $\boldsymbol{Q}$ on $T^\perp$ is given by

$$\boldsymbol{Q}_{T^\perp} = (\boldsymbol{I}_N - \boldsymbol{X})\boldsymbol{Q}(\boldsymbol{I}_N - \boldsymbol{X}) = z(\boldsymbol{I}_N - \boldsymbol{X}) + \boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp} \succeq \boldsymbol{0}$$

which gives the first expression in (6.14). For the second one in (6.14), we have $\boldsymbol{Q}_T = \boldsymbol{0}$ since $\boldsymbol{Q}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{Q} = \boldsymbol{0}$ and thus $\boldsymbol{Q}^{(a,b)}\boldsymbol{1}_{n_b} = \boldsymbol{0}$ for all pairs of (a, b). For $\boldsymbol{Q}^{(a,a)}$ with $1 \le a \le k$, $\boldsymbol{Q}^{(a,a)}\boldsymbol{1}_{n_a} = \boldsymbol{0}$ holds automatically by the definition of $\boldsymbol{Q}$ in (6.8). For $a \neq b$, straightforward calculations lead to

$$\boldsymbol{Q}^{(a,b)}\boldsymbol{1}_{n_b} = -\frac{n_b z}{2}\left(\frac{1}{n_a} + \frac{1}{n_b}\right)\boldsymbol{1}_{n_a} + \boldsymbol{M}^{(a,b)}\boldsymbol{1}_{n_b} - \boldsymbol{B}^{(a,b)}\boldsymbol{1}_{n_b} = \boldsymbol{0}. \quad (6.15)$$

Thus, one has $\frac{1}{n_b}\boldsymbol{B}^{(a,b)}\boldsymbol{J}_{n_b\times n_b} = \frac{1}{n_b}\boldsymbol{M}^{(a,b)}\boldsymbol{J}_{n_b\times n_b} - \frac{z}{2}\left(\frac{1}{n_a} + \frac{1}{n_b}\right)\boldsymbol{J}_{n_a\times n_b}$ for all $a \neq b$, which implies $\boldsymbol{B}_{T}^{(a,b)} = \boldsymbol{M}_{T}^{(a,b)} - \frac{z(n_a+n_b)}{2n_a n_b}\boldsymbol{J}_{n_a\times n_b}$. The third formula in (6.14) satisfies automatically.

(6.14)$\Longrightarrow$(6.11): It suffices to prove $\boldsymbol{Q}$ in (6.8) is positive semidefinite. By definition, the matrix $\boldsymbol{E}^{(a,b)}$ is equal to $\frac{1}{2}\left(\frac{1}{n_a} + \frac{1}{n_b}\right)\boldsymbol{J}_{n_a\times n_b}$ and $\mathcal{P}_{T^\perp}(\boldsymbol{I}_N) = \boldsymbol{I}_N - \boldsymbol{X}$. Adding the first two formulas in (6.14) blockwisely over all (a, b) gives

$$z(\boldsymbol{I}_N - \boldsymbol{X}) + \boldsymbol{M} - \boldsymbol{B} - z(\boldsymbol{E} - \boldsymbol{X}) = \underbrace{z(\boldsymbol{I}_N - \boldsymbol{E}) + \boldsymbol{M} - \boldsymbol{B}}_{\boldsymbol{Q}} \succeq \boldsymbol{0}$$

where we have used the following facts: $\boldsymbol{X}^{(a,a)} = \boldsymbol{E}^{(a,a)}$, $\boldsymbol{X}^{(a,b)} = \boldsymbol{0}$ when $a \neq b$, $\boldsymbol{M}_{T}^{(a,a)} = \boldsymbol{0}$ which follows from (6.10), and $\boldsymbol{B}_{T}^{(a,a)} = \boldsymbol{0}$ due to $\boldsymbol{B}^{(a,a)} = \boldsymbol{0}$. This shows $\boldsymbol{Q} \succeq \boldsymbol{0}$. □

According to (6.14), $\boldsymbol{B}_{T}^{(a,b)}$ is determined by $\boldsymbol{M}^{(a,b)}$ and z. So the only free variables are z and $\boldsymbol{B}_{T^\perp}^{(a,b)}$ for $a \neq b$. To determine z, we replace $z\mathcal{P}_{T^\perp}(\boldsymbol{I}_N) + \boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp} \succeq \boldsymbol{0}$

by a stronger condition $z \geq \|\boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp}\|$ which clearly implies the former one. To choose $\boldsymbol{B}_{T^\perp}^{(a,b)}$ for any $a \neq b$, notice that

$$\boldsymbol{B}^{(a,b)} > \boldsymbol{0} \Longleftrightarrow \boldsymbol{B}_{T^\perp}^{(a,b)} + \boldsymbol{B}_T^{(a,b)} > \boldsymbol{0} \Longleftrightarrow \boldsymbol{B}_{T^\perp}^{(a,b)} > \frac{z(n_a+n_b)}{2n_a n_b}\boldsymbol{J}_{n_a\times n_b} - \boldsymbol{M}_T^{(a,b)},$$

where we have used a substitution for $\boldsymbol{B}_T^{(a,b)}$. To sum up, we have derived a replacement sufficient condition which guarantees $\boldsymbol{X}$ as the unique global minimum of (6.1):

$$\begin{aligned}
&z \geq \|\boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp}\|,\\
&\boldsymbol{B} = \boldsymbol{B}^\top,\\
&\boldsymbol{B}^{(a,a)} = \boldsymbol{0}, \quad \forall\, 1 \leq a \leq k,\\
&\boldsymbol{B}_T^{(a,b)} = \boldsymbol{M}_T^{(a,b)} - \frac{z(n_a+n_b)}{2n_a n_b}\boldsymbol{J}_{n_a\times n_b}, \quad \forall\, a \neq b,\\
&\boldsymbol{B}_{T^\perp}^{(a,b)} > \frac{z(n_a+n_b)}{2n_a n_b}\boldsymbol{J}_{n_a\times n_b} - \boldsymbol{M}_T^{(a,b)}, \quad \forall\, a \neq b.
\end{aligned} \tag{6.16}$$

6.4 Proof of Theorem 2

Now we are ready to prove the main theorem, which follows directly from the proposition below.

Proposition 5 *Assume the proximity condition (1.2) holds for the partition $\{\Gamma_a\}_{a=1}^k$. We can choose z and $\boldsymbol{B}$ such that*

$$z = \|\boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp}\|, \quad \boldsymbol{B}_{T^\perp}^{(a,b)} = 4\boldsymbol{u}_{a,b}\boldsymbol{u}_{b,a}^\top, \quad \forall\, a \neq b,$$

and the sufficient condition in (6.16) is satisfied. Therefore, whenever the proximity condition holds, $\boldsymbol{X} = \sum_{a=1}^k \frac{1}{|\Gamma_a|}\mathbf{1}_{\Gamma_a}\mathbf{1}_{\Gamma_a}^\top$ is the unique minimizer of the Peng–Wei relaxation of k-means.

Lemma 3 *For any $1 \leq a, b \leq k$, $\boldsymbol{M}_{T^\perp}^{(a,b)} = \boldsymbol{D}_{T^\perp}^{(a,b)} = -2\overline{\boldsymbol{X}}_a\overline{\boldsymbol{X}}_b^\top$.*

Proof Let $\boldsymbol{x}_{a,i}$ and $\boldsymbol{x}_{b,j}$ be the ith and jth points in the ath and bth clusters, respectively. Then,

$$\|\boldsymbol{x}_{a,i} - \boldsymbol{x}_{b,j}\|^2 = \|\boldsymbol{x}_{a,i}\|^2 - 2\langle \boldsymbol{x}_{a,i}, \boldsymbol{x}_{b,j}\rangle + \|\boldsymbol{x}_{b,j}\|^2.$$

Denote by $\boldsymbol{\phi}_a \in \mathbb{R}^{n_a}$ and $\boldsymbol{\phi}_b \in \mathbb{R}^{n_b}$ the column vectors consisted of $\|\boldsymbol{x}_{a,i}\|^2$ and $\|\boldsymbol{x}_{b,j}\|^2$, respectively. Then,

$$\boldsymbol{D}^{(a,b)} = \boldsymbol{\phi}_a\mathbf{1}_{n_b}^\top - 2\boldsymbol{X}_a\boldsymbol{X}_b^\top + \mathbf{1}_{n_a}\boldsymbol{\phi}_b^\top.$$

$$\begin{aligned}\boldsymbol{D}_{T^\perp}^{(a,b)} &= (\boldsymbol{I}_{n_a} - \frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a})\boldsymbol{D}^{(a,b)}(\boldsymbol{I}_{n_b} - \frac{1}{n_b}\boldsymbol{J}_{n_b\times n_b}) \\ &= -2(\boldsymbol{I}_{n_a} - \frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a})\boldsymbol{X}_a\boldsymbol{X}_b^\top(\boldsymbol{I}_{n_b} - \frac{1}{n_b}\boldsymbol{J}_{n_b\times n_b}) \\ &= -2\overline{\boldsymbol{X}}_a\overline{\boldsymbol{X}}_b^\top.\end{aligned}$$

The matrix $\boldsymbol{M}$ is defined in (6.9), and it is easy to check that $\boldsymbol{M}_{T^\perp}^{(a,b)} = \boldsymbol{D}_{T^\perp}^{(a,b)}$. □

Lemma 4 *The operator norm of $\boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp}$ is bounded by $2\sum_{l=1}^k \|\overline{\boldsymbol{X}}_l\|^2$, i.e.,*

$$z = \|\boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp}\| \le 2\sum_{l=1}^k \|\overline{\boldsymbol{X}}_l\|^2.$$

Proof Note that $\boldsymbol{u}_{a,b} = \overline{\boldsymbol{X}}_a\boldsymbol{w}_{a,b}$ and by Lemma 3, $\boldsymbol{M}_{T^\perp}^{(a,b)} = -2\overline{\boldsymbol{X}}_a\overline{\boldsymbol{X}}_b^\top$. Hence, $\boldsymbol{B}_{T^\perp} - \boldsymbol{M}_{T^\perp} = 2\widehat{\boldsymbol{X}}\boldsymbol{W}\widehat{\boldsymbol{X}}^\top$, where $\widehat{\boldsymbol{X}} \in \mathbb{R}^{N\times mk}$ is defined as

$$\widehat{\boldsymbol{X}}^{(a,b)} = \boldsymbol{0}, \quad \widehat{\boldsymbol{X}}^{(a,a)} = \overline{\boldsymbol{X}}_a, \quad \forall a \ne b,$$

and $\boldsymbol{W} \in \mathbb{R}^{mk\times mk}$ is given by

$$\boldsymbol{W}^{(a,b)} = \boldsymbol{I}_m - 2\boldsymbol{w}_{a,b}\boldsymbol{w}_{a,b}^\top, \quad \boldsymbol{W}^{(a,a)} = \boldsymbol{I}_m, \quad \forall a \ne b.$$

Note that each $\boldsymbol{W}^{(a,b)}$ is an orthogonal matrix and thus $\|\boldsymbol{W}^{(a,b)}\| = 1$. Let $\boldsymbol{y}$ be a vector of length N, and denote by $\boldsymbol{y}_a$ the ath block of $\boldsymbol{y}$, $1 \le a \le k$. There holds,

$$\begin{aligned}\left|\boldsymbol{y}^\top(\boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp})\boldsymbol{y}\right| &\le 2\sum_{a=1}^k\sum_{b=1}^k \left|\boldsymbol{y}_a^\top\overline{\boldsymbol{X}}_a\boldsymbol{W}^{(a,b)}\overline{\boldsymbol{X}}_b\boldsymbol{y}_b^\top\right| \\ &\le 2\sum_{a=1}^k\sum_{b=1}^k \|\overline{\boldsymbol{X}}_a\|\|\boldsymbol{y}_a\|\|\overline{\boldsymbol{X}}_b\|\|\boldsymbol{y}_b\| \\ &\le 2\left(\sum_{l=1}^k \|\overline{\boldsymbol{X}}_l\|\|\boldsymbol{y}_l\|\right)^2 \\ &\le 2\left(\sum_{l=1}^k \|\overline{\boldsymbol{X}}_l\|^2\right)\left(\sum_{l=1}^k \|\boldsymbol{y}_l\|^2\right).\end{aligned}$$

Therefore, the operator norm of $\boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp}$ is bounded by $2\sum_{l=1}^k \|\overline{\boldsymbol{X}}_l\|^2$. □

It only remains to check whether (1.1) implies the second inequality in (6.16):

$$\boldsymbol{B}_{T^\perp}^{(a,b)} = 4\boldsymbol{u}_{a,b}\boldsymbol{u}_{b,a}^\top > \frac{z(n_a+n_b)}{2n_an_b}\boldsymbol{J}_{n_a\times n_b} - \boldsymbol{M}_T^{(a,b)}, \quad \forall\, a \ne b. \tag{6.17}$$

To show this, we first derive an explicit expression for $\boldsymbol{M}_T^{(a,b)}$.

