Sample Complexity Bounds for Learning High-dimensional Simplices in Noisy Regimes

Amir Hossein Saberi [1] **Amir Najafi** [2] **Seyed Abolfazl Motahari** [3] **Babak H. Khalaj** [1]

Abstract

In this paper, we find a sample complexity bound for learning a simplex from noisy samples. Assume a dataset of size n is given which includes i.i.d. samples drawn from a uniform distribution over an unknown simplex in $\mathbb{R}^K$, where samples are assumed to be corrupted by a multi-variate additive Gaussian noise of an arbitrary magnitude. We prove the existence of an algorithm that with high probability outputs a simplex having a ℓ_2-distance of at most ε from the true simplex (for any $\varepsilon > 0$). Also, we theoretically show that in order to achieve this bound, it is sufficient to have $n \geq \left(K^2/\varepsilon^2\right) e^{\Omega\left(K/\mathrm{SNR}^2\right)}$ samples, where SNR stands for the signal-to-noise ratio. This result solves an open problem in this area of research and shows as long as $\mathrm{SNR} \geq \Omega\left(K^{1/2}\right)$, the sample complexity of the noisy regime has the same order to that of the noiseless case. Our proofs are a combination of the so-called sample compression technique from (Ashtiani et al., 2018), mathematical tools from high-dimensional geometry, and Fourier analysis. In particular, we have proposed a general Fourier-based technique for recovery of a more general class of distribution families from additive Gaussian noise, which can be further used in a variety of other related problems.

1. Introduction

Many practical problems in machine learning and data science can be naturally modeled by learning of high-dimensional geometric shapes from a given set of unlabeled data points. In particular, learning of simplices from randomly scattered points arises in many fields ranging from bioinformatics to remote sensing (Chan et al., 2009; Schwartz & Shackney, 2010; Satas & Raphael, 2017). A simplex is defined as the set of all convex combinations of $K+1$ points in $\mathbb{R}^K$, for $K \in \mathbb{N}$. Formally speaking, our problem can be stated as follows: assume n i.i.d. samples are drawn from a uniform measure over an unknown simplex in $\mathbb{R}^K$. Also, each sample is assumed to be corrupted by an additive multi-variate Gaussian noise with a covariance of $\sigma^2 \boldsymbol{I}_{K\times K}$, for an unknown σ. The main question that we try to address is: How large n needs to be in terms of parameters K, σ, SNR and etc., such that the true simplex can be consistently recovered with high probability?

Learning of simplices is a well-studied problem. Due to its practical importance, several heuristic algorithms have been proposed for different scenarios such as noisy and sparse cases. In this paper, we focus on the theoretical aspects of this problem which still has several open questions. There exists at least one efficient algorithm for learning of simplices that comes with a theoretical guarantee (see Section 1.1), but still suffers from a very large sample complexity, i.e., requires $n \geq \Omega\left(K^{22}\right)$. Concurrently, researchers work on deriving information-theoretic sample complexity bounds for this task irrespective time and memory requirements. Recently, the optimal sample complexity bound is derived for the noiseless case which is achieved by Maximum Likelihood Estimator (MLE). The MLE for this problem is the minimum volume simplex that contains all the samples, which runs in exponential time[1]. Najafi et al. (2021) proved that MLE in this case is a PAC-learning algorithm with an asymptotically decreasing Total Variation (TV) distance from the true simplex. They showed that $\tilde{\Omega}\left(K^2/\varepsilon\right)$ (Logarithmic terms are ignored) samples are sufficient to estimate the true simplex up to a TV distance of $\varepsilon > 0$. However, the noiseless case is unrealistic since we do not have access to clean data in real world. On the other hand, the minimum-volume inclusive simplex proposed by Najafi et al. (2021) is no longer a 'proper' solution when samples

[1]Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran. [2]School of Mathematics, Institute for Research in Fundamental Sciences (IPM), P.O. Box: 19395-5746, Tehran, Iran. [3]Department of Computer Engineering, Sharif University of Technology, Tehran, Iran.. Correspondence to: Seyed Abolfazl Motahari <motahari@sharif.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

[1]Finding a polynomial algorithm with provable guarantees that comes with a reasonable dependency on K is still an open problem.

are noisy, since the corrupted samples are not necessarily contained inside the true simplex.

In this work, we aim at finding the information-theoretic sample complexity of learning simplices in noisy regimes. Mathematically speaking, we assume samples are generated according to the following model:

$$\boldsymbol{y}_i = \boldsymbol{V}\boldsymbol{\phi}_i + \boldsymbol{z}_i \quad , \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{V}$ is a $K \times (K+1)$ matrix including the vertices of the true simplex $\mathcal{S}_T$ as its columns. Each $\boldsymbol{\phi}_i \in \mathbb{R}^K$ is drawn from a uniform Dirichlet distribution and $\boldsymbol{z}_i \in \mathbb{R}^K$ are sampled from the multivariate Gaussian distribution $\mathcal{N}\left(0, \sigma^2 \boldsymbol{I}_K\right)$. We tackle the task of estimating the true simplex via an exponential-time estimator $\hat{\mathcal{S}} = \hat{\mathcal{S}}\left(\boldsymbol{y}_{1:n}\right)$. In particular, we prove that having at most $\tilde{\Omega}\left(K^2/\varepsilon^2\right) e^{\Omega\left(K/\mathrm{SNR}^2\right)}$ noisy samples from a K-simplex is sufficient to estimate the simplex with a ℓ_2-distance of at most ε, for any $\epsilon > 0$. Here, SNR denotes the ratio of the signal (maximum standard deviation of the noiseless measure over $\mathcal{S}_T$) to the component-wise standard deviation of the noise, i.e., σ. This result solves one of the open problems in this area and might justify a mysterious observation in practice: Almost all heuristic algorithms perform close the noiseless case as long as SNR is larger than a presumably sub-linear function of dimension K. However, they experience a sudden decline in performance as SNR is decreased further from that point (Najafi et al., 2021; Agathos et al., 2014; Lin et al., 2013).

Our approach to proving the above has four main steps: 1) We find a ball in $\mathbb{R}^K$ which with high probability contains the true simplex. 2) We quantize this ball by an ε-cover of isolated points. Each $K+1$ combination of such points forms a candidate simplex for approximating $\mathcal{S}_T$. By assuming that the covering has a sufficient precision, we prove the existence of at least one candidate that has a small TV-distance from the main simplex. 3) Then, we choose a noise-corrupted density, i.e., the convolution of the uniform measure over a candidate simplex with a Gaussian distribution, from the candidate set such that, with high probability, it has a small TV distance from the noisy version of the true simplex. This can be done using some aspects of the "sample compression" technique in (Ashtiani et al., 2018). 4) In the final step, we propose a novel Fourier-based technique to show that as long as the noise-smoothed versions of any two simplices are close in TV-distance, their underlying simplices are also close to each other in the sense of ℓ_2-distance.

1.1. Related Works

We categorize the existing works based on their focus, which could be the efficiency of the proposed algorithms, or the fundamental sample complexity of the problem irrespective of its time complexity. In the former, the main purpose to provide a heuristic solution with promising results in practice. However, papers in the latter category aim to find fundamental and information-theoretic limitations for the problem. In the remainder of this part, we review a number of works from both categories.

Anderson et al. (2013) proved that by having $\Omega\left(K^{22}\right)$ noiseless samples, one can estimate the true simplex via a polynomial-time algorithm. The core idea is to utilize the third moment and local search techniques from Independent Component Analysis (ICA) research. To the best of our knowledge, finding polynomial-time algorithms with a more relaxed dependence to K is still an open problem. Concurrently, Najafi et al. (2021) proved that sample complexity of the MLE in the noiseless case is $\tilde{\Omega}\left(K^2/\varepsilon\right)$, where ε is the permissible TV-distance of the output of the algorithm from the true simplex. Since MLE in this case runs in exponential time, they also provide an alternative heuristic approach as a practical surrogate to MLE, which still does not come with a rigid theoretical guarantee.

Theoretical attempts to tackle this problem in the noisy setting is limited to the work of Bhattacharyya & Kannan (2020). Their work is also based on the sample compression technique, originally used in the seminal work of Ashtiani et al. (2018) in learning high-dimensional Gaussian mixture models. Bhattacharyya & Kannan (2020) prove that a noisy simplex can be learned using $\tilde{\Omega}\left(K^2\right)$ samples only if there exists at least one sample near each vertex of the simplex. This assumptions is very strong, and a back of the envelope calculation shows that one needs around $\tilde{\Omega}\left(1/\varepsilon^K\right)$ samples to hold with high probability.

From a practical point of view, several heuristic methods have been introduced so far in order to deal with real-world problems in bioinformatics, hyper-spectral imaging, and etc. (Piper et al., 2004; Bioucas-Dias et al., 2012; Lin et al., 2013). In hyper-spectral imaging, one aims at finding the distribution of the constituent elements in an area by examining remote hyper-spectral images. Each pixel in such an image can be thought as a random convex combination of a finite set of fixed prototypes which correspond to pure elements that exist in that region. Therefore, the problem of finding constituent minerals would naturally translate into estimating an unknown simplex from a set of presumably uniform samples (Ambikapathi et al., 2011; Agathos et al., 2014; Zhang et al., 2017). Learning of simplices in bioinformatics usually arises in the study of complex tissues. A complex tissue is composed of multiple cell-types –a group of cells with similar characteristics– such as blood, brain, or even tumor cells (Tolliver et al., 2010; Zuckerman et al., 2013). Bulk data from complex tissues, such as gene expression level vectors, can again be modeled as convex combinations of its constituent cell-type profiles. In this

line of work, researchers aim at finding the structure of tissues through studying the state of cell-types, which again translates into learning a high-dimensional simplex (Shoval et al., 2012; Korem et al., 2015).

With respect to using Fourier-based techniques for distribution learning, a recent body of work has mainly focused on the special case of Gaussian mixture models (Regev & Vijayaraghavan, 2017; Chakraborty & Narayanan, 2020; Chen & Moitra, 2021; Liu & Li, 2022; Qiao et al., 2022). In particular, Qiao et al. (2022) have utilized the techniques developed by Candès & Fernandez-Granda (2014) for "super resolution" research in recovering the parameters of mixtures of well-separated Gaussians. Even though there are similarities between our noise analysis tools and those proposed by the above researchers in the application stage, the theoretical nature of what we propose in Section 3.4 is completely distinct from existing works.

The rest of the paper is organized as follows: In Section 2, we formally define the problem and present our notation and definitions. Our main theoretical results as well as the algorithm that achieves our bounds are discussed in Section 3. Finally, Section 4 concludes the paper and presents some suggestions for future works.

2. Notation and Definitions

We use the same notations as Najafi et al. (2021). A K-simplex is defined as the set of all convex combination of $K+1$ affinely independent points in $\mathbb{R}^K$. Let $\boldsymbol{V} = [\boldsymbol{v}_0|\boldsymbol{v}_1|\cdots|\boldsymbol{v}_k] \in \mathbb{R}^{K\times(K+1)}$ be a matrix whose columns represent vertices of the simplex, then K-simplex $\mathcal{S}$ can be defined as

$$\mathcal{S} = \left\{ \boldsymbol{V}\boldsymbol{\phi} \,\middle|\, \boldsymbol{\phi} \in [0,1]^{K+1},\ \boldsymbol{\phi}^T\mathbf{1} = 1 \right\}.$$

Also, let us denote the set of all possible K-simplices in $\mathbb{R}^K$ by $\mathbb{S}_K$. We denote the uniform probability measure over a simplex $\mathcal{S}$ by $\mathbb{P}_{\mathcal{S}}$, and its probability density function by $f_{\mathcal{S}}(x)$. Thus, $f_{\mathcal{S}}(x)$ can be written as

$$f_{\mathcal{S}}(x) = \frac{\mathbf{1}(\boldsymbol{x} \in \mathcal{S})}{\mathrm{Vol}(\mathcal{S})}, \quad \forall \boldsymbol{x} \in \mathbb{R}^K,$$

where $\mathrm{Vol}(\mathcal{S})$ denotes the Lebesgue measure (or volume) of $\mathcal{S}$. The noisy simplex family, i.e., the class of distributions formed by *convolving* $f_{\mathcal{S}}$ (for $\mathcal{S} \in \mathbb{S}_K$) by $G_\sigma \triangleq \mathcal{N}\left(\mathbf{0}, \sigma^2\boldsymbol{I}\right)$, is denoted by $\mathbb{G}_{K,\sigma}$. Mathematically speaking,

$$\mathbb{G}_{K,\sigma} \triangleq \{ f_{\mathcal{S}} * G_\sigma |\, \mathcal{S} \in \mathbb{S}_K \}, \tag{2}$$

where $*$ denotes the convolution operator. Obviously, the distribution of input data points $\boldsymbol{y}_i$ lie in $\mathbb{G}_{K,\sigma}$. With a little abuse of notation, we use the term "class of simplices" to both refer to $\mathbb{S}_K$ and also $\{f_{\mathcal{S}} |\, \mathcal{S} \in \mathbb{S}_K\}$ whenever the difference is clear from the context. In a similar fashion, we refer to $\mathbb{G}_{K,\sigma}$ as the class of "noisy simplices".

In order to measure the difference between two distributions, we use both ℓ_2 and total variation distances. Consider two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$, with respective density functions f_1 and f_2, which are defined over $\mathbb{R}^K$. Then, TV distance between $\mathbb{P}_1$ and $\mathbb{P}_2$ can be defined as

$$\mathrm{TV}(\mathbb{P}_1, \mathbb{P}_2) \triangleq \sup_{\mathrm{A}\in\mathcal{B}} |\mathbb{P}_1(\mathrm{A}) - \mathbb{P}_2(\mathrm{A})| = \frac{1}{2}\|f_1 - f_2\|_1,$$

where $\mathcal{B}$ is the set of all Borel sets in $\mathbb{R}^K$.

Definition 2.1 (PAC-Learnability of a distribution class in realizable setting)**.** We say a class of distributions $\mathcal{F}$ is PAC learnable in realizable setting, if there exists a learning method which for any distribution $g \in \mathcal{F}$ and any $\epsilon, \delta > 0$, outputs an estimator $\hat{g}$ using $n \geq \mathrm{poly}(1/\epsilon, 1/\delta)$ i.i.d. samples from g, which with probability at least $1-\delta$ satisfies

$$\|\hat{g} - g\|_{\mathrm{TV}} \leq \epsilon. \tag{3}$$

We also need to define a series of geometric restrictions for the simplex in noisy cases. In fact, those simplices that are significantly stretched toward on particular direction can be shown to be more prune to noise than those with some minimum levels of geometric regularity. We discuss the necessity of such definitions in later stages. Similar to Najafi et al. (2021), let us denote the volume of the largest facet of a K-simplex (here, volume needs be calculated in $\mathbb{R}^{K-1}$) by $\mathcal{A}_{\max}(\mathcal{S})$, and the length of the largest line segment inside the simplex (diameter) by $\mathcal{L}_{\max}(\mathcal{S})$. In this regard, we define the isoperimetricity of a K-simplex as follow:

Definition 2.2 ($(\underline{\theta}, \bar{\theta})$-isoperimetricity of simplices)**.** A K-simplex $\mathcal{S} \in \mathbb{S}_K$ is defined to be $(\underline{\theta}, \bar{\theta})$-isoperimetric if the following inequalities hold:

$$\begin{aligned} \mathcal{A}_{\max}(\mathcal{S}) &\leq \bar{\theta}\mathrm{Vol}(\mathcal{S})^{\frac{K-1}{K}}, \\ \mathcal{L}_{\max}(\mathcal{S}) &\leq \underline{\theta}K\mathrm{Vol}(\mathcal{S})^{\frac{1}{K}}. \end{aligned}$$

The importance of isoperimetricity arises from the fact that for a simplex to be (even partially) recoverable from noisy data, it must not be stretched too much in any direction or having highly acute angles. In other words, sample complexity in noisy regimes is also affected by the geometric shape of the underlying simplex.

Definition 2.3 (ϵ-representative set)**.** For any $\epsilon > 0$, we say that a finite set of distributions $\mathcal{G}$ is an ϵ-representative set for a distribution class $\mathcal{F}$, if for any distribution $f \in \mathcal{F}$, there exists at least one $g \in \mathcal{G}$ that satisfies

$$\|f - g\|_{\mathrm{TV}} \leq \epsilon.$$

Throughout the paper, and for the sake of simplicity in notation, for any fixed $\bar{\theta}, \underline{\theta} > 0$, whenever we say the class of simplices we actually mean the class of $\left(\underline{\theta}, \bar{\theta}\right)$-isoperimetric simplices.

2.1. Problem Definition

In the first three parts of the paper, we aim at proving the PAC-learnability of the noisy simplex family $\mathbb{G}_{K,\sigma}$. The final part is devoted to showing that if two members in $\mathbb{G}_{K,\sigma}$ are close to each other, e.g., the true noisy simplex and its consistent estimator, then their underlying noiseless simplices are also consistently close.

In this regard and based on Definition 2.1, and in order to show that the class of noisy K-simplices is PAC-learnable, one should find an algorithm that for all $f_{\mathcal{S}} * G_\sigma \in \mathbb{G}_{K,\sigma}$, any positive $\varepsilon, \delta > 0$, and given a dataset $\boldsymbol{D} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n\}$ of i.i.d. samples drawn from $f_{\mathcal{S}} * G_\sigma$, as long as $n \geq \text{poly}\left(1/\varepsilon, 1/\delta\right)$, outputs a noisy simplex $f_{\widehat{\mathcal{S}}} * G_\sigma$ such that

$$\mathbb{P}\left[\left\|\left(f_{\widehat{\mathcal{S}}} - f_{\mathcal{S}}\right) * G_\sigma\right\|_{\text{TV}} \geq \varepsilon\right] \leq \delta. \tag{4}$$

Our ultimate goal in Sections 3.1, 3.2, and 3.3 is the following: To derive explicit polynomial forms for the above-mentioned sample complexity $n \geq \text{poly}\left(1/\varepsilon, \log\left(1/\delta\right)\right)$, which (as we show in Theorem 3.5) turns out to be

$$n \geq \tilde{\Omega}\left[\frac{K^2}{\varepsilon^2} \log \frac{1}{\delta}\right].$$

The final part is dedicated to show that as long as (with high probability) $\left(f_{\widehat{\mathcal{S}}} - f_{\mathcal{S}}\right) * G_\sigma \leq \varepsilon$, then we also have

$$\left\|f_{\widehat{\mathcal{S}}} - f_{\mathcal{S}}\right\|_2 \leq \varepsilon e^{\Omega\left(\frac{K}{\text{SNR}^2}\right)},$$

where $\text{SNR} \triangleq \mathcal{L}_{\max}\left(\mathcal{S}\right) / \left(K\sigma\right)$ denotes the signal-to-noise ratio, since based on the definition of a uniform Dirichlet distribution, $\mathcal{L}_{\max}/K$ represents the maximum component-wise standard deviation associated to $f_{\mathcal{S}}$. This will prove our claims in Section 1.

3. Statistical Learning of Noisy Simplices

Before going through the details, let us first present an sketch of proof for PAC-learnability of noisy simplices.

