A CONNECTIONS TO ADVERSARIAL CLUSTERING

We consider clustering problems that arise naturally from stochastic mixture models such as Gaussian, Mallows, categorical and so on (Sanjeev & Kannan, 2001; Vempala & Wang, 2004; Lu & Boutilier, 2011; Charikar et al., 2017; Diakonikolas et al., 2018; Liu & Moitra, 2018). We can then formulate such a clustering problem in the Latent Simplex Model as follows: Given n data points $\mathbf{A}_{*,1}, \mathbf{A}_{*,2}, \ldots, \mathbf{A}_{*,n} \in \mathbb{R}^d$, such that the data is a mixture of k distinct clusters, $\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_k$, with means $\mathbf{M}_{*,1}, \mathbf{M}_{*,2}, \ldots, \mathbf{M}_{*,k}$, the goal is to approximately learn the means. Further, we can set each of the n latent vectors $\mathbf{P}_{*,j}$ to denote the mean of the cluster that the point $\mathbf{A}_{*,j}$ belongs to, and thus $\mathbf{P}_{*,j} \in \{\mathbf{M}_{*,1}, \mathbf{M}_{*,2}, \ldots, \mathbf{M}_{*,k}\}$. Prior work of (Kumar & Kannan, 2010) and (Awasthi & Sheffet, 2012) shows that if the minimum cluster size is δn and for all $\ell \neq \ell'$, $\|\mathbf{M}_{*,\ell} - \mathbf{M}_{*,\ell'}\| \geq ck\frac{\sigma}{\sqrt{\delta}}$ the $\mathbf{M}_{*,\ell}$ can be found within error $O(\sqrt{k}\sigma/\sqrt{\delta})$.

However, the aforementioned algorithms are not robust to adversarial perturbations. Therefore, we describe the perturbations we can handle in the Latent Simplex Model. The adversarial model is the same as the one considered in (Bhattacharyya & Kannan, 2020). The adversary is allowed to select a subset S_ℓ of each cluster $\mathbf{C}_\ell$ of cardinality at most δn and perturb each point $\mathbf{A}_{*,j}$ for $j \in S_\ell$ by Δ_j such that :

- $\mathbf{P}_{*,j} + \Delta_j$ is still in the convex hull of $\mathbf{M}_{*,1}, \mathbf{M}_{*,2}, \ldots, \mathbf{M}_{*,k}$
- The norm of the perturbation is bounded, i.e., $|\Delta_j|_2 \leq 4\sigma/\sqrt{\delta}$.

Intuitively, the adversary can move a $1-\delta$ fraction of the data points in each cluster an arbitrary amount towards the convex hull of the means of the remaining clusters. For the remaining δn points, the perturbation should have norm at most $O(\sigma/\sqrt{\delta})$. The goal is to still learn the means $\mathbf{M}_{*,\ell}$ approximately. (Bhattacharyya & Kannan, 2020) shows that the aforementioned model satisfies Well-Separateness, Proximate Latent Points and Spectrally Bounded Perturbations assumptions. The proof for the Significant Singular Values assumption follows from Lemma 2.1.

B FULL ANALYSIS

We first give the proof of Lemma 2.1.

Proof. Assumptions (2) and (3) follow from Lemma 7.1 in (Bhattacharyya & Kannan, 2020). By Claim 8.1 in (Bhattacharyya & Kannan, 2020), $\sigma_k(\mathbf{A}) \geq c\alpha\sqrt{\delta/k}\min_\ell \mathbf{M}_{*,\ell}$. Each column of $\mathbf{A}$ sums to 1, so $\|\mathbf{A}\|_F^2 = O(n)$ and $\sigma_k(\mathbf{A}) \geq \alpha\sqrt{\delta/k}\|\mathbf{A}\|_F$. Since $\|\mathbf{A} - \mathbf{P}\|_2 \leq \sigma\sqrt{n}$ by definition of σ, and $\mathbf{P}$ consists of n point in the convex hull of k points and thus $\sigma_{k+1}(\mathbf{P}) = 0$, we have $\sigma_{k+1}(\mathbf{A}) \leq \sigma_{k+1}(\mathbf{P}) + \|\mathbf{A} - \mathbf{P}\|_2 \leq \sigma\sqrt{n} \leq \sigma\|\mathbf{A}\|_F$. Thus if $\sigma \leq \alpha\sqrt{\delta}/\text{poly}(k)$ for a large enough $\text{poly}(k)$, our Significant Singular Values assumption holds. □

In the remainder of this section, we analyze Algorithm 1 and show that it outputs a set of k vectors that approximate the vertices of the latent simplex K. Formally, the main theorem we prove is as follows:

Theorem 1.2 *(Restated.)* Given input data $\mathbf{A}$ from the Latent Simplex Model, there exists an algorithm that takes $\widetilde{O}\left(\mathsf{nnz}(\mathbf{A}) + (n+d)\text{poly}(k)\right)$ time to output k vectors $\mathcal{R}_1, \ldots, \mathcal{R}_k$ such that upon permuting the columns of $\mathbf{M}$, for all $\ell \in [k]$, we have

$$\|\mathcal{R}_\ell - \mathbf{M}_{*,\ell}\|_2 \leq \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}},$$

with probability at least $1 - \frac{1}{\Omega(\sqrt{k})}$.

We show that the subspace $\mathbf{Y}$ obtained via spectral low-rank approximation is a good approximation to the subspace $\mathbf{U}_k$ in angular distance. The appropriate measure of angular distance between subspaces can be formalized as the principal angle between the subspaces and the corresponding

$\sin\Theta$ function. (Wedin, 1972) bounded the $\sin\Theta$ between the SVD subspace of a matrix and the SVD subspace of a slight perturbation of the matrix.

Theorem B.1 (Wedin's $\sin\Theta$ theorem (Wedin, 1972))**.** *Let $\mathbf{R}, \mathbf{S} \in \mathbb{R}^{d\times n}$ and $0 < m \leq \ell$ be integers. Let $\mathbf{R}_m$ and $\mathcal{S}\sigma_\ell$ denote the subspaces spanned by the top m singular vectors of $\mathbf{R}$ and top ℓ singular vectors of $\mathbf{S}$, respectively. Suppose $\gamma = \sigma_m(\mathbf{R}) - \sigma_{\ell+1}(\mathbf{S})$. Then*

$$\sin\Theta(\mathbf{R}_m, \mathcal{S}\sigma_\ell) \leq \frac{\|\mathbf{R}-\mathbf{S}\|_2}{\gamma}.$$

(Bhattacharyya & Kannan, 2020) use Wedin's $\sin\Theta$ theorem to measure the distance between the subspace $\mathbf{U}_k$ spanned by the top k left singular vectors of $\mathbf{A}$ and the subspace returned by their iterative subspace power method. Since we create the sketch $\mathbf{Y}$ for $\mathbf{U}_k$, we would instead like to argue that $\mathbf{Y}$ and $\mathbf{U}_k$ are close in $\sin\Theta$ distance.