Lemma 5 *For any $a \neq b$, there holds*

$$\frac{1}{n_b}\boldsymbol{D}^{(a,b)}\mathbf{1}_{n_b} - \frac{1}{n_a}\boldsymbol{D}^{(a,a)}\mathbf{1}_{n_a} = \left(h_{a,b}^2 + \frac{1}{n_b}\|\overline{\boldsymbol{X}}_b\|_F^2 - \frac{1}{n_a}\|\overline{\boldsymbol{X}}_a\|_F^2\right)\mathbf{1}_{n_a} - 2h_{a,b}\boldsymbol{u}_{a,b}.$$

Proof The ith entry of the left hand side is

$$\begin{aligned}
(LHS)_i &= \frac{1}{n_b}\sum_{l=1}^{n_b}\|\boldsymbol{x}_{a,i} - \boldsymbol{x}_{b,l}\|^2 - \frac{1}{n_a}\sum_{l=1}^{n_a}\|\boldsymbol{x}_{a,i} - \boldsymbol{x}_{a,l}\|^2 \\
&= \|\boldsymbol{c}_a - \boldsymbol{c}_b\|^2 - 2\langle \boldsymbol{x}_{a,i} - \boldsymbol{c}_a, \boldsymbol{c}_b - \boldsymbol{c}_a\rangle + \frac{1}{n_b}\sum_{l=1}^{n_b}\|\boldsymbol{x}_{b,l} - \boldsymbol{c}_b\|^2 - \frac{1}{n_a}\sum_{l=1}^{n_a}\|\boldsymbol{x}_{a,l} - \boldsymbol{c}_a\|^2 \\
&= h_{a,b}^2 - 2h_{a,b}(\overline{\boldsymbol{X}}_a\boldsymbol{w}_{b,a})_i + \frac{1}{n_b}\|\overline{\boldsymbol{X}}_b\|_F^2 - \frac{1}{n_a}\|\overline{\boldsymbol{X}}_a\|_F^2 \\
&= (RHS)_i.
\end{aligned}$$

□

Lemma 6 *For any $a \neq b$, there holds*

$$\boldsymbol{M}_T^{(a,b)} = h_{a,b}^2\boldsymbol{J}_{n_a\times n_b} - 2h_{a,b}\boldsymbol{u}_{a,b}\mathbf{1}_{n_b}^\top - 2h_{a,b}\mathbf{1}_{n_a}\boldsymbol{u}_{b,a}^\top.$$

Proof By the definition of $\boldsymbol{M}^{(a,b)}$ in (6.9),

$$\begin{aligned}
\boldsymbol{M}_T^{(a,b)} &= \boldsymbol{D}_T^{(a,b)} - \frac{1}{n_a}\boldsymbol{D}^{(a,a)}\boldsymbol{J}_{n_a\times n_b} - \frac{1}{n_b}\boldsymbol{J}_{n_a\times n_b}\boldsymbol{D}^{(b,b)} \\
&\quad + \frac{1}{2}\left(\frac{1}{n_a^2}\langle \boldsymbol{D}^{(a,a)}, \boldsymbol{J}_{n_a\times n_a}\rangle + \frac{1}{n_b^2}\langle \boldsymbol{D}^{(b,b)}, \boldsymbol{J}_{n_b\times n_b}\rangle\right)\boldsymbol{J}_{n_a\times n_b} \\
&= \underbrace{\frac{1}{n_b}\boldsymbol{D}^{(a,b)}\boldsymbol{J}_{n_b\times n_b} - \frac{1}{n_a}\boldsymbol{D}^{(a,a)}\boldsymbol{J}_{n_a\times n_b}}_{\Pi_1} + \underbrace{\frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a}\boldsymbol{D}^{(a,b)} - \frac{1}{n_b}\boldsymbol{J}_{n_a\times n_b}\boldsymbol{D}^{(b,b)}}_{\Pi_2} \\
&\quad + \underbrace{\left(\frac{1}{2n_a^2}\langle \boldsymbol{D}^{(a,a)}, \boldsymbol{J}_{n_a\times n_a}\rangle + \frac{1}{2n_b^2}\langle \boldsymbol{D}^{(b,b)}, \boldsymbol{J}_{n_b\times n_b}\rangle - \frac{1}{n_a n_b}\langle \boldsymbol{D}^{(a,b)}, \boldsymbol{J}_{n_a\times n_b}\rangle\right)\boldsymbol{J}_{n_a\times n_a}}_{\Pi_3},
\end{aligned}$$

where we have used

$$\boldsymbol{D}_T^{(a,b)} = \frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a}\boldsymbol{D}^{(a,b)} + \frac{1}{n_b}\boldsymbol{D}^{(a,b)}\boldsymbol{J}_{n_b\times n_b} - \frac{1}{n_a n_b}\langle \boldsymbol{D}^{(a,b)}, \boldsymbol{J}_{n_a\times n_b}\rangle\boldsymbol{J}_{n_a\times n_b}.$$

By Lemma 5, we have

$$\begin{aligned}\Pi_1 &= \left(\frac{1}{n_b}\boldsymbol{D}^{(a,b)}\mathbf{1}_{n_b} - \frac{1}{n_a}\boldsymbol{D}^{(a,a)}\mathbf{1}_{n_a}\right)\mathbf{1}_{n_b}^\top \\ &= \left(h_{a,b}^2 + \frac{1}{n_b}\|\overline{\boldsymbol{X}}_b\|_F^2 - \frac{1}{n_a}\|\overline{\boldsymbol{X}}_a\|_F^2\right)\boldsymbol{J}_{n_a\times n_b} - 2h_{a,b}\boldsymbol{u}_{a,b}\mathbf{1}_{n_b}^\top.\end{aligned}$$

Similarly,

$$\begin{aligned}\Pi_2 &= \mathbf{1}_{n_a}\left(\frac{1}{n_a}\boldsymbol{D}^{(b,a)}\mathbf{1}_{n_a} - \frac{1}{n_b}\boldsymbol{D}^{(b,b)}\mathbf{1}_{n_b}\right)^\top \\ &= \left(h_{a,b}^2 + \frac{1}{n_a}\|\overline{\boldsymbol{X}}_a\|_F^2 - \frac{1}{n_b}\|\overline{\boldsymbol{X}}_b\|_F^2\right)\boldsymbol{J}_{n_a\times n_b} - 2h_{a,b}\mathbf{1}_{n_a}\boldsymbol{u}_{b,a}^\top.\end{aligned}$$

Moreover, the (i, j)-entry of Π_3 is

$$\begin{aligned}(\Pi_3)_{i,j} &= \frac{1}{2n_a^2}\sum_{i=1}^{n_a}\sum_{j=1}^{n_a}\|\boldsymbol{x}_{a,i} - \boldsymbol{x}_{a,j}\|^2 + \frac{1}{2n_b^2}\sum_{i=1}^{n_b}\sum_{j=1}^{n_b}\|\boldsymbol{x}_{b,i} - \boldsymbol{x}_{b,j}\|^2 \\ &\quad - \frac{1}{n_a n_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\|\boldsymbol{x}_{a,i} - \boldsymbol{x}_{b,j}\|^2 \\ &= \frac{1}{n_a}\sum_{i=1}^{n_a}\|\boldsymbol{x}_{a,i} - \boldsymbol{c}_a\|^2 + \frac{1}{n_b}\sum_{i=1}^{n_b}\|\boldsymbol{x}_{b,i} - \boldsymbol{c}_b\|^2 - \frac{1}{n_a}\sum_{i=1}^{n_a}\|\boldsymbol{x}_{a,i} - \boldsymbol{c}_a\|^2 \\ &\quad - \frac{1}{n_b}\sum_{j=1}^{n_b}\|\boldsymbol{x}_{b,j} - \boldsymbol{c}_b\|^2 - \|\boldsymbol{c}_a - \boldsymbol{c}_b\|^2 = -h_{a,b}^2.\end{aligned}$$

Adding up $(\Pi_1)_{i,j}$, $(\Pi_2)_{i,j}$ and $(\Pi_3)_{i,j}$ leads to the desired identity. □

Proof of Proposition 5 Combined with the explicit expression of $\boldsymbol{M}_T^{(a,b)}$, (6.17) is equivalent to

$$-4\boldsymbol{u}_{a,b}\boldsymbol{u}_{b,a}^\top + \left(\frac{z(n_a+n_b)}{2n_a n_b} - h_{a,b}^2\right)\boldsymbol{J}_{n_a\times n_b} + 2h_{(a,b)}(\boldsymbol{u}_{a,b}\mathbf{1}_{n_b}^\top + \mathbf{1}_{n_a}\boldsymbol{u}_{b,a}^\top) < 0. \tag{6.18}$$

By definition of $\tau_{a,b}$, we have

$$\tau_{a,b} \geq \max(\boldsymbol{u}_{a,b}), \quad \tau_{a,b} \geq \max(\boldsymbol{u}_{b,a}).$$

Define

$$f(x, y) := -4xy - 2h_{(a,b)}(x + y) + \frac{z(n_a+n_b)}{2n_a n_b} - h_{(a,b)}^2.$$

Let $u_{a,b,i}$ and $u_{b,a,j}$ be the ith and jth entry of $\boldsymbol{u}_{a,b}$ and $\boldsymbol{u}_{b,a}$ respectively. One can easily see that $f(-u_{a,b,i}, -u_{b,a,j})$ is equal to the (i, j)th entry of the matrix on the left hand side of (6.18). Therefore, in order to prove (6.18), it suffices to show that $f(x, y) < 0$ for all $x, y \geq -\tau_{a,b}$. Note that if the proximity condition (1.1) holds, then $2\tau_{a,b} \leq \|\boldsymbol{c}_a - \boldsymbol{c}_b\|$. Therefore, $x, y \geq -\tau_{a,b} \geq -\frac{1}{2}h_{a,b}$.