(*Bounding the candidate set*): In order to estimate the true simplex from noisy i.i.d. samples, we first split our data in half and use the first half to restrict the set of all K-simplices in $\mathbb{S}_K$ to a compact subset of $\mathbb{R}^K$, denoted by $\mathbb{S}_K^{\mathcal{D}}$, which consists of all the simplices that happen to entirely fall within a bounded ball. We propose a way to construct $\mathbb{S}_K^{\mathcal{D}}$ based on data such that it includes (with high probability) the true underlying simplex. This way, we can eliminate very far candidates and thus focus on simplices that are placed near the data samples.

(*Quantization*): We quantize this bounded set and create a finite ϵ-representative set of K-simplices denoted by $\widehat{\mathbb{S}}_K^{\mathcal{D}} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_M\}$ for $M \in \mathbb{N}$, such that for each simplex $\mathcal{S} \in \mathbb{S}_K^{\mathcal{D}}$, there exists some $i \in \{1, 2, \cdots, M\}$ where $\|\mathbb{P}_{\mathcal{S}_i} - \mathbb{P}_{\mathcal{S}}\|_{TV} \leq \epsilon$.

(*Density selection*): In this part, we use the second half of data and try to choose the best simplex in $\widehat{\mathbb{S}}_K^{\mathcal{D}}$. By the best simplex, we mean the one with the minimum TV-distance from the true simplex. We show that as long as the proposed sample complexity bound is satisfied, the output of this selection procedure falls within a ε TV-distance of $f_{\mathcal{S}_T} * G_\sigma$.

(*Noise analysis*): Finally, we show that estimating a noisy version of a simplex leads to consistent estimation for the simplex as well. This completes our proof.

3.1. Bounding The Candidate Set

In this part, we show how the first half of the dataset can be utilized in order to bound the set of all candidate simplices into a ball with a finite radius in $\mathbb{R}^K$. This procedure is crucial for later stages of the proof. In this regard, first let us review the generation process of noisy samples $\boldsymbol{y}_i$ for $i \in [n]$:

$$\begin{aligned} \boldsymbol{y} \sim f_{\mathcal{S}_T} * G_\sigma \implies & \boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}, \\ & \boldsymbol{x} = \boldsymbol{V}_{\mathcal{S}} \boldsymbol{\phi}, \quad \boldsymbol{\phi} \sim \text{Dir}\left(1, 1, \cdots, 1\right), \\ & \boldsymbol{z} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2 \mathbf{I}\right), \end{aligned} \tag{5}$$

where $\boldsymbol{V}_{\mathcal{S}}$ denotes the vertex matrix for $\mathcal{S}$.

The following lemma shows that having enough samples from a noisy simplex $f_{\mathcal{S}} * G_\sigma$, one can find a hyper-sphere in $\mathbb{R}^K$ which (with high probability) contains $\mathcal{S}$.

Lemma 3.1 (Creating $\mathbb{S}_K^{\mathcal{D}}$)**.** *Suppose that we have a set of i.i.d. samples $\boldsymbol{D} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_{2m}\}$ from $f_{\mathcal{S}} * G_\sigma$ for $m \in \mathbb{N}$. If $m \geq 1000(K+1)(K+2)\log\frac{6}{\delta}$, then the true simplex $\mathcal{S}$ with probability at least $1 - \delta$ is confined in a K-dimensional sphere with center point $\boldsymbol{p}$ and radius R, where R and $\boldsymbol{p}$ are defined as follows:*

$$\text{R} = 8\sqrt{(K+1)(K+2)D} \quad , \quad \boldsymbol{p} = \frac{1}{2m}\sum_{i=1}^{2m} \boldsymbol{y}_i, \tag{6}$$

and the variance of the noise can be upper-bounded as $\sigma^2 \leq \frac{D}{K-2} = \text{R}_n$, where D is defined as

$$\text{D} = \frac{1}{2m}\sum_{i=1}^{m} \|\boldsymbol{y}_{2i} - \boldsymbol{y}_{2i-1}\|_2^2.$$

Proof can be found in Section B of the Appendix. This way, the set $\mathbb{S}_K^{\mathcal{D}}$ can be fixed. Next, we discuss how to

properly quantize this set in order to choose an appropriate final candidate for the true simplex, denoted by $\widehat{\mathcal{S}}$.

3.2. Quantization

Let us call the hyper-sphere of Lemma 3.1 as $\mathrm{C}^K(\boldsymbol{p}, R)$, which with high probability contains the true simplex. In this part, we choose a set of points $\mathrm{T}_\epsilon^l(\mathrm{C}^K(\boldsymbol{p}, R)) = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_l\}$ inside this sphere such that for any point $\boldsymbol{x} \in \mathrm{C}^K(\boldsymbol{p}, R)$, there exists some $i \in \{1, 2, \cdots, l\}$ such that $\|\boldsymbol{x} - \boldsymbol{p}_i\|_2 \leq \epsilon$. We call $\mathrm{T}_\epsilon^l(\mathrm{C}^K(\boldsymbol{p}, R))$ a covering set for $\mathrm{C}^K(\boldsymbol{p}, R)$. One way to construct such covering set is to uniformly draw points from the sphere. We show that if the number of drawn random points exceeds $O\left(\left(1 + \frac{2R}{\epsilon}\right)^{2K}\right)$, then with high probability, they form an ϵ-covering set for $\mathrm{C}^K(\boldsymbol{p}, R)$.

Suppose that we construct a random covering set as described above and denote it with $\mathrm{T}_\epsilon(\mathrm{C}^K(\boldsymbol{p}, R))$. Now, using each $K+1$ distinct points in $\mathrm{T}_\epsilon(\mathrm{C}^K(\boldsymbol{p}, R))$ we can build a K-simplex. Assume we collect all such simplices in a set called $\widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R))$, i.e.,

$$\widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R)) = \left\{\mathcal{S}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{K+1}) \middle| \boldsymbol{x}_i \in \mathrm{T}_\epsilon(\mathrm{C}^K(\boldsymbol{p}, R)),\ i \in [K+1]\right\}.$$

Obviously, there exist at most $\binom{|\mathrm{T}_\epsilon(\mathrm{C}^K(\boldsymbol{p}, R))|}{K+1}$ simplices in $\widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R))$. The following lemma states that this set is a sufficiently good representative for all $(\underline{\theta}, \bar{\theta})$-isoperimetric K-simplices inside $\mathrm{C}^K(\boldsymbol{p}, R)$.

Lemma 3.2. *For any $\epsilon \in (0, 1)$, denote the set of all possible K-simplices with the vertices in $\mathrm{T}_{\frac{\alpha\epsilon}{K+1}}(\mathrm{C}^K(\boldsymbol{p}, R))$ as $\widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R))$. Then, the resulting set is an ϵ-representative set for all $(\underline{\theta}, \bar{\theta})$-isoperimetric K-simplices in $\mathrm{C}^K(\boldsymbol{p}, R)$, as long as we have:*

$$\alpha = \frac{\mathrm{Vol}\left(\mathcal{S}\right)^{1/K}}{5\bar{\theta}}.$$

The proof can be found in Section B of the Appendix. This way, the finite candidate set $\mathbb{S}_K^{\mathcal{D}}$ can be formed and we can jump to the next stage of our algorithm for finding a "good" candidate for the true simplex.

3.3. Density Selection

We take advantage of the fundamental result in Devroye & Lugosi (2012) which plays a central role in the remainder of our derivations in this subsection. At this stage, we have already created a finite set of representative simplices $\widehat{\mathbb{S}}_K^{\mathcal{D}}$ and set out to find the "best" simplex in this set. This can be done using the following theorem:

Theorem 3.3 (Theorem 6.3 of (Devroye & Lugosi, 2012))**.** *Let $\mathcal{F}$ be a finite class of distributions consisting of M distinct members $\{f_1, f_2, \cdots, f_M\}$. Also, suppose we have $n \geq \frac{\log(3M^2/\delta)}{2\epsilon^2}$ i.i.d. samples from an arbitrary distribution g for some $\epsilon, \delta > 0$. Then, there exist a deterministic algorithm $\mathscr{A}$, which outputs a number $j \in \{1, \ldots, M\}$ satisfying the following inequality with probability at least $1 - \delta$:*

$$\|f_j - g\|_{TV} \leq 3 \min_{i \in \{1, 2, \cdots, M\}} \|f_i - g\|_{TV} + 4\epsilon. \quad (7)$$

Proof can be found inside the reference. The above theorem is also the core of the "sample compression" technique used by Ashtiani et al. (2018) for learning mixtures of Gaussian distributions. By combining the results of Theorem 3.3 and Lemmas 3.1 and 3.2, we can present our first main results as follows: The set of all shape-restricted simplices in $\mathbb{R}^K$ which entirely fall inside a sphere with radius R is PAC-learnable:

Theorem 3.4 (PAC-Learnability of Simplices in $\mathrm{C}^K(\boldsymbol{p}, R)$ Corrupted by Gaussaim Noise with Bounded Variance)**.** *Assume $\boldsymbol{p} \in \mathbb{R}^K$, and $R, R_n > 0$. Then, the class of $(\underline{\theta}, \bar{\theta})$-isoperimetric K-simplices contained in the K-dimensional hyper-sphere $\mathrm{C}^K(\boldsymbol{p}, R)$ and convolved with an isotropic Gaussian noise $\mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$, with $\sigma \leq \mathrm{R}_n$, is PAC-learnable. Specifically, for some $\epsilon, \delta > 0$, assume we have at least n i.i.d. samples from a distribution $f_{\mathcal{S}} * G_\sigma$ with $\sigma \leq \mathrm{R}_n$ and $\mathcal{S} \in \mathrm{C}^K(\boldsymbol{p}, R)$, where the the following bound is satisfied:*

$$n \geq 50 \frac{\log \frac{30R_n\sqrt{K}}{\delta\epsilon} + 2(K+1)^2 \log\left(1 + \frac{100R\bar{\theta}(K+1)}{\epsilon \mathrm{Vol}(\mathcal{S})^{\frac{1}{K}}}\right)}{\epsilon^2}. \quad (8)$$

Then, there exists an algorithm $\mathscr{A}$ whose outputs $\mathcal{S}_{\mathscr{A}}$ and $\sigma_{\mathscr{A}}$ satisfy the following inequality with probability at least $1 - \delta$:

$$\|f_{\mathcal{S}} * G_\sigma - f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}}}\|_{\mathrm{TV}} \leq \epsilon. \quad (9)$$

The proof can be found in Section A of the Appendix. Let $\mathbb{G}_{K,\sigma}(\boldsymbol{p}, R)$ represent the distribution set

$$\left\{f_{\mathcal{S}} * G_\sigma | \mathcal{S} \subseteq C^K(\boldsymbol{p}, R)\right\}.$$

In an agnostic setting, so far we have actually proved that for any $\epsilon, R, R_n > 0$, $\boldsymbol{p} \in \mathbb{R}^K$, and having $n \geq \tilde{\Omega}\left(K^2 \log R/\epsilon^2\right)$ samples from "any" distribution in

$$\bigcup_{\sigma \leq R_n} \mathbb{G}_{K,\sigma}(\boldsymbol{p}, R),$$

there exists an algorithm $\mathscr{A}$ that outputs a simplex $\mathcal{S}_{\mathscr{A}}$ and noise standard deviation $\sigma_{\mathscr{A}}$ such that with a positive probability we have

$$\|f_{\mathcal{S}_{\mathcal{A}}} * G_{K,\sigma_{\mathscr{A}}} - f_{\mathcal{S}} * G_\sigma\|_{\mathrm{TV}} \leq 4 \min_{\substack{\mathcal{S}^* \subseteq \mathrm{C}^K(\boldsymbol{p}, R) \\ \sigma^* \leq R_n}} \|f_{\mathcal{S}^*} * G_{\sigma^*} - f_{\mathcal{S}} * G_\sigma\|_{\mathrm{TV}} + \epsilon, \quad (10)$$

while the first term in the r.h.s. becomes zero in the realizable case. Also, note that $\bar{\theta}$ which represents the regularity level in the shape of simplices is also present inside the bounds. In fact, trying to learn a highly stretched simplex from noisy data can become very challenging since even a small amount of noise can shoot almost all the samples outside of the simplex. Speaking in a more rigorous fashion, suppose that we put no constraints on the shape of the simplices. Then, there may exist a direction $\boldsymbol{u}$ such that the diameter of the simplex $\mathcal{S}_T$ along $\boldsymbol{u}$ is small, say $\leq \epsilon$. Let us shift the whole simplex $\mathcal{S}_T$ towards $\boldsymbol{u}$ by 2ϵ and call this new simplex by $\mathcal{S}'_T$. Then, one can show that $\mathrm{TV}\left(\mathcal{S}_T, \mathcal{S}'_T\right) = 1$ since they do not overlap. However, the total variation distance between the convolution of $\mathcal{S}_T$ and $\mathcal{S}'_T$ with a Gaussian distribution can be shown to go to 0, when $\epsilon \to 0$ and $\sigma = \Omega\left(1\right)$.

Another limitation of Theorem 9 is that it requires the simplices to be inside an sphere of radius R. Using Lemma 3.1, we show that this is not needed. The following theorem completes our first main result in this paper by showing PAC-learnability of the whole class of noisy simplices:

Theorem 3.5 (PAC-Learnability of Noisy Simplices in General)**.** *The class of $\left(\underline{\theta}, \bar{\theta}\right)$-isoperimetric K-simplices which are corrupted with Gaussian white noise, i.e.,*

$$\bigcup_{\sigma>0} \mathbb{G}_{K,\sigma}$$

is PAC-learnable. In other words, assume $n \geq \tilde{O}\left(K^2/\varepsilon^2\right)$ i.i.d. samples from a noisy simplex $f_{\mathcal{S}} * G_\sigma$, for any $\mathcal{S} \in \mathbb{S}_K$ and $\sigma > 0$, are given. Then, there exists an algorithm $\mathscr{A}$ which outputs a noisy simplex $f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}}}$ which with high probability satisfies

$$\|f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}}} - f_{\mathcal{S}} * G_\sigma\|_{\mathrm{TV}} \leq \varepsilon. \tag{11}$$

Proof of the above theorem can be found in Section A of the Appendix. Obviously, the $\log R$ dependence inside the sample complexity of Theorem 9 has been disappeared in that of Theorem 3.5.

3.4. Noise Analysis

So far, we have shown that for a simplex $\mathcal{S} \in \mathbb{S}_K$ with some levels of geometric regularity, one can learn $f_{\mathcal{S}} * G_\sigma$ up to an arbitrarily small error as long as n satisfies the bound in Theorem 3.5. What remains to prove is that a consistent estimation of $f_{\mathcal{S}} * G_\sigma$ leads to the same type of estimate for $f_{\mathcal{S}}$ as well. Our approach is based on showing the following three important properties:

- The difference function between any two distinct and geometrically regular members of $\{f_{\mathcal{S}} |\, \mathcal{S} \in \mathbb{S}_K\}$ corresponds to a *low-frequency* function in the Fourier domain. In other words, it means that the majority of the *energy* of the difference function concentrates near the origin.
- Corruption by Gaussian additive noise is equivalent to convolving each $f_{\mathcal{S}}$ with G_σ, which transforms into the much simpler form of point-wise multiplication in the Fourier domain.
- Finally, convolution with a Gaussian kernel G_σ preserves the low-frequency parts of the difference function. Therefore, the two simplices remain distinguishable even after corruption via additive Gaussian noise.

Mathematically speaking, assume $\mathcal{S}_1, \mathcal{S}_2 \in \mathbb{S}_K$ are distinct and have a minimum degree of geometric regularity, e.g., $\mathcal{L}_{\max}\left(\mathcal{S}_i\right)$ is bounded for $i = 1, 2$. Our aim is to show that if $f_{\mathcal{S}_1} * G_\sigma$ and $f_{\mathcal{S}_2} * G_\sigma$ have a total variation distance of at least $\epsilon > 0$, then the ℓ_2-distance between $f_{\mathcal{S}_1}$ and $f_{\mathcal{S}_2}$ is also bounded away from zero according to a function of ϵ, σ and the geometric regularity of simplices $\mathcal{S}_1$ and $\mathcal{S}_2$. The theoretical core behind our method is stated in the following general theorem.

Theorem 3.6 (Recovery of Low-Frequency Objects from Additive Noise)**.** *For $K \in \mathbb{N}$, consider a probability density function family $\mathscr{F}$ which is supported over $\mathbb{R}^K$. Assume for sufficiently large $\alpha > 0$, the following bound holds for all $f, g \in \mathscr{F}$:*

$$\frac{1}{(2\pi)^K} \int_{\|\boldsymbol{\omega}\|_\infty \geq \alpha} \left|\mathcal{F}\left\{f - g\right\}\left(\boldsymbol{\omega}\right)\right|^2 \leq \zeta\left(\alpha^{-1}\right) \int_{\mathbb{R}^K} \left|f - g\right|^2,$$

where $\mathcal{F}\left\{\cdot\right\}$ denotes the Fourier transform and ζ is an increasing function with $\zeta\left(0\right) = 0$ and continuity at 0. Also, assume the probability density function Q (also supported over $\mathbb{R}^K$) has the following property:

$$\inf_{\|\boldsymbol{\omega}\|_\infty \leq \alpha} \left|\mathcal{F}\left\{Q\right\}\left(\boldsymbol{\omega}\right)\right| \geq \eta\left(\alpha\right),$$

where $\eta\left(\cdot\right)$ is a non-negative decreasing function. Then, there exists a non-negative constant C where for any $\sigma, \varepsilon > 0$ and $f, g \in \mathscr{F}$ with $\|f - g\|_2 \geq \varepsilon$, we have

$$\|(f - g) * Q\|_2 \geq \frac{\varepsilon}{(2\pi)^{K/2}} \left(\sup_{\alpha \geq 0} \eta\left(\alpha\right) \sqrt{1 - \zeta\left(\alpha^{-1}\right)}\right),$$

Proof. For the sake of simplicity in notations, let $\mathcal{F}, \mathcal{G}, \mathcal{Q} : \mathbb{R}^K \to \mathbb{C}$ denote the Fourier transforms of f, g and Q, respectively. Due to Parseval's theorem, we have

$$\|f - g\|_2^2 = \frac{1}{(2\pi)^K} \|\mathcal{F} - \mathcal{G}\|_2^2.$$

Also, due to the properties of the Fourier transform, which is the transformation of convolution into direct multiplication, one can write

$$\mathcal{F}\left\{(f - g) * Q\right\} = \mathcal{Q}\left(\mathcal{F} - \mathcal{G}\right).$$

Thus, for sufficiently large α we have:

$$\begin{aligned}(2\pi)^K \|(f-g)*Q\|_2^2 &= \int_{\mathbb{R}^K} |\mathcal{Q}(\boldsymbol{\omega})(\mathcal{F}(\boldsymbol{\omega})-\mathcal{G}(\boldsymbol{\omega}))|^2 \\ &\geq \int_{\|\boldsymbol{\omega}\|_\infty \leq \alpha} |\mathcal{Q}(\boldsymbol{\omega})|^2 |\mathcal{F}(\boldsymbol{\omega})-\mathcal{G}(\boldsymbol{\omega})|^2 \\ &\geq \eta^2(\alpha) \int_{\|\boldsymbol{\omega}\|_\infty \leq \alpha} |\mathcal{F}(\boldsymbol{\omega})-\mathcal{G}(\boldsymbol{\omega})|^2 \\ &\geq \varepsilon^2 \eta^2(\alpha)\left[1-\zeta\left(\alpha^{-1}\right)\right]. \end{aligned} \tag{12}$$

