Supplementary Material for the paper
“Learning single-index models via harmonic decomposition"

Contents

A Additional discussions and details from the main text

A.1 Discussion on learning spherical SIMs

Likelihood ratio. In our Definition 1, we consider spherical SIMs such that $\nu_d \ll \nu_{d,0}$ and the associated Radon-Nikodym derivative satisfies $\|\frac{\mathrm{d}\nu_d}{\mathrm{d}\nu_{d,0}}\|_{L^2(\nu_{d,0})} < \infty$. We can expand the likelihood ratio in $L^2(\nu_{d,0})$ into the Gegenbauer basis:

$$\frac{\mathrm{d}\nu_d}{\mathrm{d}\nu_{d,0}}(y,r,z) \overset{L^2(\nu_{d,0})}{=} 1 + \sum_{\ell=1}^{\infty} \xi_{d,\ell}(y,r) Q_\ell(z)\,, \tag{18}$$

where

$$\xi_{d,\ell}(Y,R) := \mathbb{E}_{Z\sim\tau_{d,1}}\left[\frac{\mathrm{d}\nu_d}{\mathrm{d}\nu_{d,0}}(Y,R,Z)Q_\ell(Z)\right] = \mathbb{E}_{\nu_d}\left[Q_\ell(Z)|Y,R\right]\,.$$

The mutual χ^2-mutual information divergence of $((Y,R),Z)$ is given by

$$I_{\chi^2}[\nu_d] = \mathbb{E}_{\nu_{d,0}}\left[\left(\frac{\mathrm{d}\nu_d}{\mathrm{d}\nu_{d,0}}\right)^2\right] - 1 = \sum_{\ell=1}^{\infty} \|\xi_{d,\ell}\|_{L^2}\,.$$

Link function ν_d unknown. When ν_d is unknown, similar to [15], one can still hope to learn the planted direction $\boldsymbol{w}_*$, as long as it is possible to approximate the non-linearity $\mathcal{T}_\ell(y,r)$ using random linear combinations of the first few orthogonal functions in a basis of $L^2(\nu_{d,Y,R})$. Similar to [15, Assumption 4.1] this requires assuming that the expansion of $\mathcal{T}_\ell$ has non-vanishing mass on these first few basis functions of $L^2(\nu_{d,Y,R})$. Intuitively, this amounts to a 'smoothness' condition on the link function ν_d. We leave this as a direction for future work.

Weak to strong recovery. In our framework, it is only meaningful to restrict ourselves to the sequence $\{\nu_d\}_{d\geq 1}$ such that $I_{\chi^2}[\nu_d]$ is non-vanishing and there exists a component $\ell \geq 1$ (independent of d) such that $\|\xi_{d,\ell}\|_{L^2} > c > 0$. For example, in Gaussian SIMs with the generative exponent $\mathsf{k}_\star$, such an $\ell = \mathsf{k}_\star$ always (both with or without using the norm). Under this mild assumption, we can carry out the online SGD algorithm similar to the final phase algorithms of [2, 14, 49] but now on the frequency Q_ℓ. As we have non-vanishing signal $\|\xi_{d,\ell}\| = \Omega_d(1)$, we can achieve strong recovery $|\langle \hat{\boldsymbol{w}}, \boldsymbol{w}_*\rangle| \geq 1-\varepsilon$ using $O(d/\varepsilon)$ samples, and so $O(d^2/\varepsilon)$ runtime hiding constants in ℓ.

Spherical SIMs with sample-runtime trade-off. Below we show how to construct examples of spherical single-index models with $\mathsf{l}_{\mathsf{m},\star} > \mathsf{l}_{\mathsf{T},\star}$. We construct ν_d such that it is a mixture of two spherical SIMs. Consider $\nu_d^{(1)}$ and $\nu_d^{(2)}$ associated to two Gaussian SIMs with generative exponents k_1 and k_2, where we marginalized over the norm (that is, $R=1$). In particular, as shown in [15, Theorem 5.1], we can choose the Gaussian SIMs to be $y^{(j)} = \sigma_j(R\cdot Z) + \tau\mathsf{N}(0,1)$, with $\|\sigma_j\|_\infty < \infty$ and τ sufficiently large such that $C^{-1} \leq \nu_d^{(1)}(y)/\nu_d^{(2)}(y) \leq C$.

Assume that $(Y,Z)\sim\nu_d$ (recall $R=1$ here, and we remove it for clarity) is drawn with probability $d^{-\alpha}$ from $\nu_d^{(1)}$ and with probability $1-d^{-\alpha}$ from $\nu_d^{(2)}$, where $\alpha > 0$ is a constant chosen later. In this model, we have

$$\xi_{d,\ell}(Y) = \mathbb{E}_{\nu_d}[Q_\ell(Z)|Y] = d^{-\alpha} C_\alpha^{(1)}(Y)\xi_{d,\ell}^{(1)}(Y) + C_\alpha^{(2)}(Y)\xi_{d,\ell}^{(2)}(Y),$$

where

$$\xi_{d,\ell}^{(1)}(Y) = \mathbb{E}_{\nu_d^{(1)}}[Q_\ell(Z)|Y], \qquad \xi_{d,\ell}^{(2)}(Y) = \mathbb{E}_{\nu_d^{(2)}}[Q_\ell(Z)|Y],$$

and

$$C_\alpha^{(1)}(Y) = \frac{1}{d^{-\alpha} + (d^\alpha - 1)\nu_d^{(2)}(Y)/\nu_d^{(1)}(Y)}, \qquad C_\alpha^{(2)}(Y) = \frac{1}{d^{-\alpha}\nu_d^{(1)}(Y)/\nu_d^{(2)}(Y) + 1 - d^{-\alpha}}.$$

From our choice of $\nu_d^{(j)}$, there exists a constant C, such that for $d \geq C$, we have $C^{-1} \leq C_\alpha^{(1)}(y), C_\alpha^{(2)}(y) \leq C$ for all $y \in \mathbb{R}$. We deduce that there exist a constant $\tilde{C}$ sufficiently large but independent of d such that

$$\begin{aligned}
&\|\xi_{d,\ell}\|_{L^2}^2 \leq \tilde{C}\max(d^{-2\alpha}\|\xi_{d,\ell}^{(1)}\|_{L^2}^2, \|\xi_{d,\ell}^{(2)}\|_{L^2}^2),\\
&\|\xi_{d,\ell}\|_{L^2}^2 \geq \tilde{C}^{-1}\max\Big(d^{-2\alpha}\|\xi_{d,\ell}^{(1)}\|_{L^2}^2 - \tilde{C}^2\|\xi_{d,\ell}^{(2)}\|_{L^2}^2, \|\xi_{d,\ell}^{(2)}\|_{L^2}^2 - \tilde{C}^2 d^{-2\alpha}\|\xi_{d,\ell}^{(1)}\|_{L^2}^2\Big),
\end{aligned} \tag{19}$$

where the L^2-norms are with respect to the associated nulls $\nu_{d,0}$, $\nu_{d,0}^{(1)}$, and $\nu_{d,0}^{(2)}$.

Consider $\mathsf{k}_\star$ a multiple of 10 for simplicity, and set $k_2 = \mathsf{k}_\star$, $k_1 = 2\mathsf{k}_\star/5$, and $\alpha = \mathsf{k}_\star/5$. Using Lemma 2, we can bound the contributions from $\nu_d^{(1)}$ and $\nu_d^{(2)}$:

- Consider the contributions of $\nu_d^{(1)}$ to the sample and runtime complexity:
  - Sample complexity:

$$\begin{aligned}
\ell \leq k_1,\ \ell \equiv k_1[2]: &\qquad d^{2\alpha}\frac{\sqrt{n_{d,\ell}}}{\|\xi_{d,\ell}^{(1)}\|_{L^2}^2} \asymp d^{2k_1 - \ell/2} = d^{4\mathsf{k}_\star/5 - \ell/2},\\
\ell \leq k_1,\ \ell \not\equiv k_1[2]: &\qquad d^{2\alpha}\frac{\sqrt{n_{d,\ell}}}{\|\xi_{d,\ell}^{(1)}\|_{L^2}^2} \asymp d^{2k_1 - \ell/2 + 1} = d^{4\mathsf{k}_\star/5 - \ell/2 + 1},\\
\ell > k_1: &\qquad d^{2\alpha}\frac{\sqrt{n_{d,\ell}}}{\|\xi_{d,\ell}^{(1)}\|_{L^2}^2} \gtrsim d^{k_1 + \ell/2} = d^{2\mathsf{k}_\star/5 + \ell/2}.
\end{aligned} \tag{20}$$

    Thus the optimal sample complexity is achieved at degree $\ell = k_1 = 2\mathsf{k}_\star/5$ with $d^{3\mathsf{k}_\star/5}$ lower bound.
  - Runtime complexity:

$$\begin{aligned}
\ell \leq k_1,\ \ell \equiv k_1[2]: &\qquad d^{2\alpha}\frac{n_{d,\ell}}{\|\xi_{d,\ell}^{(1)}\|_{L^2}^2} \asymp d^{2k_1} = d^{4\mathsf{k}_\star/5},\\
\ell \leq k_1,\ \ell \not\equiv k_1[2]: &\qquad d^{2\alpha}\frac{n_{d,\ell}}{\|\xi_{d,\ell}^{(1)}\|_{L^2}^2} \asymp d^{2k_1 + 1} = d^{4\mathsf{k}_\star/5 + 1},\\
\ell > k_1: &\qquad d^{2\alpha}\frac{n_{d,\ell}}{\|\xi_{d,\ell}^{(1)}\|_{L^2}^2} \gtrsim d^{k_1 + \ell} = d^{2\mathsf{k}_\star/5 + \ell}.
\end{aligned} \tag{21}$$

    Thus the optimal runtime complexity is achieved at degrees $\ell \leq k_1 = 2\mathsf{k}_\star/5$, $\ell \equiv k_1[2]$ with $d^{4\mathsf{k}_\star/5}$ lower bound.
- Consider the contributions of $\nu_d^{(2)}$ to the sample and runtime complexity:
  - Sample complexity:

$$\begin{aligned}
\ell \leq k_2,\ \ell \equiv k_2[2]: &\qquad \frac{\sqrt{n_{d,\ell}}}{\|\xi_{d,\ell}^{(2)}\|_{L^2}^2} \asymp d^{k_2 - \ell/2} = d^{\mathsf{k}_\star - \ell/2},\\
\ell \leq k_2,\ \ell \not\equiv k_2[2]: &\qquad \frac{\sqrt{n_{d,\ell}}}{\|\xi_{d,\ell}^{(2)}\|_{L^2}^2} \asymp d^{k_1 - \ell/2 + 1} = d^{\mathsf{k}_\star - \ell/2 + 1},\\
\ell > k_2: &\qquad \frac{\sqrt{n_{d,\ell}}}{\|\xi_{d,\ell}^{(2)}\|_{L^2}^2} \gtrsim d^{\ell/2}.
\end{aligned} \tag{22}$$

Thus the optimal sample complexity is achieved at degree $\ell = k_2 = \mathsf{k}_\star$ with $d^{\mathsf{k}_\star/2}$ lower bound.

– Runtime complexity:

$$\begin{aligned} \ell \leq k_2,\ \ell \equiv k_2[2] : \qquad & \frac{n_{d,\ell}}{\|\xi^{(2)}_{d,\ell}\|^2_{L^2}} \asymp d^{k_2} = d^{\mathsf{k}_\star}, \\ \ell \leq k_2,\ \ell \not\equiv k_2[2] : \qquad & \frac{n_{d,\ell}}{\|\xi^{(2)}_{d,\ell}\|^2_{L^2}} \asymp d^{k_2+1} = d^{\mathsf{k}_\star+1}, \\ \ell > k_2 : \qquad & \frac{n_{d,\ell}}{\|\xi^{(2)}_{d,\ell}\|^2_{L^2}} \gtrsim d^{\ell}. \end{aligned} \tag{23}$$

Thus the optimal runtime complexity is achieved at degrees $\ell \leq k_2 = \mathsf{k}_\star$, $\ell \equiv k_2[2]$ with $d^{\mathsf{k}_\star}$ lower bound.

From the bounds (19), the sample or runtime complexity for each ℓ is the minimum of the two contributions associated to $\nu_d^{(1)}$ and $\nu_d^{(2)}$. We deduce that in this model:

- Optimal sample complexity is achieved at degree $\mathsf{l}_{\mathsf{m},\star} = \mathsf{k}_\star$, with a matching algorithm that succeeds with $\mathsf{m} = \Theta_d(d^{\mathsf{k}_\star/2})$ samples and $\mathsf{T} = \Theta(d^{\mathsf{k}_\star})$ runtime (thanks to contribution $\nu_d^{(2)}$).
- Optimal runtime is achieved at degrees $\mathsf{l}_{\mathsf{T},\star} = \ell$ with $\ell \leq 2\mathsf{k}_\star/5$ with $\ell \equiv 2\mathsf{k}_\star/5[2]$. For example, choosing $\mathsf{l}_{\mathsf{T},\star} = 2\mathsf{k}_\star/5$, we have a matching algorithm that succeeds with $\mathsf{T} = \Theta_d(d^{4\mathsf{k}_\star/5})$ runtime and $\mathsf{m} = \Theta_d(d^{3\mathsf{k}_\star/5})$ samples (thanks to contribution $\nu_d^{(1)}$).

We conjecture that for this distribution ν_d, no algorithm exists that achieves both optimal sample complexity $\mathsf{m} = \Theta_d(d^{\mathsf{k}_\star/2})$ and optimal runtime complexity $\mathsf{T} = \Theta_d(d^{4\mathsf{k}_\star/5})$. Further note, that by choosing intermediary degrees ℓ, one can trade-off sample and runtime complexity.

A.2 Discussion on learning Gaussian SIMs

Consider learning a Gaussian single-index model (1) with link function $\rho \in \mathcal{P}(\mathbb{R}^2)$ and generative exponent $\mathsf{k}_\star := \mathsf{k}_\star(\rho)$. Recall the definition from [15]:

$$\mathsf{k}_\star(\rho) = \arg\min\{k \geq 1 \ :\ \|\zeta_k\|_{L^2(\rho)} > 0 \quad \text{where} \quad \zeta_k(Y) := \mathbb{E}_\rho[\mathrm{He}_k(G)|Y]\}. \tag{24}$$

In light of our results on learning general spherical SIMs via harmonic decomposition (Section 3), we revisit the three algorithms for learning Gaussian SIMs mentioned in the introduction [2, 14, 15], and reinterpret their behavior through the lens of spherical harmonics.