Lemma 3.6 *(Proximity of Subspace Projections, Restated.) Let $\mathbf{Y}$ be obtained from Lemma 3.4 and let $\mathbf{U}_k$ be the subspace spanned by the top k left singular vectors of $\mathbf{A}$. Let $\mathbf{P}_\mathbf{Y}$ and $\mathbf{P}_{\mathbf{U}_k}$ be the $d \times d$ projection matrices onto the row span of $\mathbf{Y}$ and $\mathbf{U}_k$. Then $\|\mathbf{P}_\mathbf{Y} - \mathbf{P}_{\mathbf{U}_k}\|_2 \leq \frac{1}{1000k^{10}}$.*

Proof. Suppose by way of contradiction that $\|\mathbf{P}_\mathbf{Y} - \mathbf{P}_{\mathbf{U}_k}\|_2 \geq \frac{1}{1000k^{10}}$. Note that since $\mathbf{Y}$ and $\mathbf{U}_k$ are each orthonormal matrices with rank k, then

$$\|\mathbf{U}_k\mathbf{U}_k^T - \mathbf{Y}\mathbf{Y}^T\|_F^2 \geq \|\mathbf{U}_k\mathbf{U}_k^T - \mathbf{Y}\mathbf{Y}^T\|_2^2 \geq \frac{1}{(1000k^{10})^2}$$

so that

$$\begin{aligned}\|\mathbf{U}_k\mathbf{U}_k^T - \mathbf{Y}\mathbf{Y}^T\|_F^2 &= \|\mathbf{U}_k\|_F^2 + \|\mathbf{Y}\|_F^2 - 2\|\mathbf{U}_k\mathbf{Y}^T\|_F^2 \\ &= 2k - 2\|\mathbf{U}_k\mathbf{Y}^T\|_F^2 \geq \frac{1}{(1000k^{10})^2}\end{aligned}$$

Hence, $\|\mathbf{U}_k\mathbf{Y}^T\|_F^2 \leq k - \frac{1}{(1000k^{10})^2}$. Now we would like to show for the sake of contradiction that $\|\mathbf{A} - \mathbf{P}_\mathbf{Y}\mathbf{A}\|_2$ is large. Thus, for the singular value decomposition $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$, we write

$$\begin{aligned}\|\mathbf{A} - \mathbf{P}_\mathbf{Y}\mathbf{A}\|_2 &= \|\mathbf{U}^T\Sigma - \mathbf{Y}\mathbf{Y}^T\mathbf{U}^T\Sigma\|_2 \\ &\geq \|\mathbf{U}_k\mathbf{U}^T\Sigma - \mathbf{U}_k\mathbf{Y}\mathbf{Y}^T\mathbf{U}^T\Sigma\|_2\end{aligned}$$

since $\|\mathbf{U}_k\|_2 \leq \|\mathbf{U}\|_2 \leq 1$. Thus, there exist matrices $\mathbf{C}_1, \mathbf{C}_2$ such that

$$\mathbf{U}_k\mathbf{U}^T\Sigma - \mathbf{U}_k\mathbf{Y}\mathbf{Y}^T\mathbf{U}^T\Sigma = \begin{bmatrix}\mathbf{C}_1 & \mathbf{C}_2\end{bmatrix}\begin{bmatrix}\Sigma_k & 0 \\ 0 & \Sigma_{n-k}\end{bmatrix},$$

where Σ_k is the diagonal matrix consisting of the top k singular values of $\mathbf{A}$ and Σ_{n-k} is the diagonal matrix consisting of the bottom $n-k$ singular values of $\mathbf{A}$. Now we know that one of the top k eigenvalues of $\mathbf{U}_k^T\mathbf{Y}\mathbf{Y}^T\mathbf{U}_k$ is at most $1 - \frac{1}{(1000k^{10})^2}$. Thus, one of the top k eigenvalues of $\mathbf{I}_k - \mathbf{C}_1$ is at least $\frac{1}{(1000k^{10})^2}$. In particular, let λ be such an eigenvalue and let $\mathbf{x}$ be the corresponding unit eigenvector of $\mathbf{I} - \mathbf{C}_1$. Then we have

$$\|\mathbf{U}_k\mathbf{U}^T\Sigma - \mathbf{U}_k\mathbf{Y}\mathbf{Y}^T\mathbf{U}^T\Sigma\|_2 \geq \|(\mathbf{I}-\mathbf{C}_1)\Sigma_k\mathbf{x}\|_2 \geq \sigma_k(\mathbf{A})\lambda \geq \frac{1}{(1000k^{10})^2}\sigma_k(\mathbf{A}).$$

Since the Significant Singular Values assumption implies that $\frac{1}{(1000k^{10})^2}\sigma_k(\mathbf{A}) > (1+\epsilon)\sigma_{k+1}(\mathbf{A})$, this implies that $\|\mathbf{A} - \mathbf{P}_\mathbf{Y}\mathbf{A}\|_2 > (1+\epsilon)\sigma_{k+1}(\mathbf{A})$, which contradicts the assumption that $\mathbf{Y}$ is a good low-rank approximation to $\mathbf{A}$. Thus we have $\|\mathbf{P}_\mathbf{Y} - \mathbf{P}_{\mathbf{U}_k}\|_2 \leq \frac{1}{1000k^{10}}$, as desired. □

(Bhattacharyya & Kannan, 2020) showed lower bounds on the k-th singular values of $\mathbf{P}$ and $\mathbf{M}$, given the Well-Separateness and Spectrally Bounded Perturbations assumptions.

Lemma B.2. *(Bhattacharyya & Kannan, 2020) If the underlying points $\mathbf{M}$ follow the Well-Separateness and Spectrally Bounded Perturbation assumptions, then*

$$\sigma_k(\mathbf{M}) \geq \frac{1000k^{8.5}}{\alpha^2}\frac{\sigma}{\sqrt{\delta}}, \qquad \sigma_k(\mathbf{P}) \geq \frac{995k^{8.5}\sqrt{n}}{\alpha^2}\sigma.$$

We can then show a low $\sin\Theta$ distance between $\mathbf{Y}$ and $\mathbf{U}_k$.

Corollary B.3. *Let $\mathbf{Y}$ be defined as in Algorithm 1 and let $\mathbf{U}_k$ be the subspace spanned by the top k left singular vectors of $\mathbf{A}$. Then $\sin\Theta(\mathbf{Y}, \mathbf{U}_k) \leq \frac{1}{1000k^{10}}$.*

Proof. By setting $m = k = \ell$ in Theorem B.1, we have

$$\begin{aligned}\sin\Theta(\mathbf{Y}, \mathbf{U}_k) &= \sin\Theta(\mathbf{P}_{\mathbf{Y}}, \mathbf{P}_{\mathbf{U}_k}) \\ &\leq \frac{\|\mathbf{P}_{\mathbf{Y}} - \mathbf{P}_{\mathbf{U}_k}\|_2}{\sigma_k(\mathbf{Y}) - \sigma_{k+1}(\mathbf{U}_k)}.\end{aligned}$$

By definition of σ, we have that $\|\mathbf{A} - \mathbf{P}\|_2 \leq \sigma\sqrt{n}$. Thus, Lemma B.2 implies that $\sigma_k(\mathbf{A}) \gg 1$. Since $\mathbf{Y}$ has rank k, we have $\sigma_{k+1}(\mathbf{Y}) = 0$. By Lemma 3.6, $\sin\Theta(\mathbf{Y}, \mathbf{U}_k) \leq \|\mathbf{P}_{\mathbf{Y}} - \mathbf{P}_{\mathbf{U}_k}\|_2 \leq \frac{1}{1000k^{10}}$. □

They also showed that vectors in $\mathbf{U}_k$ are close to the subspace $\mathbf{M}$:

Lemma B.4. *(Bhattacharyya & Kannan, 2020) Let $\mathbf{U}_k$ be the subspace spanned by the top k left singular vectors of $\mathbf{A}$ and let $\mathbf{R}$ be any k-dimensional subspace of $\mathbb{R}^d$ with*

$$\sin\Theta(\mathbf{U}_k, \mathbf{R}) \leq \frac{\alpha^2}{1001k^9}.$$

*Let $\mathbf{M}$ be the underlying latent k-simplex. Then for each unit vector $\mathbf{x} \in \mathbf{R}$, there exists a vector $\mathbf{y} \in$ **Span** $(\mathbf{M})$ with $\|\mathbf{x} - \mathbf{y}\|_2 \leq \frac{\alpha^2}{500k^{8.5}}$.*