The above chain of inequalities hold for all $\alpha > C$, therefore we have:

$$\|(f-g)*Q\|_2 \geq \frac{\varepsilon}{(2\pi)^{K/2}} \sup_\alpha \eta(\alpha)\sqrt{1-\zeta(\alpha^{-1})},$$

which completes the proof. □

Theorem 3.6 presents a general approach to prove the recoverability of latent functions (or objects, which are the main focus in this work) from a certain class of independent additive noise distributions. This approach works as long as the function class as well as the noise distribution are mostly comprised of low-frequency components in the Fourier domain. For example, the Gaussian noise hurts low-frequency parts of a geometric object far less than its high-frequency details. More specifically, we prove the following corollary for Theorem 3.6:

Corollary 3.7 (Recoverability from Additive Gaussian Noise $\mathcal{N}\left(\mathbf{0},\sigma^2\boldsymbol{I}\right)$)**.** *Consider the setting in Theorem 3.6, and assume the noise distribution follows $Q \triangleq \mathcal{N}\left(\mathbf{0},\sigma^2\boldsymbol{I}\right)$ for $\sigma > 0$. Then, as long as for $f, g \in \mathscr{F}$ we have $\|f-g\|_2 \geq \varepsilon$ for some $\varepsilon \geq 0$, we also have*

$$\|(f-g)*Q\|_2 \geq \frac{\varepsilon}{(2\pi)^{K/2}}\left(\sup_{\alpha>C}\sqrt{1-\zeta\left(\frac{1}{\alpha}\right)}e^{-K(\sigma\alpha)^2/2}\right) \tag{13}$$

Proof. The Fourier transform of $Q = \mathcal{N}\left(\mathbf{0},\sigma^2\boldsymbol{I}\right)$ can be computed as follows:

$$\mathcal{F}\{Q\}(\boldsymbol{\omega}) = \prod_{i=1}^K \mathcal{F}\left\{\mathcal{N}\left(0,\sigma^2\right)\right\}(\omega_i) = e^{-\sigma^2\|\boldsymbol{\omega}\|_2^2/2}.$$

Also, it can be easily checked that

$$\inf_{\|\boldsymbol{\omega}\|_\infty \leq \alpha} e^{-\sigma^2\|\boldsymbol{\omega}\|_2^2/2} = e^{-\sigma^2/2\left(\alpha^2+\ldots+\alpha^2\right)} = e^{-K(\alpha\sigma)^2/2}.$$

By substitution into the end result of Theorem 3.6, the claimed bounds can be achieved and the proof is complete. □

In this regard, our main explicit theoretical contribution in this section with respect to simplices has been stated in the following theorem:

Theorem 3.8 (Recoverability of Simplices from Additive Noise)**.** *For any two $\left(\bar{\theta},\underline{\theta}\right)$-isoperimetric simplices $\mathcal{S}_1, \mathcal{S}_2 \in \mathbb{S}_K$, given that*

$$\|f_{\mathcal{S}_1} * G_\sigma - f_{\mathcal{S}_2} * G_\sigma\|_{\mathrm{TV}} \leq \varepsilon$$

for some $\varepsilon \geq 0$, we have

$$\|f_{\mathcal{S}_1} - f_{\mathcal{S}_2}\|_2 \leq \varepsilon e^{\mathcal{O}\left(\frac{K}{\mathrm{SNR}^2}\right)},$$

where SNR *denotes the effective signal-to-noise ratio which is defined earlier.*

Proof is given in Section C of the supplementary document. The core idea is to first showing that simplices, in general, are low-frequency objects in the Fourier domain which holds due to their convexity and simple geometric shape. More specifically, Lemma C.3 shows that for a geometrically regular simplex $\mathcal{S} \in \mathbb{S}_K$, we have

$$\frac{1}{(2\pi)^K}\int_{\|\boldsymbol{\omega}\|_\infty \geq \alpha} |\mathcal{F}\{f_{\mathcal{S}}\}(\boldsymbol{\omega})|^2 \leq \frac{1}{\mathrm{Vol}(\mathcal{S})}\mathcal{O}\left(\frac{K}{\alpha}\right),$$

for a sufficiently large $\alpha > 0$, where constants in $\mathcal{O}(\cdot)$ only depend on regularity parameters of $\mathcal{S}$. This inequality is trivial when $K = 1$, since a one-dimensional simplex is a pulse function whose Fourier transform is a $\mathrm{sinc}(x) = \frac{\sin ax}{ax}$ for some $a > 0$. However, showing this relation for larger values of K requires more mathematical effort which is already carried out in the proof of Theorem 3.8. Furthermore, we show a similar inequality holds for the difference of simplices as well. This would ultimately enable us to use the result of Corollary 3.7 to prove the main claim of the paper.

So far in the noise analysis section, we have treated the parameter σ as if it is a known entity. However, the output of algorithm $\mathscr{A}$ from Theorem 3.5 returns only an estimated value $\sigma_{\mathscr{A}}$ with no direct guarantees. In Section C.1 of Appendix, we have theoretically dealt with this issue in details.

4. Conclusions

We present the first sample complexity bounds for consistent learning of high-dimensional simplices in noisy regimes. Formally speaking, we prove that given a sufficient amount of noisy data, one can estimate the true simplex up to an arbitrarily small ℓ_2-error. Also, a presumably optimal polynomial dependence on a number of parameters, such as dimension has been achieved which matches those of the already-solved noiseless case. More interestingly, we have found theoretical justification for an already observed phenomenon in practice: The performance of most heuristic

methods in learning of high-dimensional simplices undergo a rather sharp phase transition w.r.t. the noise power, where for $\mathrm{SNR} \geq \Omega\left(K^{1/2}\right)$ they behave almost identical to the noiseless case, but their performance severely degrades when SNR becomes smaller. Our proofs are based on a number of recent techniques which have been previously used on learning of Gaussian mixture models, a number of tools from high-dimensional geometry and a novel Fourier-based technique for denoising. The latter technique has applications a broader spectra of problems where general recovery of distribution families with certain low-frequency properties from additive Gaussian noise is under question. For future directions in this area, one can think of analyzing this problem under a broader noise and distortion model which might match the practice even more. Also, finding the first non-trivial and amenable polynomial-time algorithms for this problem is an interesting line of work which has remained open until this day.

5. Acknowledgements

This research was partially supported by a grant from IPM.

References

Agathos, A., Li, J., Bioucas-Dias, J. M., and Plaza, A. Robust minimum volume simplex analysis for hyperspectral unmixing. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, pp. 1582–1586. IEEE, 2014.

Ambikapathi, A., Chan, T.-H., Ma, W.-K., and Chi, C.-Y. Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11): 4194–4209, 2011.

Anderson, J., Goyal, N., and Rademacher, L. Efficient learning of simplices. In *Conference on Learning Theory*, pp. 1020–1045, 2013.

Ashtiani, H., Ben-David, S., Harvey, N., Liaw, C., Mehrabian, A., and Plan, Y. Nearly tight sample complexity bounds for learning mixtures of gaussians via sample compression schemes. *Advances in Neural Information Processing Systems*, 31, 2018.

Bhattacharyya, C. and Kannan, R. Near-optimal sample complexity bounds for learning latent $k-$ polytopes and applications to ad-mixtures. In *International Conference on Machine Learning*, pp. 854–863. PMLR, 2020.

Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., and Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):354–379, 2012.

Candès, E. J. and Fernandez-Granda, C. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6):906–956, 2014.

Chakraborty, S. and Narayanan, H. Learning mixtures of spherical gaussians via fourier analysis. *arXiv preprint arXiv:2004.05813*, 2020.

Chan, T.-H., Chi, C.-Y., Huang, Y.-M., and Ma, W.-K. A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing. *IEEE Transactions on Signal Processing*, 57(11):4418–4432, 2009.

Chen, S. and Moitra, A. Algorithmic foundations for the diffraction limit. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 490–503, 2021.

Devroye, L. and Lugosi, G. *Combinatorial methods in density estimation*. Springer Science & Business Media, 2012.

Devroye, L., Mehrabian, A., and Reddad, T. The total variation distance between high-dimensional gaussians. *arXiv preprint arXiv:1810.08693*, 2018.

Korem, Y., Szekely, P., Hart, Y., Sheftel, H., Hausser, J., Mayo, A., Rothenberg, M. E., Kalisky, T., and Alon, U. Geometry of the gene expression space of individual cells. *PLoS computational biology*, 11(7):e1004224, 2015.

Lin, C.-H., Ambikapathi, A., Li, W.-C., and Chi, C.-Y. On the endmember identifiability of craig's criterion for hyperspectral unmixing: A statistical analysis for three-source case. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 2139–2143. IEEE, 2013.

Liu, A. and Li, J. Clustering mixtures with almost optimal separation in polynomial time. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1248–1261, 2022.

Najafi, A., Ilchi, S., Saberi, A. H., Motahari, S. A., Khalaj, B. H., and Rabiee, H. R. On statistical learning of simplices: Unmixing problem revisited. *The Annals of Statistics*, 49(3):1626–1655, 2021.

Piper, J., Pauca, V. P., Plemmons, R. J., and Giffin, M. Object characterization from spectral data using nonnegative factorization and information theory. In *Proceedings of AMOS technical conference*, 2004.

Qiao, M., Guruganesh, G., Rawat, A. S., Dubey, K. A., and Zaheer, M. A fourier approach to mixture learning. *arXiv preprint arXiv:2210.02415*, 2022.

Regev, O. and Vijayaraghavan, A. On learning mixtures of well-separated gaussians. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 85–96. IEEE, 2017.

Satas, G. and Raphael, B. J. Tumor phylogeny inference using tree-constrained importance sampling. *Bioinformatics*, 33(14):i152–i160, 2017.

Schwartz, R. and Shackney, S. E. Applying unmixing to gene expression data for tumor phylogeny inference. *BMC bioinformatics*, 11(1):42, 2010.

Shoval, O., Sheftel, H., Shinar, G., Hart, Y., Ramote, O., Mayo, A., Dekel, E., Kavanagh, K., and Alon, U. Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space. *Science*, 336(6085):1157–1160, 2012.

Tolliver, D., Tsourakakis, C., Subramanian, A., Shackney, S., and Schwartz, R. Robust unmixing of tumor states in array comparative genomic hybridization data. *Bioinformatics*, 26(12):i106–i114, 2010.

Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Zhang, S., Agathos, A., and Li, J. Robust minimum volume simplex analysis for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11): 6431–6439, 2017.

Zuckerman, N. S., Noam, Y., Goldsmith, A. J., and Lee, P. P. A self-directed method for cell-type identification and separation of gene expression microarrays. *PLoS computational biology*, 9(8):e1003189, 2013.

A. Proof of Theorems

proof of Theorem 3.4. Suppose that $\mathcal{S}$ is a $\left(\underline{\theta}, \bar{\theta}\right)$-isoperimetric simplex in $\mathbb{R}^K$ which is confined in a K-dimensional hyper-sphere $\mathrm{C}^K(\boldsymbol{p}, R)$. From Lemma 3.2, we know how to build an ϵ-representative set, for any $\epsilon > 0$, for all $\left(\underline{\theta}, \bar{\theta}\right)$-isoperimetric K-simplices in $\mathrm{C}^K(\boldsymbol{p}, R)$. Let us denote this set with $\widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R))$.

Now Consider the class of all isotropic Gaussian noise, $\mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$, with $\sigma \leq R_n$. We denote this class of distributions with $\mathfrak{N}^K\left(R_n\right)$. Consider a set $C_{R_n}^{\epsilon/\sqrt{K}} = \left\{0, \frac{\epsilon}{\sqrt{K}}, \frac{2\epsilon}{\sqrt{K}}, \cdots, R_n\right\}$, which $\frac{\epsilon}{\sqrt{K}}$-covers the interval $[0, R_n]$. Now for all $\sigma_i \in C_{R_n}^{\epsilon/\sqrt{K}}$ we put $\mathcal{N}\left(\mathbf{0}, \sigma_i^2 \boldsymbol{I}\right)$ in a set called $\widehat{\mathfrak{N}}^K\left(R_n\right)$. From Theorem 1.1 in (Devroye et al., 2018) it can be shown that $\widehat{\mathfrak{N}}^K\left(R_n\right)$ is an ϵ-representative set for $\mathfrak{N}^K\left(R_n\right)$. According to the way we build $\widehat{\mathfrak{N}}^K\left(R_n\right)$, we have $|\widehat{\mathfrak{N}}^K\left(R_n\right)| \leq \frac{R_n\sqrt{K}}{\epsilon}$. Now we build a set of noisy simplices as follows:

$$\widehat{\mathbb{G}}_{K,\sigma}(R, R_n) \triangleq \left\{f_{\mathcal{S}}\left(\boldsymbol{x}\right) * G_\sigma\left(\boldsymbol{x}\right) | \mathcal{S} \in \widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R)), G_\sigma \in \widehat{\mathfrak{N}}^K\left(R_n\right)\right\}. \tag{14}$$

Now for any density function $f_{\mathcal{S}}\left(\boldsymbol{x}\right) * G_\sigma\left(\boldsymbol{x}\right)$, where $\mathcal{S} \in \mathrm{C}^K(\boldsymbol{p}, R)$ and $\sigma \leq R_n$, we can find some $\mathcal{S}^\star \in \widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R))$, and $G_{\sigma^\star} \in \widehat{\mathfrak{N}}^K\left(R_n\right)$ such that

$$\begin{aligned}
&\|f_{\mathcal{S}} - f_{\mathcal{S}^\star}\|_{\mathrm{TV}} \leq \epsilon, \\
&\|G_\sigma - G_{\sigma^\star}\|_{\mathrm{TV}} \leq \epsilon, \\
&f_{\mathcal{S}^\star} * G_{\sigma^\star} \in \widehat{\mathbb{G}}_{K,\sigma}(R, R_n).
\end{aligned}$$

Now for the distance between $f_{\mathcal{S}^\star}^{\sigma^\star}$ and $f_{\mathcal{S}}^{\sigma}$ we have:

$$\begin{aligned}
\|f_{\mathcal{S}^\star} * G_{\sigma^\star} - f_{\mathcal{S}} * G_\sigma\|_{\mathrm{TV}} \leq& \|f_{\mathcal{S}^\star} * G_{\sigma^\star} - f_{\mathcal{S}^\star} * G_\sigma\|_{\mathrm{TV}} + \|f_{\mathcal{S}^\star} * G_\sigma - f_{\mathcal{S}} * G_\sigma\|_{\mathrm{TV}} \\
\leq& \|G_{\sigma^\star} - G_\sigma\|_{\mathrm{TV}} + \|f_{\mathcal{S}^\star} - f_{\mathcal{S}}\|_{\mathrm{TV}} \\
\leq& 2\epsilon.
\end{aligned} \tag{15}$$

From the above inequalities it can be seen that for any density function $f_{\mathcal{S}} * G_\sigma$, where $\mathcal{S} \in \mathrm{C}^K(\boldsymbol{p}, R)$ and $\sigma \leq R_n$, there exist some density function $f^\star \in \widehat{\mathbb{S}}_n(\mathrm{C}^K(\boldsymbol{p}, R), R_n)$ where $\|f^\star - f_{\mathcal{S}} * G_\sigma\|_{\mathrm{TV}} \leq 2\epsilon$. Therefore the set $\widehat{\mathbb{G}}_{K,\sigma}(R, R_n)$ is a 2ϵ-representative set for the class of K-simplices confined in a hyper-sphere with radius R which convolved with a Gaussian noise with variance $\sigma \leq R_n$. We show this class with

$$\mathbb{G}_{K,\sigma}\left(R, R_n\right) \triangleq \left\{f_{\mathcal{S}} * G_\sigma | \, \mathcal{S} \in \mathbb{S}_K, \mathcal{S} \in \mathrm{C}^K(\boldsymbol{p}, R), \sigma \leq R_n\right\} \tag{16}$$

Assume that we have a set of i.i.d. samples from some distribution $f_{\mathcal{S}} * G_\sigma \in \mathbb{G}_{K,\sigma}\left(R, R_n\right)$. From Theorem 3.3, we know that there exists a deterministic algorithm $\mathscr{A}$ such that given

$$n \geq \frac{\log\left(3|\widehat{\mathbb{G}}_{K,\sigma}(R, R_n)|^2/\delta\right)}{2\epsilon^2}$$

i.i.d. samples from $f_{\mathcal{S}} * G_\sigma$, the output of the algorithm denoted by $f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}}}$, with probability at least $1 - \delta$, satisfies:

$$\begin{aligned}
\|f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}}} - f_{\mathcal{S}} * G_\sigma\|_{\mathrm{TV}} &\leq 3 \min_{f \in \widehat{\mathbb{G}}_{K,\sigma}(R, R_n)} \|f - f_{\mathcal{S}} * G_\sigma\|_{\mathrm{TV}} + 4\epsilon \\
&\leq 6\epsilon + 4\epsilon = 10\epsilon.
\end{aligned} \tag{17}$$

For the cardinality of $\widehat{\mathbb{G}}_{K,\sigma}(R, R_n)$ we have:

$$\begin{aligned}
|\widehat{\mathbb{G}}_{K,\sigma}(R, R_n)| =& |\widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R))||\widehat{\mathfrak{N}}^K\left(R_n\right)| \\
\leq& |\widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R))| \frac{R_n\sqrt{K}}{\epsilon}.
\end{aligned} \tag{18}$$

And, according to Lemma 3.2 the followings hold for the cardinality of $\widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R))$:

$$\begin{aligned}\left|\widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R))\right| &= \binom{\left|\mathrm{T}_{\frac{\alpha\epsilon}{K+1}}(\mathrm{C}^K(\boldsymbol{p}, R))\right|}{K+1} \\ &\leq \left|\mathrm{T}_{\frac{\alpha\epsilon}{K+1}}(\mathrm{C}^K(\boldsymbol{p}, R))\right|^{K+1} \\ &\leq \left(\left(1+\frac{2(K+1)R}{\alpha\epsilon}\right)^K\right)^{K+1} \\ &= \left(1+\frac{2(K+1)R}{\alpha\epsilon}\right)^{K(K+1)}. \end{aligned} \tag{19}$$

Now from 19 and 18 we have:

$$|\widehat{\mathbb{G}}_{K,\sigma}(R, R_n)| \leq \frac{R_n\sqrt{K}}{\epsilon}\left(1+\frac{2(K+1)R}{\alpha\epsilon}\right)^{K(K+1)}. \tag{20}$$