We claim that the maximum of $f(x, y)$ over $\{(x, y) \in \mathbb{R}^2 : x \geq -\tau_{a,b}, y \geq -\tau_{a,b}\}$ is attained at $x = y = -\tau_{a,b}$ due to bilinearity of $f(x, y)$. More precisely, this follows from $2\tau_{a,b} \leq h_{a,b}$ and

$$\begin{aligned}\frac{\partial f}{\partial x} &= -4y - 2h_{(a,b)} \leq 4\tau_{a,b} - 2h_{a,b} \leq 0,\\ \frac{\partial f}{\partial y} &= -4x - 2h_{(a,b)} \leq 4\tau_{a,b} - 2h_{a,b} \leq 0\end{aligned}$$

over $\{(x, y) \in \mathbb{R}^2 : x \geq -\tau_{a,b}, y \geq -\tau_{a,b}\}$.

Therefore, (6.18) holds if

$$\max_{\{x,y\geq -\tau_{a,b}\}} f(x, y) = -4\tau_{a,b}^2 + 4h_{a,b}\tau_{a,b} - h_{a,b}^2 + \frac{z(n_a + n_b)}{2n_a n_b} < 0.$$

Since $2\tau_{a,b} \leq h_{a,b}$, the inequality above is equivalent to

$$h_{a,b} - 2\tau_{a,b} > \sqrt{\frac{z(n_a + n_b)}{2n_a n_b}}.$$

Meanwhile, the proximity condition implies

$$h_{(a,b)} - 2\tau_{a,b} > \sqrt{\frac{\sum_{l=1}^{k} \|\overline{\boldsymbol{X}}_l\|^2 (n_a + n_b)}{n_a n_b}} \geq \sqrt{\frac{z(n_a + n_b)}{2n_a n_b}}.$$

Hence, we have $-4\tau_{a,b}^2 + 4h_{a,b}\tau_{a,b} - h_{a,b}^2 + \frac{z(n_a+n_b)}{2n_a n_b} < 0$ and (6.18) holds. □

6.5 Proof of Theorem 3

This subsection is devoted to proving Theorem 3, the necessary lower bound of $\frac{1}{2}h_{a,b} - \tau_{a,b}$ for $\boldsymbol{X} = \sum_{a=1}^{k} \frac{1}{|\Gamma_a|}\mathbf{1}_{\Gamma_a}\mathbf{1}_{\Gamma_a}^\top$ to be a global minimum of the Peng–Wei relaxation of k-means. We will use the necessary condition established in Proposition 2 for the proof which states that, if $\boldsymbol{X}$ is global minimizer, then there exist a number z and a matrix $\boldsymbol{B}$ obeying $\boldsymbol{B} \geq 0$, $\boldsymbol{B}^{(a,a)} = \mathbf{0}$ for all $1 \leq a \leq k$, and $\boldsymbol{Q} = z(\boldsymbol{I}_N - \boldsymbol{E}) + \boldsymbol{M} - \boldsymbol{B} \succeq 0$.

Proof of Theorem 3 The proof is partitioned into three steps:

Step One: We first show that for any $a \neq b$, there holds

$$h_{a,b}^2 \mathbf{1}_{n_a} - 2h_{a,b}\boldsymbol{u}_{a,b} = \frac{z(n_a + n_b)}{2n_a n_b}\mathbf{1}_{n_a} + \frac{1}{n_b}\boldsymbol{B}^{(a,b)}\mathbf{1}_{n_b}. \tag{6.19}$$

Note that $\langle \boldsymbol{D}^{(a,a)}, \boldsymbol{J}_{n_a\times n_a}\rangle = 2n_a\|\overline{\boldsymbol{X}}_a\|_F^2$. By Lemma 5 and the definition of $\boldsymbol{M}^{(a,b)}$ in (6.9), we have

$$\begin{aligned}\boldsymbol{M}^{(a,b)}\mathbf{1}_{n_b} &= n_b\left(\frac{1}{n_b}\boldsymbol{D}^{(a,b)}\mathbf{1}_{n_b} - \frac{1}{n_a}\boldsymbol{D}^{(a,a)}\mathbf{1}_{n_a}\right)\\ &\quad + \frac{n_b}{2}\left(\frac{1}{n_a^2}\langle \boldsymbol{D}^{(a,a)}, \boldsymbol{J}_{n_a\times n_a}\rangle - \frac{1}{n_b^2}\langle \boldsymbol{D}^{(b,b)}, \boldsymbol{J}_{n_b\times n_b}\rangle\right)\mathbf{1}_{n_a}\\ &= n_b(h_{a,b}^2\mathbf{1}_{n_b} - 2h_{a,b}\boldsymbol{u}_{a,b})\\ &= \frac{n_b z}{2}\left(\frac{1}{n_a}+\frac{1}{n_b}\right)\mathbf{1}_{n_a} + \boldsymbol{B}^{(a,b)}\mathbf{1}_{n_b},\end{aligned}$$

where the last equation follows from (6.15).

Step Two: Next we establish a lower bound for z and show that $z \geq 2\max\|\overline{\boldsymbol{X}}_a\|^2$. Combining $\boldsymbol{Q} = z(\boldsymbol{I}_N - \boldsymbol{E}) + \boldsymbol{M} - \boldsymbol{B} \succeq \mathbf{0}$ with $\boldsymbol{B}^{(a,a)} = \mathbf{0}$ results in

$$\boldsymbol{Q}^{(a,a)} = z\left(\boldsymbol{I}_{n_a} - \frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a}\right) + \boldsymbol{M}^{(a,a)} \succeq \mathbf{0}$$

for all $1 \leq a \leq k$. Also, Lemma 3 and (6.10) imply $\boldsymbol{M}^{(a,a)} = \boldsymbol{M}_{T^\perp}^{(a,a)} = -2\overline{\boldsymbol{X}}_a\overline{\boldsymbol{X}}_a^\top$. Therefore, z cannot be negative and

$$z\boldsymbol{I}_{n_a} \succeq z\left(\boldsymbol{I}_{n_a} - \frac{1}{n_a}\boldsymbol{J}_{n_a\times n_a}\right) \succeq -\boldsymbol{M}^{(a,a)} = 2\overline{\boldsymbol{X}}_a\overline{\boldsymbol{X}}_a^\top,$$

which gives $z \geq 2\max_{1\leq a\leq k}\|\overline{\boldsymbol{X}}_a\|^2$.

Step Three: By applying $\boldsymbol{B} \succeq \mathbf{0}$ and $z \geq 2\max_{1\leq a\leq k}\|\overline{\boldsymbol{X}}_a\|^2$ to (6.19), we get

$$h_{a,b}^2\mathbf{1}_{n_a} - 2h_{a,b}\boldsymbol{u}_{a,b} \geq \frac{z(n_a+n_b)}{2n_an_b}\mathbf{1}_{n_a} \geq \frac{\max\|\overline{\boldsymbol{X}}_a\|^2(n_a+n_b)}{n_an_b}\mathbf{1}_{n_a}.$$

Similarly, we have

$$h_{a,b}^2\mathbf{1}_{n_b} - 2h_{a,b}\boldsymbol{u}_{b,a} \geq \frac{\max\|\overline{\boldsymbol{X}}_a\|^2(n_a+n_b)}{n_an_b}\mathbf{1}_{n_b}.$$

Together they imply

$$h_{a,b}^2 - 2h_{a,b}\tau_{a,b} \geq \frac{\max\|\overline{\boldsymbol{X}}_a\|^2(n_a+n_b)}{n_an_b},$$

where $\tau_{a,b} = \max\{\max(\boldsymbol{u}_{a,b}), \max(\boldsymbol{u}_{b,a})\}$. □

6.6 Proof of Proposition 1

Proof of Proposition 1 It suffices to prove $\min_{1\le i\le n_a}\left\langle \boldsymbol{x}_{a,i}-\frac{\boldsymbol{c}_a+\boldsymbol{c}_b}{2}, \boldsymbol{w}_{b,a}\right\rangle = \frac{1}{2}h_{a,b}-\tau_{a,b}$. For any $1\le i\le n_a$, there holds

$$\begin{aligned}\left\langle \boldsymbol{x}_{a,i}-\frac{\boldsymbol{c}_a+\boldsymbol{c}_b}{2}, \boldsymbol{w}_{b,a}\right\rangle &= \left\langle \boldsymbol{x}_{a,i}-\boldsymbol{c}_a+\frac{\boldsymbol{c}_a-\boldsymbol{c}_b}{2}, \boldsymbol{w}_{b,a}\right\rangle\\ &= \langle \boldsymbol{x}_{a,i}-\boldsymbol{c}_a, \boldsymbol{w}_{b,a}\rangle+\frac{1}{2}\|\boldsymbol{c}_a-\boldsymbol{c}_b\|\\ &= (\overline{\boldsymbol{X}}_a\boldsymbol{w}_{b,a})_i+\frac{1}{2}\|\boldsymbol{c}_a-\boldsymbol{c}_b\|\\ &= -(\boldsymbol{u}_{a,b})_i+\frac{1}{2}\|\boldsymbol{c}_a-\boldsymbol{c}_b\|.\end{aligned}$$

Similarly, for any $1\le j\le n_b$, we have,

$$\left\langle \boldsymbol{x}_{b,j}-\frac{\boldsymbol{c}_a+\boldsymbol{c}_b}{2}, \boldsymbol{w}_{b,a}\right\rangle = -(\boldsymbol{u}_{b,a})_j+\frac{1}{2}\|\boldsymbol{c}_a-\boldsymbol{c}_b\|.$$

Combining those two identities gives

$$\min_{a\ne b}\left\{\frac{1}{2}h_{a,b}-\tau_{a,b}\right\} = \min_{a\ne b}\min_{1\le i\le n_a}\left\langle \boldsymbol{x}_{a,i}-\frac{\boldsymbol{c}_a+\boldsymbol{c}_b}{2}, \boldsymbol{w}_{b,a}\right\rangle,$$

which completes the proof. □

7 Proof for Section 3.2

In this section, we provide concise proofs for Theorem 5 and Theorem 6. The proofs for the balanced case is parallel to the general case to a large extent. To avoid redundancy, we skip proofs and calculations that are basically the same as those in Sect. 6. Also, we adopt similar notation as in Sect. 6 to emphasize the close relation between these two SDP relaxations of k-means.

7.1 Proof of Theorem 5

Amini and Levina's relaxation is equivalent to the following optimization problem:

$$\begin{aligned}\min\quad & \langle \boldsymbol{Z}, \boldsymbol{D}\rangle\\ \text{s.t.}\quad & \boldsymbol{Z}\succeq 0,\quad \boldsymbol{Z}\ge 0,\quad \frac{1}{2}(\boldsymbol{Z}+\boldsymbol{Z}^\top)\mathbf{1}_N=\mathbf{1}_N,\quad \operatorname{diag}(\boldsymbol{Z})=\frac{1}{n}\mathbf{1}_N.\end{aligned}\tag{7.1}$$

In the standard form of a conic program, the optimization takes the form

$$\min \quad \langle \boldsymbol{Z}, \boldsymbol{D} \rangle, \quad \text{s.t.} \quad \mathcal{A}(\boldsymbol{Z}) = \begin{bmatrix} \frac{1}{n}\mathbf{1}_N \\ \mathbf{1}_N \end{bmatrix}, \quad \boldsymbol{Z} \in \mathcal{K}, \tag{7.2}$$

where $\mathcal{K} = \mathcal{S}_+^N \cap \mathbb{R}_+^{N\times N}$ and the linear operator $\mathcal{A}$ is given by

$$\mathcal{A}(\boldsymbol{Z}): \quad \boldsymbol{Z} \to \begin{bmatrix} \operatorname{diag}(\boldsymbol{Z}) \\ \frac{1}{2}(\boldsymbol{Z} + \boldsymbol{Z}^\top)\mathbf{1}_N \end{bmatrix}.$$

Thus, it is effortless to derive the dual problem of Amini and Levina's relaxation using the duality theory of conic programming. The dual program reads

$$\max \quad -\left\langle \frac{1}{n}\boldsymbol{z} + \boldsymbol{\alpha}, \mathbf{1}_N \right\rangle, \quad \text{s.t.} \quad \boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}) \in \mathcal{K}^*, \tag{7.3}$$

where $\boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{\alpha} \end{bmatrix} \in \mathbb{R}^{2N}$ is the dual variable with respect to the affine constraints, $\mathcal{K}^* = \mathcal{S}_+^N + \mathbb{R}_+^{N\times N}$ is the dual cone and

$$\mathcal{A}^*(\boldsymbol{\lambda}) := \frac{1}{2}(\boldsymbol{\alpha}\mathbf{1}_N^\top + \mathbf{1}_N\boldsymbol{\alpha}^\top) + \operatorname{diag}(\boldsymbol{z}) \tag{7.4}$$

is the adjoint operator of $\mathcal{A}$ under the canonical inner product over $\mathbb{R}^{N\times N}$, where $\operatorname{diag}(\boldsymbol{z})$ is the diagonal matrix whose diagonal is given by $\boldsymbol{z}$.