Since we have $\sin\Theta(\mathbf{Y}, \mathbf{U}_k) \leq \frac{1}{1000k^{10}}$ from Corollary B.3, then it follows from Lemma B.4 and the triangle inequality of $\sin\Theta$ distance that vectors in $\mathbf{Y}_k$ are close to the subspace $\mathbf{M}$:

Corollary B.5. *Let $\mathbf{Y}$ be defined as in Algorithm 1 and let $\mathbf{R}$ be any k-dimensional subspace of $\mathbb{R}^d$ with*

$$\sin\Theta(\mathbf{Y}, \mathbf{R}) \leq \frac{\alpha^2}{1000k^9}.$$

*Let $\mathbf{M}$ be the underlying latent k-simplex. Then for each unit vector $\mathbf{x} \in \mathbf{R}$, there exists a vector $\mathbf{y} \in$ **Span** $(\mathbf{M})$ with $\|\mathbf{x} - \mathbf{y}\|_2 \leq \frac{\alpha^2}{500k^{8.5}}$.*

(Bhattacharyya & Kannan, 2020) then show the following structural result between the first r points selected by Algorithm 1 and the closest r points in the latent k-simplex $\mathbf{M}$.

Lemma B.6. *For $r \in [k]$ let $\mathcal{R}_1, \ldots, \mathcal{R}_k \in \mathbb{R}^d$ be points such that there exist distinct $\ell_1, \ldots, \ell_r \subseteq [n]$ with*

$$\|\mathcal{R}_i - \mathbf{M}_{*,\ell_i}\|_2 \leq \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}}$$

Let $\widehat{\mathbf{A}} = \mathcal{R}_1 \circ \ldots \circ \mathcal{R}_t$ and $\widehat{\mathbf{M}} = \mathbf{M}_{*,\ell_1} \circ \ldots \circ \mathbf{M}_{*,\ell_r}$. Then

$$\|\widehat{\mathbf{M}} - \widehat{\mathbf{A}}\|_2 \leq \frac{k^{4.5}}{\alpha}\frac{\sigma}{\sqrt{\delta}}.$$

Proof. Note that the claim follows immediately from the hypothesis and applying the Cauchy-Schwarz inequality. □

We first bound the $\sin\Theta$ distance between **Span** $(\mathbf{M}) \cap \text{Null}(\widehat{\mathbf{M}})$ and $\mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)$. This essentially says that we can work in the subspace $\mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)$ rather than **Span** $(\mathbf{M}) \cap \text{Null}(\widehat{\mathbf{M}})$ and we will not incur too much error.

Lemma 3.8 *(Angular Distance between Subspaces, Restated.) Let $\widehat{\mathbf{M}} = \mathbf{M}_{*,\ell_1} \circ \ldots \circ \mathbf{M}_{*,\ell_r}$ be the r points in the latent k-simplex $\mathbf{M}$ closest to $\mathcal{R}_1, \ldots, \mathcal{R}_r$, the first r points selected by our algorithm,*

respectively. Suppose $\|\mathcal{R}_i - \mathbf{M}_{*,\ell_i}\|_2 \leq \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}}$ *for each* $i \in [r]$. *Let* $\mathbf{P}_r$ *be the projection matrix orthogonal to* $\mathcal{R}_1, \ldots, \mathcal{R}_r$. *Then*

$$\sin\Theta\left(\mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r), \textbf{\textit{Span}}\,(\mathbf{M}) \cap \textit{Null}(\widehat{\mathbf{M}})\right) \leq \frac{\alpha}{100k^4}$$
$$\sin\Theta\left(\textbf{\textit{Span}}\,(\mathbf{M}) \cap \textit{Null}(\widehat{\mathbf{M}}), \mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)\right) \leq \frac{\alpha}{100k^4}.$$

Proof. Let $\mathbf{y} \in \mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)$ be a unit vector. By Corollary B.5, there exists $\mathbf{x} \in \textbf{Span}\,(\mathbf{M})$ with

$$\|\mathbf{x} - \mathbf{y}\|_2 \leq \frac{\alpha^2}{500k^{8.5}}. \tag{B.1}$$

Let $\mathbf{z} = \mathbf{x} - \widehat{\mathbf{M}}\widehat{\mathbf{M}}^\dagger\mathbf{x}$ be the component of $\mathbf{x}$ in Null$(\widehat{\mathbf{M}})$. Note that $\widehat{\mathbf{M}}\widehat{\mathbf{M}}^\dagger$ is a projection matrix and thus $\|\widehat{\mathbf{M}}\widehat{\mathbf{M}}^\dagger\|_2 \leq 1$. Then we have

$$\begin{aligned}\|\mathbf{x} - \mathbf{z}\|_2 &\leq \|\widehat{\mathbf{M}}\widehat{\mathbf{M}}^\dagger(\mathbf{x} - \mathbf{y})\|_2 + \|\widehat{\mathbf{M}}\widehat{\mathbf{M}}^\dagger\mathbf{y}\|_2 \\ &\leq \|\mathbf{x} - \mathbf{y}\|_2 + \|\widehat{\mathbf{M}}(\widehat{\mathbf{M}}^T\widehat{\mathbf{M}})^{-1}(\widehat{\mathbf{M}}^T - \widehat{\mathbf{A}}^T)\mathbf{y}\|_2\end{aligned}$$

where $\widehat{\mathbf{A}} = \mathcal{R}_1 \circ \ldots \circ \mathcal{R}_t$ so that $\widehat{\mathbf{A}}^T\mathbf{y} = 0$ since $\mathbf{P}_r$ projects away from $\widehat{\mathbf{A}}$. We also have $\|\widehat{\mathbf{M}}(\widehat{\mathbf{M}}^T\widehat{\mathbf{M}})^{-1}\|_2 = \frac{1}{\sigma_r(\widehat{\mathbf{M}})}$. Thus by (B.1) and Lemma B.6, we have

$$\begin{aligned}\|\mathbf{x} - \mathbf{z}\|_2 &\leq \|\mathbf{x} - \mathbf{y}\|_2 + \frac{1}{\sigma_r(\widehat{\mathbf{M}})}\|(\widehat{\mathbf{M}}^T - \widehat{\mathbf{A}}^T)\mathbf{y}\|_2 \\ &\leq \frac{\alpha^2}{500k^{8.5}} + \frac{k^{4.5}\sigma}{\alpha\sqrt{\delta}\sigma_k(\widehat{\mathbf{M}})}.\end{aligned}$$

Hence by the triangle inequality and Lemma B.2, we have $\|\mathbf{y} - \mathbf{z}\|_2 \leq \frac{\alpha}{100k^4}$. Since $\mathbf{y} \in \mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)$ and $\mathbf{z} \in \textbf{Span}\,(\mathbf{M}) \cap \text{Null}(\widehat{\mathbf{M}})$, then by definition of the $\sin\Theta$ distance, it follows that $\sin\Theta\left(\mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r), \textbf{Span}\,(\mathbf{M}) \cap \text{Null}(\widehat{\mathbf{M}})\right) \leq \frac{\alpha}{100k^4}$, proving the first part of the claim.