Throughout, let $\mathcal{T}_* : \mathbb{R} \to \mathbb{R}$ be a transformation of the label satisfying:

$$\|\mathcal{T}_*\|_{L^2} = 1, \qquad \|\mathcal{T}_*\|_\infty \leq C, \qquad \Gamma_{\mathsf{k}_\star} := \mathbb{E}_\rho[\mathcal{T}_*(Y)\mathrm{He}_{\mathsf{k}_\star}(G)] \geq \frac{1}{C}\|\zeta_{\mathsf{k}_\star}\|_{L^2}.$$

Such transformations always exist (see [15, Lemma F.2]). Informally, one can construct $\mathcal{T}_*$ by truncating $\zeta_{\mathsf{k}_\star}/\|\zeta_{\mathsf{k}_\star}\|_{L^2}$ (with truncation at large enough value as to approximately preserve the correlation with $\mathrm{He}_{\mathsf{k}_\star}$).

Online SGD with Hermite neuron. In a seminal paper, Ben Arous et al. [2] studied online SGD on a non-convex loss over $\boldsymbol{w} \in \mathbb{S}^{d-1}$, with planted signal $\boldsymbol{w}_*$ and a $\mathsf{k}_\star$-order saddle at the equator $\langle \boldsymbol{w}, \boldsymbol{w}_* \rangle = 0$. Adapting their results to the task of learning Gaussian SIMs, their algorithm performs online SGD on the population loss

$$\min_{\boldsymbol{w} \in \mathbb{S}^{d-1}} \mathcal{L}(\boldsymbol{w}) := \mathbb{E}_{(y,\boldsymbol{x})\sim\mathbb{P}_{\boldsymbol{w}_*}}\left[\Big(\mathcal{T}_*(y) - \mathrm{He}_{\mathsf{k}_\star}(\langle \boldsymbol{w}, \boldsymbol{x}\rangle)\Big)^2\right], \tag{HeSGD}$$

and succeeds with suboptimal $\mathsf{m} = \widetilde{\Theta}_d(d^{\mathsf{k}_\star-1})$ samples and $\mathsf{T} = \widetilde{\Theta}_d(d^{\mathsf{k}_\star})$.

We now reinterpret this result through the lens of harmonic decomposition. The following informal observations aim to build intuition rather than make formal statements:

Observation 1 (Informal, harmonic structure of the loss)**.** *When $|\langle \boldsymbol{w}, \boldsymbol{w}_* \rangle| \gtrsim d^{-1/2}$, the loss landscape $\mathcal{L}(\boldsymbol{w})$ is dominated by degree-$\mathsf{k}_\star$ spherical harmonics. The resulting SGD dynamics behaves similarly to online SGD using $Q_{\mathsf{k}_\star}(\langle \boldsymbol{w}, \boldsymbol{z} \rangle)$ in place of $\mathrm{He}_{\mathsf{k}_\star}(\langle \boldsymbol{w}, \boldsymbol{x} \rangle)$.*

This suggests that algorithm (HeSGD) effectively restricts itself to the degree-$\mathsf{k}_\star$ subspace $V_{d,\mathsf{k}_\star}$. As a consequence, we expect its performance to be constrained by the query complexity lower bound $\Omega_d(d^{\mathsf{k}_\star})$, and its behavior to be similar to the degree-$\mathsf{k}_\star$ online SGD estimator (SGD-Alg) in Section 3. We further provide an alternative perspective that highlights the role of the norm $\|\boldsymbol{x}\|_2$ of the input data in learning Gaussian SIMs:

Observation 2 (Informal, norm-invariance of dynamics)**.** *The SGD dynamics of [2] remains essentially the same if the input $\boldsymbol{x}$ is replaced by $\tilde{r} \cdot \boldsymbol{x}/\|\boldsymbol{x}\|_2$, where $\tilde{r} \sim \chi_d$ is sampled independently.*

This indicates that the algorithm does not exploit the radial component $\|\boldsymbol{x}\|_2$, and effectively operates on the normalized direction $\boldsymbol{z} = \boldsymbol{x}/\|\boldsymbol{x}\|_2$. From our theory (Section 4), any such estimator incurs a query complexity of $\Omega_d(d^{\mathsf{k}_\star})$. In this sense, algorithm (HeSGD) is runtime-optimal among methods that ignore radial information.

Landscape smoothing. To address the suboptimality of (HeSGD), Damian et al. [14] introduced a landscape smoothing operator that averages the loss on a sphere around each parameter $\boldsymbol{w} \in \mathbb{S}^{d-1}$:

$$\min_{\boldsymbol{w} \in \mathbb{S}^{d-1}} \mathbb{E}_{\boldsymbol{u} \sim \mathrm{Unif}(\mathbb{S}^{d-1})} \left[\mathcal{L}\left(\frac{\boldsymbol{w} + \lambda \boldsymbol{u}}{\|\boldsymbol{w} + \lambda \boldsymbol{u}\|_2} \right) \right]. \tag{SmLD}$$

This modification achieves near-optimal complexities: $\mathsf{m} = \widetilde{\Theta}_d(d^{\mathsf{k}_\star/2})$ and $\mathsf{T} = \widetilde{\Theta}_d(d^{\mathsf{k}_\star/2+1})$.

Observation 3 (Informal, low-pass filtering effect)**.** *Landscape smoothing suppresses high-frequency components of the loss, effectively amplifying lower-degree harmonics. The initial phase of SGD dynamics behaves essentially like optimization over $Q_1(\langle \boldsymbol{w}, \boldsymbol{z} \rangle)$ and $Q_2(\langle \boldsymbol{w}, \boldsymbol{z} \rangle)$.*

Thus, smoothing can be interpreted as projecting the dynamics onto low-degree harmonic components—specifically, the statistics of the spectral algorithm (SP-Alg) associated to spherical harmonics of degree $\ell \in \{1, 2\}$. In this sense, the first phase of the dynamics on (SmLD) essentially corresponds to running SGD on the optimal spectral estimator (SP-Alg).

Partial trace estimator. In a subsequent work, Damian et al. [15] proposed an estimator based on the partial trace of a Hermite tensor, inspired by techniques from tensor PCA. Their construction begins with the empirical Hermite tensor:

$$\hat{\boldsymbol{T}} := \frac{1}{m} \sum_{i \in [m]} \mathcal{T}_\star(y_i) \mathbf{He}_{\mathsf{k}_\star}(\boldsymbol{x}_i) \in (\mathbb{R}^d)^{\otimes k}, \tag{25}$$

where $\mathbf{He}_{\mathsf{k}_\star}(\boldsymbol{x})$ denotes the rank-$\mathsf{k}_\star$ multivariate Hermite tensor, and $\mathcal{T}_\star$ is the transformation defined earlier. The expectation $\mathbb{E}[\hat{\boldsymbol{T}}]$ is proportional to $\boldsymbol{w}_*^{\otimes \mathsf{k}_\star}$. To extract this principal component, they compute a *partial trace* of the empirical tensor by contracting $\hat{\boldsymbol{T}}$ with identity tensors. This results in an empirical vector or matrix, depending on whether $\mathsf{k}_\star$ is odd or even. The resulting estimator is

$$\begin{aligned} &\mathsf{k}_\star \text{ odd:} \quad \hat{\boldsymbol{w}}_0 = \frac{\hat{\boldsymbol{v}}}{\|\hat{\boldsymbol{v}}\|_2}, && \hat{\boldsymbol{v}} := \frac{1}{m} \sum_{i \in [m]} \mathcal{T}_\star(y_i) P_{\mathsf{k}_\star}(\|\boldsymbol{x}_i\|_2) \boldsymbol{x}_i, \\ &\mathsf{k}_\star \text{ even:} \quad \hat{\boldsymbol{w}} = \arg\max_{\boldsymbol{w} \in \mathbb{S}^{d-1}} \boldsymbol{w}^\mathsf{T} \hat{\boldsymbol{M}} \boldsymbol{w}, && \hat{\boldsymbol{M}} := \frac{1}{m} \sum_{i=1}^{m} \mathcal{T}_\star(y_i) P_{\mathsf{k}_\star}(\|\boldsymbol{x}_i\|_2) \left[\boldsymbol{x}_i \boldsymbol{x}_i^\mathsf{T} - \mathbf{I}_d \right], \end{aligned} \tag{PrTR}$$

where $P_{\mathsf{k}_\star}$ is a univariate polynomial derived from the contraction of the Hermite tensor. In the odd case, a second refinement phase (see Section D) is used to boost $\hat{\boldsymbol{w}}_0$ from $d^{-1/4}$ to constant correlation with $\boldsymbol{w}_*$. This estimator achieves the optimal sample complexity $\Theta_d(d^{\mathsf{k}_\star/2})$ and optimal runtime $\Theta_d(d^{\mathsf{k}_\star/2+1})$, matching the lower bounds for learning Gaussian SIMs (see Eq. (3)).

Importantly, estimator (PrTR) corresponds precisely to the general spectral estimator (SP-Alg), associated to the optimal degree $\ell \in \{1, 2\}$ harmonic subspaces, using

$$\mathcal{T}_\ell(y, r) = \mathcal{T}_\star(y) P_{\mathsf{k}_\star}(r) \quad \text{with } \ell = 1 \text{ if } \mathsf{k}_\star \text{ is odd}, \quad \text{and } \ell = 2 \text{ if } \mathsf{k}_\star \text{ is even}.$$

Observation 4 (Informal, lower-frequency projection)**.** *Partial trace effectively projects the high-degree Hermite tensor onto lower-degree spherical harmonic subspaces ($\ell = 1$ or 2 for partial trace over all but 1 or 2 coordinates).*

Although our estimator recovers (PrTR) in the Gaussian case, we emphasize that its derivation is very different (including constructing the non-linearity $\mathcal{T}_\ell(y, r)$, see Appendix G). We construct it directly from rotational invariance and harmonic decomposition, without relying on prior knowledge of tensor PCA, contractions of Hermite tensors, or Gaussian-specific identities. We believe this alternative, transparent derivation of (PrTR) highlights the advantages of the harmonic perspective when learning single-index models.

Remark A.1. Both landscape smoothing [7] and partial trace estimators [25] originate in the tensor PCA literature, where a similar gap between 'local' and optimal algorithms arises—with $d^{\mathsf{k}_\star/2}$ versus $d^{\mathsf{k}_\star - 1}$ gap in signal strength. It is intriguing to connect the phenomena observed in SIMs (Observations 3 and 4) to analogous behaviors in tensor PCA. We leave this direction for future work.

Next, we provide quick computations to justify the above observations.

Observation 1: harmonic decomposition of the loss. First, note that

$$\mathcal{L}(\boldsymbol{w}) = 2 - 2\beta_{\mathsf{k}_\star} \mathbb{E}_{\boldsymbol{x}}[\mathrm{He}_{\mathsf{k}_\star}(\langle \boldsymbol{w}_*, \boldsymbol{x}\rangle)\mathrm{He}_{\mathsf{k}_\star}(\langle \boldsymbol{w}, \boldsymbol{x}\rangle)] = 2 - 2\Gamma_{\mathsf{k}_\star}\langle \boldsymbol{w}_*, \boldsymbol{w}\rangle^k,$$

and it is enough to consider the correlation loss. Let's decompose the landscape into contributions from the different harmonic subspaces: using the Hermite to Gegenbauer polynomial decomposition in Eq. (17), we get

$$\begin{aligned} \mathbb{E}_{\boldsymbol{x}}[\mathrm{He}_{\mathsf{k}_\star}(\langle \boldsymbol{w}_*, \boldsymbol{x}\rangle)\mathrm{He}_{\mathsf{k}_\star}(\langle \boldsymbol{w}, \boldsymbol{x}\rangle)] &= \sum_{\substack{\ell \leq k \\ \ell \equiv \mathsf{k}_\star[2]}} \mathbb{E}[\beta_{\mathsf{k}_\star,\ell}(r)^2]\mathbb{E}[Q_\ell(\langle \boldsymbol{w}_*, \boldsymbol{z}\rangle)Q_\ell(\langle \boldsymbol{w}, \boldsymbol{z}\rangle)] \\ &= \sum_{\substack{\ell \leq k \\ \ell \equiv k[2]}} \frac{\|\beta_{\mathsf{k}_\star,\ell}\|_{L^2}^2}{\sqrt{n_{d,\ell}}} Q_\ell(\langle \boldsymbol{w}, \boldsymbol{w}_*\rangle), \end{aligned} \tag{26}$$

where $\|\beta_{\mathsf{k}_\star,\ell}\|_{L^2}^2/\sqrt{n_{d,\ell}} = \Theta_d(d^{-\mathsf{k}_\star/2})$ (see Appendix B). For $|\langle \boldsymbol{w}_*, \boldsymbol{w}\rangle| \geq C_{\mathsf{k}_\star} d^{-1/2}$, the leading contribution in the loss (and its gradient) is $\ell = \mathsf{k}_\star$ (recall that the leading term in $Q_\ell(\langle \boldsymbol{w}, \boldsymbol{w}_*\rangle)$ is $\Theta_d(d^{\ell/2})\langle \boldsymbol{w}, \boldsymbol{w}_*\rangle^\ell$). Informally, this implies that we could have replaced $\mathrm{He}_{\mathsf{k}_\star}(\langle \boldsymbol{w}_*, \boldsymbol{x}\rangle)$ by $Q_{\mathsf{k}_\star}(\langle \boldsymbol{w}_*, \boldsymbol{z}\rangle)$ in the above loss.