We proceed to find the sufficient condition for $\boldsymbol{X} = \sum_{a=1}^k \frac{1}{n}\mathbf{1}_{\Gamma_a}\mathbf{1}_{\Gamma_a}^\top$ to be the global minimum. Thanks to the conic duality theorem (Theorem 8), we can prove the following lemma using the same construction as in Lemma 1

Lemma 7 *$(\boldsymbol{X}, \boldsymbol{\lambda})$ is a pair of primal/dual optima if and only if the complementary slackness holds: $\langle \boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}), \boldsymbol{X} \rangle = 0$ where $\boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}) \in \mathcal{K}^*$.*

Proof It is easy to verify that $\widetilde{\boldsymbol{Z}} = \frac{1-\lambda}{N}\mathbf{1}_N\mathbf{1}_N^\top + \lambda \boldsymbol{I}_N$ is strictly feasible for (7.2), where $\lambda = \frac{k-1}{N-1} > 0$ for $k \geq 2$. As for the dual problem, we take $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{z} = z\mathbf{1}_N$ where z is a sufficiently large positive number, then $\boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}) = \boldsymbol{J}_{N\times N} + (\boldsymbol{D} + z\boldsymbol{I}_N - \boldsymbol{J}_{N\times N})$ is inside the interior of $\mathcal{K}^*$. □

The task is to find $\boldsymbol{z}$ and $\boldsymbol{\alpha}$ such that the complementary slackness $\langle \boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}), \boldsymbol{X} \rangle = 0$ is true. By definition, $\boldsymbol{D} + \mathcal{A}^*(\boldsymbol{\lambda}) = \boldsymbol{B} + \boldsymbol{Q}$, where $\boldsymbol{B} \geq \mathbf{0}$ and $\boldsymbol{Q} \succeq \mathbf{0}$. We choose $\boldsymbol{z}$ such that

$$\boldsymbol{z}_a = z_a\mathbf{1}_n, \quad \forall 1 \leq a \leq k,$$

where $z_1, \ldots, z_k$ are variables to be determined. In a similar fashion to Lemma 2, the complementary slackness gives

$$\boldsymbol{\alpha}_a = -\frac{2}{n}\boldsymbol{D}^{(a,a)}\mathbf{1}_n + \frac{1}{n^2}\langle \boldsymbol{D}^{(a,a)}, \boldsymbol{J}_{n\times n}\rangle\mathbf{1}_n - \frac{z_a}{n}\mathbf{1}_n.$$

As a result, matrix $\boldsymbol{B}$ must satisfy

$$\boldsymbol{B}^{(a,b)} > \mathbf{0}, \quad \boldsymbol{B}^{(a,a)} = \mathbf{0} \quad \forall a \neq b.$$

The matrix $\boldsymbol{Q}$ is rewritten as

$$\boldsymbol{Q} = \boldsymbol{F} + \boldsymbol{M} - \boldsymbol{B}, \tag{7.5}$$

where $\boldsymbol{M}$ is defined the same as before:

$$\boldsymbol{M}^{(a,b)} = \boldsymbol{D}^{(a,b)} - \frac{1}{n}\left[\boldsymbol{D}^{(a,a)}\boldsymbol{J}_{n\times n} + \boldsymbol{J}_{n\times n}\boldsymbol{D}^{(b,b)}\right] + \frac{1}{2n^2}\langle \boldsymbol{D}^{(a,a)} + \boldsymbol{D}^{(b,b)}, \boldsymbol{J}_{n\times n}\rangle \boldsymbol{J}_{n\times n}.$$

and the matrix $\boldsymbol{F}$ is given by:

$$\boldsymbol{F}^{(a,b)} = -\frac{z_a + z_b}{2n}\boldsymbol{J}_{n\times n}, \quad \boldsymbol{F}^{(a,a)} = z_a\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_{n\times n}\right) \quad \forall a \neq b.$$

Just the same as Proposition 2, the following optimality condition is not enough to guarantee that $\boldsymbol{X}$ is a unique global minimum of (7.1): $\boldsymbol{Q} \succeq \mathbf{0}$ and $\boldsymbol{B} \geq \mathbf{0}$ where $\boldsymbol{Q}$ has the form of (7.5) and $\boldsymbol{B}^{(a,a)} = \mathbf{0}$ for all $1 \leq a \leq k$. However, by following exactly the logic of the proof of Proposition 3, one can show its counterpart for the balanced case is still true:

Proposition 6 (A sufficient condition for the uniqueness of global minimum) *Any feasible pair of $\boldsymbol{Q} \succeq \mathbf{0}$ and $\boldsymbol{B} \geq \mathbf{0}$, where $\boldsymbol{Q}$ has the form of (7.5), $\boldsymbol{B}^{(a,a)} = \mathbf{0}$ for all $1 \leq a \leq k$, and $\boldsymbol{B}^{(a,b)} > \mathbf{0}$ for all $a \neq b$, certifies $\boldsymbol{X}$ to be a unique global minimum of (6.1).*

By following the argument of Proposition 4, we can transform the condition for the uniqueness of global minimum into a more useful form.

Proposition 7 *The optimality condition with uniqueness in Proposition 6 is equivalent to*

$$\begin{aligned}
&\boldsymbol{F}_{T^\perp} + \boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp} \succeq \mathbf{0},\\
&\boldsymbol{M}_T^{(a,b)} - \boldsymbol{B}_T^{(a,b)} - \frac{z_a + z_b}{2n}\boldsymbol{J}_n = \mathbf{0}, \quad \forall a \neq b,\\
&\boldsymbol{B}^{(a,b)} = (\boldsymbol{B}^{(b,a)})^\top, \quad \boldsymbol{B}^{(a,a)} = \mathbf{0}, \quad \boldsymbol{B}^{(a,b)} > \mathbf{0}, \quad \forall a \neq b.
\end{aligned} \tag{7.6}$$

Here, T and $T^\perp$ are subspaces of $\mathbb{R}^{N\times N}$ defined in Sect. 6.3. The only free variables remained in (7.6) are z_a and $\boldsymbol{B}_{T^\perp}^{(a,b)}$. We choose them as

$$z_a = 2k\|\overline{\boldsymbol{X}}_a\|^2, \quad \boldsymbol{B}_{T^\perp}^{(a,b)} = 4\boldsymbol{u}_{a,b}\boldsymbol{u}_{b,a}^\top, \quad \forall a \neq b. \tag{7.7}$$

Now we show that with such a construction leads to Theorem 5. In fact, Theorem 5 follows immediately from the proposition below as an implication of Proposition 6.

Proposition 8 *Assume the proximity condition for balanced clusters* (3.4) *holds for the partition* $\{\Gamma_a\}_{a=1}^k$. *We can choose* z_a *and* $\boldsymbol{B}$ *such that both the sufficient condition* (7.6) *and* (7.7) *are satisfied.*

Proof It remains to prove $\boldsymbol{B}^{(a,b)} > \boldsymbol{0}$ for all $a \neq b$ and $\boldsymbol{F}_{T^\perp} + \boldsymbol{M}_{T^\perp} - \boldsymbol{B}_{T^\perp} \succeq \boldsymbol{0}$. Notice that for all $a \neq b$

$$\begin{aligned}\boldsymbol{B}_T^{(a,b)} &= -\frac{z_a + z_b}{2n}\boldsymbol{J}_{n\times n} + \boldsymbol{M}_T^{(a,b)},\\ \boldsymbol{B}_{T^\perp}^{(a,b)} &= 4\boldsymbol{u}_{a,b}\boldsymbol{u}_{b,a}^\top,\end{aligned}$$

where $\boldsymbol{M}_T^{(a,b)}$ is given by Lemma 6. Then

$$\boldsymbol{B}^{(a,b)} = 4\boldsymbol{u}_{a,b}\boldsymbol{u}_{b,a}^\top + \left(-\frac{z_a + z_b}{2n} + h_{a,b}^2\right)\boldsymbol{J}_{n\times n} - 2h_{a,b}(\boldsymbol{u}_{a,b}\boldsymbol{1}_n^\top + \boldsymbol{1}_n\boldsymbol{u}_{b,a}^\top).$$

As with the proof of (6.18) in Sect. 6.4, it suffices to require

$$h_{a,b} - 2\tau_{a,b} > \sqrt{\frac{z_a + z_b}{2n}} = \sqrt{\frac{k}{n}\left(\|\overline{\boldsymbol{X}}_a\|^2 + \|\overline{\boldsymbol{X}}_b\|^2\right)},$$

which is equivalent to the proximity condition for balanced clusters thanks to Proposition 1.

Next we show $\boldsymbol{F}_{T^\perp} \succeq \boldsymbol{B}_{T^\perp} - \boldsymbol{M}_{T^\perp}$. Based on the proof of Lemma 4, we have $\boldsymbol{M}_{T^\perp}^{(a,b)} = -2\overline{\boldsymbol{X}}_a\overline{\boldsymbol{X}}_b^\top$. Hence, $\boldsymbol{B}_{T^\perp} - \boldsymbol{M}_{T^\perp} = 2\widehat{\boldsymbol{X}}\boldsymbol{W}\widehat{\boldsymbol{X}}^\top$, where $\widehat{\boldsymbol{X}} \in \mathbb{R}^{N\times mk}$ and $\boldsymbol{W} \in \mathbb{R}^{mk\times mk}$ are given by

$$\begin{aligned}&\widehat{\boldsymbol{X}}^{(a,b)} = \boldsymbol{0}, \quad \widehat{\boldsymbol{X}}^{(a,a)} = \overline{\boldsymbol{X}}_a, \quad \forall a \neq b,\\ &\boldsymbol{W}^{(a,b)} = \boldsymbol{I}_m - 2\boldsymbol{w}_{a,b}\boldsymbol{w}_{a,b}^\top, \quad \boldsymbol{W}^{(a,a)} = \boldsymbol{I}_m, \quad \forall a \neq b.\end{aligned}$$

Note that each $\boldsymbol{W}^{(a,b)}$ is an orthogonal matrix and thus $\|\boldsymbol{W}^{(a,b)}\| = 1$. Let $\boldsymbol{y} \in \mathbb{R}^N$ be a unit vector, and denote by $\boldsymbol{y}_a = \{y_i\}_{i\in\Gamma_a}$, $1 \le a \le k$. There holds,

$$\boldsymbol{y}^\top\boldsymbol{W}\boldsymbol{y} \le \sum_{a=1}^k\sum_{b=1}^k \boldsymbol{y}_a^\top\boldsymbol{W}^{(a,b)}\boldsymbol{y}_b^\top \le \left(\sum_{l=1}^k\|\boldsymbol{y}_l\|\right)^2 \le k\left(\sum_{l=1}^k\|\boldsymbol{y}_l\|^2\right) = k.$$