To prove the second half of the claim, it suffices to show that the dimension of $\mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)$ is $k - r$, since $\textbf{Span}\,(\mathbf{M}) \cap \text{Null}(\widehat{\mathbf{M}})$ has dimension $k - r$ and the $\sin\Theta$ distance is symmetric between two subspaces of the same dimension. By construction, $\mathbf{Y}$ has dimension k so that $\mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)$ has dimension at least $k - r$. But if $\mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)$ has dimension larger than $k - r$, then there exists a set of orthonormal vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{k-r+1} \in \mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)$. By the first part of the claim and the definition of the $\sin\Theta$ distance, there exists a set of corresponding vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{k-r+1} \in \textbf{Span}\,(\mathbf{M}) \cap \text{Null}(\widehat{\mathbf{M}})$ such that $\|\mathbf{u}_i - \mathbf{v}_j\|_2 < \frac{\alpha}{100k^4}$. But then for $a \neq b$, we have by the triangle inequality and the fact that $\mathbf{u}_a \cdot \mathbf{u}_b = 0$,

$$\begin{aligned}|\mathbf{v}_a \cdot \mathbf{v}_b| &\leq |\mathbf{u}_a \cdot \mathbf{u}_b| + |(\mathbf{v}_a - \mathbf{u}_a) \cdot \mathbf{u}_b| + |\mathbf{v}_a \cdot (\mathbf{v}_b - \mathbf{u}_b)| \\ &\leq \frac{\alpha}{50k^4}\end{aligned}$$

Similarly, since $\mathbf{u}_a \cdot \mathbf{u}_a = 1$, we have

$$\begin{aligned}|\mathbf{v}_a \cdot \mathbf{v}_a| &\geq |\mathbf{u}_a \cdot \mathbf{u}_a| - |(\mathbf{v}_a - \mathbf{u}_a) \cdot \mathbf{u}_a| - |\mathbf{v}_a \cdot (\mathbf{v}_a - \mathbf{u}_a)| \\ &\geq 1 - \frac{\alpha}{50k^4}.\end{aligned}$$

Thus if $\mathbf{V} = \mathbf{v}_1 \circ \ldots \circ \mathbf{v}_{k-r+1} \in \mathbb{R}^{d \times k-r+1}$ is formed by concatenating the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{k-r+1}$, then $\mathbf{V}^T\mathbf{V}$ is diagonally-dominant. Hence, $\mathbf{V}^T\mathbf{V}$ is nonsingular, so $\mathbf{v}_1, \ldots, \mathbf{v}_{k-r+1}$ must be linearly independent vectors in $\textbf{Span}\,(\mathbf{M}) \cap \text{Null}(\widehat{\mathbf{M}})$, which contradicts the fact that its dimension is $k - r$. Therefore, the dimension of $\mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)$ must be $k - r$, and so $\sin\Theta\left(\textbf{Span}\,(\mathbf{M}) \cap \text{Null}(\widehat{\mathbf{M}}), \mathbf{Y}(\mathbf{I}_d - \mathbf{P}_r)\right) \leq \frac{\alpha}{100k^4}$. □

We now recall a structural lemma from (Bhattacharyya & Kannan, 2020).

Lemma B.7 (Claim 10.1 in (Bhattacharyya & Kannan, 2020)). *Let $a, b \notin \{\ell_1, \ldots, \ell_r\}$ be distinct indices. Then*

$$\|\textit{Proj}(\mathbf{M}_{*,a} - \mathbf{M}_{*,b}, \textit{Null}(\widehat{\mathbf{M}})\|_2 \geq \alpha \max_{\ell} \|\mathbf{M}_{*,\ell}\|_2.$$

We now show that if $\mathbf{u} = g\mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r) \in \mathbb{R}^d$, then $|\mathbf{u} \cdot \mathbf{x}|$ has a clear optimum over $\mathbf{x}$ chosen from the k vertices of the simplex $\mathbf{M}$. This shows that our optimization procedure is well-defined.

Lemma 3.9 *(Optimization is Well-Defined) Let $\widehat{\mathbf{M}} = \mathbf{M}_{*,\ell_1} \circ \ldots \circ \mathbf{M}_{*,\ell_r}$ be the r points in the latent k-simplex $\mathbf{M}$ closest to the first r points selected by Algorithm 1, $\mathcal{R}_1, \ldots, \mathcal{R}_r$, respectively. Suppose*

$$\|\mathcal{R}_i - \mathbf{M}_{*,\ell_i}\|_2 \leq \frac{300k^4}{\alpha} \frac{\sigma}{\sqrt{\delta}}$$

for each $i \in [r]$. Let $\mathbf{u} \in \mathbb{R}^d$ be a random unit vector in the space of $\mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r)$, where $\mathbf{P}_r$ is the orthogonal projection to $\mathcal{R}_1, \ldots, \mathcal{R}_r$. Then there exists a constant $c > 0$ so that with probability at least $1 - \frac{c}{k^{1.5}}$:

1. *For all distinct $a, b \notin \{\ell_1, \ldots, \ell_r\}$, then $|\mathbf{u} \cdot (\mathbf{M}_{*,a} - \mathbf{M}_{*,b})| \geq \frac{0.097}{k^4} \alpha \max_\ell \|\mathbf{M}_{*,\ell}\|_2$.*
2. *For all $a \notin \{\ell_1, \ldots, \ell_r\}$, then $|\mathbf{u} \cdot \mathbf{M}_{*,a}| \geq \frac{0.0989}{k^4} \alpha \max_\ell \|\mathbf{M}_{*,\ell}\|_2$.*

Proof. For $a \notin \{\ell_1, \ldots, \ell_r\}$, let $\mathbf{p}_a$ be the projection of $\mathbf{M}_{*,a}$ onto Null$(\widehat{\mathbf{M}})$ and $\mathbf{q}_a$ be the projection of $\mathbf{M}_{*,a}$ onto **Span** $\left(\widehat{\mathbf{M}}\right)$. By the Well-Separateness assumption, we have $\|\mathbf{p}_a\|_2 \geq \alpha \max_\ell \|\mathbf{M}_{*,\ell}\|_2$. Let $\mathbf{w}_a$ be defined so that $\mathbf{q}_a = \widehat{\mathbf{M}}\mathbf{w}_a$. Since $\|\mathbf{q}_a\|_2 \leq \|\mathbf{M}_{*,a}\|_2$ and $\sigma_r(\widehat{\mathbf{M}}) \leq \sigma_k(\mathbf{M})$, then Lemma B.2 gives

$$\|\mathbf{w}_a\|_2 \leq \frac{\|\mathbf{q}_a\|_2}{\sigma_r(\widehat{\mathbf{M}})} \leq \frac{\|\mathbf{M}_{*,a}\|_2 \alpha^2}{1000k^{8.5}} \frac{\sqrt{\delta}}{\sigma}. \tag{B.2}$$

Since $\widehat{\mathbf{A}}\mathbf{u} = 0$, we can also write

$$\begin{aligned}\mathbf{u} \cdot \mathbf{M}_{*,a} &= \mathbf{u} \cdot \mathbf{p}_a + \mathbf{u} \cdot \mathbf{q}_a \\ &= \mathbf{u} \cdot \text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r)) + \mathbf{u}^T(\widehat{\mathbf{M}} - \widehat{\mathbf{A}})\mathbf{w}_a.\end{aligned}$$

By Lemma B.6, (B.2), and normalizing so that $\|\mathbf{u}\|_2 = 1$, we have

$$\begin{aligned}|\mathbf{u} \cdot (\mathbf{M}_{*,a} - \cdot\text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r)))| &\leq \|\widehat{\mathbf{M}} - \widehat{\mathbf{A}}\|_2 \|\mathbf{w}_a\|_2 \\ &\leq \frac{\alpha \|\mathbf{M}_{*,a}\|_2}{1000k^4}.\end{aligned} \tag{B.3}$$

The same holds for $\mathbf{u} \cdot (\mathbf{M}_{*,a} - \mathbf{M}_{*,b})$, so that

$$\begin{aligned}&|\mathbf{u} \cdot (\mathbf{M}_{*,a} - \mathbf{M}_{*,b}) - \mathbf{u} \cdot \text{Proj}(\mathbf{p}_a - \mathbf{p}_b, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))| \\ &\leq \frac{\|\mathbf{M}_{*,a} - \mathbf{M}_{*,b}\|_2 \alpha}{1000k^4}.\end{aligned} \tag{B.4}$$

Let $\mathcal{E}$ be the event that:

1. For all a, $|\mathbf{u} \cdot \text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))| \geq \frac{1}{10k^4} \|\text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))\|_2$.
2. For all $a \neq b$, $|\mathbf{u} \cdot \text{Proj}(\mathbf{p}_a - \mathbf{p}_b, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))| \geq \frac{1}{10k^4} \|\text{Proj}(\mathbf{p}_a - \mathbf{p}_b, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))\|_2$.