Then using 20 and 17, we can say that for any $\epsilon, \delta > 0$, there exists a PAC-learning algorithm $\mathscr{A}$ for the class of noisy $(\underline{\theta}, \bar{\theta})$-isoperimetric K-simplices in $\mathbb{G}_{K,\sigma}(R, R_n)$, whose sample complexity is bounded as follows:

$$\begin{aligned} n &\geq 50\frac{\log\frac{10R_n\sqrt{K}}{\epsilon} + 2(K+1)^2\log\left(1+\frac{20(K+1)R}{\alpha\epsilon}\right) + \log\frac{3}{\delta}}{\epsilon^2} \\ &= 50\frac{\log\frac{30R_n\sqrt{K}}{\delta\epsilon} + 2(K+1)^2\log\left(1+\frac{100R\bar{\theta}(K+1)}{\epsilon\mathrm{Vol}(\mathcal{S})^{\frac{1}{K}}}\right)}{\epsilon^2}. \end{aligned} \tag{21}$$

In other words, given that the number of samples n satisfies the above lower-bound, then with probability at least $1-\delta$ we have

$$\|f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}}} - f_{\mathcal{S}} * G_\sigma\|_{\mathrm{TV}} \leq \epsilon.$$

This completes the proof. Using simple algebra, the bound can be further simplified into $n \geq O\left(\frac{K^2}{\epsilon^2}\log\frac{K}{\epsilon}\right)$ which completes the proof. □

proof of Theorem 3.5. From Theorem 3.4, we know that the class of $(\underline{\theta}, \bar{\theta})$-isoperimetric K-simplices contained in the K-dimensional hyper-sphere $\mathrm{C}^K(\boldsymbol{p}, R)$ and convolved with an isotropic Gaussian noise $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, with $\sigma \leq \mathrm{R}_n$, is PAC-learnable, with sample complexity $O\left(\frac{K^2}{\epsilon^2}\log\frac{K}{\epsilon}\right)$, where ϵ is defined accordingly. And from Lemma 3.1, we know that if we have $O(K^2)$ samples from a noisy simplex $f_{\mathcal{S}} * \mathbb{G}_\sigma$, we can find a K-dimensional sphere which with probability at least $1-\delta/2$ contains the true simplex as long as the radius R of the sphere and the upper-bound of the noise variance R_n satisfy

$$R \leq 4\sqrt{K+1}\left(1+\frac{K+2}{d_{\mathrm{S}}/\sigma}\right)d_{\mathrm{S}} \tag{22}$$

$$R_n \leq \frac{K+2}{K-3}\left(1+\frac{d_{\mathrm{S}}/\sigma}{K+2}\right)\sigma. \tag{23}$$

Therefore, it can be shown that there exists an algorithm $\mathscr{A}$ such that given n i.i.d samples from $\mathbb{G}_{\mathcal{S}}$ with

$$\begin{aligned} n &\geq 50\frac{\log\frac{30R_n\sqrt{K}}{\delta\epsilon} + 2(K+1)^2\log\left(1+\frac{100R\bar{\theta}(K+1)}{\epsilon\mathrm{Vol}(\mathcal{S})^{\frac{1}{K}}}\right)}{\epsilon^2} + 2000(K+1)(K+2)\log\frac{6}{\delta} \\ &= 100\frac{\log 6/\delta + (K+2)^2\log\left(1+\frac{100\bar{\theta}(K+1)^{3/2}}{\epsilon\mathrm{Vol}(\mathcal{S})^{\frac{1}{K}}}\left(1+\frac{K+2}{d_{\mathrm{S}}/\sigma}\right)d_{\mathrm{S}}\right)}{\epsilon^2} \\ &= O\left(\frac{K^2}{\epsilon^2}\log\frac{K}{\epsilon}\right), \end{aligned} \tag{24}$$

the output of the algorithm $f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}}}$ with probability at least $1-\delta$ satisfies

$$\|f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}}} - f_{\mathcal{S}} * \mathbb{G}_{\sigma}\|_{\mathrm{TV}} \leq \epsilon.$$

It should be noted that R and R_n in (24) is already replaced with the bound in the r.h.s of (22) and (23). This way, the proof is completed. □

B. Proof of Lemmas

proof of Lemma 3.1. To find a K-dimensional sphere containing the true simplex, it suffices to find a point $\boldsymbol{p}$ inside the simplex and an upper-bound R for its diameter, i.e., the maximum distance between two points in the simplex. Obviously, this ensures that the K-dimensional sphere with center point $\boldsymbol{p}$ and radius R contains the true simplex. For the maximum distance between any two points inside $\mathcal{S} \in \mathbb{S}_K$, denoted by $d_{\mathcal{S}}$, we have

$$\begin{aligned} d_{\mathcal{S}} &= \max_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{S}} \|\boldsymbol{x}-\boldsymbol{y}\|_2 \\ &= \max_{\boldsymbol{\phi}_{\boldsymbol{x}},\boldsymbol{\phi}_{\boldsymbol{y}}\in\mathcal{S}_{\mathrm{S}}^K} \|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{\boldsymbol{x}} - \boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{\boldsymbol{y}}\|_2 \\ &= \max_{\boldsymbol{\phi}_{\boldsymbol{x}},\boldsymbol{\phi}_{\boldsymbol{y}}\in\mathcal{S}_{\mathrm{S}}^K} \|\boldsymbol{V}_{\mathcal{S}}\left(\boldsymbol{\phi}_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{y}}\right)\|_2, \end{aligned} \tag{25}$$

where $\mathcal{S}_{\mathrm{S}}^K$ represents the standard simplex in $\mathbb{R}^K$ and $\boldsymbol{V}_{\mathcal{S}}$ denotes the vertex matrix of $\mathcal{S}$ (Equation 4 in the main paper). Let $\sigma_{\max}(\mathcal{S})$ denote the maximum singular-value for the matrix $\boldsymbol{V}_{\mathcal{S}}$, and assume $\boldsymbol{v}_{\max}$ be the unitary eigenvector that corresponds to $\sigma_{\max}(\mathcal{S})$. To be more precise, let $\boldsymbol{V}_{\mathcal{S}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ be the singular value decomposition of the vertex matrix. Then, $\sigma_{\max}(\mathcal{S})$ refers to the largest singular value on the main diagonal of $\boldsymbol{\Sigma}$, and $\boldsymbol{v}_{\max}$ represents the corresponding column in $\boldsymbol{V}$. Then, for any point $\boldsymbol{x} \in \mathbb{R}^{K+1}$, one can write

$$\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{x}\|_2^2 = \sum_i \Sigma_{i,i}^2 \left\|\boldsymbol{U}_i\boldsymbol{V}_i^T\boldsymbol{x}\right\|_2^2 = \sum_i \Sigma_{i,i}^2 \left(\boldsymbol{V}_i^T\boldsymbol{x}\right)^2,$$

which results in the following useful inequality:

$$\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{x}\|_2 \geq \sigma_{\max}(\mathcal{S}) \left|\boldsymbol{v}_{\max}^T\boldsymbol{x}\right|.$$

For $\boldsymbol{\phi}_{\boldsymbol{x}} \neq \boldsymbol{\phi}_{\boldsymbol{y}}$, let $\boldsymbol{r} \triangleq \left(\boldsymbol{\phi}_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{y}}\right) / \left\|\boldsymbol{\phi}_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{y}}\right\|_2$, and $\ell \triangleq \left\|\boldsymbol{\phi}_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{y}}\right\|_2$. In other words, let $\boldsymbol{r}$ represent the unitary direction vector for the difference, and ℓ to denote the corresponding length. In this regard, we have

$$\max_{\boldsymbol{\phi}_{\boldsymbol{x}},\boldsymbol{\phi}_{\boldsymbol{y}}\in\mathcal{S}_{\mathrm{S}}^K} \left\|\boldsymbol{V}_{\mathcal{S}}\left(\boldsymbol{\phi}_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{y}}\right)\right\|_2 \geq \sigma_{\max}(\mathcal{S}) \max_{\boldsymbol{r},\ell} \ell \left|\boldsymbol{v}_{\max}^T\boldsymbol{r}\right| \geq \sigma_{\max}(\mathcal{S}) \max_{\ell|\ \boldsymbol{r}=\boldsymbol{v}_{\max}} \ell. \tag{26}$$

Now we should find a lower-bound for maximum possible ℓ that can be achieved for some fixed direction $\boldsymbol{r} = \boldsymbol{v}_{\max}$. To this aim, assume the difference vector $\boldsymbol{\phi}_{\boldsymbol{x}} - \boldsymbol{\phi}_{\boldsymbol{y}}$ starts from any arbitrary vertex, i.e., $\boldsymbol{\phi}_{\boldsymbol{x}}$ is a one-hot vector with K components equal to zero and the remaining one equal to 1. In this regard, the minimum possible ℓ that can be achieved (minimum is taken w.r.t. direction of the difference vector) occurs when the difference vector becomes perpendicular to its front facet. Since all $\boldsymbol{\phi}$ vectors belong to the standard simplex, such difference vectors can be easily shown to be of the form $\boldsymbol{\Delta}$ with

$$\Delta_i = 1 \quad \text{and} \quad \Delta_{j|j\neq i} = -1/K,$$

for any $i \in 0, \ldots, K$. The ℓ_2-norm of all such vectors equals to $\sqrt{1+1/K}$. Thus, we have

$$d_{\mathcal{S}} \geq \sqrt{\frac{K+1}{K}}\sigma_{\max}(S) \geq \sigma_{\max}(S). \tag{27}$$

Assume we have $2m$ i.i.d. samples drawn from a noisy simplex $f_{\mathcal{S}} * \mathbb{G}_{\sigma}$, denoted by $\{\boldsymbol{y}\}_{i=1}^{2m}$. This way, we have $\boldsymbol{y}_i = \boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_i + \boldsymbol{z}_i$ for $i = 1, \ldots, 2m$, where $\boldsymbol{\phi}_i$ represents the weight vector for the ith sample (drawn from a uniform Dirichlet distribution), while $\boldsymbol{z}_i \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}\right)$ is an independently drawn noise vector. Let us define D as

$$\mathrm{D} = \frac{1}{2m}\sum_{i=1}^{m} \|\boldsymbol{y}_{2i} - \boldsymbol{y}_{2i-1}\|_2^2. \tag{28}$$

To find an upper-bound for $d_{\mathcal{S}}$, it is enough to find a high probability lower-bound for D in terms of $d_{\mathcal{S}}$. To this aim, we rewrite D as

$$\begin{aligned}
\mathrm{D} &= \frac{1}{2m}\sum_{i=1}^{m}\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}+\boldsymbol{z}_{2i}-\boldsymbol{z}_{2i-1}\|_2^2 \qquad (29)\\
&= \frac{1}{2m}\sum_{i=1}^{m}\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\|_2^2+\frac{1}{2m}\sum_{i=1}^{m}\|\boldsymbol{z}_{2i}-\boldsymbol{z}_{2i-1}\|_2^2+\frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\right)^T(\boldsymbol{z}_{2i}-\boldsymbol{z}_{2i-1})\\
&= \frac{1}{2m}\sum_{i=1}^{m}\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\|_2^2+\frac{2\sigma^2}{2m}\sum_{i=1}^{m}\|\tilde{\boldsymbol{z}}_{2i}-\tilde{\boldsymbol{z}}_{2i-1}\|_2^2+\frac{\sigma\sqrt{2}}{m}\sum_{i=1}^{m}\left(\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\right)^T(\tilde{\boldsymbol{z}}_{2i}-\tilde{\boldsymbol{z}}_{2i-1}).
\end{aligned}$$

Let us denote the three terms in r.h.s. of (29) as I, II and III, respectively. Also, we have $\tilde{\boldsymbol{z}}_i=\boldsymbol{z}_i/\sigma$. To find a lower-bound for D, we should find respective lower-bounds for I, II and III. First, let us discuss about (I):

$$\mathrm{I}=\frac{1}{2m}\sum_{i=1}^{m}\left\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\right\|_2^2=f_{\boldsymbol{V}_{\mathcal{S}}}\left(\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\cdots,\boldsymbol{\phi}_{2m}\right). \qquad (30)$$

For the term inside of the summation, we have

$$\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\|_2^2 \overset{a.s.}{\leq} d_{\mathcal{S}}^2, \qquad (31)$$

$$\mathrm{Var}\left(\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\|_2^2\right) \leq \frac{12\sigma_{\max}^4(\mathcal{S})}{K^3} \leq \frac{12d_{\mathcal{S}}^4}{K^3} \qquad (32)$$

From the above inequalities we can see that $\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\|_2^2$ satisfies the Bernstein's condition with $b=2d_{\mathcal{S}}^2$ (see eq. 2.15 in Wainwright (2019)). Then, from proposition 2.10 in Wainwright (2019), the following inequality holds with probability at least $1-\delta/6$, for any $\delta\in(0,1)$:

$$f_{\boldsymbol{V}_{\mathcal{S}}}\left(\boldsymbol{\phi}_1,\cdots,\boldsymbol{\phi}_{2m}\right)\geq\mathbb{E}\left[f_{\boldsymbol{V}_{\mathcal{S}}}\left(\boldsymbol{\phi}_1,\cdots,\boldsymbol{\phi}_{2m}\right)\right]-2\sqrt{\max\left\{\frac{4\log\frac{6}{\delta}}{m},\frac{3}{K^3}\right\}\cdot\frac{d_{\mathcal{S}}^4}{m}\log\frac{6}{\delta}}, \qquad (33)$$

where for the expected value of the function f we have:

$$\begin{aligned}
\mathbb{E}\left[f_{\boldsymbol{V}_{\mathcal{S}}}\left(\boldsymbol{\phi}_1,\cdots,\boldsymbol{\phi}_{2m}\right)\right] &= \frac{1}{2m}\sum_{i=1}^{m}\mathbb{E}\left[\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\|_2^2\right]\\
&= \frac{1}{2}\mathbb{E}\left[\|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}'\|_2^2\right]\\
&= \frac{1}{2}\mathbb{E}\left[\left(\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}'\right)^T\left(\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}'\right)\right]\\
&= \frac{1}{2}\mathbb{E}\left[\left(\boldsymbol{\phi}-\boldsymbol{\phi}'\right)^T\boldsymbol{V}_{\mathcal{S}}^T\boldsymbol{V}_{\mathcal{S}}\left(\boldsymbol{\phi}-\boldsymbol{\phi}'\right)\right]\\
&= \frac{1}{2}\mathbb{E}\left[\mathrm{Tr}\left(\boldsymbol{V}_{\mathcal{S}}^T\boldsymbol{V}_{\mathcal{S}}\left(\boldsymbol{\phi}-\boldsymbol{\phi}'\right)\left(\boldsymbol{\phi}-\boldsymbol{\phi}'\right)^T\right)\right]\\
&= \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{V}_{\mathcal{S}}^T\boldsymbol{V}_{\mathcal{S}}\mathbb{E}\left[\left(\boldsymbol{\phi}-\boldsymbol{\phi}'\right)\left(\boldsymbol{\phi}-\boldsymbol{\phi}'\right)^T\right]\right)\\
&= \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{V}_{\mathcal{S}}^T\boldsymbol{V}_{\mathcal{S}}\frac{1}{(K+1)^2(K+2)}\begin{bmatrix}K & -1 & \dots & -1\\ -1 & K & \dots & -1\\ \vdots & \vdots & \ddots & \vdots\\ -1 & -1 & \dots & K\end{bmatrix}\right)\\
&= \frac{1}{2(K+1)^2(K+2)}\mathrm{Tr}\left(\boldsymbol{V}_{\mathcal{S}}^T\boldsymbol{V}_{\mathcal{S}}\left((K+1)\boldsymbol{I}-\mathbf{1}\mathbf{1}^T\right)\right)\\
&= \frac{1}{2(K+1)^2(K+2)}\left((K+1)\,\mathrm{Tr}\left(\boldsymbol{V}_{\mathcal{S}}^T\boldsymbol{V}_{\mathcal{S}}\right)-\mathrm{Tr}\left(\mathbf{1}^T\boldsymbol{V}_{\mathcal{S}}^T\boldsymbol{V}_{\mathcal{S}}\mathbf{1}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2(K+2)} \left(\sum_{i=1}^{K} \frac{1}{K+1} \sum_{j=1}^{K+1} \boldsymbol{V}_{\mathcal{S}}^2(i,j) - \sum_{i=1}^{K} \left(\frac{1}{K+1} \sum_{j=1}^{K+1} \boldsymbol{V}_{\mathcal{S}}(i,j) \right)^2 \right) \\
&= \frac{1}{2(K+2)} \sum_{i=1}^{K} \frac{1}{K+1} \sum_{j=1}^{K+1} \left(\boldsymbol{V}_{\mathcal{S}}(i,j) - \frac{1}{K+1} \sum_{j=1}^{K+1} \boldsymbol{V}_{\mathcal{S}}(i,j) \right)^2 \\
&= \frac{1}{2(K+2)} \sum_{i=1}^{K} \frac{1}{K+1} \sum_{j=1}^{K+1} \left(\widehat{\boldsymbol{V}}_{\mathcal{S}}(i,j) - \frac{1}{K+1} \sum_{j=1}^{K+1} \widehat{\boldsymbol{V}}_{\mathcal{S}}(i,j) \right)^2 \\
&= \frac{1}{2(K+2)} \sum_{i=1}^{K} \frac{1}{K+1} \sum_{j=1}^{K+1} \widetilde{\boldsymbol{V}}_{\mathcal{S}}(i,j)^2 \\
&= \frac{1}{2(K+2)(K+1)} \sum_{j=1}^{K+1} \|\widetilde{\boldsymbol{V}}_{\mathcal{S}}^j\|_2^2 \qquad (34)
\end{aligned}
$$

where $\widetilde{\boldsymbol{V}}_{\mathcal{S}}$ and $\widehat{\boldsymbol{V}}_{\mathcal{S}}$ are defined as

$$
\begin{aligned}
\boldsymbol{V}_{\mathcal{S}} &= [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{K+1}] \qquad (35) \\
\widehat{\boldsymbol{V}}_{\mathcal{S}} &= [\boldsymbol{0}, \boldsymbol{v}_2 - \boldsymbol{v}_1, \boldsymbol{v}_3 - \boldsymbol{v}_1, \cdots, \boldsymbol{v}_{K+1} - \boldsymbol{v}_1] = [\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \cdots, \hat{\boldsymbol{v}}_{K+1}] \\
\widetilde{\boldsymbol{V}}_{\mathcal{S}} &= \left[\hat{\boldsymbol{v}}_1 - \frac{\sum_1^K \hat{\boldsymbol{v}}_i}{K+1}, \hat{\boldsymbol{v}}_2 - \frac{\sum_1^K \hat{\boldsymbol{v}}_i}{K+1}, \cdots, \hat{\boldsymbol{v}}_K - \frac{\sum_1^K \hat{\boldsymbol{v}}_i}{K+1} \right] = [\tilde{\boldsymbol{v}}_1, \tilde{\boldsymbol{v}}_2, \cdots, \tilde{\boldsymbol{v}}_K, \tilde{\boldsymbol{v}}_{K+1}], \qquad (36)
\end{aligned}
$$

And $\widetilde{\boldsymbol{V}}_{\mathcal{S}}^j$ is the jth column of $\widetilde{\boldsymbol{V}}_{\mathcal{S}}$. We know that the maximum distance between any two points inside a simplex is equal to the length of the biggest edge of the simplex. Now without loss of generality suppose that $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ are the vertices related to this edge (if more than one edge have the maximum length, suppose that, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ are belong to one of them). Therefore the length of the vector $\boldsymbol{v}_3 - \boldsymbol{v}_2$ associated with the maximum distance satisfies the following

$$
\begin{aligned}
d_{\mathcal{S}}^2 &= \|\boldsymbol{v}_3 - \boldsymbol{v}_2\|_2^2 \\
&= \|\tilde{\boldsymbol{v}_3} - \tilde{\boldsymbol{v}_2}\|_2^2 \\
&\leq \|\tilde{\boldsymbol{v}_3}\|_2^2 + \|\tilde{\boldsymbol{v}_2}\|_2^2 + 2\|\tilde{\boldsymbol{v}_3}\|_2 \|\tilde{\boldsymbol{v}_2}\|_2 \\
&\leq 2\left(\|\tilde{\boldsymbol{v}}_3\|_2^2 + \|\tilde{\boldsymbol{v}}_2\|_2^2 \right). \qquad (37)
\end{aligned}
$$

Now the following inequality can be deduced from 34 and 37

$$
\mathbb{E}\left[f_{\boldsymbol{V}_{\mathcal{S}}}\left(\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{2m} \right) \right] \geq \frac{1}{4(K+1)(K+2)} d_{\mathcal{S}}^2. \qquad (38)
$$

From 33 and 38, it is easy to show that if we have $2m \geq 2000(K+1)(K+2)\log\frac{6}{\delta}$ i.i.d. samples from $\mathbb{P}_{\mathcal{S}}$, then with probability at least $1 - \frac{\delta}{6}$ (for any $0 < \delta \leq 1$), we have

$$
f\left(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_{2m} \right) \geq \frac{d_{\mathcal{S}}^2}{8(K+1)(K+2)}. \qquad (39)
$$

Next, we should find a lower-bound for the second term in the r.h.s. of 29, i.e., II.