This implies $\boldsymbol{W} \preceq k\boldsymbol{I}_{mk}$, which further implies

$$\boldsymbol{B}_{T^\perp} - \boldsymbol{M}_{T^\perp} \preceq 2k\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}^\top \preceq \boldsymbol{G}, \tag{7.8}$$

where $\boldsymbol{G}$ stands for

$$\boldsymbol{G}^{(a,b)} = \boldsymbol{0}, \quad \boldsymbol{G}^{(a,a)} = z_a\boldsymbol{I}_n, \quad \forall a \neq b.$$

By the definition of $\boldsymbol{F}$, it is easy to verify that $\boldsymbol{F}_{T^\perp} = \boldsymbol{G}_{T^\perp}$. Applying $\mathcal{P}_{T^\perp}$ to both sides of (7.8) yields

$$\boldsymbol{B}_{T^\perp} - \boldsymbol{M}_{T^\perp} \preceq \boldsymbol{F}_{T^\perp}.$$

□

7.2 Proof of Theorem 6

Proof of Theorem 6 Lemma 3 and (6.10) imply $\boldsymbol{M}^{(a,a)} = \boldsymbol{M}_{T^\perp}^{(a,a)} = -2\overline{\boldsymbol{X}}_a\overline{\boldsymbol{X}}_a^\top$. Since $\boldsymbol{Q} \succeq \boldsymbol{0}$, $\boldsymbol{Q}^{(a,a)} \succeq \boldsymbol{0}$ for any a. Using (7.5), we have

$$\boldsymbol{Q}^{(a,a)} = \boldsymbol{F}^{(a,a)} + \boldsymbol{M}^{(a,a)} - \boldsymbol{B}^{(a,a)} = z_a\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_{n\times n}\right) - 2\overline{\boldsymbol{X}}_a\overline{\boldsymbol{X}}_a^\top \succeq \boldsymbol{0}.$$

Thus,

$$z_a\boldsymbol{I}_n \succeq z_a\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_{n\times n}\right) \succeq 2\overline{\boldsymbol{X}}_a\overline{\boldsymbol{X}}_a^\top,$$

which gives $z_a \geq 2\|\overline{\boldsymbol{X}}_a\|^2$. According to Lemma 6, there holds

$$\boldsymbol{M}_T^{(a,b)} = h_{a,b}^2\boldsymbol{J}_{n\times n} - 2h_{a,b}\boldsymbol{u}_{a,b}\boldsymbol{1}_n^\top - 2h_{a,b}\boldsymbol{1}_n\boldsymbol{u}_{b,a}^\top,$$

since for the balanced case $n_a = n$ for any a. Hence,

$$\boldsymbol{M}^{(a,b)}\boldsymbol{1}_n = \boldsymbol{M}_T^{(a,b)}\boldsymbol{1}_n = n(h_{a,b}^2\boldsymbol{1}_n - 2h_{a,b}\boldsymbol{u}_{a,b}).$$

On the other hand, by (7.6), we have

$$\boldsymbol{M}^{(a,b)}\boldsymbol{1}_n = \boldsymbol{M}_T^{(a,b)}\boldsymbol{1}_n = \boldsymbol{B}^{(a,b)}\boldsymbol{1}_n - \frac{z_a + z_b}{2}\boldsymbol{1}_n.$$

Combining the above two equations with the fact that $\boldsymbol{B} \succeq \boldsymbol{0}$, we obtain the following estimation

$$n(h_{a,b}^2\boldsymbol{1}_n - 2h_{a,b}\boldsymbol{u}_{a,b}) = \boldsymbol{B}^{(a,b)}\boldsymbol{1}_n + \frac{z_a + z_b}{2}\boldsymbol{1}_n \geq (\|\overline{\boldsymbol{X}}_a\|^2 + \|\overline{\boldsymbol{X}}_b\|^2)\boldsymbol{1}_n.$$

This is equivalent to

$$h_{a,b}^2 - 2h_{a,b}\tau_{a,b} \geq \frac{(\|\overline{\boldsymbol{X}}_a\|^2 + \|\overline{\boldsymbol{X}}_b\|^2)}{n}.$$

□

8 Proofs for Section 4

In this section, we apply the deterministic guarantee to two typical random models and prove Corollaries 2 and 4. Each of the two models inherits a partition structure from how the data are sampled, which gives a ground truth of the underlying clusters. We will discuss the sufficient condition for the exact recovery of the Peng–Wei relaxation based on the minimal separation between cluster centers.

8.1 Key lemmas

The main mathematical tools for the analysis are various concentration inequalities of random matrices as discussed in [23,26].

Theorem 9 (Matrix Bernstein inequality, Theorem 1.6 in [23]) *Let* $\{\boldsymbol{Z}_i\}_{i=1}^n$ *be a sequence of real* $d_1 \times d_2$ *random matrices. Assume that*

$$\mathbb{E}\boldsymbol{Z}_i = 0, \quad \|\boldsymbol{Z}_i\| \le R, \quad \forall\, 1 \le i \le n.$$

Consider the sum $\boldsymbol{S} = \sum_{i=1}^n \boldsymbol{Z}_i$*, and denote*

$$\sigma^2(\boldsymbol{S}) = \max\left\{\left\|\sum_{i=1}^n \mathbb{E}[\boldsymbol{Z}_i \boldsymbol{Z}_i^\top]\right\|, \left\|\sum_{i=1}^n \mathbb{E}[\boldsymbol{Z}_i^\top \boldsymbol{Z}_i]\right\|\right\}.$$

Then for all $t \ge 0$*,*

$$\mathbb{P}\left(\|\boldsymbol{S}\| \ge t\right) \le (d_1 + d_2) \cdot \exp\left(\frac{-t^2}{2\sigma^2(\boldsymbol{S}) + 2Rt/3}\right).$$

Lemma 8 (Generalized stochastic ball model) *Let* $\{\boldsymbol{a}_i\}_{i=1}^n$ *be a sequence of i.i.d. random vectors in* $\mathbb{R}^m$ *and assume each* $\boldsymbol{a}_i$ *is a zero mean vector supported on the unit ball in* $\mathbb{R}^m$ *with the covariance matrix given by* $\boldsymbol{\Sigma}$*.*

1. *Denote* $\overline{\boldsymbol{a}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{a}_i$*. We have*

$$\mathbb{P}(\|\overline{\boldsymbol{a}}\| \ge t) \le (m+1) \cdot \exp\left(-\frac{nt^2}{2 + 2t/3}\right). \tag{8.1}$$

2. *Let* $\boldsymbol{A}$ *be an* $n \times m$ *matrix whose* i*th row is* $\boldsymbol{a}_i^\top$*. Then*

$$\mathbb{P}(\|\boldsymbol{A}\| \ge \sqrt{n(\|\boldsymbol{\Sigma}\| + t)}) \le 2m \exp\left(-\frac{nt^2}{2 + 4t/3}\right). \tag{8.2}$$

Proof Note that the distribution of each $\boldsymbol{a}_i$ is supported on the unit ball with the covariance matrix given by $\boldsymbol{\Sigma}$. Thus,

$$\sigma^2\left(\sum_{i=1}^{n}\boldsymbol{a}_i\right) = n\max\{\|\boldsymbol{\Sigma}\|, \mathrm{Tr}(\boldsymbol{\Sigma})\} \le n,$$

which follows from $\|\mathbb{E}(\boldsymbol{a}_i\boldsymbol{a}_i^\top)\| = \|\boldsymbol{\Sigma}\|$ and $\|\mathbb{E}(\boldsymbol{a}_i^\top\boldsymbol{a}_i)\| = \mathrm{Tr}(\boldsymbol{\Sigma}) \le 1$. Moreover, there holds $\|\boldsymbol{a}_i\| \le 1$ and thus $R = \max_{1\le i\le n}\|\boldsymbol{a}_i\| = 1$. Therefore, applying Theorem 9 immediately results in

$$\mathbb{P}(\|\overline{\boldsymbol{a}}\| \ge t) \le (m+1)\cdot\exp\left(-\frac{nt^2}{2+2t/3}\right).$$

For the second part, first note that $\|\boldsymbol{A}\|^2 = \|\boldsymbol{A}^\top\boldsymbol{A}\| = \left\|\sum_{i=1}^{n}\boldsymbol{a}_i\boldsymbol{a}_i^\top\right\|$. Let $\boldsymbol{Z}_i = \boldsymbol{a}_i\boldsymbol{a}_i^\top - \boldsymbol{\Sigma}$ be a centered random matrix and its operator norm is controlled by

$$R = \max_{1\le i\le n}\|\boldsymbol{Z}_i\| \le \max_{1\le i\le n}\|\boldsymbol{a}_i\|^2 + \|\boldsymbol{\Sigma}\| \le 2.$$

For the variance of $\boldsymbol{Z}_i$, since $\mathbb{E}(\boldsymbol{Z}_i\boldsymbol{Z}_i^\top) = \mathbb{E}(\boldsymbol{Z}_i^\top\boldsymbol{Z}_i) = \mathbb{E}(\|\boldsymbol{a}_i\|^2\boldsymbol{a}_i\boldsymbol{a}_i^\top) - \boldsymbol{\Sigma}^2$, we have $-\boldsymbol{\Sigma}^2 \preceq \mathbb{E}(\boldsymbol{Z}_i\boldsymbol{Z}_i^\top) \preceq \boldsymbol{\Sigma}$. Therefore,

$$\|\mathbb{E}(\boldsymbol{Z}_i\boldsymbol{Z}_i^\top)\| \le \max\{\|\boldsymbol{\Sigma}\|^2, \|\boldsymbol{\Sigma}\|\} = \|\boldsymbol{\Sigma}\| \le 1$$

and $\sigma^2(\sum_{i=1}^{n}\boldsymbol{Z}_i) \le n$. Applying Theorem 9 again gives

$$\begin{aligned}\mathbb{P}\left(\left\|\sum_{i=1}^{n}\boldsymbol{Z}_i\right\| \ge nt\right) &\le 2m\cdot\exp\left(-\frac{n^2t^2}{2\sigma^2(\boldsymbol{S}) + 2Rnt/3}\right)\\ &\le 2m\cdot\exp\left(-\frac{nt^2}{2+4t/3}\right).\end{aligned}$$

Therefore, since $\|\boldsymbol{A}\|^2 \le \|\sum_{i=1}^{n}\boldsymbol{Z}_i\| + n\|\boldsymbol{\Sigma}\|$, we have

$$\|\boldsymbol{A}\| \le \sqrt{n(\|\boldsymbol{\Sigma}\| + t)}$$

with probability at least $1 - 2m\exp\left(-\frac{nt^2}{2+4t/3}\right)$. □

Lemma 9 (Gaussian mixture model) *Let $\{\boldsymbol{a}_i\}_{i=1}^n$ be a sequence of i.i.d. random vectors in $\mathbb{R}^m$ sampled from multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$.*

1. *Denote $\overline{\boldsymbol{a}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{a}_i$. There holds*

$$\mathbb{P}\left(\|\overline{\boldsymbol{a}}\| \ge \sqrt{\frac{m(1+t)\|\boldsymbol{\Sigma}\|}{n}}\right) \le \max\{e^{-mt/8}, e^{-mt^2/8}\}, \quad \forall t \ge 0. \tag{8.3}$$

2. *Let $\boldsymbol{A}$ be $n \times m$ matrix whose ith row is $\boldsymbol{a}_i^\top$, then for any $t \geq 0$*

$$\mathbb{P}(\|\boldsymbol{A}\| \geq \sqrt{\|\boldsymbol{\Sigma}\|}(\sqrt{n} + \sqrt{m} + t)) \leq 2e^{-t^2/2}. \tag{8.4}$$

3. *Let $\sigma_{\min}$ be the smallest singular value of $\boldsymbol{\Sigma}$, then for any $t \geq 0$*

$$\mathbb{P}((\|\boldsymbol{A}\| \leq \sigma_{\min}(\sqrt{n} - \sqrt{m} - t)) \leq 2e^{-t^2/2}, \tag{8.5}$$

Proof Obviously, the sample mean $\overline{\boldsymbol{a}}$ is a random vector satisfying $\mathcal{N}(\boldsymbol{0}, \frac{1}{n}\boldsymbol{\Sigma})$. Due to the rotational invariance, it can be rewritten as $\overline{\boldsymbol{a}} = \frac{1}{\sqrt{n}}\boldsymbol{\Sigma}^{1/2}\boldsymbol{w}$ where $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_m)$. Note that $\|\boldsymbol{w}\|^2$ is a χ^2_m random variable with $\mathbb{E}(\|\boldsymbol{w}\|^2) = m$ and

$$\mathbb{P}(\|\boldsymbol{w}\|^2 - m \geq t) \leq \exp\left(-\frac{t^2}{8m}\right) \vee \exp\left(-\frac{t}{8}\right).$$

It is easy to see that $\|\overline{\boldsymbol{a}}\| \leq \sqrt{\frac{m(1+t)\|\boldsymbol{\Sigma}\|}{n}}$ holds with probability at least $1 - \max\{e^{-mt/8}, e^{-mt^2/8}\}$.