Note that $|\mathbf{u} \cdot \text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))| \geq \frac{1}{10k^4} \|\text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))\|_2$ holds as long as $\mathbf{u} \cdot \text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r)) \neq 0$. Since the volume of the set $\{\mathbf{x} \in \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r) : \mathbf{u} \cdot \mathbf{x} = 0\}$ is at most $\sqrt{k}$ times the volume of the unit ball $\{\mathbf{x} \in \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r) : \|\mathbf{x}\|_2 = 1\}$, then by taking a union bound over at most k^2 indices, it follows that $\mathcal{E}$ holds with probability at least $1 - \frac{1}{k^{1.5}}$.

By Lemma 3.8, there exists $\mathbf{p}'_a \in \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r)$ such that $\|\mathbf{p}'_a - \mathbf{p}_a\|_2 \le \frac{\alpha\|\mathbf{p}_a\|_2}{100k^4}$. Hence for $k \ge 2$, $\|\mathbf{p}_a - \text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))\|_2 \le \frac{\alpha\|\mathbf{p}_a\|_2}{100k^4} \le \frac{\|\mathbf{p}_a\|_2}{1600}$. This implies $\|\text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))\|_2 \ge 0.999\|\mathbf{p}_a\|_2$. Then conditioning on $\mathcal{E}$,

$$\begin{aligned}|\mathbf{u} \cdot \text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))| &\ge \frac{\|\text{Proj}(\mathbf{p}_a, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))\|_2}{10k^4} \\ &\ge \frac{0.999\|\mathbf{p}_a\|_2}{10k^4} \\ &\ge \frac{0.999\max_\ell \|\mathbf{M}_{*,\ell}\|_2}{10k^4},\end{aligned}$$

where the last inequality follows since $\|\mathbf{p}_a\|_2 \ge \|\text{Proj}(\mathbf{M}_{*,a}, \text{Null}(\mathbf{M} \setminus \mathbf{M}_{*,a}))\|_2 \ge \alpha\max_\ell \mathbf{M}_{*,\ell}$ by the Well-Separateness assumption. Hence by (B.3), it follows that for all $a \notin \{\ell_1, \dots, \ell_r\}$,

$$\begin{aligned}|\mathbf{u} \cdot \mathbf{M}_{*,a}| &\ge |\mathbf{u} * \text{Proj}(\mathbf{u}, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r)) - \frac{\alpha\|\mathbf{M}_{*,a}\|_2}{1000k^4} \\ &\ge \frac{0.0989\alpha\max_\ell \|\mathbf{M}_{*,\ell}\|_2}{k^4},\end{aligned}$$

which proves the second half of the claim.

To prove the first half of the claim, note that conditioned on $\mathcal{E}$, then (B.4) implies

$$\begin{aligned}|\mathbf{u} \cdot (\mathbf{M}_{*,a} - \mathbf{M}_{*,b})| &\ge |\mathbf{u} \cdot \text{Proj}(\mathbf{p}_a - \mathbf{p}_b, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))| \\ &\quad - \frac{\|\mathbf{M}_{*,a} - \mathbf{M}_{*,b}\|_2\alpha}{1000k^4} \\ &\ge \frac{\|\text{Proj}(\mathbf{p}_a - \mathbf{p}_b, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))\|_2}{10k^4} \\ &\quad - \frac{\|\mathbf{M}_{*,a} - \mathbf{M}_{*,b}\|_2\alpha}{1000k^4}.\end{aligned}$$

By Lemma 3.8, there exists $\mathbf{v} \in \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r)$ such that $\|\mathbf{v} - (\mathbf{p}_a - \mathbf{p}_b)\|_2 \le \frac{\alpha\|\mathbf{p}_a - \mathbf{p}_b\|_2}{100k^4}$. Thus, $\|\text{Proj}(\mathbf{p}_a - \mathbf{p}_b, \mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r))\|_2 \ge 0.99\|\mathbf{p}_a\|_2 \ge 0.99\alpha\max_\ell \|\mathbf{M}_{*,\ell}\|_2$, by Lemma B.7. Since $\frac{\|\mathbf{M}_{*,a} - \mathbf{M}_{*,b}\|_2\alpha}{1000k^4} \le \frac{2\alpha\max_\ell \|\mathbf{M}_{*,\ell}\|_2}{1000k^4}$, it follows that $|\mathbf{u}\cdot(\mathbf{M}_{*,a} - \mathbf{M}_{*,b})| \ge \frac{0.097}{k^4}\alpha\max_\ell \|\mathbf{M}_{*,\ell}\|_2$. □

We next show that the selected index is not among the previously selected indices. Thus, we obtain a new index at each iteration, which implies that we only need k iterations.

Lemma B.8. *Let $\widehat{\mathbf{M}} = \mathbf{M}_{*,\ell_1} \circ \ldots \circ \mathbf{M}_{*,\ell_r}$ be the r points in the latent k-simplex $\mathbf{M}$ closest to the first r points selected by Algorithm 1, $\mathcal{R}_1, \dots, \mathcal{R}_r$, respectively. Suppose*

$$\|\mathcal{R}_i - \mathbf{M}_{*,\ell_i}\|_2 \le \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}}$$

for each $i \in [r]$. Let $\mathbf{u} \in \mathbb{R}^d$ be a random unit vector in the space of $\mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r)$, where $\mathbf{P}_r$ is the orthogonal projection to $\mathcal{R}_1, \dots, \mathcal{R}_r$. Let

$$\ell_{r+1} = \begin{cases}\text{argmax}_\ell\, \mathbf{u} \cdot \mathbf{M}_{*,\ell} & \text{if } \mathbf{u} \cdot \mathcal{R}_{r+1} \ge 0 \\ \text{argmin}_\ell\, \mathbf{u} \cdot \mathbf{M}_{*,\ell} & \text{if } \mathbf{u} \cdot \mathcal{R}_{r+1} < 0\end{cases}.$$

Then $\ell_{r+1} \notin \{\ell_1, \dots, \ell_r\}$.