Observation 2: dynamics with independent norm. Let's consider the loss (26) when we have independent norms between the input and the signal:

$$\mathbb{E}[\mathrm{He}_{\mathsf{k}_\star}(r \cdot \langle \boldsymbol{w}_*, \boldsymbol{z}\rangle)\mathrm{He}_{\mathsf{k}_\star}(\tilde{r} \cdot \langle \boldsymbol{w}, \boldsymbol{z}\rangle)] = \sum_{\substack{\ell \leq k \\ \ell \equiv \mathsf{k}_\star[2]}} \frac{\mathbb{E}[\beta_{\mathsf{k}_\star,\ell}(r)]^2}{\sqrt{n_{d,\ell}}} Q_\ell(\langle \boldsymbol{w}, \boldsymbol{w}_*\rangle), \tag{27}$$

where $\mathbb{E}[\beta_{\mathsf{k}_\star,\ell}(r)]^2/\sqrt{n_{d,\ell}} = \Theta_d(d^{-\mathsf{k}_\star+\ell/2})$ (see Appendix B). In particular, the leading term $\ell = \mathsf{k}_\star$ remain the same between Eq. (27) and Eq. (26). Following the proof in [2] (see Section E), the dynamics with same hyperparameters behaves similarly between the two losses (27) and (26).

Observation 3: low-pass filtering of landscape smoothing. Again, it is enough to directly consider the correlation term. Let's decompose

$$\mathbb{E}_{\boldsymbol{u}\sim\tau_d}\mathbb{E}_{\boldsymbol{x}}\left[\mathrm{He}_{\mathsf{k}_\star}(\langle \boldsymbol{w}_*, \boldsymbol{x}\rangle)\mathrm{He}_{\mathsf{k}_\star}\left(\frac{\boldsymbol{w} + \lambda\boldsymbol{u}}{\|\boldsymbol{w} + \lambda\boldsymbol{u}\|_2} \cdot \boldsymbol{x}\right)\right] = \sum_{\substack{\ell \leq k \\ \ell \equiv \mathsf{k}_\star[2]}} m_\ell(\lambda)\frac{\|\beta_{\mathsf{k}_\star,\ell}\|_{L^2}^2}{\sqrt{n_{d,\ell}}} Q_\ell(\langle \boldsymbol{w}, \boldsymbol{w}_*\rangle),$$

where each frequency (26) in the original loss $\mathcal{L}(\boldsymbol{w})$ is now reweighted by

$$m_\ell(\lambda) = \frac{1}{\sqrt{n_{d,\ell}}}\mathbb{E}_{\boldsymbol{u}}\left[Q_\ell\left(\frac{\boldsymbol{w} + \lambda\boldsymbol{u}}{\|\boldsymbol{w} + \lambda\boldsymbol{u}\|_2} \cdot \boldsymbol{w}\right)\right] = \frac{1}{\sqrt{n_{d,\ell}}}\mathbb{E}_{Z\sim\tau_{d,1}}\left[Q_\ell\left(\frac{1 + \lambda Z}{\sqrt{1 + 2\lambda Z + \lambda^2}}\right)\right].$$

When $\lambda = 0$, we indeed have $m_\ell(0) = Q_\ell(1)/\sqrt{n_{d,\ell}} = 1$. For $\lambda \gg 1$, we have $m_\ell(\lambda) \asymp 1/\lambda^\ell$, and as long as $|\langle \boldsymbol{w}, \boldsymbol{w}_*\rangle| \ll \lambda^{-1}$, the loss (and its gradient) are dominated by frequencies $\ell \in \{1, 2\}$.

A.3 Correlation queries and the information exponent

In the main text, we focused on the *generative exponent* introduced by [15]: this notion tightly capture the optimal complexity of learning Gaussian single-index models among Statistical Query and Low-Degree Polynomial algorithms. An earlier notion—the *information exponent*—was proposed in [21, 2]. Specifically, for scalar labels $\mathcal{Y} \subseteq \mathbb{R}$, the *information exponent (IE)* of ρ is defined by

$$\mathsf{k}_\mathsf{I}(\rho) = \arg\min\{k \geq 1 \ : \ \mathbb{E}_\rho[Y \mathrm{He}_k(G)] \neq 0\}. \tag{28}$$

This exponent captures the complexity of learning with so-called *correlation statistical query* (CSQ) algorithms, which only access labels through correlation statistics $y\phi(\boldsymbol{x})$. In other words, using the terminology introduced in Appendix C.1, it captures the complexity of $\mathcal{Q}$-restricted SQ algorithms, with

$$\mathcal{Q} = \mathcal{Q}_{\mathsf{CSQ}} := \{\phi(y, \boldsymbol{x}) = y\tilde{\phi}(x) : \ \tilde{\phi} \text{ measurable function}\}.$$

We denote $\mathsf{CSQ}(q, \tau) := \mathcal{Q}_{\mathsf{CSQ}}\text{-}\mathsf{SQ}(q, \tau)$ this restricted class of SQ algorithms.

For CSQ algorithm, Damian et al. [14] showed lower bounds within the $\mathcal{Q}$-SQ framework of

$$\mathsf{m} = \Theta_d(d^{\mathsf{k}_\mathsf{I}(\rho)/2}), \qquad \mathsf{T} = \Theta_d(d^{\mathsf{k}_\mathsf{I}(\rho)/2+1}). \tag{29}$$

Note however, that only the generative exponent reflects the fundamental hardness of the learning task: indeed, we always have $\mathsf{k}_\star(\rho) \leq \mathsf{k}_\mathsf{I}(\rho)$, with $\mathsf{k}_\star(\rho)$ always one or two for all y polynomial function of $\boldsymbol{x}$ (while $\mathsf{k}_\mathsf{I}(\rho) = k$ if $y = \mathrm{He}_k(G)$). In the case $\mathsf{k}_\star(\rho) < \mathsf{k}_\mathsf{I}(\rho)$, the complexity predicted by the information exponent can be improved upon by using non-correlation queries, such as using a non-correlation loss [15] or by reusing samples [18]. Nonetheless, IE remains relevant in several natural settings, such as online stochastic gradient descent on the squared or cross-entropy loss.

Below, we discuss how to recover this information exponent from our harmonic framework when considering $\mathcal{Q}_{\mathsf{CSQ}}$-SQ algorithms. Introduce the CSQ query complexity

$$\mathsf{Q}_\star^{\mathsf{CSQ}}(\nu_d) = \min_{\ell \geq 1} \frac{n_{d,\ell}}{\|\xi_{d,\ell}^{\mathsf{CSQ}}\|_{L^2}^2},$$

where we defined

$$\xi_{d,\ell}^{\mathsf{CSQ}}(Y, R) := Y q_{\star,\ell}(R), \qquad q_{\star,\ell}(R) := \frac{1}{\|Y\|_{L^2}} \mathbb{E}_{\nu_d}[Y Q_\ell(Z) | R].$$

Adapting the proofs in Appendix C.1, we obtain the following query complexity lower bound:

Proposition 1 (CSQ lower bound)**.** *Fix $\nu_d \in \mathfrak{L}_d$. If an algorithm $\mathcal{A} \in \mathsf{CSQ}(q, \tau)$ succeeds at distinguishing $\mathbb{P}_{\nu_d, \boldsymbol{w}}$ from $\mathbb{P}_{\nu_d, 0}$, then we must have*

$$q/\tau^2 \geq \mathsf{Q}_\star^{\mathsf{CSQ}}(\nu_d). \tag{30}$$

Using the non-linearity $\mathcal{T}_\ell(Y, R) := Y q_{\star,\ell}(R)$ in our algorithms (SP-Alg), (SGD-Alg) and (TU-Alg) described in Section 3, we can prove the same Theorem 2 with sample complexities replaced by $\sqrt{n_{d,\ell}}/\|\xi_{d,\ell}^{\mathsf{CSQ}}\|_{L^2}^2$ and runtime complexities replaced by $n_{d,\ell}/\|\xi_{d,\ell}^{\mathsf{CSQ}}\|_{L^2}^2$ (one simply plug these nonlinearities in the theorems in Appendices D, E and F).

Specializing to the Gaussian case, one recover the exact same result as in Section 4, but now with $\mathsf{k}_\star$ (generative exponent) replaced by k_I (information exponent) of the Gaussian SIM ρ. In particular, for all $\ell \leq \mathsf{k}_\mathsf{I}$, we have

$$\|\xi_{d,\ell}^{\mathsf{CSQ}}\|_{L^2}^2 \asymp d^{-(\mathsf{k}_\mathsf{I} - \ell)/2} \text{ for } \ell \equiv \mathsf{k}_\mathsf{I} \bmod 2 \quad \text{and} \quad \|\xi_{d,\ell}^{\mathsf{CSQ}}\|_{L^2}^2 \lesssim d^{-(\mathsf{k}_\mathsf{I} - \ell + 1)/2} \text{ for } \ell \not\equiv \mathsf{k}_\mathsf{I} \bmod 2.$$

Similarly to the generative exponent case (and general SQ), the optimal degrees for learning Gaussian SIMs with CSQ algorithms are always achieves at $\mathsf{l}_{\mathsf{m},\star} = \mathsf{l}_{\mathsf{T},\star} \in \{1, 2\}$, with the spectral estimator (SP-Alg) achieving

$$\mathsf{m} = \Theta_d(d^{\mathsf{k}_\mathsf{I}(\rho)/2}), \qquad \mathsf{T} = \Theta_d(d^{\mathsf{k}_\mathsf{I}(\rho)/2+1}).$$

Similar results as in Section 4 hold for learning with CSQ algorithms without using the norm $\|\boldsymbol{x}\|_2$. We note, however, that here, non-CSQ algorithms can achieve much better performance (attaining the complexity predicted by the generative exponent).

B Harmonic analysis on the sphere

In this section, we mention several facts about spherical harmonics, Gegenbauer polynomials, and their relationship. We will then discuss the harmonic decomposition of Hermite polynomials into Gegenbauer polynomials, which would be critical in all our analyses for the isotropic Gaussian input measure.

B.1 Spherical harmonics, Gegenbauer and Hermite polynomials

Spherical Harmonics. Consider the d-dimensional sphere $\mathbb{S}^{d-1} := \{\boldsymbol{z} : \|\boldsymbol{z}\|_2 = 1\}$ and the uniform probability measure $\tau_d \equiv \mathrm{Unif}(\mathbb{S}^{d-1})$ on it. This induces a Hilbert space $L^2(\mathbb{S}^{d-1}, \tau_d)$ equipped with the inner product:

$$\langle f, g \rangle_{L^2(\tau_d)} = \int_{\boldsymbol{z} \in \mathbb{S}^{d-1}} f(\boldsymbol{z}) g(\boldsymbol{z}) \, \tau_d(\mathrm{d}\boldsymbol{z}), \quad \text{for any } f, g \in L^2(\tau_d)$$

For $\ell \in \mathbb{N}$, consider $\tilde{V}_{d,\ell}$ be the space of degree ℓ homogeneous harmonic polynomials (i.e. homogeneous polynomial $q : \mathbb{R}^d \to \mathbb{R}$ with $\Delta q(\cdot) \equiv 0$). Let $V_{d,\ell}$ be the space of functions by restricting the domain to $\mathbb{S}^{d-1}$ of functions in $\tilde{V}_{d,\ell}$. A classical result from harmonic analysis (see [42, 11, 12] for additional background) is that we can decompose

$$L^2(\mathbb{S}^{d-1}, \tau_d) = \bigoplus_{\ell=0}^{\infty} V_{d,\ell} \,. \tag{31}$$

The dimension of each subspace is given by:

$$\dim(V_{d,\ell}) = n_{d,\ell} = \frac{2\ell + d - 2}{d - 2} \binom{d + \ell - 3}{\ell}.$$

For each $\ell \in \mathbb{N}$, the orthonormal bases for $V_{d,\ell}$ is given by $\{Y_{\ell i}^{(d)} : i \in n_{d,\ell}\}$. We then have the following orthogonality properties:

$$\langle Y_{\ell i}^{(d)}, Y_{kj}^{(d)} \rangle_{\nu_d} = \delta_{\ell k} \delta_{ij}.$$

Remark B.1. If one considers the unitary representation $\rho : \mathcal{O}_d \to U(L^2(\mathbb{S}^{d-1}))$ of the orthogonal group $\mathcal{O}_d = \{\boldsymbol{R} \in \mathbb{R}^{d \times d} : \boldsymbol{R}^\mathsf{T} \boldsymbol{R} = \mathbf{I}_d\}$ given by

$$\rho(\boldsymbol{R}) \, f(\boldsymbol{z}) = f(\boldsymbol{R}^\mathsf{T} \boldsymbol{z}) \,,$$

then Eq. (31) is also a decomposition of $L^2(\mathbb{S}^{d-1}, \tau_d)$ into a direct sum of irreducible representations of $\mathcal{O}_d$ (see [12] for a detailed treatment on the subject).

Gegenbauer Polynomials. Let $\tau_{d,1}$ denote the marginal distribution of the first coordinate $\langle \boldsymbol{z}, \boldsymbol{e}_1 \rangle$ with $\boldsymbol{z} \sim \tau_d$. We consider the family of Gegenbauer polynomials on $L^2([-1,1], \tau_{d,1})$, denoted by $\{Q_\ell^{(d)} : \ell \in \mathbb{N}\}$, where $Q_\ell^{(d)}$ is the degree-ℓ polynomial satisfying

$$\int_{-1}^{1} Q_\ell^{(d)}(z) Q_k^{(d)}(z) \, \tau_{d,1}(\mathrm{d}z) = \int_{\mathbb{S}^{d-1}} Q_\ell^{(d)}(\langle \boldsymbol{z}, \boldsymbol{e}_1 \rangle) Q_k^{(d)}(\langle \boldsymbol{z}, \boldsymbol{e}_1 \rangle) \, \tau_d(\mathrm{d}\boldsymbol{z}) = \delta_{\ell k} \,.$$

A relationship between the spherical harmonics and Gegenbauer polynomials is as follows:

$$Q_\ell^{(d)}(\langle \boldsymbol{z}, \boldsymbol{z}' \rangle) = \frac{1}{\sqrt{n_{d,\ell}}} \sum_{s \in [n_{d,\ell}]} Y_{\ell i}^{(d)}(\boldsymbol{z}) Y_{\ell i}^{(d)}(\boldsymbol{z}'), \text{ for all } \boldsymbol{z}, \boldsymbol{z}' \in \mathbb{S}^{d-1} \,. \tag{32}$$

Another important relationship is for any $\boldsymbol{w}, \boldsymbol{v} \in \mathbb{S}^{d-1}$:

$$\left\langle Q_\ell^{(d)}(\langle \cdot, \boldsymbol{w} \rangle), Q_k^{(d)}(\langle \cdot, \boldsymbol{v} \rangle) \right\rangle_{\tau_{d,1}} = \frac{\delta_{\ell k} \cdot Q_\ell^{(d)}(\langle \boldsymbol{w}, \boldsymbol{v} \rangle)}{Q_\ell^{(d)}(1)} \,, \tag{33}$$

where $Q_\ell^{(d)}(1) = \sqrt{n_{d,\ell}}$ as derived in Eq. (43) in Appendix I. We note that the normalization of Gegenbauer polynomials considered here is such that $\|Q_\ell^{(d)}\|_{L^2(\tau_{d,1})} = 1$ holds. Another popular choice is such that the value at 1 evaluates to 1 which we shall explicitly refer by the family of polynomials $\{P_\ell^{(d)}\}_{\ell\in\mathbb{N}}$ and The normalizing factor is such that $P_\ell^{(d)}(\cdot) = Q_\ell^{(d)}(\cdot)/\sqrt{n_{d,\ell}}$.