For the second and the third part, we use similar techniques by first rewriting $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{W}\boldsymbol{\Sigma}^{1/2}$ where $\boldsymbol{W}$ is an $n \times m$ standard Gaussian random matrix. Corollary 5.35 in [26] implies that $\sqrt{n} - \sqrt{m} - t \leq \|\boldsymbol{W}\| \leq \sqrt{n} + \sqrt{m} + t$ holds with probability at least $1 - e^{-t^2/2}$. Therefore,

$$\sigma_{\min}(\sqrt{n} - \sqrt{m} - t) \leq \|\boldsymbol{A}\| \leq \sqrt{\|\boldsymbol{\Sigma}\|}(\sqrt{n} + \sqrt{m} + t)$$

holds with probability at least $1 - 2e^{-t^2/2}$. □

Lemma 10 *For two independent standard Gaussian random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^m$, there holds*

$$\mathbb{P}(\boldsymbol{x}^\top\boldsymbol{\mu} \geq t\|\boldsymbol{\mu}\|) \leq e^{-t^2/2}, \quad \forall t \geq 0, \tag{8.6}$$

for a fixed deterministic vector $\boldsymbol{\mu}$. Also, we have

$$\mathbb{P}(\boldsymbol{x}^\top\boldsymbol{\Psi}\boldsymbol{y} \geq m\sqrt{t(1+t)}\|\boldsymbol{\Psi}\|) \leq 2\max\{e^{-mt/8}, e^{-mt^2/8}\}, \quad \forall t \geq 0, \tag{8.7}$$

for a fixed matrix $\boldsymbol{\Psi}$ and $t \geq 1$. Moreover,

$$\mathbb{P}(\boldsymbol{x}^\top\boldsymbol{\Sigma}\boldsymbol{x} - \mathrm{Tr}(\boldsymbol{\Sigma}) \geq t) \leq \exp\left(-\frac{t^2}{8\|\boldsymbol{\Sigma}\|_F^2}\right) \vee \exp\left(-\frac{t}{8\|\boldsymbol{\Sigma}\|}\right), \quad \forall t \geq 0, \tag{8.8}$$

for a fixed positive semidefinite matrix $\boldsymbol{\Sigma}$.

Proof Note that $\boldsymbol{x}^\top\boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ is a standard Gaussian random variable. For a standard Gaussian random variable g, we have $\mathbb{P}(g \geq t) \leq \frac{1}{2}e^{-t^2/2}$, which can be easily

verified as follows:

$$\begin{aligned}\mathbb{P}(g \geq t) &= \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2} dx \\ &= e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{(x+t)(x-t)}{2}} dx \\ &\leq e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{(x-t)^2}{2}} dx = \frac{1}{2} e^{-t^2/2}.\end{aligned}$$

For (8.7), first note that $\|\boldsymbol{y}\|^2$ is a chi-squared variable with m degree of freedom, hence

$$\|\boldsymbol{\Psi}\boldsymbol{y}\| \leq \|\boldsymbol{\Psi}\| \|\boldsymbol{y}\| \leq \sqrt{m(1+t)} \|\boldsymbol{\Psi}\|$$

holds with probability at least $1 - \max\{e^{-mt/8}, e^{-mt^2/8}\}$. Conditioned on the event $\{\|\boldsymbol{\Psi}\boldsymbol{y}\| \leq \sqrt{m(1+t)}\|\boldsymbol{\Psi}\|\}$, $\boldsymbol{x}^\top \boldsymbol{\Psi}\boldsymbol{y}$ is a Gaussian random variable with variance at most $m(1+t)\|\boldsymbol{\Psi}\|^2$. As a result,

$$\mathbb{P}(\boldsymbol{x}^\top \boldsymbol{\Psi}\boldsymbol{y} \geq m\sqrt{t(1+t)}\|\boldsymbol{\Psi}\|) \leq e^{-mt/2}$$

and $\boldsymbol{x}^\top \boldsymbol{\Psi}\boldsymbol{y} \geq m\sqrt{t(1+t)}$ holds with probability at least $1 - 2\max\{e^{-mt/8}, e^{-mt^2/8}\}$.

For (8.8), we use the rotational invariance as well as the eigen-decomposition of $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\Sigma} = \boldsymbol{U}^\top \operatorname{diag}(\lambda_1, \ldots, \lambda_m)\boldsymbol{U}$ with $\lambda_i \geq 0$ for $1 \leq i \leq m$. Therefore, $\boldsymbol{x}^\top \boldsymbol{\Sigma}\boldsymbol{x}$ is the sum of weighted χ_1^2 random variables where

$$\boldsymbol{x}^\top \boldsymbol{\Sigma}\boldsymbol{x} = \sum_{i=1}^m \lambda_i \xi_i^2, \quad \xi_i = (\boldsymbol{U}\boldsymbol{x})_i, \quad \mathbb{E}(\boldsymbol{x}^\top \boldsymbol{\Sigma}\boldsymbol{x}) = \operatorname{Tr}(\boldsymbol{\Sigma}).$$

After applying Bernstein inequality, we get the desired result where $\max_i \lambda_i = \|\boldsymbol{\Sigma}\|$ and $\sum_{i=1}^m \lambda_i^2 = \|\boldsymbol{\Sigma}\|_F^2$. □

8.2 Stochastic ball model

In this subsection, we prove Corollary 2 for the generalized stochastic ball model. It extends the results in [5,12,13] where the probability distributions are assumed to the same and isotropic for all the clusters. The question is how large the minimal separation $\Delta = \min_{a \neq b} \|\boldsymbol{\mu}_a - \boldsymbol{\mu}_b\|$ should be in order to to ensure the exact recovery of the Peng–Wei relaxation with high probability. An outline of the proof of Corollary 3 is also given at the end of the subsection.

Proof of Corollary 2 It suffices to estimate $\|\overline{\boldsymbol{X}}_a\|$, $h_{a,b}$ and $\tau_{a,b}$ for all $a \neq b$. We will bound those quantities on the premise that (8.1) and (8.2), i.e.,

$$\boldsymbol{X}_a - \boldsymbol{1}_{n_a} \boldsymbol{\mu}_a^\top \| \leq \sqrt{n_a(\|\boldsymbol{\Sigma}_a\| + t)} \quad \text{and} \quad \|\boldsymbol{c}_a - \boldsymbol{\mu}_a\| \leq t, \tag{8.9}$$

hold for all $1 \le a \le k$ with probability for all $1 \le a \le k$, at least $1 - 4km\exp(-\frac{Nw_{\min}t^2}{2+4t/3})$. Estimation of $\|\overline{\boldsymbol{X}}_a\|$: By the triangle inequality, the operator norm of $\overline{\boldsymbol{X}}_a$ can be bounded from above as

$$\begin{aligned}\|\overline{\boldsymbol{X}}_a\| &= \|\boldsymbol{X}_a - \mathbf{1}_{n_a}\boldsymbol{c}_a^\top\| \\ &\le \|\boldsymbol{X}_a - \mathbf{1}_{n_a}\boldsymbol{\mu}_a^\top\| + \sqrt{n_a}\|\boldsymbol{c}_a - \boldsymbol{\mu}_a\| \\ &\le \sqrt{n_a(\|\boldsymbol{\Sigma}_a\| + t)} + t\sqrt{n_a}\end{aligned}$$

for all $1 \le a \le k$ with probability at least $1 - 4km\exp\left(-\frac{Nw_{\min}t^2}{2+4t/3}\right)$.

Estimation of $\tau_{a,b}$ and $h_{a,b}$: Recall that $\tau_{a,b} = \max\{\max\{\overline{\boldsymbol{X}}_a\boldsymbol{w}_{a,b}\}, \max\{\overline{\boldsymbol{X}}_b\boldsymbol{w}_{b,a}\}\}$. For each entry of $\overline{\boldsymbol{X}}_a\boldsymbol{w}_{a,b}$, we have

$$(\overline{\boldsymbol{X}}_a\boldsymbol{w}_{a,b})_i \le \|\boldsymbol{x}_{a,i} - \boldsymbol{\mu}_a\| + \|\boldsymbol{c}_a - \boldsymbol{\mu}_a\| \le 1 + t$$

which follows from $\|\boldsymbol{x}_{a,i} - \boldsymbol{\mu}_a\| \le 1$ and (8.9). A similar bound holds for $\overline{\boldsymbol{X}}_b\boldsymbol{w}_{b,a}$ and thus under the event where (8.9) holds, $\tau_{a,b} \le 1 + t$ holds for all $a \ne b$ with probability at least $1 - 4km\exp(-\frac{Nw_{\min}t^2}{2+4t/3})$.

For $h_{a,b}$, it has a simple lower bound:

$$h_{a,b} = \|\boldsymbol{c}_a - \boldsymbol{c}_b\| \ge \|\boldsymbol{\mu}_a - \boldsymbol{\mu}_b\| - \|\boldsymbol{c}_a - \boldsymbol{\mu}_a\| - \|\boldsymbol{c}_b - \boldsymbol{\mu}_b\| \ge \Delta - 2t.$$

Therefore, a lower bound of $\frac{1}{2}h_{a,b} - \tau_{a,b}$ is

$$\frac{1}{2}h_{a,b} - \tau_{a,b} \ge \frac{1}{2}\Delta - t - (1+t) = \frac{1}{2}\Delta - 2t - 1,$$

which holds uniformly over all (a, b) with probability at least $1-4km\exp(-\frac{Nw_{\min}t^2}{2+4t/3})$. Proximity condition for stochastic ball model: Now we wrap up our discussion and apply the proximity condition (1.2). For each a, it follows from $\|\overline{\boldsymbol{X}}_a\| \le (\sqrt{\|\boldsymbol{\Sigma}_a\| + t} + t)\sqrt{n_a}$ that

$$\begin{aligned}\sum_{a=1}^{k}\|\overline{\boldsymbol{X}}_l\|^2 &\le \sum_{a=1}^{k}(\|\boldsymbol{\Sigma}_a\| + t + 2t\sqrt{\|\boldsymbol{\Sigma}_a\| + t} + t^2)n_a \\ &\le (\sigma_{\max}^2 + t + 2t(\sigma_{\max} + \sqrt{t}) + t^2)N \\ &\le \left[(\sigma_{\max} + t)^2 + t + 2t^{3/2}\right]N,\end{aligned}$$

where the second line follows from $\|\boldsymbol{\Sigma}_a\| \le \sigma_{\max}^2$ and $\sqrt{\|\boldsymbol{\Sigma}_a\| + t} \le \sqrt{\|\boldsymbol{\Sigma}_a\|} + \sqrt{t}$.