Proof. We consider the case $\mathbf{u} \cdot \mathcal{R}_{r+1} \ge 0$ as the analysis for the case $\mathbf{u} \cdot \mathcal{R}_{r+1} < 0$ is symmetric. Let $\ell_{r+1} = \text{argmax}_\ell\, \mathbf{u} \cdot \mathbf{M}_{*,\ell}$. Suppose by way of contradiction that $\ell_{r+1} \in \{\ell_1, \dots, \ell_r\}$. Without loss of generality, let $\ell_{r+1} = \ell_1$. Since $\|\mathcal{R}_1 - \mathbf{M}_{*,\ell_1}\|_2 \le \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}}$ and $\mathbf{u} \cdot \mathcal{R}_1$, then

$$\mathbf{u} \cdot \mathbf{M}_{*,\ell_i}\|_2 \le \mathbf{u} \cdot \mathcal{R}_i + \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}} = \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}}.$$

Since $\ell_1 = \text{argmax}_\ell \, \mathbf{u} \cdot \mathbf{M}_{*,\ell}$, then $\mathbf{u} \cdot \mathbf{M}_{*,\ell} \leq \mathbf{u} \cdot \mathbf{M}_{*,\ell_1}$ for all ℓ. Thus $\mathbf{u} \cdot \mathbf{P}_{*,S} \leq \frac{300k^4}{\alpha} \frac{\sigma}{\sqrt{\delta}}$ for any set of indices $S \subseteq [n]$ inside the convex hull of $\mathbf{M}$. In conjunction, Lemma B.9 implies

$$\mathbf{u} \cdot \mathbf{A}_{*,\mathcal{R}_{r+1}} \leq \mathbf{u} \cdot \mathbf{P}_{*,\mathcal{R}_{r+1}} + \frac{\sigma}{\sqrt{\delta}} \leq \left(\frac{300k^4}{\alpha} + 1 \right) \frac{\sigma}{\sqrt{\delta}}. \tag{B.5}$$

Recall that by Lemma 3.4, $\|\mathbf{A} - \mathbf{Y}\mathbf{Z}^T\|_2^2 \leq (1+\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2^2 + \frac{\epsilon}{k}\|\mathbf{A} - \mathbf{A}_k\|_F^2$ and thus $\|\mathbf{A} - \mathbf{Y}\mathbf{Z}^T\| \leq (1 + 2\epsilon)\|\mathbf{A} - \mathbf{A}_k\|_2$, given the Significant Singular Values assumption. Since $\mathbf{A}_{*,\mathcal{R}_{r+1}}$ is a subset of δn columns of $\mathbf{A}$ and $\mathcal{R}_{r+1}$ is a subset of δn columns of $\mathbf{Y}$, then for $\epsilon < 1$,

$$\begin{aligned} \mathbf{u} \cdot \mathcal{R}_{r+1} &\leq \mathbf{u} \cdot \mathbf{A}_{*,\mathcal{R}_{r+1}} + \mathbf{u} \cdot (\mathcal{R}_{r+1} - \mathbf{A}_{*,\mathcal{R}_{r+1}}) \\ &\leq \left(\frac{300k^4}{\alpha} + 1 \right) \frac{\sigma}{\sqrt{\delta}} + \frac{3}{\sqrt{\delta n}} \|\mathbf{A} - \mathbf{A}_k\|_2, \end{aligned}$$

where the last step follows from (B.5) and applying the Cauchy-Schwarz inequality and the fact that $\mathbf{u}$ is a unit vector. Since $\mathbf{P}$ has rank k and $\mathbf{A}_k$ is the best rank k approximation to $\mathbf{A}$, then $\|\mathbf{A} - \mathbf{A}_k\|_2 \leq \|\mathbf{A} - \mathbf{P}\|_2$ so that

$$\begin{aligned} \mathbf{u} \cdot \mathcal{R}_{r+1} &\leq \left(\frac{300k^4}{\alpha} + 1 \right) \frac{\sigma}{\sqrt{\delta}} + \frac{3}{\sqrt{\delta n}} \|\mathbf{A} - \mathbf{P}\|_2, \\ &\leq \left(\frac{300k^4}{\alpha} + 1 \right) \frac{\sigma}{\sqrt{\delta}} + \frac{3\sigma}{\sqrt{\delta}} \qquad \text{(B.6)} \\ &= \left(\frac{300k^4}{\alpha} + 4 \right) \frac{\sigma}{\sqrt{\delta}}, \qquad \text{(B.7)} \end{aligned}$$

since $\|\mathbf{A} - \mathbf{P}\|_2 \leq \sigma\sqrt{n}$ by definition of σ. However for $t \notin \{\ell_1, \ldots, \ell_r\}$, Lemma B.9 and the Proximate Latent Points assumption imply the existence of a set σ_t of δn columns such that

$$\begin{aligned} |\mathbf{u} \cdot \mathbf{A}_{*,\sigma_t}| &\geq |\mathbf{u} \cdot \mathbf{P}_{*,\sigma_t}| - \frac{\sigma}{\sqrt{\delta}} \\ &\geq |\mathbf{u} \cdot \mathbf{M}_{*,t}| - \frac{5\sigma}{\sqrt{\delta}} \\ &\geq \frac{0.0989}{k^4} \alpha \max_\ell \|\mathbf{M}_{*,\ell}\|_2 - \frac{5\sigma}{\sqrt{\delta}}, \end{aligned} \tag{B.8}$$

where the last step follows from Lemma 3.9. Moreover, σ_t has δn columns, so again by applying the Cauchy-Schwarz inequality and the fact that $\mathbf{u}$ is a unit vector, we have

$$\begin{aligned} |\mathbf{u} \cdot (\mathbf{A}_{*,\sigma_t} - \mathbf{Y}_{*,\sigma_t})| &\leq \frac{1}{\sqrt{\delta n}} \|\mathbf{A} - \mathbf{A}_k\|_2 \\ &\leq \frac{1}{\sqrt{\delta n}} \|\mathbf{A} - \mathbf{P}\|_2 \leq \frac{3\sigma}{\sqrt{\delta}}. \end{aligned} \tag{B.9}$$

where the last two inequalities come from the fact that $\mathbf{P}$ has rank k and $\|\mathbf{A} - \mathbf{P}\|_2 \leq \sigma\sqrt{n}$ by definition of σ.

Thus from (B.8) and (B.9),

$$\begin{aligned} |\mathbf{u} \cdot \mathbf{Y}_{*,\sigma_t}| &\geq |\mathbf{u} \cdot \mathbf{A}_{*,\sigma_t}| - |\mathbf{u} \cdot (\mathbf{A}_{*,\sigma_t} - \mathbf{Y}_{*,\sigma_t})| \\ &\geq \frac{0.0989}{k^4} \alpha \max_\ell \|\mathbf{M}_{*,\ell}\|_2 - \frac{8\sigma}{\sqrt{\delta}}. \end{aligned}$$

However by the Spectrally Bounded Perturbation assumption, we have $|\mathbf{u} \cdot \mathbf{Y}_{*,\sigma_t}| \geq \frac{2400k^5}{\alpha} \frac{\sigma}{\sqrt{\delta}} - \frac{8\sigma}{\sqrt{\delta}}$, which contradicts the maximality of $\mathcal{R}_{r+1}$ in (B.7). Therefore, it holds that $\ell_{r+1} \notin \{\ell_1, \ldots, \ell_r\}$. □

Before showing that the selected index completes the inductive step, we recall the following:

Lemma B.9 (Lemma 3.1 in (Bhattacharyya & Kannan, 2020)). *For a subset $S \subseteq [n]$, let $\mathbf{A}_{*,S} = \frac{1}{|S|} \sum_{i \in S} \mathbf{A}_{*,i}$. For all $S \subseteq [n]$, $|\mathbf{A}_{*,S} - \mathbf{P}_{*,S}| \leq \sigma\sqrt{n/|S|}$.*

Finally, we show that the selected index completes the inductive step.