In 29, $\boldsymbol{z}_i$s are drawn from a multivariate Gaussian distribution with mean vector $\boldsymbol{0}$ and covariance matrix $\sigma^2\mathbf{I}$. Then, $\boldsymbol{z}_i' = \tilde{\boldsymbol{z}}_{2i} - \tilde{\boldsymbol{z}}_{2i-1}$ is also a zero-mean Gaussian random vector with covariance matrix $\mathbf{I}$, and as a result $\|\boldsymbol{z}_i'\|_2^2$ is a Chi-squared random variable with K degrees of freedom. We know that a Chi-squared random variable with K degrees of freedom is sub-exponential with parameters $\left(2\sqrt{K}, 4\right)$ and therefore $\frac{1}{2m}\sum_{i=1}^{m} \|\boldsymbol{z}_i'\|_2^2$ is a sub-exponential random variable

with parameters $\left(\sqrt{K/m}, 2/m\right)$. Then the following holds with probability at least $1-\delta/6$:

$$\begin{aligned}\frac{1}{2m}\sum_{i=1}^{m}\|\tilde{\boldsymbol{z}}_{2i}-\tilde{\boldsymbol{z}}_{2i-1}\|_2^2 &\geq \frac{1}{2m}\sum_{i=1}^{m}\mathbb{E}\left[\|\tilde{\boldsymbol{z}}_{2i}-\tilde{\boldsymbol{z}}_{2i-1}\|_2^2\right]-\sqrt{\frac{2K}{m}\log\frac{6}{\delta}}\\ &\geq \frac{K}{2}-\sqrt{\frac{2K}{m}\log\frac{6}{\delta}}\\ &\geq \frac{K}{2}-\sqrt{\frac{K}{1000(K+1)(K+2)}}\\ &\geq \frac{K}{2}-1=\frac{K-2}{2}. \end{aligned} \tag{40}$$

Based on the above inequalities, we can give a lower-bound for the second term in the r.h.s. of 29:

$$\mathrm{II}=\frac{2\sigma^2}{2m}\sum_{i=1}^{m}\|\tilde{\boldsymbol{z}}_{2i}-\tilde{\boldsymbol{z}}_{2i-1}\|_2^2\geq (K-2)\sigma^2. \tag{41}$$

Now to find a lower bound for $\mathrm{D}\left(\mathrm{S}\right)$, we only need to give a lower bound for III, in the last inequality of 29. For this part we have:

$$\begin{aligned}\mathrm{III} &= \frac{\sigma\sqrt{2}}{m}\sum_{i=1}^{m}\left(\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\right)^T\left(\tilde{\boldsymbol{z}}_{2i}-\tilde{\boldsymbol{z}}_{2i-1}\right)\\ &= \frac{\sigma\sqrt{2}}{m}\sum_{i=1}^{m}\left(\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\right)^T\boldsymbol{z}_i'\\ &= \frac{\sigma\sqrt{2}}{m}\sum_{i=1}^{m}g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(z_i'^1,z_i'^2,\cdots,z_i'^m\right), \end{aligned} \tag{42}$$

where in the above equations we have:

$$\mathbb{E}\left(g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(z_i'^1,z_i'^2,\cdots,z_i'^m\right)\right)=\mathbb{E}\left(\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\right)^T\mathbb{E}\left(\boldsymbol{z}_i'\right)=0. \tag{43}$$

To find a lower bound for III we try to bound $g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(z_i'^1,z_i'^2,\cdots,z_i'^m\right)$ from below. To do so, we write the concentration inequality for this function as follow:

$$\begin{aligned}\mathbb{P}\left(g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(z_i'^1,z_i'^2,\cdots,z_i'^m\right)\geq\epsilon\right) &\leq \frac{\mathbb{E}_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}},\boldsymbol{z}_i'}\left[e^{\lambda g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(z_i'^1,z_i'^2,\cdots,z_i'^m\right)}\right]}{e^{\lambda\epsilon}}\\ &\leq \frac{\mathbb{E}_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}},\boldsymbol{z}_i'}\left[e^{\lambda g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(z_i'^1,z_i'^2,\cdots,z_i'^m\right)}\right]}{e^{\lambda\epsilon}}\\ &= \frac{\mathbb{E}_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left[\mathbb{E}_{\boldsymbol{z}_i'}\left[e^{\lambda g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(z_i'^1,z_i'^2,\cdots,z_i'^m\right)}\right]\right]}{e^{\lambda\epsilon}}, \end{aligned} \tag{44}$$

where $g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(\boldsymbol{z}\right)$ is a lipschitz function with respect to $\boldsymbol{z}'$:

$$\begin{aligned}|g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(\boldsymbol{z}_1\right)-g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(\boldsymbol{z}_2\right)| &\leq \|\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\|_2\|\boldsymbol{z}_1-\boldsymbol{z}_2\|_2\\ &\leq 2d_{\mathcal{S}}\|\boldsymbol{z}_1-\boldsymbol{z}_2\|_2. \end{aligned} \tag{45}$$

From Lemma (2.27) in Wainwright (2019) we know that any L-lipschitz function of a Gaussian R.V. is sub-Gaussian. As a result, we have

$$\mathbb{E}_{\boldsymbol{z}_i'}\left[e^{\lambda g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(z_i'^1,z_i'^2,\cdots,z_i'^m\right)}\right]\leq e^{\frac{4\pi^2 d_{\mathcal{S}}^2\lambda^2}{8}}. \tag{46}$$

From inequalities 46, 44 and equation 42 we have:

$$\mathbb{P}\left(\frac{\sigma\sqrt{2}}{2m}\sum_{i=1}^{m} g_{\boldsymbol{\phi_{2i}},\boldsymbol{\phi_{2i-1}}}\left(\boldsymbol{z}_i\right) \leq -\epsilon\right) \leq \min_{\lambda\geq 0} \frac{\mathbb{E}_{\boldsymbol{\phi_1},\cdots\boldsymbol{\phi_{2m}}}\left[e^{\frac{\lambda^2\pi^2 d_{\mathcal{S}}^2\sigma^2}{2m}\cdot}\right]}{e^{\lambda\epsilon}} \leq e^{-\frac{m\epsilon^2}{2\pi^2 d_{\mathcal{S}}^2\sigma^2}}. \tag{47}$$

Then if we have $2m \geq 2000(K+1)(K+2)\log\frac{6}{\delta}$, i.i.d samples, then with probability at least $1-\frac{\delta}{6}$ we have:

$$\frac{1}{2m}\sum_{i=1}^{m}\left(\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i}-\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_{2i-1}\right)^T\left(\boldsymbol{z}_{2i}-\boldsymbol{z}_{2i-1}\right) \geq -\frac{\pi\sigma d_{\mathcal{S}}}{20\sqrt{(K+2)(K+1)}}. \tag{48}$$

Now from inequalities 39,41, 48 and equation 29 we have:

$$\begin{aligned}
\mathrm{D} \geq & \frac{d_{\mathcal{S}}^2}{8(K+1)(K+2)} + (K-2)\sigma^2 - \frac{\pi\sigma d_{\mathcal{S}}}{20\sqrt{(K+2)(K+1)}} \\
= & \frac{d_{\mathcal{S}}^2}{8(K+1)(K+2)} + \left(\sigma\sqrt{K-2} - \frac{d_{\mathcal{S}}\sqrt{\pi}}{40\sqrt{(K+1)(K+2)(K-2)}}\right)^2 - \\
& \frac{d_{\mathcal{S}}^2\pi}{1600(K+1)(K+2)(K-2)} \qquad (49) \\
= & (K-2)\sigma^2 + \left(\frac{\pi\sigma}{10\sqrt{2}} - \frac{d_{\mathcal{S}}}{2\sqrt{2}\sqrt{(K+1)(K+2)}}\right)^2 - \frac{\pi^2\sigma^2}{200}. \qquad (50)
\end{aligned}$$

From the inequalities 50 and 49 it can be seen that if we have $2m \geq 2000(K+1)(K+2)\log\frac{6}{\delta}$ then the following inequalities with probability at least $1-\delta/2$ give upper-bounds for the maximum distance between two points in the simplex $\mathcal{S}$, and the noise variance σ:

$$d_{\mathcal{S}} \leq 4\sqrt{(K+1)(K+2)D} = R. \tag{51}$$

$$\sigma^2 \leq \frac{D}{K-3} = R_n \tag{52}$$

In the same way we could also give an upper bound for D. The following inequalities hold with probability more than $1-\delta/2$:

$$\begin{aligned}
\mathrm{D} \leq & \frac{d_{\mathrm{S}}^2}{K+2} + \sigma^2(K+2) + \frac{\pi\sigma d_{\mathcal{S}}}{20\sqrt{(K+2)(K+1)}} \\
\leq & \left(\frac{d_{\mathrm{S}}}{\sqrt{K+2}} + \sigma\sqrt{K+2}\right)^2 \\
= & \frac{d_{\mathrm{S}}^2}{K+2}\left(1+\frac{K+2}{d_{\mathrm{S}}/\sigma}\right)^2 \qquad (53) \\
= & \sigma^2(K+2)\left(1+\frac{d_{\mathrm{S}}/\sigma}{K+2}\right)^2. \qquad (54)
\end{aligned}$$

Based on equation 51 and inequations in 53 and 54, we can give an upper bound for the radius R in 51 and noise radius R_n. The following inequality holds with probability at least $1-\delta$:

$$R \leq 4\sqrt{K+1}\left(1+\frac{K+2}{d_{\mathrm{S}}/\sigma}\right)d_{\mathrm{S}} \tag{55}$$

$$R_n \leq \frac{K+2}{K-3}\left(1+\frac{d_{\mathrm{S}}/\sigma}{K+2}\right)\sigma. \tag{56}$$

Now the only thing we should do is to find a point inside of the simplex. To do this we define a new statistic $\boldsymbol{p}$ as follows:

$$\boldsymbol{p} = \frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{x}_i = \frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_i + \boldsymbol{z}_i = \boldsymbol{V}_{\mathcal{S}}\left(\frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{\phi}_i\right) + \frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{z}_i = \mathrm{I} + \mathrm{II} \tag{57}$$

It is clear that ''I" is placed inside of the true simplex. So the distance between $\boldsymbol{p}$ and a point inside the main simplex can be calculated as follows:

$$\min_{\boldsymbol{x}\in\mathcal{S}} \|\boldsymbol{p}-\boldsymbol{x}\|_2 \leq \left\|\frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{z}_i\right\|_2. \tag{58}$$

In the above equation $\frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{z}_i$, is a zero mean Gaussian random vector with covariance matrix $\frac{\sigma^2}{2m}\boldsymbol{I}$, therefore $\|\frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{z}_i\|_2^2$ is a chi-squared random variable with K degrees of freedom, and so we can write concentration inequality for it:

$$\mathrm{P}\left[\left|\left\|\frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{z}_i\right\|_2^2 - \frac{K\sigma^2}{2m}\right| \geq \epsilon\right] \leq 2e^{-\frac{m^2\epsilon^2}{2K\sigma^4}}. \tag{59}$$

Then with probability at least $1-\delta$ we have the following inequality:

$$\begin{aligned}\min_{\boldsymbol{x}\in\mathcal{S}} \|\boldsymbol{p}-\boldsymbol{x}\|_2^2 \leq\ & \frac{K\sigma^2}{2m} + \frac{\sigma^2}{m}\sqrt{2K\log\frac{2}{\delta}} \\ \leq\ & \frac{K\sigma^2}{m}\log\frac{2}{\delta}\end{aligned} \tag{60}$$

Now if we have $2m \geq 2000(K+1)(K+2)\log\frac{6}{\delta}$, then we can rewrite the above inequality as follows:

$$\min_{\boldsymbol{x}\in\mathcal{S}} \|\boldsymbol{p}-\boldsymbol{x}\|_2 \leq \frac{\sigma}{10\sqrt{10}\sqrt{K+2}}. \tag{61}$$

As we mentioned earlier, the first part in the right hand side of the inequality 57 , $\frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_i$, belongs to the interior of the simplex. Then a K-dimensional sphere with radius R and centered at this point, with probability at least $1-\delta$ contains the simplex $\mathcal{S}$. However, in the process of learning we do not have access to the noiseless data and thus cannot have such a point as the center of the sphere. Let us show the distance between $\boldsymbol{p}$ and $\frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{V}_{\mathcal{S}}$ with d. Then, it is clear to see that the K-dimensional sphere with radius $R+d$ and center point $\boldsymbol{p}$, contains the sphere with center point at $\frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_i$ and radius R. So any simplex which is placed in sphere $\mathrm{C}^K(\frac{1}{2m}\sum_{i=1}^{2m}\boldsymbol{V}_{\mathcal{S}}\boldsymbol{\phi}_i, R)$ is also placed in sphere $\mathrm{C}^K(\boldsymbol{p}, R+d)$.

So we conclude that if we have $2m \geq 2000(K+1)(K+2)\log\frac{6}{\delta}$ i.i.d. samples from $\mathbb{G}_{\mathcal{S}}$ then the main simplex $\mathcal{S}$ with probability more than or equal to $1-\delta$ will be confined in a K-dimensional sphere with center point at $\boldsymbol{p}$ and with radius R:

$$\begin{aligned}R =\ & 4\sqrt{(K+1)(K+2)D} + \frac{\sqrt{D}}{40\sqrt{K+2}} \\ \leq\ & 4\sqrt{(K+1)(K+2)D}\left(1 + \frac{1}{160(K+2)\sqrt{K+1}}\right) \\ \leq\ & 8\sqrt{(K+1)(K+2)D}.\end{aligned} \tag{62}$$

This completes the proof. □

proof of Lemma 3.2. Consider a $(\underline{\theta}, \bar{\theta})$-isoperimetric K-simplex $\mathcal{S}$, which is bounded in a sphere $\mathrm{C}^K(\boldsymbol{p}, R)$. It is clear that all vertices of this simplex $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{K+1}\}$ are placed in $\mathrm{C}^K(\boldsymbol{p}, R)$. From the definition of $\mathrm{T}_{\frac{\alpha\epsilon}{K+1}}(\mathrm{C}^K(\boldsymbol{p}, R))$, we know

that for each vertex of the simplex $\boldsymbol{v}_i$, there exists some $\boldsymbol{v}'_i \in \mathrm{T}_{\frac{\alpha\epsilon}{K+1}}(\mathrm{C}^K(\boldsymbol{p}, R))$ such that $\|\boldsymbol{v}_i - \boldsymbol{v}'_i\|_2 \leq \frac{\alpha\epsilon}{K+1}$. Assume for each vertex $\boldsymbol{v}_i$ of $\mathcal{S}$, we denote its closest point in $\mathrm{T}_{\frac{\alpha\epsilon}{K+1}}(\mathrm{C}^K(\boldsymbol{p}, R))$ as

$$\hat{\boldsymbol{v}}_i = \operatorname{argmin}\left\{\|\boldsymbol{v}_i - \hat{\boldsymbol{v}}\|_2 \,\middle|\, \hat{\boldsymbol{v}} \in \mathrm{T}_{\frac{\alpha\epsilon}{K+1}}(\mathrm{C}^K(\boldsymbol{p}, R))\right\}.$$

Using these points we make a new simplex $\widehat{\mathcal{S}}$. It is clear that $\widehat{\mathcal{S}}$ belongs to $\widehat{\mathbb{S}}\left(\mathrm{C}^K(\boldsymbol{p}, R)\right)$. Assume $f_{\mathcal{S}}$ and $f_{\widehat{\mathcal{S}}}$ denote the probability density functions that correspond to $\mathcal{S}$ and $\widehat{\mathcal{S}}$, respectively. Then, the TV-distance between $f_{\mathcal{S}}$ and $f_{\widehat{\mathcal{S}}}$ can be written and bounded as

$$\begin{aligned}
\mathrm{TV}\left(\mathbb{P}_{\mathcal{S}}, \mathbb{P}_{\widehat{\mathcal{S}}}\right) &= \sup_{A} \mathbb{P}_{\mathcal{S}}(A) - \mathbb{P}_{\widehat{\mathcal{S}}}(A) \\
&= \int_{\boldsymbol{x} \in \{x' : f_{\mathcal{S}}(\boldsymbol{x}') \geq f_{\widehat{\mathcal{S}}}(\boldsymbol{x}')\}} f_{\mathcal{S}}(\boldsymbol{x}) - f_{\widehat{\mathcal{S}}}(\boldsymbol{x}) \\
&= \int_{\boldsymbol{x} \in \mathcal{S} - \widehat{\mathcal{S}}} f_{\mathcal{S}}(\boldsymbol{x}) - f_{\widehat{\mathcal{S}}}(\boldsymbol{x}) + \left(\int_{\boldsymbol{x} \in \mathcal{S} \cap \widehat{\mathcal{S}}} f_{\mathcal{S}}(\boldsymbol{x}) - f_{\widehat{\mathcal{S}}}(\boldsymbol{x})\right) \mathbf{1}\left(\mathrm{Vol}(\mathcal{S}) \leq \mathrm{Vol}\left(\widehat{\mathcal{S}}\right)\right) \\
&\leq \int_{\boldsymbol{x} \in \mathcal{S} - \widehat{\mathcal{S}}} f_{\mathcal{S}}(\boldsymbol{x}) - f_{\widehat{\mathcal{S}}}(\boldsymbol{x}) + \int_{\boldsymbol{x} \in \mathcal{S} \cap \widehat{\mathcal{S}}} \left| f_{\mathcal{S}}(\boldsymbol{x}) - f_{\widehat{\mathcal{S}}}(\boldsymbol{x}) \right| \\
&= \frac{\mathrm{Vol}\left(\mathcal{S} - \widehat{\mathcal{S}}\right)}{\mathrm{Vol}(S)} + \mathrm{Vol}\left(\mathcal{S} \cap \widehat{\mathcal{S}}\right) \left| \frac{1}{\mathrm{Vol}(\mathcal{S})} - \frac{1}{\mathrm{Vol}(\widehat{\mathcal{S}})} \right|,
\end{aligned} \tag{63}$$

where $\mathcal{S} - \widehat{\mathcal{S}}$ shows the set difference between $\mathcal{S}$ and $\widehat{\mathcal{S}}$. Let us denote the two terms in the r.h.s. of 63 as I and II, respectively. To find an upper-bound for the TV-distance in 63 we find respective upper-bounds for I and II. In order to do so, first let us discuss about I:

$$\mathrm{I} = \frac{\mathrm{Vol}\left(\mathcal{S} - \widehat{\mathcal{S}}\right)}{\mathrm{Vol}(S)}. \tag{64}$$