Therefore, for all pairs of a and b, the proximity condition (1.2) for the generalized stochastic ball model is guaranteed if

$$\Delta \ge 2 + 4t + \sqrt{\frac{2\left((\sigma_{\max} + t)^2 + t + 2t^{3/2}\right)}{w_{\min}}}, \tag{8.10}$$

which holds with probability at least $1 - 4km\exp(-\frac{Nw_{\min}t^2}{2+4t/3})$. Now we choose $t = \sqrt{\frac{4\log(4kmN^\gamma)}{Nw_{\min}}}$. We further assume that $N \geq \frac{4}{w_{\min}}\log(4kmN^\gamma)$, then $t \leq 1$ and (8.10) holds with probability at least

$$1 - 4km\exp\left(-\frac{Nw_{\min}\cdot t^2}{2+4t/3}\right) \geq 1 - 4km\exp\left(-\frac{1}{4}Nw_{\min}\cdot t^2\right) \geq 1 - N^{-\gamma}.$$

Note that $w_{\min} \leq \frac{1}{k} \leq \frac{1}{2}$ and $t \leq 1$. By enlarging the right hand side of (8.10) as the following,

$$\begin{aligned} 2 + 4t + \sqrt{\frac{2\left((\sigma_{\max}+t)^2 + t + 2t^{3/2}\right)}{w_{\min}}} &\leq 2 + \sqrt{\frac{2}{w_{\min}}}\sigma_{\max} + \sqrt{\frac{t}{w_{\min}}} \\ + (4 + \sqrt{\frac{2}{w_{\min}}})t + \sqrt{\frac{2t^{3/2}}{w_{\min}}} &\leq 2 + \sqrt{\frac{2}{w_{\min}}}\sigma_{\max} + 7\sqrt{\frac{t}{w_{\min}}}, \end{aligned}$$

we derive a sufficient condition of (8.10) which guarantees the proximity condition (1.2) for the stochastic ball models with probability at least $1 - N^{-\gamma}$:

$$\Delta \geq 2 + \sqrt{\frac{2}{w_{\min}}}\sigma_{\max} + 7\sqrt{\frac{t}{w_{\min}}}.$$

In particular, if $n_a = n$ for all a and each $\mathcal{D}_a$ is the uniform distribution over $\mathbb{R}^m$, there holds $\sigma_{\max}^2 = \|\mathbf{\Sigma}_a\| = \frac{1}{m+2}$ and (8.10) can be simplified into

$$\Delta \geq 2 + \sqrt{\frac{2k}{m+2}} + 7\sqrt{tk}$$

which completes the proof. □

The necessary lower bound (Theorem 3) can also be applied to the generalized stochastic ball model. For the sake of simplicity, we restrict our discussion to the special case where distributions are all uniform distributions over the unit balls and clusters are balanced, i.e., $n_a = n$, $\forall 1 \leq a \leq k$.

Proof outline of Corollary 3 For each pair of a and b, $\tau_{a,b} > 1-\epsilon$ with high probability for any $\epsilon > 0$, provided that N is large. As for the operator norms, Theorem 5.41 in [26] implies that $\|\overline{\boldsymbol{X}}_a\| \geq (1-\epsilon)\sqrt{\frac{n}{m+2}}$ with high probability. Simple calculations show that the necessary lower bound (3.1) is equivalent to

$$h_{a,b} \geq \tau_{a,b} + \sqrt{\tau_{a,b}^2 + \frac{2}{n}\max\|\overline{\boldsymbol{X}}_a\|^2}, \quad \forall a \neq b. \tag{8.11}$$

Adding up all these together, we yield the necessary lower bound for the special case as in Corollary 3. □

8.3 Gaussian mixture model

In this subsection, we prove Corollary 4 for the Gaussian mixture model. We still focus on the minimal separation condition for the exactness of the Peng–Wei relaxation. Denote $p(t) = \max\{e^{-mt/8}, e^{-mt^2/8}\}$.

Proof of Corollary 4 Let N be the number of points drawn from the Gaussian mixture model and n_a be the number of points belonging to $\mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$. To simplify our analysis, we assume $n_a = w_a N$ and $\boldsymbol{x}_{a,i} \sim \mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$ for all $1 \le a \le k$.

Estimation of $\|\overline{\boldsymbol{X}}_a\|$: Let $\boldsymbol{X}_a \in \mathbb{R}^{n_a \times m}$ be the data drawn from $\mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$. Lemma 9 states that the sample mean $\boldsymbol{c}_a = \frac{1}{n_a}\sum_{i=1}^{n_a} \boldsymbol{x}_{a,i}$ satisfies $\|\boldsymbol{c}_a - \boldsymbol{\mu}_a\| \le \sqrt{\frac{m(1+t)\|\boldsymbol{\Sigma}_a\|}{n_a}}$ for all a with probability at least $1 - k \cdot p(t)$. Considering $\|\overline{\boldsymbol{X}}_a\|$, it obeys

$$\begin{aligned}\|\overline{\boldsymbol{X}}_a\| &\le \|\boldsymbol{X}_a - \mathbf{1}_{n_a}\boldsymbol{\mu}_a^\top\| + \sqrt{n_a}\|\boldsymbol{c}_a - \boldsymbol{\mu}_a\| \\ &\le \sqrt{\|\boldsymbol{\Sigma}_a\|}(\sqrt{n_a} + \sqrt{m} + \sqrt{mt} + \sqrt{m(1+t)}) \\ &\le \sqrt{\|\boldsymbol{\Sigma}_a\|}(\sqrt{n_a} + 2\sqrt{m}(1+\sqrt{t}))\end{aligned}$$

for all $1 \le a \le k$ with probability at least $1 - 2ke^{-mt/2}$, where we have used (8.4) in the second line. It follows that

$$\begin{aligned}\frac{\sum_{l=1}^k \|\overline{\boldsymbol{X}}_l\|^2(n_a+n_b)}{4n_a n_b} &\le \frac{1}{2N}\left(\sum_{l=1}^k \|\boldsymbol{\Sigma}_l\|\,(n_a + 8m(1+t))\right)\left(\frac{1}{w_a}+\frac{1}{w_b}\right) \\ &\le \frac{\sigma_{\max}^2}{N w_{\min}}\,(N + 8km(1+t)) \\ &\le \frac{\sigma_{\max}^2}{w_{\min}}\left(1 + \frac{8km(1+t)}{N}\right),\end{aligned}$$

where $w_{\min} = \frac{1}{N}\min_{1\le l\le k} n_l$ and $w_{\min} \le \frac{1}{k}$.

Therefore, for all $a \ne b$ and all $t \ge 0$, the right hand side of (1.2) is bounded from above by

$$\begin{aligned}\sqrt{\frac{\sum_{l=1}^k \|\overline{\boldsymbol{X}}_l\|^2(n_a+n_b)}{4n_a n_b}} &\le \sqrt{\frac{\sigma_{\max}^2}{w_{\min}}\left(1 + \frac{8km(1+t)}{N}\right)} \\ &\le \frac{\sigma_{\max}}{\sqrt{w_{\min}}}\left(1 + \sqrt{\frac{8km(1+t)}{N}}\right) \qquad (8.12)\end{aligned}$$

with probability at least $1 - k \cdot p(t) - 2ke^{-mt/2}$, which is greater than $1 - 3k \cdot p(t)$.

Estimation of $\tau_{a,b}$ and $h_{a,b}$: For $h_{a,b}$, it follows from Lemma 9 that

$$\begin{aligned} h_{a,b} = \|\boldsymbol{c}_a - \boldsymbol{c}_b\| &\geq \|\boldsymbol{\mu}_a - \boldsymbol{\mu}_b\| - \|\boldsymbol{c}_a - \boldsymbol{\mu}_a\| - \|\boldsymbol{c}_b - \boldsymbol{\mu}_b\| \\ &\geq \|\boldsymbol{\mu}_a - \boldsymbol{\mu}_b\| - \sqrt{m(1+t)\sigma_{\max}^2}\left(\frac{1}{\sqrt{n_a}} + \frac{1}{\sqrt{n_b}}\right) \\ &\geq \|\boldsymbol{\mu}_a - \boldsymbol{\mu}_b\| - 2\sigma_{\max}\sqrt{\frac{m(1+t)}{N w_{\min}}} \end{aligned} \tag{8.13}$$

holds with probability at least $1 - 2ke^{-mt/8}$ for any a and b. Further assume $N \geq \frac{16\sigma_{\max}^2 m(1+t)}{\Delta^2 w_{\min}}$, then

$$h_{a,b} \geq \frac{\|\boldsymbol{\mu}_a - \boldsymbol{\mu}_b\|}{2}. \tag{8.14}$$

Note that $\boldsymbol{u}_{a,b}$ is defined as $\boldsymbol{u}_{a,b} = \overline{\boldsymbol{X}}_a \boldsymbol{w}_{a,b}$ and each entry of $\boldsymbol{u}_{a,b}$ is given by $(\boldsymbol{u}_{a,b})_i = \frac{1}{h_{a,b}}(\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top(\boldsymbol{c}_a - \boldsymbol{c}_b)$. To get an upper bound for $\boldsymbol{u}_{a,b}$, it suffices to bound $(\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top(\boldsymbol{c}_a - \boldsymbol{c}_b)$, which can be partitioned into three terms:

$$(\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top(\boldsymbol{c}_a - \boldsymbol{c}_b) = \underbrace{(\boldsymbol{x}_{a,i} - \boldsymbol{\mu}_a)^\top(\boldsymbol{c}_a - \boldsymbol{\mu}_a)}_{J_1} + \underbrace{(\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top(\boldsymbol{\mu}_a - \boldsymbol{c}_b)}_{J_2} - \|\boldsymbol{c}_a - \boldsymbol{\mu}_a\|^2.$$

1. For J_1, note that $\boldsymbol{x}_{a,i} - \boldsymbol{\mu}_a$ and $\boldsymbol{c}_a - \boldsymbol{\mu}_a$ are not completely independent from each other. Thus we further decompose J_1 into

$$(\boldsymbol{x}_{a,i} - \boldsymbol{\mu}_a)^\top(\boldsymbol{c}_a - \boldsymbol{\mu}_a) = \frac{1}{n_a}\|\boldsymbol{x}_{a,i} - \boldsymbol{\mu}_a\|^2 + \frac{1}{n_a}(\boldsymbol{x}_{a,i} - \boldsymbol{\mu}_a)^\top\left(\sum_{j\neq i}(\boldsymbol{x}_{a,j} - \boldsymbol{\mu}_a)\right).$$

For the first term above, (8.8) implies $\|\boldsymbol{x}_{a,i} - \boldsymbol{\mu}_a\|^2 \leq m(1+t)\|\boldsymbol{\Sigma}_a\|$ with probability at least $1 - e^{-mt/8}$. For the second term, we can reformulate it as

$$\frac{1}{n_a}(\boldsymbol{x}_{a,i} - \boldsymbol{\mu}_a)^\top\left(\sum_{j\neq i}(\boldsymbol{x}_{a,j} - \boldsymbol{\mu}_a)\right) = \left\langle \boldsymbol{w}, \frac{1}{n_a}\boldsymbol{\Sigma}_a^{1/2}\sum_{j\neq i}(\boldsymbol{x}_{a,j} - \boldsymbol{\mu}_a)\right\rangle$$

where $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_m)$ and $\boldsymbol{w}$ is independent of $\frac{1}{n_a}\sum_{j\neq i}(\boldsymbol{x}_{a,j} - \boldsymbol{\mu}_a) \sim \mathcal{N}\left(\boldsymbol{0}, \frac{n_a - 1}{n_a^2}\boldsymbol{\Sigma}_a\right)$. Applying (8.7) implies

$$(\boldsymbol{x}_{a,i} - \boldsymbol{\mu}_a)^\top\left(\sum_{j\neq i}(\boldsymbol{x}_{a,j} - \boldsymbol{\mu}_a)\right) \leq m\|\boldsymbol{\Sigma}_a\|\sqrt{\frac{t(1+t)}{n_a}}$$

with probability at least $1 - 2 \cdot p(t)$. So we can conclude that

$$J_1 \leq m\|\boldsymbol{\Sigma}_a\| \left(\frac{1+t}{n_a} + \sqrt{\frac{t(1+t)}{n_a}} \right)$$

for all a with probability at least $1 - 3N \cdot p(t)$, for all $t \geq 0$.