Lemma 3.10 *(Recovery Guarantees, Restated). Let $\widehat{\mathbf{M}} = \mathbf{M}_{*,\ell_1} \circ \ldots \circ \mathbf{M}_{*,\ell_r}$ be the r points in the latent k-simplex $\mathbf{M}$ closest to the first r points selected by Algorithm 1, $\mathcal{R}_1, \ldots, \mathcal{R}_r$, respectively. Suppose*

$$\|\mathcal{R}_i - \mathbf{M}_{*,\ell_i}\|_2 \leq \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}}$$

for each $i \in [r]$. Let $\mathbf{u} \in \mathbb{R}^d$ be a random unit vector in the space of $\mathbf{Y}^T(\mathbf{I}_d - \mathbf{P}_r)$, where $\mathbf{P}_r$ is the orthogonal projection to $\mathcal{R}_1, \ldots, \mathcal{R}_r$. Let

$$\ell_{r+1} = \begin{cases} \operatorname{argmax}_\ell \mathbf{u} \cdot \mathbf{M}_{*,\ell} & \text{if } \mathbf{u} \cdot \mathcal{R}_{r+1} \geq 0 \\ \operatorname{argmin}_\ell \mathbf{u} \cdot \mathbf{M}_{*,\ell} & \text{if } \mathbf{u} \cdot \mathcal{R}_{r+1} < 0 \end{cases}.$$

Then

$$\|\mathcal{R}_{r+1} - \mathbf{M}_{*,\ell_{r+1}}\|_2 \leq \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}}.$$

Proof. We consider the case $\mathbf{u} \cdot \mathcal{R}_{r+1} \geq 0$ as the analysis for the case $\mathbf{u} \cdot \mathcal{R}_{r+1} < 0$ is symmetric. Let $\ell_{r+1} = \operatorname{argmax}_\ell \mathbf{u} \cdot \mathbf{M}_{*,\ell}$. By Lemma B.8, we have $\ell_{r+1} \notin \{\ell_1, \ldots, \ell_r\}$. Thus applying Lemma 3.9,

$$\mathbf{u} \cdot \mathbf{M}_{*,\ell_{r+1}} \geq \frac{0.0989}{k^4}\alpha \max_\ell \|\mathbf{M}_{*,\ell}\|. \tag{B.10}$$

By the Proximate Latent Points assumption, there exists a set $\sigma_{\ell_{r+1}}$ of size δn so that $\|\mathbf{P}_{*,j} - \mathbf{M}_{*,\ell_{r+1}}\|_2 \leq \frac{4\sigma}{\sqrt{\delta}}$ for all $j \in \sigma_{\ell_{r+1}}$ so that $\|\mathbf{P}_{*,\sigma_{\ell_{r+1}}} - \mathbf{M}_{*,\ell_{r+1}}\|_2 \leq \frac{4\sigma}{\sqrt{\delta}}$. Then by Lemma B.9,

$$\mathbf{u} \cdot \mathbf{A}_{*,\sigma_{\ell_{r+1}}} \geq \mathbf{u} \cdot \mathbf{P}_{*,\sigma_{\ell_{r+1}}} - \frac{\sigma}{\sqrt{\delta}} \geq \mathbf{u} \cdot \mathbf{M}_{*,\ell_{r+1}} - \frac{5\sigma}{\sqrt{\delta}}.$$

By the same reasoning as B.9, we have $\|\mathcal{R}_{r+1} - \mathbf{A}_{*,\sigma_{\ell_{r+1}}}\|_2 \leq \frac{3\sigma}{\sqrt{\delta}}$ and thus,

$$\mathbf{u} \cdot \mathcal{R}_{r+1} \geq \mathbf{u} \cdot \mathbf{M}_{*,\ell_{r+1}} - \frac{8\sigma}{\sqrt{\delta}}. \tag{B.11}$$

Now for any $a \notin \{\ell_1, \ldots, \ell_{r+1}\}$, Lemma 3.9 says

$$\mathbf{u} \cdot \mathbf{M}_{*,a} \leq \mathbf{u} \cdot \mathbf{M}_{*,\ell_{r+1}} - \frac{0.097}{k^4}\alpha \max_\ell \|\mathbf{M}_{*,\ell}\|_2. \tag{B.12}$$

Similarly, for $a \in \{\ell_1, \ldots, \ell_r\}$, we have $\|\mathcal{R}_a - \mathbf{M}_{*,a}\| \leq \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}}$ by the inductive hypothesis. Since $\mathbf{u} \cdot \mathcal{R}_a = 0$, then

$$\begin{aligned}\mathbf{u} \cdot \mathbf{M}_{*,a} &\leq \mathbf{u} \cdot \mathcal{R}_a + \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}} = \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}} \\ &\leq \mathbf{u} \cdot \mathbf{M}_{*,\ell_{r+1}} - \frac{0.0989}{k^4}\alpha \max_\ell \|\mathbf{M}_{*,\ell}\| \\ &\quad + \frac{300k^4}{\alpha}\frac{\sigma}{\sqrt{\delta}}\end{aligned}$$

by (B.10). Thus by the Spectrally Bounded Perturbation assumption,

$$\mathbf{u} \cdot \mathbf{M}_{*,a} \leq \mathbf{u} \cdot \mathbf{M}_{*,\ell_{r+1}} - \frac{0.097}{k^4}\alpha \max_\ell \|\mathbf{M}_{*,\ell}\| \tag{B.13}$$

Since $\mathbf{P}_{*,\mathcal{R}_{r+1}}$ is a convex combination of the columns of $\mathbf{M}$, there exists a vector $\mathbf{w}$ such that $\mathbf{P}_{*,\mathcal{R}_{r+1}} = \mathbf{M}\mathbf{w}$. Then by the same reasoning as B.9 and Lemma B.9,

$$\begin{aligned}\mathbf{u} \cdot \mathcal{R}_{r+1} &\leq \mathbf{u} \cdot \mathbf{A}_{*,\mathcal{R}_{r+1}} + \frac{3\sigma}{\sqrt{\delta}} \leq \mathbf{u} \cdot \mathbf{P}_{*,\mathcal{R}_{r+1}} + \frac{3\sigma}{\sqrt{\delta}} + \frac{4\sigma}{\sqrt{\delta}} \\ &\leq w_{\ell_{r+1}}(\mathbf{u} \cdot \mathbf{M}_{*,\ell_{r+1}}) + \\ &\quad \sum_{a \neq \ell_{r+1}} w_a \left((\mathbf{u} \cdot \mathbf{M}_{*,\ell_{r+1}} - \frac{0.097}{k^4}\alpha \max_\ell \|\mathbf{M}_{*,\ell}\|_2 \right) \\ &\quad + \frac{4\sigma}{\sqrt{\delta}},\end{aligned}$$

where the last line follows from decomposing $\mathbf{M}$ and applying (B.12) and (B.13) to $\mathbf{M}_{*,a}$ for $a \neq \ell_{r+1}$. Hence,

$$\mathbf{u} \cdot \mathcal{R}_{r+1} \leq \mathbf{u} \cdot \mathbf{M}_{*,\ell_{r+1}} - \frac{0.097\alpha \max_\ell \|\mathbf{M}_{*,\ell}\|_2 (1 - w_{\ell_{r+1}})}{k^4} + \frac{4\sigma}{\sqrt{\delta}}.$$

Combining with (B.11), we have

$$(1 - w_{\ell_{r+1}}) \max_\ell \|\mathbf{M}_{*,\ell}\|_2 \leq \frac{12\sigma}{\sqrt{\delta}} \frac{k^4}{0.097\alpha} \leq \frac{124k^4}{\alpha} \frac{\sigma}{\sqrt{\delta}}.$$