The distance between vertices of $\mathcal{S}$ and $\widehat{\mathcal{S}}$ are less than $\frac{\alpha\epsilon}{K+1}$, then the maximal difference set (in terms of volume) between $\mathcal{S}$ and $\widehat{\mathcal{S}}$ can be bounded as follows: simplex has $K+1$ facets, and the difference set that can occur from altering each of them is upper-bounded as $\leq \mathcal{A}_i(\mathcal{S}) \times \alpha\epsilon/(K+1)$, where $i = 1, \ldots, K+1$ denotes the facet index. This way, for I we have:

$$\begin{aligned}
\mathrm{I} &\leq \sum_{i=1}^{K+1} \frac{\alpha\epsilon}{K+1} \frac{\mathcal{A}_i(\mathcal{S})}{\mathrm{Vol}(\mathcal{S})} \\
&\leq \alpha\epsilon \mathcal{A}_{max}(\mathcal{S}) \\
&\leq \alpha\epsilon\bar{\theta}\mathrm{Vol}^{\frac{1}{K}},
\end{aligned} \tag{65}$$

where $\mathcal{A}_{max}(\mathcal{S})$ is the volume of the largest facet of $\mathcal{S}$. In the final inequality of 65, we take advantage of the isoperimetricity property of $\mathcal{S}$. Next, we should find an upper-bound for the second term in the r.h.s. of 63, i.e., II:

$$\mathrm{II} = \mathrm{Vol}\left(\mathcal{S} \cap \widehat{\mathcal{S}}\right) \left| \frac{1}{\mathrm{Vol}(\mathcal{S})} - \frac{1}{\mathrm{Vol}(\widehat{\mathcal{S}})} \right|. \tag{66}$$

To find an upper-bound for II we should find an upper-bound for $\mathrm{Vol}\left(\widehat{\mathcal{S}}\right)$. To do so, we create a "rounded" simplex $\mathcal{S}^r$ which forms by adding a K-dimensional sphere with radius $r = \frac{\alpha\epsilon}{K+1}$ to the original simplex $\mathcal{S}$. It can be shown that $\widehat{\mathcal{S}}$ is

definitely placed inside $\mathcal{S}^r$ and therefore $\mathrm{Vol}(\hat{\mathcal{S}}) \leq \mathrm{Vol}(\mathcal{S}^r)$. In this regard, we have

$$\begin{aligned}
\mathrm{II} &= \mathrm{Vol}\left(\mathcal{S} \cap \widehat{\mathcal{S}}\right) \left| \frac{1}{\mathrm{Vol}(\mathcal{S})} - \frac{1}{\mathrm{Vol}(\hat{\mathcal{S}})} \right| \\
&\leq \left| \frac{\mathrm{Vol}(\hat{\mathcal{S}}) - \mathrm{Vol}(\mathcal{S})}{\mathrm{Vol}(\mathcal{S})} \right| \\
&\leq \left| \frac{\mathrm{Vol}(\mathcal{S}^r) - \mathrm{Vol}(\mathcal{S})}{\mathrm{Vol}(\mathcal{S})} \right| \\
&\leq \left(\sum_{i=1}^{K+1} \frac{\alpha\epsilon}{K+1} \mathcal{A}_i\left(\mathcal{S}\right) + (K+1)\mathcal{C}_\alpha \left(\frac{\alpha\epsilon}{K+1}\right)^K \right) \frac{1}{\mathrm{Vol}(\mathcal{S})} \\
&\leq \left((K+1)\frac{\alpha\epsilon}{K+1} \mathcal{A}_{\max}\left(\mathcal{S}\right) + (K+1)\mathcal{C}_\alpha \left(\frac{\alpha\epsilon}{K+1}\right)^K \right) \frac{1}{\mathrm{Vol}(\mathcal{S})} \\
&\leq 2\alpha\epsilon\bar{\theta}\mathrm{Vol}^{\frac{1}{K}},
\end{aligned} \tag{67}$$

where $\mathcal{C}_\alpha \left(\frac{\alpha\epsilon}{K+1}\right)^K$ is the volume of a K-dimensional sphere with radius $r = \frac{\alpha\epsilon}{K+1}$. Now, using 65 and 67 we have

$$\mathrm{TV}\left(\mathbb{P}_{\mathcal{S}}, \mathbb{P}_{\widehat{\mathcal{S}}}\right) \leq 3\alpha\epsilon\bar{\theta}\mathrm{Vol}^{\frac{1}{K}} \leq \frac{3}{5}\epsilon \leq \epsilon. \tag{68}$$

From 68, it can be seen that for any $\left(\underline{\theta}, \bar{\theta}\right)$-isoperimetric simplex $\mathcal{S}$ which is bounded in a K-dimensional sphere $\mathrm{C}^K(\boldsymbol{p}, R)$, there exists at least one simplex in $\widehat{\mathbb{S}}(\mathrm{C}^K(\boldsymbol{p}, R))$ within a TV-distance of ϵ from $\mathcal{S}$. Thus the proof is complete.

In the end, it is worth mentioning one possible method to build an ϵ-covering set for a K-dimensional sphere $\mathrm{C}^K(\boldsymbol{p}, R)$. Although there exists deterministic ways to build this set, an alternative approach is by uniformly sampling points from the sphere. Consider an $\epsilon/2$-packing set with size L for the sphere. Based on the result of the "coupon collector problem" we know that having $L\log(L)$ uniform samples from the sphere guarantees with high probability that there exists at least one sample within a distance of $\epsilon/2$ from each element of the packing set. Therefore, the resulting samples are an ϵ-covering set for the sphere. Hence, the cardinality of an ϵ-covering set built in this way is at most $\left(1 + \frac{4R}{\epsilon}\right)^{2K}$. □

C. Consistency Analysis

Assume $\mathcal{S}_1, \mathcal{S}_2 \in \mathbb{S}_K$ represent two arbitrary simplices in $\mathbb{R}^K$. In this regard, let $\mathbb{P}_{\mathcal{S}_1}$ and $\mathbb{P}_{\mathcal{S}_2}$ denote the probability measures, and $f_{\mathcal{S}_1}$ and $f_{\mathcal{S}_2}$ represent the probability density functions associated to $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. Let us assume that $\mathcal{S}_1, \mathcal{S}_2$ have a minimum degree of geometric regularity in the following sense.

Definition C.1. For a simplex $\mathcal{S} \in \mathbb{S}_K$ with vertices $\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_K \in \mathbb{R}^K$, we say $\mathcal{S}$ is $\left(\bar{\lambda}, \underline{\lambda}\right)$-regular if

$$\underline{\lambda} \leq \lambda_{\min}\left(\boldsymbol{\Theta}\right) \leq \lambda_{\max}\left(\boldsymbol{\Theta}\right) \leq \bar{\lambda},$$

where $\lambda_{\max}\left(\cdot\right)$ and $\lambda_{\min}\left(\cdot\right)$ denotes the largest and smallest eigenvalues of a matrix, respectively. Here, $\boldsymbol{\Theta}$ represents the zero-centerd vertex matrix of $\mathcal{S}$, i.e.,

$$\boldsymbol{\Theta} \triangleq \left[\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0 | \cdots | \boldsymbol{\theta}_K - \boldsymbol{\theta}_0\right].$$

This definition has tight connections to the previously used notion of $\left(\bar{\theta}, \underline{\theta}\right)$-isoperimetricity which is already used in the main body of the manuscript. In fact, it can be shown that $\mathcal{L}_{\max} = \mathcal{O}\left(\bar{\lambda}\right)$ and vice versa.

Our aim is to show that if the noisy versions of $\mathcal{S}_1$ and $\mathcal{S}_2$, which we denote by $f_{\mathcal{S}_1} * G_\sigma$ and $f_{\mathcal{S}_2} * G_\sigma$, respectively, have a maximum total variation distance of at least $\epsilon > 0$, then the TV distance between $\mathbb{P}_{\mathcal{S}_1}$ and $\mathbb{P}_{\mathcal{S}_2}$ is also bounded away from zero according to a function of ϵ, σ and the geometric reqularity of simplices $\mathcal{S}_1$ and $\mathcal{S}_2$. The theoretical core behind our method is stated in the following general theorem.

Theorem C.2 (Recovery of Low-Frequency Objects from Additive Noise)**.** *For $K \in \mathbb{N}$, consider a probability density function family $\mathscr{F} \subseteq \mathcal{M}\left(\mathbb{R}^K\right)$, i.e., a subset of distributions supported over $\mathbb{R}^K$. Assume for sufficiently large $\alpha > 0$, the*

following bound holds for all $f, g \in \mathscr{F}$:

$$\frac{1}{(2\pi)^K} \int_{\|\boldsymbol{\omega}\|_\infty \geq \alpha} |\mathcal{F}\{f\}(\boldsymbol{\omega}) - \mathcal{F}\{g\}(\boldsymbol{\omega})|^2 \leq \zeta\left(\alpha^{-1}\right) \int_{\mathbb{R}^K} |f-g|^2 ,$$

where $\mathcal{F}\{\cdot\}$ denotes the Fourier transform, and ζ is an increasing fuction with $\zeta(0) = 0$ and continuity at 0. Also, assume the probability density function $Q \in \mathcal{M}\left(\mathbb{R}^K\right)$ again for a sufficiently large $\alpha > 0$ has the following property:

$$\inf_{\|\boldsymbol{\omega}\|_\infty \leq \alpha} \|\mathcal{F}\{Q\}(\boldsymbol{\omega})\| \geq \eta(\alpha) ,$$

where $\eta(\cdot)$ is a non-negative decreasing function. Then, there exists a non-negative constant C where for any $\sigma, \varepsilon > 0$ and $f, g \in \mathscr{F}$ with $\|f-g\|_2 \geq \varepsilon$, we have

$$\|(f-g) * Q\|_2 \geq \frac{\varepsilon}{(2\pi)^K} \left(\sup_{\alpha > C} \eta(\alpha) \sqrt{1 - \zeta\left(\alpha^{-1}\right)} \right),$$

with $*$ denoting the multi-dimensional convolution operator.

Proof. For the sake of simplicity in notations, let $\mathcal{F}, \mathcal{G}, \mathcal{Q} : \mathbb{R}^K \to \mathbb{C}$ denote the Fourier transforms of f, g and Q, respectively. Due to Parseval's theorem, we have

$$\|f-g\|_2^2 = \frac{1}{(2\pi)^K} \|\mathcal{F} - \mathcal{G}\|_2^2 .$$

Also, due to the properties of the Fourier transform, which is the transofrmation of convolution into direct multiplication, one can write

$$\mathcal{F}\{(f-g) * Q\} = \mathcal{Q}(\mathcal{F} - \mathcal{G}) .$$

Thus, there exists universal constant $C > 0$ such that for any $\alpha > C$:

$$\begin{aligned} (2\pi)^K \|(f-g) * Q\|_2^2 &= \int_{\mathbb{R}^K} |\mathcal{Q}(\boldsymbol{\omega})(\mathcal{F}(\boldsymbol{\omega}) - \mathcal{G}(\boldsymbol{\omega}))|^2 \\ &\geq \int_{\|\boldsymbol{\omega}\|_\infty \leq \alpha} |\mathcal{Q}(\boldsymbol{\omega})|^2 |\mathcal{F}(\boldsymbol{\omega}) - \mathcal{G}(\boldsymbol{\omega})|^2 \\ &\geq \eta^2(\alpha) \int_{\|\boldsymbol{\omega}\|_\infty \leq \alpha} |\mathcal{F}(\boldsymbol{\omega}) - \mathcal{G}(\boldsymbol{\omega})|^2 \\ &\geq \varepsilon^2 \eta^2(\alpha) \left[1 - \zeta\left(\alpha^{-1}\right)\right] . \end{aligned} \tag{69}$$

The above chain of inequalities hold for all $\alpha > C$, therefore we have:

$$\|(f-g) * Q\|_2^2 \geq \frac{\varepsilon}{(2\pi)^K} \sup_{\alpha > C} \eta(\alpha) \sqrt{1 - \zeta(\alpha^{-1})}, \tag{70}$$

which completes the proof. □

Theorem C.2 presents a general approach to prove the recoverability of latent functions (or objects, which are the main focus in this work) from a certain class of independent additive noise. This approach works as long as the function class as well as the noise distribution are mostly comprised of low-frequency components in the Forier domain. For example, the Gaussian noise hurts low-frequency parts of a geometric object far less than its high-frequency details. More specifically, we prove the following corollary for Theorem C.2:

Corollary C.3 (Recoverability from Additive Gaussian Noise $\mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$)**.** *Consider the setting in Theorem C.2, and assume the noise distribution follows $Q \triangleq \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$ for $\sigma > 0$. Then, as long as for $f, g \in \mathscr{F}$ we have $\|f-g\|_2 \geq \varepsilon$ for some $\varepsilon \geq 0$, we also have*

$$\|(f-g) * Q\|_2 \geq \frac{\varepsilon}{(2\pi)^K} \left(\sup_{\alpha > C} \sqrt{1 - \zeta\left(\frac{1}{\alpha}\right)} e^{-K(\sigma\alpha)^2/2} \right)$$

Proof. The Fourier transform of $Q = \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$ can be computed as follows:

$$\mathcal{F}\left\{Q\right\}\left(\boldsymbol{\omega}\right) = \prod_{i=1}^{K} \mathcal{F}\left\{\mathcal{N}\left(0, \sigma^2\right)\right\}\left(\omega_i\right) = e^{-\sigma^2 \|\boldsymbol{\omega}\|_2^2/2}. \tag{71}$$

Also, it can be easily checked that

$$\inf_{\|\boldsymbol{\omega}\|_\infty \leq \alpha} e^{-\sigma^2 \|\boldsymbol{\omega}\|_2^2/2} = e^{-\sigma^2/2\left(\alpha^2 + \ldots + \alpha^2\right)} = e^{-K(\alpha\sigma)^2/2}.$$

By subsititution into the end result of Theorem C.2, the claimed bounds can be achieved and the proof is complete. □

In this regard, our main explicit theoretical contribution in this section with respect to simplices has been stated in the following theorem:

Theorem C.4 (Recoverability of Simplices from Additive Noise)**.** *For any two $\left(\bar{\lambda}, \underline{\lambda}\right)$-regular simplices $\mathcal{S}_1, \mathcal{S}_2 \in \mathbb{S}_K$ with $\bar{\lambda}, \underline{\lambda} > 0$, given that*

$$\mathcal{D}_{\mathrm{TV}}\left(\mathbb{P}_{\mathcal{S}_1 * G_\sigma}, \mathbb{P}_{\mathcal{S}_2 * G_\sigma}\right) = \frac{1}{2}\int \left|\left(f_{\mathcal{S}_1} - f_{\mathcal{S}_2}\right) * G_\sigma\right| \leq \varepsilon$$

for some $\varepsilon \geq 0$, where G_σ (for $\sigma \geq 0$) represents the density function associated to a Gaussian measure with zero mean and covariance matrix of $\sigma^2 \boldsymbol{I}_{K\times K}$ in $\mathbb{R}^K$, i.e., $\mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$. Then, we have

$$\mathcal{D}_{\mathrm{TV}}\left(\mathbb{P}_{\mathcal{S}_1}, \mathbb{P}_{\mathcal{S}_2}\right) \leq \varepsilon e^{\Omega\left(\frac{K}{\mathrm{SNR}^2}\right)},$$

where $\mathrm{SNR} \triangleq \frac{\bar{\lambda}}{K\sigma}$ denotes the effective signal-to-noise ratio, which is the ratio of the standard deviation of a scaled uniform Dirichlet distribution to that of the noise, per dimension.

Proof. Proof is based on properties of the Fourier transforms of simplices $\mathcal{S}_1$ and $\mathcal{S}_2$. For $K \in \mathbb{N}$, and any integrable function $f : \mathbb{R}^K \to \mathbb{R}$, the Fourier transform of f, denoted by $\mathcal{F}\left\{f\right\}\left(\boldsymbol{\omega}\right) : \mathbb{R}^K \to \mathbb{C}$ is defined as follows:

$$\mathcal{F}\left\{f\right\}\left(\boldsymbol{\omega}\right) \triangleq \int_{\boldsymbol{x}\in\mathbb{R}^K} f\left(\boldsymbol{x}\right) e^{-i\boldsymbol{\omega}^T\boldsymbol{x}} \mathrm{d}\boldsymbol{x}. \tag{72}$$

Throughout this proof, for $\mathcal{S} \in \mathbb{S}_K$, let us denote by $\mathcal{F}_{\mathcal{S}}$ the Fourier transform of $f_{\mathcal{S}}$, i.e., the uniform proability density function over $\mathcal{S}$. Also, the inverse Fourier transform which recovers $f_{\mathcal{S}}$ from $\mathcal{F}_{\mathcal{S}}$ can be written as

$$f_{\mathcal{S}}\left(\boldsymbol{x}\right) = \frac{1}{2\pi}\int_{\boldsymbol{\omega}\in\mathbb{R}^K} \mathcal{F}_{\mathcal{S}}\left(\boldsymbol{\omega}\right) e^{i\boldsymbol{\omega}^T\boldsymbol{x}} \mathrm{d}\boldsymbol{\omega}.$$

In this regard, let Δ_K represent the standard simplex in $\mathbb{R}^K$ which means a K-simplex with $\boldsymbol{\theta}_0 = \mathbf{0}$ and $\boldsymbol{\Theta} = \boldsymbol{I}$. We begin by deriving a number of useful properties for Δ_K through the following lemmas.