The derivative of the ℓ^{th} Gegenbauer polynomial for $\ell \geq 1$ can be expressed as

$$\frac{\mathrm{d}}{\mathrm{d}z} Q_\ell^{(d)}(z) = Q_\ell^{(d)}(z)' = \frac{\ell(\ell+d-2)\sqrt{n_{d,\ell}}}{(d-1)\sqrt{B(d+2,\ell-1)}} Q_{k-1}^{d+2}(z) = C(d,\ell)\, Q_{k-1}^{(d+2)}(z)\,, \tag{34}$$

where for a fixed constant ℓ and growing d, we have $C(d,\ell) = \Theta_d(\sqrt{d})$. Let $f \in L^2(\mathbb{S}^{d-1}, \tau_d)$ such that f is invariant by the action of $\mathcal{O}_{\boldsymbol{w}^\perp} = \{\boldsymbol{W} \in \mathcal{O}_d \colon \boldsymbol{W}^\top \boldsymbol{w} = \boldsymbol{w}\}$ which is the set of orthogonal matrices which keeps the direction $\boldsymbol{w}$ fixed, i.e. f only depends on the projection $\langle \boldsymbol{w}, \boldsymbol{z}\rangle$. Then f admits the following decomposition

$$f(\boldsymbol{z}) = \sum_{\ell=0}^{\infty} \alpha_\ell Q_\ell^{(d)}(\langle \boldsymbol{w}, \boldsymbol{z}\rangle). \tag{35}$$

Hermite polynomials. Consider the probabilist's Hermite polynomials $\{\mathrm{He}_k : k \in \mathbb{N}\}$, in the normalization that form an orthonormal basis of $L^2(\mathbb{R},\gamma)$, where $\gamma(\mathrm{d}x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$ is the standard Gaussian measure. He_k is a polynomial of degree k.

$$\mathbb{E}_{G\sim\mathsf{N}(0,1)}\left[\mathrm{He}_j(G)\mathrm{He}_k(G)\right] = \delta_{jk}\,.$$

As a consequence, for any $g \in L^2(\mathbb{R},\gamma)$, we have the following decomposition

$$g(x) = \sum_{k=0}^{\infty} \mu_k(g)\, \mathrm{He}(x), \quad \mu_k(g) = \mathbb{E}_{G\sim\mathsf{N}(0,1)}[g(G)\mathrm{He}_k(G)]\,.$$

B.2 Harmonic decomposition of Hermite into Gegenbauer polynomials

We know that $\boldsymbol{x} \sim \mathsf{N}(\boldsymbol{0}, \mathbf{I}_d)$ can be equivalently represented by

$$\boldsymbol{x} = \|\boldsymbol{x}\|_2 \cdot \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}, \quad \text{where } \|\boldsymbol{x}\|_2 =: r \sim \chi_d \text{ and } \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2} =: \boldsymbol{z} \sim \mathrm{Unif}(\mathbb{S}^{d-1}) \text{ are independent.}$$

Therefore, $x_1 = r \cdot u_1$, where $x_1 \sim \mathsf{N}(0,1)$ and $z_1 \sim \tau_{d,1}$. In what follows, we denote $x = x_1$ and $z = z_1$ for convenience. The Gegenbauer polynomials are an orthonormal basis for $\tau_{d,1}$ and Hermite polynomials are (unnormalized) orthogonal basis for $\mathsf{N}(0,1)$. Our goal is to explicitly express $\mathrm{He}_k(x) = \mathrm{He}_k(r\cdot z)$ in terms Gegenbauer polynomials $\{Q_\ell^{(d)}(z)\}$, formalized in the following proposition.

Proposition 2 (Decomposing Hermite into Gegenbauer)**.** *For any $k \in \mathbb{N}$, we have*

$$\mathrm{He}_k(r\cdot z) = \sum_{\ell=0}^{\infty} \beta_{k,\ell}(r) Q_\ell^{(d)}(z)\,, \tag{36}$$

where $\beta_{k,\ell}(r) = 0$ if $(\ell > k)$ or $(\ell \not\equiv k \mod 2)$, and otherwise

$$\beta_{k,\ell}(r) := \frac{\sqrt{k!}\sqrt{K(d,\ell)}}{(N!)\,2^N}\left(\sum_{i=0}^{N} \frac{\binom{N}{i}(-1)^{N-i}\, r^{\ell+2i}}{\prod_{j=0}^{i+\ell-1}(d+2j)}\right), \tag{37}$$

where $N = (k-\ell)/2$ and $K(d,\ell) \asymp d^\ell$ as $d \to \infty$ and ℓ is constant, i.e. $K(d,\ell) = \Theta_d(d^\ell)$.

Essentially, we are decomposing the Hermite basis into the Gegenbauer polynomials, which is the correct basis for the "directional" component z, and explicitly computing the coefficients that depend on the radial component r. Such relationship is derived by technical algebraic manipulations, and in similar spirit to [34], relating Hermite and Gegenbauer polynomials. However, the precise expression is sensitive to the normalization used for Gegenbauer polynomials. Thus, we provide explicit calculations of this decomposition in Appendix I in our choice of normalization.

For our upper and lower bound analyses, in order to measure correlation of $\mathrm{He}_k(r\cdot z)$ with $Q_\ell^{(d)}(z)$, using both type of queries (with or without norm), the asymptotic bounds on the following moments of these coefficients will play an important role.

Lemma 3. *For any fixed $\ell \leq k \in \mathbb{N}$ with same parity, i.e. $k \equiv \ell \mod 2$, we have*

$$\mathbb{E}_{r\sim\chi_d}[\beta_{k,\ell}(r)^2] \asymp d^{-\frac{(k-\ell)}{2}} \quad \textit{and} \quad \mathbb{E}_{r\sim\chi_d}[\beta_{k,\ell}(r)]^2 \asymp d^{-(k-\ell)}\,.$$

In order to show this lemma, we will use the following well-known facts.

Fact 1 (Moments of χ_d distribution)**.** *For any $p \in \mathbb{N}$, the even and odd moments of χ_d distribution are given by,*

$$\mathbb{E}_{r\sim\chi_d}[r^{2p}] = \prod_{j=0}^{p-1}(d+2j) \quad \textit{and} \quad \mathbb{E}_{r\sim\chi_d}[r^{2p+1}] = \mathbb{E}[r]\prod_{j=0}^{p-1}(d+2j+1) = \frac{\sqrt{2}\,\Gamma(\frac{d+1}{2})}{\Gamma(d/2)}\prod_{j=0}^{p-1}(d+2j+1) \tag{38}$$

Therefore, for any fixed $m \in \mathbb{N}$, the asymptotic behavior of the m^{th} moment as $d \to \infty$ is given by

$$\mathbb{E}_{r\sim\chi_d}[r^m] \asymp \left(\sqrt{d}\cdot\mathbf{1}\{m \equiv 1 \bmod 2\}\right)\prod_{j=0}^{\lfloor m/2\rfloor-1}(d+2j)\,.$$

Fact 2. *For any univariate polynomial $g : \mathbb{R} \to \mathbb{R}$, the n^{th} forward finite difference of g at any value u is given by*

$$\Delta^n g(u) := \sum_{i=0}^{n}\binom{n}{i}(-1)^{n-i}g(u+i)\,.$$

For any polynomial g, the n^{th} forward finite difference $\Delta^n g(u) \equiv 0$ if $\deg(g) < n$ and $\Delta^n g(u)$ is a non-zero constant if $\deg(g) = n$. Moreover, for any polynomial given by shifted binomial coefficient of degree n, with shift $u_0 \in \mathbb{R}$

$$g_n(u) = \frac{(u+u_0)(u+u_0-1)\cdots(u+u_0-n+1)}{n!} =: \binom{u+u_0}{n}\,,$$

the constant value of n^{th} forward finite difference is unity, i.e. $\Delta^n g(u) = 1$.

Proof of Lemma 3. We will use the above facts directly along with $K(d,\ell) = d^\ell$ and N, k, ℓ are constants (not dependent on d) from Proposition 2 throughout the proof. We start by the first part of the claim

$$\begin{aligned}
\mathbb{E}_{r\sim\chi_d}[\beta_{k,\ell}(r)^2] &= \frac{(k!)}{(N!)^2\,2^{2N}}K(d,\ell)\left(\sum_{n=0}^{N}\sum_{m=0}^{N}\frac{\binom{N}{n}\binom{N}{m}(-1)^{2N-n-m}\,\mathbb{E}[r^{2\ell+2n+2m}]}{\prod_{j=0}^{n+\ell-1}(d+2j)\prod_{j=0}^{m+\ell-1}(d+2j)}\right)\\
&\asymp d^\ell\left(\sum_{n=0}^{N}\sum_{m=0}^{N}\frac{\binom{N}{n}\binom{N}{m}(-1)^{2N-n-m}\prod_{j=0}^{\ell+m+n-1}(d+2j)}{\prod_{j=0}^{n+\ell-1}(d+2j)\prod_{j=0}^{m+\ell-1}(d+2j)}\right) \qquad \text{(using Fact 1)}\\
&\asymp \frac{d^\ell}{\prod_{j=0}^{N+\ell-1}(d+2j)}\left(\sum_{n=0}^{N}\sum_{m=0}^{N}\binom{N}{n}\binom{N}{m}(-1)^{2N-n-m}\prod_{j=m+\ell}^{\ell+m+n-1}(d+2j)\prod_{j=n+\ell}^{N+\ell-1}(d+2j)\right)\\
&\asymp d^{-N}\left(\sum_{n=0}^{N}\sum_{m=0}^{N}\binom{N}{n}\binom{N}{m}(-1)^{2N-n-m}\prod_{j=m+\ell}^{\ell+m+n-1}(d+2j)\prod_{j=n+\ell}^{N+\ell-1}(d+2j)\right)\\
&= d^{-\frac{(k-\ell)}{2}}\left(\sum_{n=0}^{N}\binom{N}{n}(-1)^{N-n}\prod_{j=n+\ell}^{N+\ell-1}(d+2j)\left(\sum_{m=0}^{N}\binom{N}{m}(-1)^{N-m}\prod_{j=m+\ell}^{\ell+m+n-1}(d+2j)\right)\right)\\
&= d^{-\frac{(k-\ell)}{2}}\left(\sum_{n=0}^{N}\binom{N}{n}(-1)^{N-n}\prod_{j=n+\ell}^{N+\ell-1}(d+2j)\left(\sum_{m=0}^{N}\binom{N}{m}(-1)^{N-m}\binom{\frac{d}{2}+m+n-1}{n}2^n n!\right)\right)
\end{aligned}$$

We are now going to show that term inside the parenthesis is some constant independent of d. Apriori it may seem that it depends on d, however, we have a sum of several terms with alternative positive and negative signs and we will show that all the terms that depend on d mutually cancel out.

To show this we first observe that $g(m) = \binom{\frac{d}{2}+m+n-1}{n}$ is a polynomial of degree n given by binomial coefficient. And, therefore, the N^{th} forward finite difference $\Delta^N g(m) \equiv 0$ for any $n < N$, or siimply 1 for $n = N$ using Fact 2. Formally,

$$\sum_{m=0}^{N} \binom{N}{m} (-1)^{N-m} \binom{\frac{d}{2}+m+n-1}{n} 2^n n! = 2^n\, n! \cdot \delta_{Nn}\,,$$

where the scaling factor of $2^n n!$ of the polynomial $g(m)$ can be taken out as the forward finite difference operator Δ^N is linear. We conclude that

$$\begin{aligned}
\mathbb{E}[\beta_{k,\ell}(r)^2] &= d^{-\frac{(k-\ell)}{2}} \left(\sum_{n=0}^{N} \binom{N}{n} (-1)^{N-n} \prod_{j=n+\ell}^{N+\ell-1} (d+2j) \left(\sum_{m=0}^{N} \binom{N}{m} (-1)^{N-m} \binom{\frac{d}{2}+m+n-1}{n} 2^n n! \right) \right) \\
&= d^{-\frac{(k-\ell)}{2}} \left(\sum_{n=0}^{N} \binom{N}{n} (-1)^{N-n} \prod_{j=n+\ell}^{N+\ell-1} (d+2j) \;\cdot\; \delta_{Nn} \;\cdot 2^n n! \right) = 2^N N!\; d^{-\frac{(k-\ell)}{2}} \\
&\asymp d^{-\frac{(k-\ell)}{2}}\,.
\end{aligned}$$

We now show the second part that

$$\mathbb{E}[\beta_{k,\ell}(r)]^2 \asymp d^{-(k-\ell)}\,.$$

Let us consider any k, ℓ such that they have the same parity $k \equiv \ell \mod 2$. We have

$$\begin{aligned}
\mathbb{E}[\beta_{k,\ell}(r)] &= \frac{\sqrt{(k!)\, K(d,\ell)}}{(N!)\, 2^N} \left(\sum_{i=0}^{N} \frac{\binom{N}{i} (-1)^{N-i}\, \mathbb{E}[r^{\ell+2i}]}{\prod_{j=0}^{\ell+i-1}(d+2j)} \right) \\
&\asymp d^{\ell/2} \left(\sqrt{d} \cdot \mathbf{1}\{\ell \equiv 1 \bmod 2\} \right) \left(\sum_{i=0}^{N} \binom{N}{i} (-1)^{N-i} \frac{\prod_{j=0}^{\lfloor \ell/2 \rfloor + i-1}(d+2j)}{\prod_{j=0}^{\ell+i-1}(d+2j)} \right) \qquad \text{(using Fact 1)} \\
&\asymp d^{\lceil \ell/2 \rceil} \left(\sum_{i=0}^{N} \binom{N}{i} (-1)^{N-i} \frac{1}{\prod_{j=\ell/2+i}^{\ell+i-1}(d+2j)} \right) \\
&\asymp d^{\lceil \ell/2 \rceil} \sum_{i=0}^{N} \binom{N}{i} (-1)^{N-i} \frac{\prod_{j=i}^{N+i-1}(d+2j)}{\prod_{j=i}^{N+i-1}(d+2j) \prod_{j=\lfloor \ell/2 \rfloor+i}^{\ell+i-1}(d+2j)} \\
&\asymp \frac{d^{\lceil \ell/2 \rceil}}{d^{\lceil k/2 \rceil}} \sum_{i=0}^{N} (1+o_d(1)) \binom{N}{i} (-1)^{N-i} \prod_{j=i}^{N+i-1} (d+2j)
\end{aligned}$$