2. For J_2, we decompose it into two terms:

$$(\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top (\boldsymbol{\mu}_a - \boldsymbol{c}_b) = (\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b) + (\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top (\boldsymbol{\mu}_b - \boldsymbol{c}_b).$$

Since $(\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b) \sim \mathcal{N}(0, \frac{n_a - 1}{n_a} (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)^\top \boldsymbol{\Sigma}_a (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b))$, (8.6) indicates

$$(\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b) \leq \sqrt{s(\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)^\top \boldsymbol{\Sigma}_a (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)}$$

for all (a, b, i) with probability at least $1 - kNe^{-s/2}$. On the other hand, (8.7) directly gives

$$(\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top (\boldsymbol{\mu}_b - \boldsymbol{c}_b) \leq m\sqrt{\frac{t(1+t)\|\boldsymbol{\Sigma}_a\|\|\boldsymbol{\Sigma}_b\|}{n_b}} \leq m\sigma_{\max}^2 \sqrt{\frac{t(1+t)}{n_b}}$$

for all (a, b, i) with probability at least $1 - 2kN \cdot p(t)$. Therefore,

$$J_2 \leq \sqrt{s(\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)^\top \boldsymbol{\Sigma}_a (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)} + m\sigma_{\max}^2 \sqrt{\frac{t(1+t)}{n_b}}$$

holds with probability at least $1 - 2kN \cdot p(t) - kNe^{-s/2}$, for all $s, t \geq 0$.

Using the estimation of J_1 and J_2, we can see that, for all (a, b, i),

$$(\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top (\boldsymbol{c}_a - \boldsymbol{c}_b) \leq \sqrt{s(\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)^\top \boldsymbol{\Sigma}_a (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)} + 3m\sigma_{\max}^2 \frac{1+t}{\sqrt{\min\{n_a, n_b\}}}$$

holds with probability at least $1 - kN(4 \cdot p(t) + e^{-s/2})$. Since $(\boldsymbol{u}_{a,b})_i = \frac{1}{h_{a,b}} (\boldsymbol{x}_{a,i} - \boldsymbol{c}_a)^\top (\boldsymbol{c}_a - \boldsymbol{c}_b)$,, if $N \geq \frac{16\sigma_{\max}^2 m(1+t)}{\Delta^2 w_{\min}}$, then by (8.14) there hold,

$$\tau_{a,b} = \max\{\max\{\boldsymbol{u}_{a,b}\}, \max\{\boldsymbol{u}_{b,a}\}\} \leq 2\sqrt{s}\sigma_{\max} + \frac{6m\sigma_{\max}^2(1+t)}{\Delta\sqrt{N w_{\min}}}. \tag{8.15}$$

Proximity condition for Gaussian mixture model By combing (8.12), (8.13) and (8.15), we have shown the proximity condition is satisfied with probability at least $1 - kN(5 \cdot p(t) + e^{-s/2})$ if

$$\Delta \geq \frac{2\sigma_{\max}}{\sqrt{w_{\min}}} + 4\sigma_{\max}\sqrt{s} + 2\sigma_{\max}(4\sqrt{k}+1)\sqrt{\frac{m(1+t)}{Nw_{\min}}} + \frac{6m\sigma_{\max}^2(1+t)}{\Delta\sqrt{Nw_{\min}}},$$

provided that $N \geq \frac{16\sigma_{\max}^2 m(1+t)}{\Delta^2 w_{\min}}$. These two inequalities are in turn implied by

$$\Delta \geq \frac{2\sigma_{\max}}{\sqrt{w_{\min}}} + 4\sigma_{\max}\sqrt{s} + 10\sigma_{\max}\sqrt{\frac{km(1+t)}{Nw_{\min}}} + \frac{6m\sigma_{\max}(1+t)}{\sqrt{N}} \tag{8.16}$$

Here by choosing $t = \max\left\{8\log(kN^{1+\gamma})/m, \sqrt{8\log(kN^{1+\gamma})/m}\right\}$ and $s = 2\log(kN^{1+\gamma})$ where $\gamma > 0$, then the proximity condition holds with probability at least

$$1 - kN(5 \cdot p(t) + e^{-s/2}) \geq 1 - 6N^{-\gamma}.$$

To simplify the expression, we assume $N = (m^2k^2\log(k)/w_{\min})u$, where $u \gg 1$. Denote $q(N; m, k, w_{\min})$ the sum of the last two terms of (8.16) divided by $\sigma_{\max}$. We have the following asymptotic analysis:

$$q(N; m, k, w_{\min}) \leq \sqrt{\mathcal{O}\left(\frac{1+\log(km)+\log(u)}{kmu}\right)} + \mathcal{O}\left(\frac{1}{\sqrt{u}} + \frac{\log(k)}{k\sqrt{u}} + \frac{\log(N)}{\sqrt{N}}\right) = o(1).$$

This completes the Proof of Corollary 4. □

References

1. Achlioptas, D., McSherry, F.: On spectral learning of mixtures of distributions. In: International Conference on Computational Learning Theory, pp. 458–469. Springer, New York (2005)
2. Aloise, D., Deshpande, A., Hansen, P., Popat, P.: NP-hardness of Euclidean sum-of-squares clustering. Mach. Learn. **75**(2), 245–248 (2009)
3. Amini, A.A., Levina, E.: On semidefinite relaxations for the block model. Ann. Stat. **46**(1), 149–179 (2018)
4. Arthur, D., Manthey, B., Röglin, H.: Smoothed analysis of the k-means method. J. ACM (JACM) **58**(5), 19 (2011)
5. Awasthi, P., Bandeira, A.S., Charikar, M., Krishnaswamy, R., Villar, S., Ward, R.: Relax, no need to round: integrality of clustering formulations. In: Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, pp. 191–200. ACM (2015)
6. Awasthi, P., Sheffet, O.: Improved spectral-norm bounds for clustering. In: APPROX-RANDOM, pp. 37–49. Springer, New York (2012)
7. Ben-Tal, A., Nemirovski, A.: Lectures on modern convex optimization: analysis, algorithms, and engineering applications. SIAM (2001)
8. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
9. Dasgupta, S.: Learning mixtures of gaussians. In: 40th Annual Symposium on Foundations of Computer Science, pp. 634–644. IEEE (1999)
10. Du, Q., Faber, V., Gunzburger, M.: Centroidal Voronoi tessellations: applications and algorithms. SIAM Rev. **41**(4), 637–676 (1999)
11. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. The Johns Hopkins University Press, Baltimore (1996)
12. Iguchi, T., Mixon, D.G., Peterson, J., Villar, S.: On the tightness of an SDP relaxation of k-means (2015). arXiv preprint arXiv:1505.04778

13. Iguchi, T., Mixon, D.G., Peterson, J., Villar, S.: Probably certifiably correct *k*-means clustering. Math. Progr. **165**(2), 605–642 (2017)
14. Kannan, R., Vempala, S.: Spectral algorithms. Found. Trends Theor. Comput. Sci. **4**(3–4), 157–288 (2009)
15. Kumar, A., Kannan, R.: Clustering with spectral norm and the k-means algorithm. In: 2010 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 299–308. IEEE (2010)
16. Ling, S., Strohmer, T.: Certifying global optimality of graph cuts via semidefinite relaxation: a performance guarantee for spectral clustering (2018). arXiv preprint arXiv:1806.11429
17. Lloyd, S.: Least squares quantization in PCM. IEEE Trans. Inf. Theory **28**(2), 129–137 (1982)
18. Lu, Y., Zhou, H.H.: Statistical and computational guarantees of Lloyd's algorithm and its variants (2016). arXiv preprint arXiv:1612.02099
19. Mahajan, M., Nimbhorkar, P., Varadarajan, K.: The planar *k*-means problem is NP-hard. In: International Workshop on Algorithms and Computation, pp. 274–285. Springer, New York (2009)
20. Mixon, D.G., Villar, S., Ward, R.: Clustering subgaussian mixtures by semidefinite programming. Inf. Inference: J. IMA **6**(4), 389–415 (2017)
21. Peng, J., Wei, Y.: Approximating k-means-type clustering via semidefinite programming. SIAM J. Opt. **18**(1), 186–205 (2007)
22. Selim, S.Z., Ismail, M.A.: k-Means-type algorithms: a generalized convergence theorem and characterization of local optimality. IEEE Trans. Pattern Anal. Mach. Intell. **6**(1), 81–87 (1984)
23. Tropp, J.A.: User-friendly tail bounds for sums of random matrices. Found. Comput. Math. **12**(4), 389–434 (2012)
24. Vattani, A.: *k*-Means requires exponentially many iterations even in the plane. Discrete Comput. Geom. **45**(4), 596–616 (2011)
25. Vempala, S., Wang, G.: A spectral algorithm for learning mixture models. J. Comput. Syst. Sci. **68**(4), 841–860 (2004)
26. Vershynin, R.: Introduction to the non-asymptotic analysis of random matrices. In: Eldar, Y.C., Kutyniok, G. (eds.) Compressed Sensing: Theory and Applications, Chapter 5. Cambridge University Press, Cambridge (2012)
27. Wright, S.J.: Primal-dual interior-point methods. SIAM (1997)
28. Yang, L., Sun, D., Toh, K.-C.: SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. Math. Progr. **7**(3), 331–366 (2015)
29. Zhao, X.-Y., Sun, D., Toh, K.-C.: A Newton-CG augmented Lagrangian method for semidefinite programming. SIAM J. Opt. **20**(4), 1737–1765 (2010)

Affiliations

Xiaodong Li[1] · Yang Li[2] · Shuyang Ling[3] · Thomas Strohmer[2] · Ke Wei[4]

✉ Yang Li
ly@math.ucdavis.edu

Xiaodong Li
xdgli@ucdavis.edu

Shuyang Ling
sling@cims.nyu.edu

Thomas Strohmer
strohmer@math.ucdavis.edu

Ke Wei
kewei@fudan.edu.cn

[1] Department of Statistics, University of California Davis, Davis, CA 95616, USA

[2] Department of Mathematics, University of California Davis, Davis, CA 95616, USA

[3] Courant Institute of Mathematical Sciences and Center for Data Science, New York, NY 10012, USA

[4] School of Data Science, Fudan University, Shanghai 200433, China