Lemma C.5. *Let $\mathscr{F}_{\Delta_k}\left(\omega_1, \ldots, \omega_k\right) : \mathbb{R}^k \to \mathbb{C}$ for $k \in [K]$ represent the Fourier transform of f_{Δ_k} in $\mathbb{R}^k$. Also, we have $\boldsymbol{\omega}_{1:k} \triangleq \left(\omega_1, \ldots, \omega_k\right)$. Then, the following recursive relation holds for $k > 1$:*

$$\mathscr{F}_{\Delta_k}\left(\boldsymbol{\omega}_{1:k}\right) = \frac{k}{i\omega_k}\left[\mathscr{F}_{\Delta_{k-1}}\left(\boldsymbol{\omega}_{1:k-1}\right) - e^{-i\omega_k}\mathscr{F}_{\Delta_{k-1}}\left(\omega_1 - \omega_k, \ldots, \omega_{k-1} - \omega_k\right)\right].$$

Proof. For any $\left(x_1, \ldots, x_k\right) \in \Delta_k$, Let $Z_i \triangleq x_1 + \ldots + x_i$ for $i \in [k]$. Then

$$\begin{aligned}\mathscr{F}_{\Delta_k}\left(\boldsymbol{\omega}_{1:k}\right) &= \int_0^1 \int_0^{1-Z_1} \cdots \int_0^{1-Z_{k-1}} k! e^{-i\left(\omega_1 x_1 + \ldots + \omega_k x_k\right)} \mathrm{d}x_1 \ldots \mathrm{d}x_k \\ &= \int_0^1 \cdots \int_0^{1-Z_{k-2}} k! e^{-i\left(\omega_1 x_1 + \ldots + \omega_{k-1} x_{k-1}\right)} \left(\int_0^{1-Z_{k-1}} e^{-i\omega_k x_k} \mathrm{d}x_k\right) \mathrm{d}x_1 \ldots \mathrm{d}x_{k-1},\end{aligned} \tag{73}$$

where a uniform probability density function over Δ_k has been assumed to be $f_{\Delta_k}(\boldsymbol{x}) = (k!)\mathbf{1}(\boldsymbol{x} \in \Delta_k)$ for all $\boldsymbol{x} \in \mathbb{R}^k$, due to the fact the Lebesgue measure (or volume) of the standard simplex is $1/k!$. Since we have

$$\int_0^{1-Z_{k-1}} e^{-i\omega_k x_k} \mathrm{d}x_k = \frac{1}{i\omega_k}\left(1 - e^{-i\omega_k(1-Z_{k-1})}\right),$$

the Fourier transform can be rewritten as follows:

$$\begin{aligned}\mathscr{F}_{\Delta_k}(\boldsymbol{\omega}_{1:k}) =& \frac{k}{i\omega_k}\int_0^1 \cdots \int_0^{1-Z_{k-2}} (k-1)! e^{-i(\omega_1 x_1 + \ldots + \omega_{k-1}x_{k-1})} \mathrm{d}x_1 \ldots \mathrm{d}x_{k-1} \\ &- \frac{k e^{-i\omega_k}}{i\omega_k}\int_0^1 \cdots \int_0^{1-Z_{k-2}} (k-1)! \left(\prod_{i=1}^{k-1} e^{-i(\omega_i - \omega_k)x_i}\right) \mathrm{d}x_1 \ldots \mathrm{d}x_{k-1} \\ =& \frac{k}{i\omega_k}\left[\mathscr{F}_{\Delta_{k-1}}(\boldsymbol{\omega}_{1:k-1}) - e^{-i\omega_k}\mathscr{F}_{\Delta_{k-1}}(\omega_1 - \omega_k, \ldots, \omega_{k-1} - \omega_k)\right], \end{aligned} \tag{74}$$

which completes the proof. □

Lemma C.5 gives us a recursive procedure to produce the Fourier transform, or derive usefull properties for K-dimensional standard simplex through tools such as induction. The following lemma establishes a relation between the Fourier transform of the standard simplex and that of an arbitrary simplex in $\mathbb{S}_K$ with a non-zero Lebesgue measure.

Lemma C.6. *For a simplex $S \in \mathbb{S}_K$ with vertices $\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_K$ and a reversible zero-translated vertex matrix $\boldsymbol{\Theta}$, we have*

$$\mathscr{F}_{\mathcal{S}}(\boldsymbol{\omega}_{1:k}) = e^{-i\boldsymbol{\omega}_{1:k}^T\boldsymbol{\theta}_0}\mathscr{F}_{\Delta_K}\left(\boldsymbol{\Theta}^T\boldsymbol{\omega}_{1:k}\right).$$

Proof. Let $f_{\mathcal{S}} : \mathbb{R}^K \to \mathbb{R}_{\geq 0}$ denote the probability density function associated to $\mathcal{S}$. Then, it can be seen that

$$f_{\mathcal{S}}(\boldsymbol{x}) = \frac{1}{\det(\boldsymbol{\Theta})} f_{\Delta_K}\left(\boldsymbol{\Theta}^{-1}(\boldsymbol{x} - \boldsymbol{\theta}_0)\right), \quad \forall \boldsymbol{x} \in \mathbb{R}^K.$$

In this regard, one just needs to write down the definition of Fourier transform for simplex $\mathcal{S}$ and utilizes the change of variables technique as follows:

$$\begin{aligned}\mathscr{F}_{\mathcal{S}}(\boldsymbol{\omega}_{1:k}) &= \int_{\mathbb{R}^K} f_{\mathcal{S}}(\boldsymbol{x}_{1:k}) e^{-i\boldsymbol{\omega}_{1:k}^T\boldsymbol{x}_{1:k}} \mathrm{d}x_1 \ldots \mathrm{d}x_k \qquad (75)\\ &= \frac{1}{\det(\boldsymbol{\Theta})}\int_{\mathbb{R}^K} f_{\Delta_K}\left(\boldsymbol{\Theta}^{-1}(\boldsymbol{x} - \boldsymbol{\theta}_0)\right) e^{-i\boldsymbol{\omega}_{1:k}^T\boldsymbol{x}_{1:k}} \mathrm{d}x_1 \ldots \mathrm{d}x_k \\ &= e^{-i\boldsymbol{\omega}_{1:k}^T\boldsymbol{\theta}_0}\int_{\mathbb{R}^K} f_{\Delta_K}(\boldsymbol{u}_{1:k}) e^{-i\boldsymbol{\omega}_{1:k}^T\boldsymbol{\Theta}\boldsymbol{u}_{1:k}} \mathrm{d}u_1 \ldots \mathrm{d}u_k \\ &= e^{-i\boldsymbol{\omega}_{1:k}^T\boldsymbol{\theta}_0}\mathscr{F}_{\Delta_K}\left(\boldsymbol{\Theta}^T\boldsymbol{\omega}_{1:k}\right). \qquad (76)\end{aligned}$$

Therefore, the proof is complete. □

In the following lemma, we show that the uniform measure over the standard simplex Δ_K corresponds to a low-frequency probability density function. This would be the first step toward using Corollary C.3, in order to prove Theorem C.4.

Lemma C.7 (Low-Pass Property of Δ_K)**.** *There exists a universal constant $C > 0$, such that for $\alpha > C$ and $K \in \mathbb{N}$ the uniform probability density function over Δ_K, i.e., $f_{\Delta_K} : \mathbb{R}^K \to \mathbb{R}_{\geq 0}$, is a low-frequency function in the following sense:*

$$\frac{\mathrm{Vol}(\Delta_K)}{(2\pi)^K}\int_{\|\boldsymbol{\omega}\|_\infty \geq \alpha} |\mathscr{F}_{\Delta_K}(\boldsymbol{\omega})|^2 \leq \mathcal{O}\left(\frac{K}{\alpha}\right), \tag{77}$$

where by $\|\boldsymbol{\omega}\|_\infty$ we simply mean $\max_{i\in[K]}|\omega_i|$.

Proof. Proof is based on the direct analysis of the Fourier transform of f_{Δ_k}, which we denoted as $\mathscr{F}_{\Delta_K}(\boldsymbol{\omega})$. We take advantage of the fact that the K-dimensional unit hypercube $[0,1]^K$ can be thought as the union of $K!$ properly rotated and translated versions of Δ_K.

Mathematically speaking, assume the ordered tuple of unit axis-aligned vectors $E = (\mathbf{1}_1, \ldots, \mathbf{1}_K)$, where $\mathbf{1}_i$ for $i \in [K]$ denotes the one-hot vector over the ith component. Then, let $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_{K!} \in \mathbb{R}^{K\times K}$ be the set of orthonormal matrices, where each matrix transforms E into one of its $K!$ possible permutations, i.e.,

$$(\boldsymbol{V}_i\mathbf{1}_1, \ldots, \boldsymbol{V}_i\mathbf{1}_K) = \left(\mathbf{1}_{p_{i,1}}, \ldots, \mathbf{1}_{p_{i,K}}\right), \quad \boldsymbol{p}_i \in \mathrm{Perm}\left([K]\right).$$

In this regard, we already know that there exist $K!$ corresponding vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{K!} \in \mathbb{R}^K$, such that the following combined probability density function

$$\frac{1}{K!}\sum_{i=1}^{K!} f_{(\boldsymbol{V}_i[\Delta_K+\boldsymbol{b}_i])} = f_{\square_K}, \tag{78}$$

where $f_{\square_K}$ denotes the pdf over the K-dimensional unit hypercube. Also, it should be noted that any two distinct summands in (78) have an empty overlap. In fact, each $\boldsymbol{V}_i\left[\Delta_K + \boldsymbol{b}_i\right]$ represents a translated and rotatated version of the standard simplex Δ_K.

Next, one can see that

$$\begin{aligned}\mathcal{F}\left\{\frac{1}{K!}\sum_{i=1}^{K!} f_{(\boldsymbol{V}_i[\Delta_K+\boldsymbol{b}_i])}\right\} &= \frac{1}{K!}\sum_{i=1}^{K!}\mathcal{F}\left\{f_{(\boldsymbol{V}_i[\Delta_K+\boldsymbol{b}_i])}\right\} \\ &= \frac{1}{K!}\sum_{i=1}^{K!} e^{-i\boldsymbol{\omega}^T\boldsymbol{b}_i}\mathscr{F}_{\Delta_K}\left(\boldsymbol{V}_i^{-1}\boldsymbol{\omega}\right),\end{aligned} \tag{79}$$

where we have used the result of Lemma C.6. Also, note that $\boldsymbol{V}_i^{-1} = \boldsymbol{V}_i$ for all i, and $\boldsymbol{V_i\omega}$ represents a permutation of the components of $\boldsymbol{\omega}$. As a result and due to the symmtery of the standard simplex Δ_K and also the ℓ_∞-norm w.r.t. the ordering of the edges, we have

$$\int_{\|\boldsymbol{\omega}\|\le\alpha}\left|\mathcal{F}\left\{f_{(\boldsymbol{V}_i[\Delta_K+\boldsymbol{b}_i])}\right\}(\boldsymbol{\omega})\right|^2 = \int_{\|\boldsymbol{\omega}\|\le\alpha}\left|\mathscr{F}_{\Delta_K}\left(\boldsymbol{V}_i\boldsymbol{\omega}\right)\right|^2 = \int_{\|\boldsymbol{\omega}\|\le\alpha}\left|\mathscr{F}_{\Delta_K}\left(\boldsymbol{\omega}\right)\right|^2, \quad \forall i \in [K]. \tag{80}$$

Now, using (78) we have:

$$\begin{aligned}\int_{\|\boldsymbol{\omega}\|\le\alpha}\left|\frac{1}{K!}\sum_{i=1}^{K!}\mathcal{F}\left\{f_{(\boldsymbol{V}_i[\Delta_K+\boldsymbol{b}_i])}\right\}\right|^2 &= \frac{1}{(K!)^2}\sum_{i,j}^{K!}\int_{\|\boldsymbol{\omega}\|\le\alpha}\left|\mathscr{F}_{\boldsymbol{V}_i^{-1}\Delta_K}\bar{\mathscr{F}}_{\boldsymbol{V}_j^{-1}\Delta_K}\right|\left|e^{-i\boldsymbol{\omega}^T(\boldsymbol{b}_i-\boldsymbol{b}_j)}\right| \\ &= \frac{1}{K!}\int_{\|\boldsymbol{\omega}\|\le\alpha}\left|\mathscr{F}_{\Delta_K}\left(\boldsymbol{\omega}\right)\right|^2 + \left(1-\frac{1}{K!}\right)\mathcal{O}\left(\frac{1}{\alpha}\right).\end{aligned} \tag{81}$$

The latter term in the r.h.s. of (81) corresponds to the $(K!)^2 - K!$ summands for which we have $i \neq j$. In fact, if $i \neq j$ and based on the fact that Fourier transform preserves inner product[2], we have

$$\int_{\mathbb{R}^K}\left|\mathscr{F}_{\boldsymbol{V}_i\Delta_K}\bar{\mathscr{F}}_{\boldsymbol{V}_j\Delta_K}e^{-i\boldsymbol{\omega}^T(\boldsymbol{b}_i-\boldsymbol{b}_j)}\right| = \int_{\mathbb{R}^K} f_{\boldsymbol{V}_i[\Delta_K+\boldsymbol{b}_i]}f_{\boldsymbol{V}_j[\Delta_K+\boldsymbol{b}_j]} = 0,$$

which holds since $\Delta_i \triangleq \boldsymbol{V}_i\left[\Delta_K + \boldsymbol{b}_i\right]$ and $\Delta_j \triangleq \boldsymbol{V}_j\left[\Delta_K + \boldsymbol{b}_j\right]$ do not overlap with each other. However, when we add the constraint $\|\boldsymbol{\omega}\|_\infty \le \alpha$, we effectively take the integral over the hypercube $[-\alpha,\alpha]^K$ instead of the whole $\mathbb{R}^K$. This procedure is equivalent to the innner product of the two "smoothed" versions of f_{Δ_i} and f_{Δ_j} in the spatial domain. Here, by "smoothed" we simply mean being convolved with a K-dimensional sinc function with parameter α, i.e.,

$$\mathrm{sinc}_{\alpha,K}\left(\boldsymbol{x}\right) \triangleq \prod_{i=1}^{K}\frac{\sin\left(\alpha x_i\right)}{\alpha x_i}.$$

[2]This is a direct result of the fact that Fourier transform is an "orthonormal" transformation.

Hence, we have

$$\int_{\|\boldsymbol{\omega}\|_\infty \le \alpha} \left|\mathscr{F}_{\Delta_i} \bar{\mathscr{F}}_{\Delta_j}\right| = \int_{\mathbb{R}^K} \left[f_{\Delta_i} * \mathrm{sinc}_{\alpha,K}\right] \left[f_{\Delta_j} * \mathrm{sinc}_{\alpha,K}\right], \tag{82}$$

which is at most $\mathcal{O}(1/\alpha)$, since only the leakages that are due to $\mathrm{sinc}_{\alpha,K}$ overlap with each other.

On the other hand, the l.h.s. of (81) represents the integration of the Fourier transform of $\square_K$ within $[-\alpha,\alpha]^K$. Therefore, we have

$$\begin{aligned}
\frac{1}{(2\pi)^K K!} \int_{\|\boldsymbol{\omega}\| \le \alpha} \left|\mathscr{F}_{\Delta_K}(\boldsymbol{\omega})\right|^2 &\ge \frac{1}{(2\pi)^K} \int_{\|\boldsymbol{\omega}\| \le \alpha} \left|\mathscr{F}_{\square_K}(\boldsymbol{\omega})\right|^2 - \mathcal{O}\left(\frac{1}{\alpha}\right) \\
&= \prod_{i=1}^K \frac{1}{2\pi} \left(\int_{-\alpha}^{\alpha} \left|\int_0^1 e^{-i\omega_i x_i}\right|^2\right) - \mathcal{O}\left(\frac{1}{\alpha}\right) \\
&= \left(1 - \frac{2}{\pi\alpha} + o\left(\alpha^{-1}\right)\right)^K - \mathcal{O}\left(\frac{1}{\alpha}\right) \\
&= 1 - \mathcal{O}\left(\frac{K}{\alpha}\right),
\end{aligned} \tag{83}$$

where we have used Laurent series expansion for the integral of sinc function to derive the bound. Finally, noting the fact that we have $\mathrm{Vol}(\Delta_K) = 1/K!$, and also the Parseval's theorem:

$$\frac{1}{(2\pi)^K} \int_{\mathbb{R}^K} \left|\mathscr{F}_{\Delta_K}(\boldsymbol{\omega})\right|^2 = \int_{\mathbb{R}^K} f_{\Delta_K}^2(\boldsymbol{x}) = \frac{\mathrm{Vol}(\Delta_K)}{\mathrm{Vol}^2(\Delta_K)} = \mathrm{Vol}^{-1}(\Delta_K)$$

completes the proof. □

So far, we have managed to show that the standard simplex Δ_K is associated to a low-frequency PDF and thus would preserve a minimum level of information even after getting corrupted by additive Gaussian noise. Using Lemma C.6, one can simply extend this notion to any $(\bar{\lambda}, \underline{\lambda})$-regular simplex. Before that, let us also extend this notion of "low-frequency" property to the difference function

$$f_{\Delta_K} - f_{\boldsymbol{A}_\varepsilon(\Delta_K + \boldsymbol{b}_\varepsilon)}, \tag{84}$$

where the linear transformation matrix $\boldsymbol{A}_\varepsilon$ and translation vector $\boldsymbol{b}_\varepsilon$ (for $\varepsilon > 0$) are ε-controlled perturbations that alter the standard simplex with the following magnitude

$$\mathcal{D}_{\mathrm{TV}}\left(\mathbb{P}_{\Delta_K}, \mathbb{P}_{\boldsymbol{A}_\varepsilon(\Delta_K + \boldsymbol{b}_\varepsilon)}\right) = \varepsilon.$$

This task is more challenging since the difference PDF function in (84) can be both positive and negative, and in fact, integrates to zero over $\mathbb{R}^K$. However, the following lemma states that such functions for any $\boldsymbol{A}_\varepsilon$ and $\boldsymbol{b}_\varepsilon$ are "band-pass" in nature: even though in the Fourier domain they become exactly zero at the origin, they also become infinitesimally small in terms of magnitude as one asymotitically increases $\|\boldsymbol{\omega}\|_\infty$.

Lemma C.8 (Low-Frequency Property for Difference of Simplices)**.** *For $\varepsilon \ge 0$, Assume the linear transformation matrix $\boldsymbol{A}_\varepsilon \in \mathbb{R}^{K\times K}$ where $A_{i,j} = \delta_{ij} + \mathcal{O}(\varepsilon)$, with δ_{ij} denoting the kroecker's delta function. Also, assume the translation vector $\boldsymbol{b}_\varepsilon \in \mathbb{R}^K$ where $b_i = \mathcal{O}(\varepsilon)$. Let $\hat{\Delta}_{K,\varepsilon}$ denote the simplex $\boldsymbol{A}_\varepsilon(\Delta_K + \boldsymbol{b}_K)$, and assume the total variation distance between $\mathbb{P}_{\Delta_K}$ and $\mathbb{P}_{\hat{\Delta}_{K,\varepsilon}}$ is ε. Then*

$$\frac{1}{(2\pi)^K} \int_{\|\boldsymbol{\omega}\|_\infty \ge \alpha} \left\|\mathscr{F}_{\Delta_K} - \mathscr{F}_{\hat{\Delta}_{K,\varepsilon}}\right\|^2 \le \mathcal{O}\left(\frac{K\varepsilon}{\alpha}\right),$$

for sufficiently large $\alpha > 0$.