In the last line, we used the fact that $k \equiv \ell \mod 2$ and $N = (k-\ell)/2$ and thus for every $0 \le i \le N$,

$$\prod_{j=i}^{N+i-1} (d+2j) \prod_{j=\ell/2+i}^{i+\ell-1} (d+2j) = d^{N+\ell/2} \prod_{j=i}^{N+i-1} \left(1 + \frac{2j}{d}\right) \prod_{j=\lfloor \ell/2 \rfloor + i}^{\ell+i-1} \left(1 + \frac{2j}{d}\right) = (1+o_d(1))\, d^{N+\ell-\lfloor \ell/2 \rfloor}\,,$$

and

$$N + \ell - \lfloor \ell/2 \rfloor = \frac{(k-\ell)}{2} + \ell - \lfloor \ell/2 \rfloor = \frac{k+\ell}{2} - \lfloor \ell/2 \rfloor = \lceil k/2 \rceil\,.$$

Continuing to simplify the original expression

$$\begin{aligned}
\mathbb{E}[\beta_{k,\ell}(r)] \asymp \frac{d^{\lceil \ell/2 \rceil}}{d^{\lceil k/2 \rceil}} &\sum_{i=0}^{N} (1+o_d(1)) \binom{N}{i} (-1)^{N-i} \prod_{j=i}^{N+i-1} (d+2j) \\
&= \frac{1}{d^{\frac{k-\ell}{2}}} \sum_{i=0}^{N} \binom{N}{i} (-1)^{N-i} \binom{\frac{d}{2}+i+N-1}{N} 2^N N! \\
&= d^{-\frac{(k-\ell)}{2}} 2^N N! \;\asymp\; d^{-\frac{(k-\ell)}{2}}\,.
\end{aligned}$$

Here the last line followed from the fact that the polynomial $g(i) = \binom{\frac{d}{2}+i+N-1}{N}$ is of degree N given by a shifted binomial coefficient, and thus, the N^{th} forward finite difference of g is constant, which in this case is just 1 by Fact 2. We finally conclude the proof by noting that $\mathbb{E}[\beta_{k,\ell}(r)]^2 \asymp d^{-(k-\ell)}$. $\square$

B.3 Proof of Proposition 2

We now return to the deferred proof of Proposition 2. First, we use the explicit expression of He_k so that $\|\mathrm{He}_k\|_{L^2} = 1$.

$$\mathrm{He}_k(r \cdot z) = \mathrm{He}_k(x) = \sqrt{k!} \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m}{m!(k-2m)!} \frac{x^{k-2m}}{2^m} = \sqrt{k!} \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m r^{k-2m}}{m!(k-2m)!} \frac{z^{k-2m}}{2^m} . \tag{39}$$

Our goal is to express z^{k-2m} in terms of $Q_\ell^{(d)}(z)$ to get the final decomposition. To this end, we use the explicit expressions computed with (a different normalization) of Gegenbauer polynomials. In particular, using [20, Eq. 18.18.17]

$$z^n = \frac{n!}{2^n} \sum_{l=0}^{\lfloor n/2 \rfloor} \frac{\alpha + n - 2l}{\alpha} \frac{1}{l!(\alpha+1)_{n-l}} C_{n-2l}^{(\alpha)}(z) , \tag{40}$$

where $(b)_a$ is the rising factorial and $C_\ell^{(\alpha)}(z)$ is an unnormalized Gegenbauer polynomial $Q_\ell^{(d)}(z)$ with $\alpha = (d-2)/2$ satisfying

$$\int_{-1}^{+1} C_\ell^{(\alpha)}(z) C_k^{(\alpha)}(z)(1-z^2)^{\alpha - \frac{1}{2}} \mathrm{d}z = \delta_{\ell k} \frac{\pi 2^{1-2\alpha}\Gamma(\ell + 2\alpha)}{\ell!(\ell+\alpha)[\Gamma(\alpha)]^2} \tag{41}$$

We can express $C_\ell^{(\alpha)}(z)/\sqrt{K(d,\ell)} = Q_\ell^{(d)}(z)$ where $K(d,\ell)$ can be computed using

$$1 = \frac{1}{K(d,\ell)} \int_{-1}^{+1} C_\ell^{(\alpha)}(z)^2 \, \tau_{d,1}(\mathrm{d}z) = \frac{1}{K(d,\ell)\mathsf{B}\left(\alpha+\frac{1}{2},\frac{1}{2}\right)} \int_{-1}^{+1} C_\ell^{(\alpha)}(z)^2 \, (1-z^2)^{\alpha-\frac{1}{2}} \tau_{d,1}(\mathrm{d}z)$$

where $\mathsf{B}(\cdot,\cdot)$ is the standard Beta function. We can use Eq. (41) to compute

$$K(d,\ell) = \frac{\pi 2^{1-2\alpha}\Gamma(\ell+2\alpha)}{\mathsf{B}\left(\alpha+\frac{1}{2},\frac{1}{2}\right) \ell!(\ell+\alpha)[\Gamma(\alpha)]^2} \tag{42}$$

It is straight-forward to simplify

$$K(d,\ell) = \frac{\alpha\Gamma(\ell+2\alpha)}{\ell!\,(\alpha+\ell)\Gamma(2\alpha)} = \frac{(d-2)\Gamma(d-2+\ell)}{\ell!\,(d-2+2\ell)\Gamma(d-2)} = \Theta_d(d^\ell) ,$$

and also substituting $C_\ell^{(\alpha)}(1)$ from [48], we obtain

$$Q_\ell^{(d)}(1) = \frac{C_\ell^{(\alpha)}(1)}{\sqrt{K(d,\ell)}} = \sqrt{\frac{\ell!\,(\alpha+\ell)\Gamma(2\alpha)}{\alpha\Gamma(2\alpha+\ell)}} \cdot \frac{\Gamma(2\alpha+\ell)}{\Gamma(2\alpha)\ell!} = \sqrt{\frac{2\ell+d-2}{d-2}\binom{d+\ell-3}{\ell}} = \sqrt{n_{d,\ell}} . \tag{43}$$

We are now ready to combine the equations derived and compute the desired decomposition Eq. (36). For any $M \in \mathbb{N}$, we let $\boxed{M} = \{m \in \mathbb{N} : m \leq M \text{ and } m \equiv M \mod 2\}$. Recall from (39)

$$
\begin{aligned}
\mathrm{He}_k(r \cdot u) &= \sqrt{k!} \sum_{m=0}^{\lfloor k/2 \rfloor} \frac{(-1)^m r^{k-2m}}{m!(k-2m)!} \frac{u^{k-2m}}{2^m} = \sqrt{k!} \sum_{m \in \boxed{k}} \frac{(-1)^{(k-m)/2} r^m}{m!((k-m)/2)!} \frac{u^m}{2^{(k-m)/2}} \\
&\hspace{20em} \text{(Change of variables)} \\
&= \sqrt{k!} \sum_{m \in \boxed{k}} \frac{(-1)^{(k-m)/2} r^m}{m!((k-m)/2)!} \frac{m!}{2^m 2^{(k-m)/2}} \sum_{l=0}^{\lfloor m/2 \rfloor} \frac{\alpha + m - 2l}{\alpha} \frac{1}{l!(\alpha+1)_{m-l}} C^{(\alpha)}_{m-2l}(u) \\
&\hspace{20em} \text{(Using (40))} \\
&= \sqrt{k!} \sum_{m \in \boxed{k}} \frac{(-1)^{(k-m)/2} r^m}{((k-m)/2)!} \frac{1}{2^{(k+m)/2}} \sum_{\ell \in \boxed{m}} \frac{\alpha+\ell}{\alpha} \frac{1}{((m-\ell)/2)!(\alpha+1)_{(m+\ell)/2}} \sqrt{K(d,\ell)}\, Q^{(d)}_\ell(u) \\
&\hspace{12em} \text{(Changing } \ell = m - 2l \text{ and } C^{(\alpha)}_\ell = \sqrt{K(d,\ell)}\, Q^{(d)}_\ell \text{)} \\
&= \sqrt{(k!)\, K(d,\ell)} \sum_{\ell \in \boxed{k}} Q^{(d)}_\ell(u) \frac{\alpha+\ell}{\alpha} \left(\sum_{\substack{m=\ell \\ m \equiv k \bmod 2}}^{k} \frac{(-1)^{(k-m)/2} r^m}{((k-m)/2)!} \frac{1}{2^{(k+m)/2}((m-\ell)/2)!(\alpha+1)_{(m+\ell)/2}} \right) \\
&:= \sum_{\ell=0}^{\infty} Q^{(d)}_\ell(u) \beta_{k,\ell}(r)\,, \text{ where if } \ell \notin \boxed{k} \text{ then } \beta_{k,\ell}(r) = 0.
\end{aligned}
$$

Otherwise, letting $N = (k-\ell)/2$

$$
\begin{aligned}
\beta_{k,\ell}(r) &= \sqrt{k!\, K(d,\ell)} \frac{\alpha+\ell}{\alpha} \left(\sum_{\substack{m=\ell \\ m \equiv k \bmod 2}}^{k} \frac{(-1)^{(k-m)/2} r^m}{((k-m)/2)!\, 2^{(k+m)/2}((m-\ell)/2)!(\alpha+1)_{(m+\ell)/2}} \right) \\
&= \sqrt{k!\, K(d,\ell)} \frac{\alpha+\ell}{\alpha} \left(\sum_{i=0}^{N} \frac{(-1)^{N-i} r^{\ell+2i}}{(N-i)!\, 2^{(k+\ell)/2+i} (i)! (\alpha+1)_{\ell+i}} \right) \quad \text{(changing } m = \ell + 2i\text{)} \\
&= \frac{\sqrt{k!\, K(d,\ell)}}{(N!)\, 2^N} \frac{\alpha+\ell}{\alpha} \left(\sum_{i=0}^{N} \binom{N}{i} \frac{(-1)^{N-i} r^{\ell+2i}}{2^{\ell+i}(\alpha+1)_{\ell+i}} \right).
\end{aligned}
$$

Recall that $\alpha = (d-2)/2$ here, and thus

$$
2^{\ell+i}(\alpha+1)_{\ell+i} = 2^{\ell+i} \prod_{j=0}^{\ell+i-1} (\alpha+1+j) = 2^{\ell+i} \prod_{j=0}^{\ell+i-1} \left(\frac{d-2}{2} + 1 + j \right) = \prod_{j=0}^{\ell+i-1} (d+2j)\,.
$$

Thus, for $\ell \equiv k \mod 2$

$$
\beta_{k,\ell}(r) = \frac{\sqrt{k!\, K(d,\ell)}}{(N!)\, 2^N} \frac{d+2\ell-2}{d-2} \left(\sum_{i=0}^{N} \binom{N}{i} \frac{(-1)^{N-i} r^{\ell+2i}}{\prod_{j=0}^{\ell+i-1}(d+2j)} \right).
$$

The lemma follows by redefining $K(d,\ell)$ with $K(d,\ell)(d+2\ell-2)^2/(d-2)^2 = \Theta_d(d^\ell)$.

C Statistical Query (SQ) and Low-Degree Polynomial (LDP) lower bounds

In this appendix, we briefly review the statistical query (SQ) and low-degree polynomial (LDP) frameworks and present the proof of Theorem 1. In particular, we provide an interpretation of our lower bounds in terms of subproblems with queries restricted to the harmonic subspace $V_{d,\ell}$.

C.1 Statistical Query lower bounds

The Statistical Query (SQ) framework, introduced by Kearns [31], models algorithms that interact with data only through expectations of query functions, rather than direct access to samples. The complexity of these algorithms is measured up to some worst-case tolerance on these expectations. While based on worst-case error rather than sampling error encountered in practice, the SQ framework has proven remarkably effective in analyzing the computational complexity of statistical problems, often yielding accurate predictions for algorithmic feasibility. We refer to [10, 39] for additional background.

Below, it will be useful to present a variant of SQ algorithms, called *query-restricted statistical query algorithms*, introduced in [28]. In this model, queries are restricted to a set $\mathcal{Q} \subseteq \mathbb{R}^{\mathcal{Y}\times\mathbb{R}^d}$ of measurable functions $\mathcal{Y} \times \mathbb{R}^d \to \mathbb{R}$. We denote $\mathcal{Q}$-$\mathsf{SQ}(q,\tau)$ this class of algorithms, with number of queries q and tolerance $\tau > 0$. We will mainly consider the standard case of *unrestricted queries*, denoted $\mathsf{SQ}(q,\tau)$, where $\mathcal{Q}$ contains all measurable functions. In Appendix A.3 we discuss the case of *correlation statistical queries* (CSQ).

Our lower bounds hold for the detection problem (hypothesis testing) of distinguishing between

$$\{\mathbb{P}_{\nu_d,\boldsymbol{w}} : \boldsymbol{w} \in \mathbb{S}^{d-1}\} \qquad \text{v.s.} \qquad \{\mathbb{P}_{\nu_d,0}\}. \tag{44}$$

Below, we describe $\mathcal{Q}$-SQ algorithms in this context.