Thus,

$$\begin{aligned}
\|\mathbf{P}_{*,\mathcal{R}_{r+1}} - \mathbf{M}_{*,\ell_{r+1}}\|_2 &= \|(w_{\ell_{r+1}} - 1)\mathbf{M}_{*,\ell_{r+1}} \\
&\quad + \sum_{a \neq \ell_{r+1}} w_a \mathbf{M}_{*,a}\| \\
&\leq \sum_{a \neq \ell_{r+1}} w_a \|\mathbf{M}_{*,\ell_{r+1}} - \mathbf{M}_{*,a}\|_2 \\
&\leq 2(1 - w_{\ell_{r+1}}) \max_\ell \|\mathbf{M}_{*,\ell}\|_2 \\
&\leq \frac{248k^4}{\alpha} \frac{\sigma}{\sqrt{\delta}}.
\end{aligned}$$

Finally from the triangle inequality and Lemma B.9, we have

$$\begin{aligned}
\|\mathcal{R}_{r+1} - \mathbf{M}_{*,\ell_{r+1}}\|_2 &\leq \|\mathcal{R}_{r+1} - \mathbf{P}_{*,\mathcal{R}_{r+1}}\|_2 \\
&\quad + \|\mathbf{P}_{*,\mathcal{R}_{r+1}} - \mathbf{M}_{*,\ell_{r+1}}\|_2 \\
&\leq \frac{3\sigma}{\sqrt{\delta}} + \frac{248k^4}{\alpha} \frac{\sigma}{\sqrt{\delta}} \\
&\leq \frac{300k^4}{\alpha} \frac{\sigma}{\sqrt{\delta}}.
\end{aligned}$$

□

C Connection to Spectral Low-Rank Approximation

In this section, we show that learning a latent simplex is closely related to computing a spectral low-rank approximation. Spectral low-rank approximation is a fundamental primitive for algorithm design and numerical linear algebra and the best known algorithm for computing a $(1+\epsilon)$-approximation is $O(\mathsf{nnz}(\mathbf{A}) \cdot k)$ (Musco & Musco, 2015). A major open question in randomized linear algebra is to determine whether the dependence on k in the running time is necessary for spectral low-rank approximation.

We show that for a candidate hard distribution over the input, determined by a Stochastic Block Model (with appropriate parameters) satisfying Well-Separateness1, Proximate Latent Points2 and Spectrally Bounded Perturbations3, an algorithm for learning a latent simplex requiring $o(\mathsf{nnz}(\mathbf{A}) \cdot k)$ time also recovers a spectral low-rank approximation for the input. One way to interpret this statement is that improving the running time for learning a latent simplex under the same assumptions as (Bhattacharyya & Kannan, 2020) would likely lead to a major algorithmic breakthrough for spectral low-rank approximation.

Theorem C.1 (Spectral LRA to Latent Simplex)**.** *Given $k \in [n]$, let $\mathcal{S}_1, \mathcal{S}_2 \ldots, \mathcal{S}_k$ be a partition of $[n]$ such that for all $\ell \in [k]$, $|\mathcal{S}_\ell| = n/k$. Consider a stochastic block model with k communities, $\mathcal{S}_1, \ldots, \mathcal{S}_k$ such that for all $i \in \mathcal{S}_\ell$ and $j \in \mathcal{S}_{\ell'}$, the probability of an edge (i, j) is $p = \text{poly}(k)/n^{1/8}$ when $\ell = \ell'$ and $q = p/10$ otherwise. Let $\mathbf{A}$ be a matrix drawn from the aforementioned model such that $\mathbf{A}_{i,j} = 1$ if there exists an edge between (i, j) and 0 otherwise. Then any algorithm that learns the simplex also recovers a rank k matrix $\mathbf{B}$ such that $\|\mathbf{A} - \mathbf{B}\|_2^2 \leq \|\mathbf{A} - \mathbf{A}_k\|_2^2 + \frac{1}{n^{1/3}} \|\mathbf{A} - \mathbf{A}_k\|_F^2$.*

Proof. Let $\mathbf{P_B}$ be the projection matrix onto the column span of the output matrix $\mathbf{B}$. We show that $\mathbf{A} - \mathbf{P_B}\mathbf{A}$ is a good mixed spectral-Frobenius low-rank approximation to $\mathbf{A}$.

$$\begin{aligned}\|\mathbf{A} - \mathbf{P_B}\mathbf{A}\|_2 &\leq \|\mathbf{A} - \mathbf{P} + \mathbf{P}\|_2 \|\mathbf{I} - \mathbf{P_B}\|_2 \\ &\leq \|\mathbf{A} - \mathbf{P}\|_2 \|\mathbf{I} - \mathbf{P_B}\|_2 + \mathbf{P}\|_2 \|\mathbf{I} - \mathbf{P_B}\|_2 \\ &\leq \|\mathbf{A} - \mathbf{P}\|_2 + \|\mathbf{P}\|_2 \|\mathbf{I} - \mathbf{P_B}\|_2.\end{aligned}$$

From the definition of σ, we have $\|\mathbf{A} - \mathbf{P}\|_2 \leq \sigma\sqrt{n}$. For the specific stochastic block model, we have $\sigma \leq \sqrt{p(1-p)}$, e.g., see (Awasthi, 2017). Moreover, the algorithm of (Bhattacharyya & Kannan, 2020) guarantees specifically in their Theorem 7.2 that $\|\mathbf{I} - \mathbf{P_B}\|_2 \leq \frac{C_1 k^{4.5} d^{1/8}}{n^{1/4}}$ for some constant $C_1 > 0$. Since $\|\mathbf{P}\|_F \geq \|\mathbf{P}\|_2$ and $\|\mathbf{P}\|_F^2 \leq C_2 p^2 nd$ for some constant $C_2 > 0$ with high probability, then we have

$$\begin{aligned}\|\mathbf{A} - \mathbf{P_B}\mathbf{A}\|_2 &\leq \sqrt{p(1-p)n} + \frac{C_1 k^{4.5} d^{1/8} \sqrt{C_2 p^2 nd}}{n^{1/4}} \\ &\leq \sqrt{pn} + C_1 p k^{4.5} d^{5/8} \sqrt{C_2} n^{1/4}.\end{aligned}$$

On the other hand, we have $\|\mathbf{A} - \mathbf{A}_k\|_F^2 \geq \|\mathbf{A}\|_F^2 - k\|\mathbf{A}\|_2^2$. As before, we have $\|\mathbf{P}\|_2 \leq p\sqrt{C_2 nd}$, so that

$$\|\mathbf{A}\|_2 \leq \|\mathbf{P}\|_2 + \|\mathbf{A} - \mathbf{P}\|_2 \leq p\sqrt{C_2 nd} + \sqrt{p(1-p)n}.$$

Moreover, we have $\|\mathbf{A}\|_F \geq C_3\sqrt{qnd}$ for some constant $C_3 > 0$ with high probability. Hence for $q > C_4 p^2$ with a sufficiently high constant C_4, we have

$$\|\mathbf{A} - \mathbf{A}_k\|_F^2 \geq C_5 qnd,$$

for some $C_5 > 0$. Let $p = O(q)$ and $d = n^{1/C}$ for some constant $C \geq 3$ so that $k^{4.5} d^{5/8} = o(n^{1/4})$. Since $\|\mathbf{A} - \mathbf{P_B}\mathbf{A}\|_2^2 \leq C_6 pn$ for some constant C_6, then

$$\begin{aligned}\|\mathbf{A} - \mathbf{P_B}\mathbf{A}\|_2^2 \leq C_6 pn \leq \frac{C_5}{n^{1/C}} qnd = O\left(\frac{1}{n^{1/C}}\right) \|\mathbf{A} - \mathbf{A}_k\|_F^2 \\ \leq \|\mathbf{A} - \mathbf{A}_k\|_2^2 + O\left(\frac{1}{n^{1/C}}\right) \|\mathbf{A} - \mathbf{A}_k\|_F^2.\end{aligned}$$

Taking $C = 3$ gives the desired claim. □