Proof. Based on Lemmas C.7 and C.6 and for sufficiently small ε, we already know that both $\mathscr{F}_{\Delta_K}$ and $\mathscr{F}_{\hat{\Delta}_{K,\varepsilon}}$ are low-frequency functions in the following sense: their normalized ℓ_2 energy outside of the hypercube $[-\alpha,\alpha]^K$ is at most $\mathcal{O}(K/\alpha)$. We also know that

$$\frac{1}{(2\pi)^K} \int_{\|\boldsymbol{\omega}\|_\infty \ge \alpha} \left\|\mathscr{F}_{\Delta_K} - \mathscr{F}_{\hat{\Delta}_{K,\varepsilon}}\right\|^2 = \mathcal{O}(\varepsilon),$$

due to the assumption of the lemma. However, in order to prove the $\leq \mathcal{O}\left(K\varepsilon/\alpha\right)$ bound, we need to prove another important property: that the energy distribution of the difference function is not significantly controlled by ε, e.g., it does not concentrate on high-frequency regions with $\|\boldsymbol{\omega}\|_\infty \geq \mathcal{O}\left(1/\varepsilon\right)$.

Before going further into the mathematical details of the above statement, let us present a better view on the Fourier transform of a simplex. We start this procedure with computing $\mathscr{F}_{\Delta_K}$. In this regard, we show the following relation holds for $k \in \mathbb{N}$:

$$\mathcal{F}_{\Delta_k} = \frac{k!}{i^k} \sum_{\ell=1}^{k} \left(1 - e^{-i\omega_\ell}\right) \left[\omega_\ell \prod_{j\neq\ell}^{k} \left(\omega_j - \omega_\ell\right)\right]^{-1}. \tag{85}$$

Proof of (85) is by induction. **Base**: First, we show the equality holds for the simple case of $k = 1$. **Step**: Then, we prove that for $k \geq 2$, if the relation holds for $k-1$, then it also holds for k.

Base For $k = 1$, it can be readily seen that a uniform measure over Δ_1 is a *unit pulse function* over the interval $[0, 1]$. Therefore, the Fourier transform of f_{Δ_1} is already known to be a sinc function, i.e.,

$$\mathcal{F}_{\Delta_1}\left(\omega\right) = \int_0^1 e^{-i\omega x} \mathrm{d}x = \frac{1 - e^{-i\omega}}{i\omega}, \tag{86}$$

which matches the formulation of (85) if one sets $k = 1$.

Induction step Assume (85) holds for $k-1$. By taking advantage of Lemma C.5, we have:

$$\begin{aligned}
\mathcal{F}_{\Delta_k} =& \frac{k}{i\omega_k} \left(\mathcal{F}_{\Delta_{k-1}}\left(\omega_1, \ldots, \omega_{k-1}\right) - e^{-i\omega_k} \mathcal{F}_{\Delta_{k-1}}\left(\omega_1 - \omega_k, \ldots, \omega_{k-1} - \omega_k\right)\right) \\
=& \frac{(k-1)!}{i^{k-1}} \frac{k}{i\omega_k} \left(\sum_{\ell=1}^{k-1} \left(1 - e^{-i\omega_\ell}\right) \left[\omega_\ell \prod_{j\neq\ell}^{k-1} \left(\omega_j - \omega_\ell\right)\right]^{-1} \right. \\
& \left. - e^{-i\omega_k} \sum_{\ell=1}^{k-1} \left(1 - e^{-i(\omega_\ell - \omega_k)}\right) \left[\left(\omega_\ell - \omega_k\right) \prod_{j\neq\ell}^{k-1} \left(\omega_j - \omega_\ell\right)\right]^{-1} \right) \\
=& \frac{k!}{i^k} \sum_{\ell=1}^{k-1} \left(\left(\omega_k - \omega_\ell\right)\left(1 - e^{-i\omega_\ell}\right) - \omega_\ell \left(e^{-i\omega_\ell} - e^{-i\omega_k}\right)\right) \left[\omega_\ell \omega_k \prod_{j\neq\ell}^{k-1} \left(\omega_j - \omega_\ell\right)\left(\omega_k - \omega_\ell\right)\right]^{-1} \\
=& \frac{k!}{i^k} \sum_{\ell=1}^{k-1} \left(\omega_k \left(1 - e^{-i\omega_\ell}\right) - \omega_\ell \left(1 - e^{-i\omega_k}\right)\right) \left[\omega_\ell \omega_k \prod_{j\neq\ell}^{k-1} \left(\omega_j - \omega_\ell\right)\left(\omega_k - \omega_\ell\right)\right]^{-1} \\
=& \frac{k!}{i^k} \sum_{\ell=1}^{k-1} \left(1 - e^{-i\omega_\ell}\right) \left[\omega_\ell \prod_{j\neq\ell}^{k} \left(\omega_j - \omega_\ell\right)\right]^{-1} - \frac{k!}{i^k} \left(\frac{1 - e^{-i\omega_k}}{\omega_k}\right) \sum_{\ell=1}^{k-1} \prod_{j\neq\ell}^{k} \frac{1}{\omega_j - \omega_\ell} \\
=& \frac{k!}{i^k} \sum_{\ell=1}^{k} \left(\frac{1 - e^{-i\omega_\ell}}{\omega_\ell}\right) \prod_{j\neq\ell}^{k} \frac{1}{\omega_j - \omega_\ell},
\end{aligned} \tag{87}$$

which again matches with the formulation of (85) and therefore completes the induction step. Here, for the last equality we have used the following mathematical identity:

$$\sum_{\ell=1}^{k} \prod_{j\neq\ell}^{k} \frac{1}{\omega_j - \omega_\ell} = 0.$$

End of induction

Based on this explicit formula, the difference function $\mathscr{F}_{\Delta_K} - \mathscr{F}_{\hat{\Delta}_{K,\varepsilon}}$ can be written as follows:

$$\mathscr{F}_{\Delta_K} - \mathscr{F}_{\hat{\Delta}_{K,\varepsilon}} = \frac{K!}{i^K} \sum_{\ell=1}^{K} \left(1 - e^{-i\omega_\ell}\right) \left[\omega_\ell \prod_{j \neq \ell}^{K} (\omega_j - \omega_\ell)\right]^{-1} \left(1 - r_\ell(\boldsymbol{\omega})\right), \tag{88}$$

where

$$\begin{aligned} r_\ell(\boldsymbol{\omega}) &\triangleq e^{-i\boldsymbol{\omega}^T \boldsymbol{b}_\varepsilon} \left(\frac{1 - e^{-i\boldsymbol{\omega}^T \boldsymbol{a}_\ell}}{1 - e^{-i\omega_\ell}}\right) \frac{\omega_\ell}{\boldsymbol{\omega}^T \boldsymbol{a}_\ell} \prod_{j \neq \ell} \frac{\omega_j - \omega_\ell}{\boldsymbol{\omega}^T (\boldsymbol{a}_j - \boldsymbol{a}_\ell)} \\ &= e^{-i\boldsymbol{\omega}^T (\boldsymbol{b}_\varepsilon - \boldsymbol{\delta}_\ell)} \frac{\sin\left(\omega_\ell/2 + \boldsymbol{\delta}_\ell^T \boldsymbol{\omega}/2\right)}{\sin(\omega_\ell/2)} \frac{1}{1 + (\boldsymbol{\omega}/\omega_\ell)^T \boldsymbol{\delta}_\ell} \prod_{j \neq \ell} \frac{1}{1 + \frac{(\boldsymbol{\delta}_j - \boldsymbol{\delta}_\ell)^T \boldsymbol{\omega}}{(\omega_j - \omega_\ell)}}, \end{aligned} \tag{89}$$

where $\boldsymbol{a}_i$ denotes the ith row of $\boldsymbol{A}_\varepsilon$. Also, we have assumed $\boldsymbol{a}_i = \mathbf{1}_i + \boldsymbol{\delta}_i$ with $\|\boldsymbol{\delta}_i\|_\infty$, $i \in [K]$ being upper bounded by $\mathcal{O}(\varepsilon)$ according to the definition of $\boldsymbol{A}_\varepsilon$. It can be readily checked that $r_\ell(\boldsymbol{\omega}) = 1 + \mathcal{O}(\varepsilon)$. Also, the multiplicative factors that form $r_\ell(\boldsymbol{\omega})$ are either not dependent on $\|\boldsymbol{\omega}\|_2$ (dependence is only on the direction, and not the magnitude) or they oscillate as a function of magnitude. This implies that $r_\ell(\boldsymbol{\omega})$ which appears solely due to the difference between the two simplices does not behave differently in regions which are far from the origin or the areas in its vicinity.

Due to the above derivations, the difference function has a similar behaviour to that of $\mathscr{F}_{\Delta_K}(\boldsymbol{\omega})$ and thus we have

$$\frac{1}{(2\pi)^K} \int_{\|\boldsymbol{\omega}\|_\infty \geq \alpha} \left\|\mathscr{F}_{\Delta_K} - \mathscr{F}_{\hat{\Delta}_{K,\varepsilon}}\right\|^2 \leq \mathcal{O}\left(\frac{K\varepsilon}{\alpha}\right),$$

which completes the proof. □

With the help of Corollary C.3 and Lemmas C.6, C.7 and C.8, we can finally prove the claimed bound. Let simplex $\mathcal{S}_1$ and $\mathcal{S}_2$ to have the vertex matrices $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, respectively. Also, we have already assumed

$$\underline{\lambda} \leq \lambda_{\min}(\boldsymbol{\Theta}_i) \quad , \quad \lambda_{\max}(\boldsymbol{\Theta}_i) \leq \bar{\lambda} \quad \text{for } i = 1, 2.$$

Then, based on Lemma C.6 and for a sufficiently large $\alpha > 0$, we have

$$\begin{aligned} \int_{\|\boldsymbol{\omega}\|_\infty \geq \alpha} |\mathscr{F}_{\mathcal{S}_i}(\boldsymbol{\omega})|^2 &= \int_{\|\boldsymbol{\omega}\|_\infty \geq \alpha} \left|\mathscr{F}_{\Delta_K}\left(\boldsymbol{\Theta}_i^T \boldsymbol{\omega}\right)\right|^2 \\ &\leq \frac{1}{\det(\boldsymbol{\Theta})} \int_{\|\boldsymbol{\omega}\|_\infty \geq \bar{\lambda}\alpha} |\mathscr{F}_{\Delta_K}(\boldsymbol{\omega})|^2 \\ &\leq \frac{(2\pi)^K}{\mathrm{Vol}(\mathcal{S}_i)} \mathcal{O}\left(\frac{K}{\bar{\lambda}\alpha}\right), \end{aligned} \tag{90}$$

for $i = 1, 2$. Therefore, Lemma C.8 would consequently imply the following relation for the difference function of the two simplices:

$$\frac{1}{(2\pi)^K} \int_{\|\boldsymbol{\omega}\|_\infty \geq \alpha} |\mathscr{F}_{\mathcal{S}_1} - \mathscr{F}_{\mathcal{S}_2}|^2 \leq \mathcal{O}\left(\frac{K}{\bar{\lambda}\alpha}\right) \int_{\mathbb{R}^K} (f_{\mathcal{S}_1} - f_{\mathcal{S}_2})^2, \tag{91}$$

which facilitates the usage of our main general result in Theorem C.2, and more specifically Corollary C.3 which leads to the following relation:

$$\begin{aligned} \|(f_{\mathcal{S}_1} - f_{\mathcal{S}_2}) * G_\sigma\|_2 &\geq \|f_{\mathcal{S}_1} - f_{\mathcal{S}_2}\|_2 \sup_\alpha \left(\sqrt{1 - \mathcal{O}\left(\frac{K}{\bar{\lambda}\alpha}\right)} e^{-K(\sigma\alpha)^2}\right) \\ &\geq \|f_{\mathcal{S}_1} - f_{\mathcal{S}_2}\|_2 \, e^{-\mathcal{O}\left(\frac{K}{\mathrm{SNR}^2}\right)}, \end{aligned} \tag{92}$$

since $\alpha \geq \Omega\left(K/\bar{\lambda}\right)$ guarantees that the l.h.s. of the bound remains positive, while gives the minimum possible exponent (at least order-wise) to the exponential term. Combining the fact ℓ_2-norm $\leq \ell_1$-norm together with theorem's assumptions, we have

$$\|(f_{\mathcal{S}_1} - f_{\mathcal{S}_2}) * G_\sigma\|_2 \leq \|(f_{\mathcal{S}_1} - f_{\mathcal{S}_2}) * G_\sigma\|_1 \leq 2\varepsilon.$$

Consequently, (92) implies the following upper-bound on the ℓ_2-norm of the difference PDF function $f_{\mathcal{S}_1} - f_{\mathcal{S}_2}$ as follows:

$$\|f_{\mathcal{S}_1} - f_{\mathcal{S}_2}\|_2 \leq \varepsilon e^{\mathcal{O}\left(\frac{K}{\mathrm{SNR}^2}\right)},$$

and completes the proof. □

C.1. The case of unknown σ

In this final part, we take care of the fact that σ might not be known at first and could be estimated during the learning procedure. We show that the outputs of the algorithm $\mathscr{A}$ from Theorem 3.5, which is an estimated simplex $\mathcal{S}_{\mathscr{A}}$ and an estimated standard variation $\sigma_{\mathscr{A}}$ are both at worst $\varepsilon e^{\mathcal{O}\left(\frac{K}{\mathrm{SNR}^2}\right)}$-close to their respective true counterparts. Recall from Theorem 3.5 that the class of $\left(\underline{\theta}, \bar{\theta}\right)$-isoperimetric K-simplices which are corrupted with Gaussian white noise, i.e.,

$$\bigcup_{\sigma > 0} \mathbb{G}_{K,\sigma}$$

is PAC-learnable in the following sense: Assume $n \geq \tilde{O}\left(K^2/\varepsilon^2\right)$ i.i.d. samples from a noisy simplex $f_{\mathcal{S}} * G_\sigma$, for any $\mathcal{S} \in \mathbb{S}_K$ and $\sigma > 0$, are given. Then, there exists an algorithm $\mathscr{A}$ which outputs a noisy simplex $f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}}}$ which with high probability satisfies

$$\|f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}}} - f_{\mathcal{S}} * G_\sigma\|_{\mathrm{TV}} \leq \varepsilon. \tag{93}$$

Without loss of generality, let us assume $\sigma_{\mathcal{A}} \geq \sigma$. Therefore, we have already proved that

$$\|(f_{\mathcal{S}} - f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}} - \sigma}) * G_\sigma\|_{\mathrm{TV}} \leq \varepsilon,$$

which holds due to the fact that Gaussian distributions are invariant w.r.t. convolution.

Using the same techniques introduced in the prior parts of this section, we already know that both $f_{\mathcal{S}}$ and $f_{\mathcal{S}_{\mathscr{A}}}$ have the desired "low frequency" property in the Fourier domain. Also, it is easy to see that $f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}} - \sigma}$ should also possess the same attribute since convolution with a Gaussian further weakens high frequency details of the function. As a result, we can apply the result of Theorem 3.8 and deduce the following bound:

$$\|f_{\mathcal{S}} - f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}} - \sigma}\|_2 \leq \varepsilon e^{\Omega\left(\frac{K}{\mathrm{SNR}^2}\right)}. \tag{94}$$

Using the parseval's theorem, we have

$$\begin{aligned}
\|f_{\mathcal{S}} - f_{\mathcal{S}_{\mathscr{A}}} * G_{\sigma_{\mathscr{A}} - \sigma}\|_2^2 &= \int_{\mathbb{R}^K} \left|\mathscr{F}_{\mathcal{S}}(\boldsymbol{\omega}) - \mathscr{F}_{\mathcal{S}_{\mathscr{A}}}(\boldsymbol{\omega}) \mathcal{F}\left\{G_{\sigma_{\mathscr{A}} - \sigma}\right\}(\boldsymbol{\omega})\right|^2 \mathrm{d}\boldsymbol{\omega} \\
&= \int_{\mathbb{R}^K} \left|\mathscr{F}_{\mathcal{S}}(\boldsymbol{\omega})\right|^2 \left|1 - \frac{\mathscr{F}_{\mathcal{S}_{\mathscr{A}}}(\boldsymbol{\omega})}{\mathscr{F}_{\mathcal{S}}(\boldsymbol{\omega})} \mathcal{F}\left\{G_{\sigma_{\mathscr{A}} - \sigma}\right\}(\boldsymbol{\omega})\right|^2 \mathrm{d}\boldsymbol{\omega} \\
&\leq \mathcal{O}\left(\varepsilon^2\right).
\end{aligned} \tag{95}$$

The above chain of inequalities imply that

$$\left|\frac{\mathscr{F}_{\mathcal{S}_{\mathscr{A}}}(\boldsymbol{\omega})}{\mathscr{F}_{\mathcal{S}}(\boldsymbol{\omega})} \mathcal{F}\left\{G_{\sigma_{\mathscr{A}} - \sigma}\right\}(\boldsymbol{\omega})\right| = \left|\frac{\mathscr{F}_{\mathcal{S}_{\mathscr{A}}}(\boldsymbol{\omega})}{\mathscr{F}_{\mathcal{S}}(\boldsymbol{\omega})}\right| \left|\mathcal{F}\left\{G_{\sigma_{\mathscr{A}} - \sigma}\right\}(\boldsymbol{\omega})\right| = 1 + \Theta(\varepsilon), \quad \forall \boldsymbol{\omega} \in \mathbb{R}^K. \tag{96}$$

According to the analysis derived in Section C, we already know that

$$\left|\frac{\mathscr{F}_{\mathcal{S}_{\mathscr{A}}}(\boldsymbol{\omega})}{\mathscr{F}_{\mathcal{S}}(\boldsymbol{\omega})}\right| = \frac{\mathrm{poly}_K(\boldsymbol{\omega})}{\mathrm{poly}_K(\boldsymbol{\omega})} \quad , \quad \left|\mathcal{F}\left\{G_{\sigma_{\mathscr{A}} - \sigma}\right\}(\boldsymbol{\omega})\right| = e^{-\omega^2(\sigma_{\mathscr{A}} - \sigma)/2},$$

where the former equality is the direct result of calculating the Fourier transform of the general simplices which we have already carried out in the previous parts. Since (96) holds for all $\boldsymbol{\omega} \in \mathbb{R}^K$, and due to the fact that a polynomial and an exponential function cannot cancel each other (in a multiplicative way) everywhere, one can deduce that

$$\left| \frac{\mathscr{F}_{\mathcal{S}_{\mathscr{A}}}(\boldsymbol{\omega})}{\mathscr{F}_{\mathcal{S}}(\boldsymbol{\omega})} \right| = 1 + \Theta(\varepsilon),$$
$$\left| \mathcal{F}\left\{G_{\sigma_{\mathscr{A}}-\sigma}\right\}(\boldsymbol{\omega}) \right| = 1 + \Theta(\varepsilon),$$

which means both $\|f_{\mathcal{S}} - f_{\mathcal{S}_{\mathscr{A}}}\|_2$ and $||$ are both at most $\varepsilon e^{\mathcal{O}\left(\frac{K}{\mathrm{SNR}^2}\right)}$. This proves our claim.