$\mathcal{Q}$-restricted SQ algorithm. For a number of queries q and tolerance $\tau > 0$, a $\mathcal{Q}$-restricted SQ algorithm $\mathcal{A} \in \mathcal{Q}$-$\mathsf{SQ}(q,\tau)$ for detecting SIMs takes an input distribution $\mathbb{P}$ from (44) and operates in q rounds where at each round $t \in \{1,\dots,q\}$, it issues a query $\phi_t \in \mathcal{Q}$, and receives a response v_t such that

$$\left|v_t - \mathbb{E}_{\mathbb{P}}[\phi_t(y,\boldsymbol{x})]\right| \leq \tau\sqrt{\mathrm{Var}_{\mathbb{P}_0}(\phi_t)}, \tag{45}$$

where we set $\mathbb{P}_0$ to be the null distribution (that is, $\mathbb{P}_{\nu_d,0}$ here). The query ϕ_t can depend on the past responses $v_1,\dots,v_{t-1}$. After issuing q queries, the learner outputs $\mathcal{A}(\mathbb{P}) \in \{0,1\}$. We say that $\mathcal{A}$ succeeds in distinguishing $\mathbb{P}_{\nu_d,\boldsymbol{w}}$ and $\mathbb{P}_{\nu_d,0}$, if $\mathcal{A}(\mathbb{P}_{\nu_d,\boldsymbol{w}}) = 1$ for all $\boldsymbol{w} \in \mathbb{S}^{d-1}$, and $\mathcal{A}(\mathbb{P}_{\nu_d,0}) = 0$.

Remark C.1. The variance scaling on the right-hand side in (45) is non-standard in the SQ literature. It is introduced here as a convenient way to normalize queries, which is necessary for τ to be meaningful. We note that other normalizations are possible and refer to [28, Remark 3.1] for a discussion.

General lower-bound. The following proposition is a simple, standard lower bound on the query complexity based on the second moment method (e.g., see [28]):

Proposition 3 (General $\mathcal{Q}$-restricted SQ lower bound)**.** *Fix $\nu_d \in \mathfrak{L}_d$. If an algorithm $\mathcal{A} \in \mathcal{Q}$-$\mathsf{SQ}(q,\tau)$ succeeds at distinguishing $\mathbb{P}_{\nu_d,\boldsymbol{w}}$ from $\mathbb{P}_{\nu_d,0}$, then we must have*

$$q/\tau^2 \geq \left[\sup_{\phi\in\mathcal{Q}} \frac{\mathrm{Var}_{\boldsymbol{w}\sim\tau_d}\{\mathbb{E}_{\mathbb{P}_{\nu_d,\boldsymbol{w}}}\phi\}}{\mathrm{Var}_{\mathbb{P}_{\nu_d,0}}\{\phi\}}\right]^{-1}. \tag{46}$$

Proof. Consider $\mathcal{A} \in \mathcal{Q}$-$\mathsf{SQ}(q,\tau)$ and denote $\phi_1,\dots,\phi_q \in \mathcal{Q}$ the sequence of queries issued by $\mathcal{A}$ when it receives responses $v_t = \mathbb{E}_{\mathbb{P}_{\nu_d,0}}[\phi_t]$, $t \in [q]$. Here the responses are fixed and deterministic, and the queries $\{\phi_t\}_{t\in[q]}$ do not depend on the source distribution $\mathbb{P}_{\nu_d,\boldsymbol{w}}$, and in particular $\boldsymbol{w}$. By union bound and Markov's inequality,

$$\begin{aligned}\mathbb{P}_{\boldsymbol{w}\sim\tau_d}\left(\exists t\in[q],\ |\mathbb{E}_{\mathbb{P}_{\nu_d,\boldsymbol{w}}}[\phi_t] - v_t| > \tau\sqrt{\mathrm{Var}_{\mathbb{P}_{\nu_d,0}}[\phi_t]}\right) &\leq \frac{q}{\tau^2}\cdot\sup_{t\in[q]}\frac{\mathrm{Var}_{\boldsymbol{w}\sim\tau_d}\{\mathbb{E}_{\mathbb{P}_{\nu_d,\boldsymbol{w}}}\phi_t\}}{\mathrm{Var}_{\mathbb{P}_{\nu_d,0}}\{\phi_t\}} \\ &\leq \frac{q}{\tau^2}\cdot\sup_{\phi\in\mathcal{Q}}\frac{\mathrm{Var}_{\boldsymbol{w}\sim\tau_d}\{\mathbb{E}_{\mathbb{P}_{\nu_d,\boldsymbol{w}}}\phi\}}{\mathrm{Var}_{\mathbb{P}_{\nu_d,0}}\{\phi\}}.\end{aligned}$$

This implies that the $v_t = \mathbb{E}_{\mathbb{P}_{\nu_d,0}}[\phi_t]$ responses are compatible for all q queries with positive probability over $\boldsymbol{w} \sim \tau_d$ whenever inequality (46) is not satisfied, and $\mathcal{A}$ fails the detection task in that case. This concludes the proof. $\square$

The query complexity bound in Theorem 1.(i) follows from Proposition 3 with unrestricted queries $\mathcal{Q}_{\mathsf{SQ}}$ and the following identity:

Lemma 4. *For $\nu_d \in \mathfrak{L}_d$ and $\mathcal{Q}_{\mathsf{SQ}}$ the class of unrestricted queries (all measurable functions), we have the identity*

$$\sup_{\phi \in \mathcal{Q}_{\mathsf{SQ}}} \frac{\mathrm{Var}_{\boldsymbol{w}\sim\tau_d}\{\mathbb{E}_{\mathbb{P}_{\nu_d,\boldsymbol{w}}}\phi\}}{\mathrm{Var}_{\mathbb{P}_{\nu_d,0}}\{\phi\}} = \sup_{\ell \geq 1} \frac{\|\xi_{d,\ell}\|_{L^2}^2}{n_{d,\ell}} =: [\mathsf{Q}_\star(\nu_d)]^{-1}. \tag{47}$$

We defer the proof of this lemma below to Section C.1.1. The identity (47) shows that the lower bound effectively decouples across the different harmonic subspaces. Below, we provide an interpretation of this result: if we restrict the queries $\mathcal{Q}$ to be in $V_{d,\ell}$, then the SQ lower bound becomes $n_{d,\ell}/\|\xi_{d,\ell}\|_{L^2}^2$. Specifically, for each $\ell \geq 1$, define $\mathcal{Q}_{\mathsf{SQ},\ell}$ to be the set of all queries $\phi(y, \boldsymbol{x})$ than can be written as

$$\phi(y, \boldsymbol{x}) = \sum_{s \in [n_{d,\ell}]} g_\ell(y, r) Y_{\ell s}(\boldsymbol{z}), \tag{48}$$

where $\{Y_{\ell s}\}_{s\in[n_{d,\ell}]}$ is a basis of $V_{d,\ell}$. Then by Proposition 3 and the proof of Lemma 4, we have:

Corollary 3. *Fix $\nu_d \in \mathfrak{L}_d$ and $\ell \geq 1$. If an algorithm $\mathcal{A} \in \mathcal{Q}_{\mathsf{SQ},\ell}$-$\mathsf{SQ}(q,\tau)$ succeeds at distinguishing $\mathbb{P}_{\nu_d,\boldsymbol{w}}$ from $\mathbb{P}_{\nu_d,0}$, then we must have*

$$q/\tau^2 \geq \left[\sup_{\phi \in \mathcal{Q}_{\mathsf{SQ},\ell}} \frac{\mathrm{Var}_{\boldsymbol{w}\sim\tau_d}\{\mathbb{E}_{\mathbb{P}_{\nu_d,\boldsymbol{w}}}\phi\}}{\mathrm{Var}_{\mathbb{P}_{\nu_d,0}}\{\phi\}}\right]^{-1} = \frac{n_{d,\ell}}{\|\xi_{d,\ell}\|_{L^2}^2}. \tag{49}$$

For each $\ell \geq 1$, we show algorithms with queries restricted to $V_{d,\ell}$ as in (48) that matches this lower bound. Thus, effectively, the problem decouples into subproblems, one for each $V_{d,\ell}$: on each harmonic subspace, we have a matching upper and lower bound on the query complexity, and the optimal algorithm is obtained by choosing the optimal degree ℓ that attains the maximum in (47).

C.1.1 Proof of Lemma 4

For clarity, we drop the subscript ν_d below and denote $\mathbb{P}_{\boldsymbol{w}} := \mathbb{P}_{\nu_d,\boldsymbol{w}}$ and $\mathbb{P}_0 := \mathbb{P}_{\nu_d,0}$. Note that a property of the null distribution is that

$$\mathbb{E}_{\boldsymbol{w}}\left[\mathbb{E}_{\mathbb{P}_{\boldsymbol{w}}}[\phi]\right] = \mathbb{E}_{\mathbb{P}_0}[\phi],$$

that is, $\mathbb{P}_0$ is the marginal distribution of $(y, \boldsymbol{x})$ under the uniform prior $\boldsymbol{w} \sim \tau_d$. Thus,

$$\mathrm{Var}_{\boldsymbol{w}}\{\mathbb{E}_{\mathbb{P}_{\boldsymbol{w}}}\phi\} = \mathbb{E}_{\boldsymbol{w}}\left[|\Delta_\phi(\boldsymbol{w})|^2\right], \quad \text{where} \quad \Delta_\phi(\boldsymbol{w}) = \mathbb{E}_{\mathbb{P}_{\boldsymbol{w}}}[\phi] - \mathbb{E}_{\mathbb{P}_0}[\phi].$$

Let's introduce the Radon-Nikodym derivative and write

$$\Delta_\phi(\boldsymbol{w}) = \mathbb{E}_{\mathbb{P}_0}\left[\left(\frac{\mathrm{d}\mathbb{P}_{\boldsymbol{w}}}{\mathrm{d}\mathbb{P}_0}(y_0, \boldsymbol{z}, r) - 1\right)\phi(y_0, \boldsymbol{z}, r)\right].$$

Recall that the likelihood ratio decomposes into Gegenbauer polynomials as (equality in $L^2(\mathbb{P}_0)$)

$$\frac{\mathrm{d}\mathbb{P}_{\boldsymbol{w}}}{\mathrm{d}\mathbb{P}_0}(y_0, \boldsymbol{z}, r) - 1 = \sum_{\ell=1}^\infty \xi_{d,\ell}(y_0, r) Q_\ell(\langle \boldsymbol{w}, \boldsymbol{z}\rangle), \qquad \xi_{d,\ell}(y_0, r) = \mathbb{E}_{\nu_d}[Q_\ell(Z)|Y = y_0, R = r].$$

Similarly, we can expand $\phi \in L^2(\mathbb{P}_0)$ as

$$\phi(y_0, \boldsymbol{z}, r) = \sum_{\ell=0}^\infty \sum_{s\in[n_{d,\ell}]} \alpha_{\ell s}(y_0, r) Y_{\ell s}(\boldsymbol{z}),$$

where $\{Y_{\ell s}\}_{\ell\geq 0, s\in[n_{d,\ell}]}$ is an orthonormal basis of spherical harmonics in $L^2(\mathbb{S}^{d-1})$. Using the identity $Q_\ell(\langle \boldsymbol{w}, \boldsymbol{z}\rangle) = n_{d,\ell}^{-1/2}\sum_{s\in[n_{d,\ell}]} Y_{\ell s}(\boldsymbol{w})Y_{\ell s}(\boldsymbol{z})$, we obtain the decomposition

$$\Delta_\phi(\boldsymbol{w}) = \sum_{\ell=1}^{\infty}\sum_{s\in[n_{d,\ell}]} Y_{\ell s}(\boldsymbol{w})\frac{\mathbb{E}_{\mathbb{P}_0}[\xi_{d,\ell}(y_0,r)\alpha_{\ell s}(y_0,r)]}{\sqrt{n_{d,\ell}}},$$

and thus,

$$\mathbb{E}_{\boldsymbol{w}}[|\Delta_\phi(\boldsymbol{w})|^2] = \sum_{\ell=1}^{\infty}\sum_{s\in[n_{d,\ell}]}\frac{\mathbb{E}_{\mathbb{P}_0}[\xi_{d,\ell}(y_0,r)\alpha_{\ell s}(y_0,r)]^2}{n_{d,\ell}}.$$

Denote $\mathsf{P}_\ell\phi = \sum_{s\in[n_{d,\ell}]}\alpha_{\ell s}Y_{\ell s}$ the projection on the degree-ℓ harmonics. We can decompose the supremum over $\phi\in\mathcal{Q}_{\mathsf{SQ}}$ as

$$\begin{aligned}
&\sup_{\phi\in\mathcal{Q}_{\mathsf{SQ}}}\frac{\mathbb{E}_{\boldsymbol{w}}[|\Delta_\phi(\boldsymbol{w})|^2]}{\mathrm{Var}_{\mathbb{P}_0}(\phi)}\\
&= \sup_{\phi\in\mathcal{Q}_{\mathsf{SQ}}}\frac{1}{\sum_{\ell\geq 1}\|\mathsf{P}_\ell\phi\|_{L^2}^2}\sum_{\ell\geq 1}\frac{\|\mathsf{P}_\ell\phi\|_{L^2}^2}{n_{d,\ell}}\left[\sum_{s\in[n_{d,\ell}]}\frac{\mathbb{E}_{\mathbb{P}_0}[\xi_{d,\ell}(y_0,r)\alpha_{\ell s}(y_0,r)]^2}{\|\mathsf{P}_\ell\phi\|_{L^2}^2}\right]\\
&= \sup_{\phi\in\mathcal{Q}_{\mathsf{SQ}}}\frac{1}{\sum_{\ell\geq 1}\|\mathsf{P}_\ell\phi\|_{L^2}^2}\sum_{\ell\geq 1}\frac{\|\mathsf{P}_\ell\phi\|_{L^2}^2}{n_{d,\ell}}\left[\sup_{\psi\in L^2(\nu_{d,Y,R})}\frac{\langle\xi_{d,\ell},\psi\rangle_{L^2}^2}{\|\psi\|_{L^2}^2}\right]\\
&= \sup_{\phi\in\mathcal{Q}_{\mathsf{SQ}}}\frac{\sum_{\ell\geq 1}\|\mathsf{P}_\ell\phi\|_{L^2}^2\frac{\|\xi_{d,\ell}\|_{L^2}^2}{n_{d,\ell}}}{\sum_{\ell\geq 1}\|\mathsf{P}_\ell\phi\|_{L^2}^2} = \sup_{\ell\geq 1}\frac{\|\xi_{d,\ell}\|_{L^2}^2}{n_{d,\ell}},
\end{aligned}$$

which concludes the proof of this lemma.

C.2 Low-Degree Polynomial lower bounds

We now consider sample complexity lower bounds within the Low-Degree Polynomial (LDP) framework—another powerful tool for studying computational hardness in statistical inference problems. We refer to [24, 32, 40] for background.

Below we follow the presentation of [15]. The planted distribution with m samples is generated by first drawing $\boldsymbol{w}\sim\tau_d$ (uniformly at random on the sphere), then sampling m points $(y_i,\boldsymbol{x}_i)\sim_{iid}\mathbb{P}_{\nu_d,\boldsymbol{w}_*}$. The null distribution corresponds to $(y_i,\boldsymbol{x}_i)\sim_{iid}\mathbb{P}_{\nu_d,0}$. The likelihood ratio in this model is given by

$$\mathcal{R}((y_i,\boldsymbol{x}_i)_{i\in[m]}) = \mathbb{E}_{\boldsymbol{w}}\left[\prod_{i\in[m]}\frac{\mathrm{d}\mathbb{P}_{\nu_d,\boldsymbol{w}}}{\mathrm{d}\mathbb{P}_{\nu_d,0}}(y_i,\boldsymbol{x}_i)\right].$$

We consider the orthogonal projection $\mathcal{P}_{\leq D}$ (in $L^2(\mathbb{P}_{\nu_d,0}^{\otimes m})$) onto degree at most D polynomial in $\boldsymbol{z}_i$, that is, we allow arbitrary degree on the scalars (y_i,r_i). We denote

$$\mathcal{R}_{\leq D}((y_i,\boldsymbol{x}_i)_{i\in[m]}) = \mathcal{P}_{\leq D}\mathcal{R}((y_i,\boldsymbol{x}_i)_{i\in[m]}). \tag{50}$$

Informally, the low-degree conjecture [24] states that for $D=\omega_d(\log d)$:

- *Weak detection hardness:* If $\|\mathcal{R}_{\leq D}\|_{L^2}^2 = 1+o_d(1)$, then no polynomial time algorithm can distinguish between $\mathbb{E}_{\boldsymbol{w}}[\mathbb{P}_{\nu_d,\boldsymbol{w}}^{\otimes m}]$ and $\mathbb{P}_{\nu_d,0}^{\otimes m}$ with non-vanishing probability.
- *Strong detection hardness:* If $\|\mathcal{R}_{\leq D}\|_{L^2}^2 = O_d(1)$, then polynomial time algorithm can only succeed with constant probability.

Below, we state our results for weak and strong detection for a sequence of spherical SIMs $\{\nu_d\}_{d\geq 1}$ with $\nu_d\in\mathfrak{L}_d$. Recall that we defined:

$$\mathsf{M}_\star(\nu_d) = \inf_{\ell\geq 1}\frac{\sqrt{n_{d,\ell}}}{\|\xi_{d,\ell}\|_{L^2}^2}$$

Without loss of generality, we will assume that $\mathsf{M}_\star(\nu_d) = O_d(\mathrm{poly}(d))$—that is, the model can be solved in polynomial time—as stated in the following assumption:

Assumption 3. *There exists $p \in \mathbb{N}$ such that the sequence $\{\nu_d\}_{d\geq 1}$ satisfies $\mathsf{M}_\star(\nu_d) = O_d(d^{p/2})$.*

We can now state our bound on the low-degree projection of the likelihood ratio in this problem.

Theorem 3. *Let $\{\nu_d\}_{d\geq 1}$ be a sequence of spherical SIMs $\nu_d \in \mathfrak{L}_d$ satisfying Assumption 3 for some integer $p \in \mathbb{N}$. Consider the detection task with m samples as defined above. There exists a constant $c > 0$ that only depends on the constants in Assumption 3 such that if $D \leq c d^{2/(p+4)}$, then*

$$\|\mathcal{R}_{\leq D}\|_{L^2}^2 - 1 \leq \sum_{s=1}^{D} \left(m \frac{D^{p/2-1}}{\mathsf{M}_\star(\nu_d)} [e(p+1)] \right)^s . \tag{51}$$

In particular,

(i) (Weak detection.) *If $m = o_d\left(\frac{\mathsf{M}_\star(\nu_d)}{D^{p/2-1}}\right)$, then $\|\mathcal{R}_{\leq D}\|_{L^2}^2 = 1 + o_d(1)$.*

(ii) (Strong detection.) *If $m = O_d\left(\frac{\mathsf{M}_\star(\nu_d)}{D^{p/2-1}}\right)$, then $\|\mathcal{R}_{\leq D}\|_{L^2}^2 = O_d(1)$.*

The proof of this theorem can be found in Section C.2.1 below.

Combining this theorem with the low-degree conjecture stated above, we conclude that no-polynomial time algorithm can detect (and thus, estimate) the spherical single-index model $\mathbb{P}_{\nu_d,\boldsymbol{w}}$ unless

$$m \gtrsim M_\star(\nu_d).$$

We further remark that we recover the tight threshold $\mathsf{M}_\star(\nu_d)/D^{p/2-1}$ from [15]. Indeed, consider the case of Gaussian SIM with information exponent $\mathsf{k}_\star$. We can set $p = \mathsf{k}_\star$, and our bound recover the (conjectured) optimal computational-statistical trade-off $d^{\mathsf{k}_\star/2}/D^{\mathsf{k}_\star/2-1}$ from [15], which matches the optimal known trade-off in tensor PCA [47].

Decoupling across harmonic subspaces. Again, we provide an interpretation of this lower bound as the optimal lower bound among subproblems indexed by $\ell \geq 1$. For each $\ell \geq 1$, we consider the task of detecting single-index models only using degree-ℓ spherical harmonics. We consider only using polynomials that are product of degree-ℓ spherical harmonics in $\boldsymbol{z}_i$. Denote $\mathcal{P}_{\leq D,\ell}$ the projection onto this subspace, that is

$$\mathcal{P}_{\leq D,\ell}\mathcal{R}((y_i,\boldsymbol{x}_i)_{i\in[m]}) = \sum_{S\subset[m],|S|\leq\lfloor D/\ell\rfloor} \mathbb{E}_{\boldsymbol{w}}\left[\prod_{i\in S} \xi_{d,\ell}(y_i,r_i) Q_\ell(\langle \boldsymbol{w}, \boldsymbol{z}_i\rangle)\right].$$

Then we have the following upper bound on the norm of this projected likelihood ratio.

Corollary 4. *Let $\{\nu_d\}_{d\geq 1}$ be a sequence of spherical SIMs $\nu_d \in \mathfrak{L}_d$ satisfying Assumption 3 for some integer $p \in \mathbb{N}$. Consider the detection task with m samples as defined above and fix $\ell_* \in \mathbb{N}$. Then for all $D \geq 1$, we have*

$$\|\mathcal{R}_{\leq D}\|_{L^2}^2 - 1 \leq \sum_{s=1}^{\lfloor D/\ell_*\rfloor} \left(m \frac{e D^{\ell_*/2-1} \|\xi_{d,\ell_*}\|_{L^2}^2}{\sqrt{n_{d,\ell_*}}} \right)^s . \tag{52}$$

By analogy with the low-degree conjecture, we expect that no polynomial-time algorithm only using degree-ℓ spherical harmonics will succeed at the detection task, unless

$$m \gtrsim \frac{\sqrt{n_{d,\ell_*}}}{\|\xi_{d,\ell_*}\|_{L^2}^2}. \tag{53}$$

It would be interesting to make this subspace-restricted low-degree polynomial statement more formal, and we leave it to future work. Our harmonic tensor unfolding estimator matches this heuristic sample lower bound (53) for each $\ell_* \geq 3$.

C.2.1 Proof of Theorem 3

Recalling the expansion of the likelihood function into Gegenbauer polynomials, we can write

$$\begin{aligned}\mathcal{R}_{\leq D}((y_i, r_i, \boldsymbol{z}_i)_{i\in[m]}) &= \mathcal{P}_{\leq D}\mathbb{E}_{\boldsymbol{w}}\left[\prod_{i\in[m]} \frac{\mathrm{d}\mathbb{P}_{\nu_d,\boldsymbol{w}}}{\mathrm{d}\mathbb{P}_{\nu_d,0}}(y_i, r_i, \boldsymbol{z}_i)\right] \\ &= \sum_{\ell_1+\ldots+\ell_m\leq D} \mathbb{E}_{\boldsymbol{w}}\left[\prod_{i\in[m]} \xi_{d,\ell_i}(y_i, r_i) Q_{\ell_i}(\langle \boldsymbol{w}, \boldsymbol{z}_i\rangle)\right].\end{aligned}$$

The norm of this projection with respect to $\mathbb{P}_0^{\otimes m}$ is then given by

$$\|\mathcal{R}_{\leq D}\|_{L^2}^2 = \sum_{\ell_1+\ldots+\ell_m\leq D} \mathbb{E}_{\boldsymbol{w},\boldsymbol{w}'}\left[\prod_{i\in[m]} \frac{\|\xi_{d,\ell_i}\|_{L^2}^2}{\sqrt{n_{d,\ell_i}}} Q_{\ell_i}(\langle \boldsymbol{w}, \boldsymbol{w}'\rangle)\right], \tag{54}$$

where we used

$$\mathbb{E}_{\boldsymbol{z}}[Q_\ell(\langle \boldsymbol{w}, \boldsymbol{z}\rangle)Q_k(\langle \boldsymbol{z}, \boldsymbol{w}'\rangle)] = \frac{\delta_{\ell=k'}}{\sqrt{n_{d,\ell}}} Q_\ell(\langle \boldsymbol{w}, \boldsymbol{w}'\rangle).$$

Let's separate the zero degrees from the non-zero degrees in (54):

$$\|\mathcal{R}_{\leq D}\|_{L^2}^2 - 1 = \sum_{s=1}^{D} \binom{m}{s} \sum_{\substack{1\leq \ell_1,\ldots,\ell_s\leq D\\ \ell_1+\ldots+\ell_s\leq D}} \mathbb{E}_{\boldsymbol{w},\boldsymbol{w}'}\left[\prod_{i\in[s]} \frac{\|\xi_{d,\ell_i}\|_{L^2}^2}{\sqrt{n_{d,\ell_i}}} Q_{\ell_i}(\langle \boldsymbol{w}, \boldsymbol{w}'\rangle)\right].$$

To upper bound the expectation, we will not be careful and simply use Hölder's inequality and the hypercontractivity (Lemma 25) of Gegenbauer polynomials,

$$\mathbb{E}_{\boldsymbol{w},\boldsymbol{w}'}\left[\prod_{i\in[s]} Q_{\ell_i}(\langle \boldsymbol{w}, \boldsymbol{w}'\rangle)\right] \leq \prod_{i\in[s]} \|Q_{\ell_i}\|_{L^s} \leq \prod_{i\in[s]} s^{\ell_i/2}.$$

We obtain the upper bound

$$\|\mathcal{R}_{\leq D}\|_{L^2}^2 - 1 \leq \sum_{s=1}^{D} \binom{m}{s} \sum_{\substack{1\leq \ell_1,\ldots,\ell_s\leq D\\ \ell_1+\ldots+\ell_s\leq D}} \prod_{i\in[s]} s^{\ell_i/2}\frac{\|\xi_{d,\ell_i}\|_{L^2}^2}{\sqrt{n_{d,\ell_i}}} \leq \sum_{s=1}^{D} \binom{m}{s} \rho(s, D)^s,$$

where

$$\rho(s, D) = \sum_{\ell=1}^{D} s^{\ell/2}\frac{\|\xi_{d,\ell}\|_{L^2}^2}{\sqrt{n_{d,\ell}}}.$$

Let's upper bound $\rho(s, D)$. By definition $\|\xi_{d,\ell}\|_{L^2}^2/n_{d,\ell} \leq 1/\mathsf{M}_\star$ for all $\ell \geq 1$ (where we denote $\mathsf{M}_\star = \mathsf{M}_\star(\nu_d)$ for simplicity). Furthermore, by Assumption 3, we have $\mathsf{M}_\star \leq C d^{p/2}$. Using $\|\xi_{d,\ell}\|_{L^2} \leq 1$, we deduce a second upper bound

$$\frac{\|\xi_{d,\ell}\|_{L^2}^2}{\sqrt{n_{d,\ell}}} \leq C\frac{d^{p/2}}{\mathsf{M}_\star\sqrt{n_{d,\ell}}}.$$

Separating $\ell \leq p$ and $\ell > p$, we get

$$\rho(s, D) \leq \sum_{\ell=1}^{p} \frac{s^{\ell/2}}{\mathsf{M}_\star} + C\sum_{\ell=p+1}^{D} \frac{s^{\ell/2}d^{p/2}}{\mathsf{M}_\star\sqrt{n_{d,\ell}}} \leq \frac{s^{p/2}}{\mathsf{M}_*}\left[p + C\sum_{\ell=p+1}^{D} \frac{D^{(\ell-p)/2}d^{p/2}}{\sqrt{n_{d,\ell}}}\right].$$

Using that for $\ell \leq D \leq \sqrt{d}$, we have $n_{d,\ell} \geq c\binom{d}{\ell} \geq c(d/\ell)^\ell$ for some constant $c > 0$, the sum simplifies to

$$\sum_{\ell=p+1}^{D} \frac{D^{(\ell-p)/2}d^{p/2}}{\sqrt{n_{d,\ell}}} \leq C'\sum_{\ell=p+1}^{d} \frac{D^{(\ell-p)/2}d^{p/2}\ell^{\ell/2}}{d^{\ell/2}} \leq C' D^{p/2}\sum_{\ell=1}^{\infty}\left(\frac{D^2}{d}\right)^\ell \leq C''\frac{D^{p/2+2}}{d}.$$

Assuming that $D \le d^{2/(p+4)}/\tilde{C}$, we deduce that

$$\rho(s, D) \le \frac{s^{p/2}}{\mathsf{M}_\star}(p+1).$$

Thus, we obtain

$$\|\mathcal{R}_{\le D}\|_{L^2}^2 - 1 \le \sum_{s=1}^{D} \binom{m}{s} \frac{s^{sp/2}}{\mathsf{M}_\star^s}(p+1)^s \le \sum_{s=1}^{D} \left(\frac{mD^{p/2-1}}{\mathsf{M}_\star}[e(p+1)] \right)^s,$$

which concludes the proof of Theorem 3.

Restricted projection. Consider the likelihood ratio projected on degree-ℓ_* spherical harmonics. The proof is particularly simple in this case:

$$\begin{aligned}
\|\mathcal{R}_{\le D,\ell_*}\|_{L^2}^2 - 1 &= \sum_{s=1}^{\lfloor D/\ell_* \rfloor} \binom{m}{s} \left(\frac{\|\xi_{d,\ell_*}\|_{L^2}^2}{\sqrt{n_{d,\ell_*}}} \right)^s \mathbb{E}_{\boldsymbol{w},\boldsymbol{w}'} \left[Q_{\ell_*}(\langle \boldsymbol{w}, \boldsymbol{w}' \rangle)^s \right] \\
&\le \sum_{s=1}^{\lfloor D/\ell_* \rfloor} \left(\frac{em}{s} \right)^s \left(\frac{\|\xi_{d,\ell_*}\|_{L^2}^2}{\sqrt{n_{d,\ell_*}}} \right)^s s^{s\ell_*/2} \\
&\le \sum_{s=1}^{\lfloor D/\ell_* \rfloor} \left(m \frac{eD^{\ell_*/2-1}\|\xi_{d,\ell_*}\|_{L^2}^2}{\sqrt{n_{d,\ell_*}}} \right)^s.
\end{aligned}$$

D Spectral estimators

In this section, we provide the spectral algorithm for Theorem 2 (part 1).

Requirement 1. *We are going to implement our algorithms on $\mathcal{T}_\ell$ satisfying the following criteria.*

1. $\|\mathcal{T}_\ell\|_2 = 1$ *and* $\mathbb{E}_{(y,r,z)\sim\nu_d}[\mathcal{T}_\ell(y,r)Q_\ell(z)] := \beta_{d,\ell} > 0$ *(w.l.o.g.).*
2. *There exits* $\kappa_\ell > 1$, $k \in \mathbb{N}$, *such that, for any* $p \ge 3$, *we have* $\|\mathcal{T}_\ell\|_p \le \kappa_\ell \, p^{k/2}$

Note that a transformation $\mathcal{T}_\ell$ satisfying Assumption 2 is a special case of this requirement with $k = 0$ and $\beta_{d,\ell} \ge \|\xi_{d,\ell}\|_{L^2}/\kappa_\ell$, and thus the theorem will follow by invoking the guarantee for this more general $\mathcal{T}_\ell$ satisfying Requirement 1. We first specify the spectral algorithm.

Algorithm 2: A spectral algorithm on the frequency $\ell = 1$ and $\ell = 2$.

Input : An example set $S = \{(\boldsymbol{x}_i, y_i) : i \in [m]\} \sim_{iid} \mathbb{P}_{\boldsymbol{w}_*}$, the frequency $\ell \in \{1, 2\}$, and a transformation $\mathcal{T}_\ell$.

Output : An estimator $\hat{\boldsymbol{w}} \in \mathbb{R}^d$.

Decompose $\boldsymbol{x}_i = (r_i, \boldsymbol{z}_i)$.

if $\ell = 1$ **then**

  Let $\hat{\boldsymbol{v}}_m := \frac{1}{m}\sum_{i\in[m]} \mathcal{T}_\ell(y_i, r_i)\sqrt{d}\,\boldsymbol{z}_i$.

  $\hat{\boldsymbol{w}} = \frac{\hat{\boldsymbol{v}}_m}{\|\hat{\boldsymbol{v}_m}\|_2}$

end

if $\ell = 2$ **then**

  Let $\boldsymbol{M}_m = \frac{1}{m}\sum_{i=1}^m \mathcal{T}_\ell(y_i, r_i)(d\boldsymbol{z}_i\boldsymbol{z}_i^\mathsf{T} - \mathbf{I}_d)$.

  Let $\hat{\boldsymbol{w}} = \boldsymbol{v}_1(\boldsymbol{M}_m)$ be the eigenvector associated with the highest magnitude eigenvalue .

end

Return $\hat{\boldsymbol{w}}$.

D.1 Analysis of $\ell = 2$

Let $\boldsymbol{M}^* := \mathbb{E}[\boldsymbol{M}_m]$ and $(\lambda_i^*, \boldsymbol{v}_i^*)_{i\in[d]}$ be eigenpairs of $\boldsymbol{w}^*$ such that $|\lambda_1^*| \ge \cdots \ge |\lambda_d^*|$. We first show that the top eigenvector $\boldsymbol{v}_1^* = \boldsymbol{w}_*$ with $\lambda_1^* = (1 + o_d(1))\beta_{d,2}$ and the other eigenvalues are of vanishing order relative to λ_1^*.

Lemma 5. *We have that $\boldsymbol{M}^*$ has top eigenvalue $\lambda_1^* = (1 + o_d(1))\, \beta_{d,2}$ with $\boldsymbol{v}_1(\boldsymbol{M}^*) = \boldsymbol{w}_*$ and for any $2 \le i \le d$, we have $|\lambda_i^*| \lesssim \frac{\lambda_1^*}{d}$.*

Proof. We have that $\mathbb{E}_{(y,r,\boldsymbol{z})\sim \mathsf{P}_{\boldsymbol{w}_*}}[\mathcal{T}(y,r) \mid \boldsymbol{z}] = \sum_{\ell=0}^{\infty} \beta_{d,\ell} Q_\ell(\langle \boldsymbol{w}_*, \boldsymbol{z}\rangle)$. We now analyze $\boldsymbol{M}^*$ through its quadratic form: for any $\boldsymbol{w} \in \mathbb{S}^{d-1}$, consider

$$\begin{aligned}
\boldsymbol{w}^\mathsf{T} \boldsymbol{M}^* \boldsymbol{w} &= \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}_{(y_i,r_i,\boldsymbol{z}_i)\sim \mathbb{P}_{\boldsymbol{w}_*}}[\mathcal{T}(y_i,r_i) d\boldsymbol{w}^\mathsf{T}(\boldsymbol{z}_i\boldsymbol{z}_i^\mathsf{T} - \mathbf{I}_d)\boldsymbol{w}] = \mathbb{E}_{(y,r,\boldsymbol{z})\sim \mathbb{P}_{\boldsymbol{w}_*}}[\mathcal{T}(y,r)(d\langle \boldsymbol{w}, \boldsymbol{z}\rangle^2 - 1)] \\
&= (1+o_d(1))\mathbb{E}_{(y,r,\boldsymbol{z})\sim \mathbb{P}_{\boldsymbol{w}_*}}[\mathcal{T}(y,r) Q_2^{(d)}(\langle \boldsymbol{w}, \boldsymbol{z}\rangle)] \\
&= (1+o_d(1))\beta_{d,2}\frac{Q_2^{(d)}(\langle \boldsymbol{w}, \boldsymbol{w}_*\rangle)}{\sqrt{n_{d,2}}}
\end{aligned}$$

where we used the fact that $Q_2^{(d)}(z) = (1+o_d(1))(dz^2-1)$. As $Q_2^{(d)}(\cdot)$ has its maximum value at 1. Clearly, the $\boldsymbol{w}^\mathsf{T}\boldsymbol{M}^*\boldsymbol{w}$ is maximized for $\boldsymbol{w} = \boldsymbol{w}_*$, and thus it is an eigenvector with the eigenvalue

$$\lambda_1^* = (1+o_d(1))\beta_{d,2} Q_2^{(d)}(1)/\sqrt{n_{d,2}} = (1+o_d(1))\beta_{d,2}\,.$$

It suffices to show that the other eigenvalues are of much lower magnitude. For any $\boldsymbol{w} \perp \boldsymbol{w}_*$, we have $Q_2^{(d)}(\langle \boldsymbol{w}, \boldsymbol{w}_*\rangle) = Q_2^{(d)}(0) = (1+o_d(1))(-1)$, and thus, for $2 \le i \le d$, we have

$$|\lambda_i^*| = (1+o_d(1))\frac{\beta_{d,2}}{\sqrt{n_{d,2}}} \lesssim \frac{\lambda_1^*}{d}\,.$$

$\square$

Our goal is to ensure that the top eigenvectors $\hat{\boldsymbol{w}} = \boldsymbol{v}_1(\boldsymbol{M}_m)$ and $\boldsymbol{w}_* = \boldsymbol{v}_1(\boldsymbol{M}^*)$ are close to each other, when m is chose sufficiently large. By Davis-Kahan's theorem, it suffices to show the concentration between the empirically estimated matrix $\boldsymbol{M}_m$, and its expectation, in the following sense.

Lemma 6. *There exists a constant $C > 0$ (only depending on k) such that, for any $\delta > 0$, with*

$$m \ge C\frac{\kappa_\ell\, d}{\beta_{d,2}^2}\left(1 + \beta_{d,2}\log^{k/2+2}\left(\frac{d}{\delta\beta_{d,2}^2}\right)\right)$$

we have that with probability $1-\delta$,

$$\|\boldsymbol{M}_m - \boldsymbol{M}^*\|_{\mathrm{op}} \le \frac{\lambda_1^*}{8}\ ,$$

where recall that λ_1^* is the top eigenvalue of $\boldsymbol{M}^*$ (see Lemma 5).

Proof. Our goal is to use Lemma 28, with $\boldsymbol{Y}_i = \frac{1}{m}\left(\mathcal{T}(y_i,r_i)(d\boldsymbol{z}_i\boldsymbol{z}_i^\mathsf{T} - \mathbf{I}_d) - \boldsymbol{M}^*\right) \in \mathbb{R}^{d\times d}$ which are zero mean, and thus, $\boldsymbol{Y} = \boldsymbol{M}_m - \boldsymbol{M}^*$. Let us bound the each quantity of interest

$$\begin{aligned}
\sigma^2 = \|\mathbb{E}[(\boldsymbol{M}_m - \boldsymbol{M}^*)^2\|_2 &= \frac{1}{m}\|\mathbb{E}[(\boldsymbol{M}_1 - \boldsymbol{M}^*)^2\|_2 \le \frac{2}{m}\|\mathbb{E}[\boldsymbol{M}_1^2]\|_2 \\
&\le \frac{2}{m}\sup_{\boldsymbol{w}\in\mathbb{S}^{d-1}} \boldsymbol{w}^\mathsf{T}\mathbb{E}[\boldsymbol{M}_1^2]\boldsymbol{w} \\
&= \frac{2}{m}\sup_{\boldsymbol{w}\in\mathbb{S}^{d-1}} \mathbb{E}_{(y,r,\boldsymbol{z})\sim \mathbb{P}_{\boldsymbol{w}_*}} \boldsymbol{w}^\mathsf{T}\left[\mathcal{T}(y,r)^2\left((d^2-2d)\boldsymbol{z}\boldsymbol{z}^\mathsf{T} + \mathbf{I}_d\right)\right]\boldsymbol{w} \\
&= \frac{2}{m}\sup_{\boldsymbol{w}\in\mathbb{S}^{d-1}} \mathbb{E}_{(y,r,\boldsymbol{z})\sim \mathbb{P}_{\boldsymbol{w}_*}}\left[\mathcal{T}(y,r)^2((d^2-2d)\langle \boldsymbol{w}, \boldsymbol{z}\rangle^2 + 1)\right]
\end{aligned}$$

We now note that $g(\boldsymbol{z}) = (d^2-2d)\langle \boldsymbol{w}, \boldsymbol{z}\rangle^2 + 1$ is a polynomial of degree 2 with $\|g(\boldsymbol{z})\|_{L^2(\tau_d)} \lesssim d$. Therefore using spherical hypercontractivity (Lemma 25), we have $\|g(\boldsymbol{z})\|_{L^p(\tau_d)} \lesssim p\cdot d$. Applying Lemma 27 then gives us

$$\mathbb{E}_{(y,r,\boldsymbol{z})\sim \mathbb{P}_{\boldsymbol{w}_*}}\left[\mathcal{T}(y,r)^2\cdot((d^2-2d)\cdot\langle \boldsymbol{w}, \boldsymbol{z}\rangle^2 + 1)\right] \lesssim d\cdot\|\mathcal{T}\|_2^2\cdot\max\left(1, \log\left(\frac{\|\mathcal{T}\|_4}{\|\mathcal{T}\|_2}\right)\right) \lesssim d\log(\kappa_\ell)$$

where we used the fact that $\|\mathcal{T}\|_2^2 = 1$ and $\|\mathcal{T}\|_4 \lesssim \log(\kappa_\ell)$. We finally obtain that

$$\sigma \lesssim \sqrt{\frac{d \log \kappa_\ell}{m}}\,.$$

We next analyze the other variance term using similar idea:

$$\begin{aligned}\sigma_*^2 &:= \sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{d-1}} \mathbb{E}\left[(\boldsymbol{u}^\mathsf{T}(\boldsymbol{M}_m - \boldsymbol{M}^*)\boldsymbol{v})^2\right] = \frac{1}{m}\sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{d-1}} \mathbb{E}\left[(\boldsymbol{u}^\mathsf{T}(\boldsymbol{M}_1 - \boldsymbol{M})\boldsymbol{v})^2\right] \le \frac{4}{m}\sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{d-1}} \mathbb{E}[(\boldsymbol{u}^\mathsf{T}\boldsymbol{M}_1\boldsymbol{v})^2] \\ &\le \frac{4}{m}\sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{d-1}} \mathbb{E}[\mathcal{T}(y,r)^2(d\langle \boldsymbol{u},\boldsymbol{z}\rangle\langle \boldsymbol{v},\boldsymbol{z}\rangle - \langle \boldsymbol{u},\boldsymbol{v}\rangle)^2]\end{aligned}$$

We would like to use Lemma 27 to bound the expectation, for which we will first obtain a tight bound on all the moments of $g(\boldsymbol{z}) := (d\langle \boldsymbol{u},\boldsymbol{z}\rangle\langle \boldsymbol{v},\boldsymbol{z}\rangle - \langle \boldsymbol{u},\boldsymbol{v}\rangle)^2$. Let us compute

$$\begin{aligned}\|g\|_{L^2(\tau_d)} &\lesssim \mathbb{E}[d^4\langle \boldsymbol{u},\boldsymbol{z}\rangle^4\langle \boldsymbol{v},\boldsymbol{z}\rangle^4 + 1]^{1/2} = (d^4\mathbb{E}[\langle \boldsymbol{u},\boldsymbol{z}\rangle^4\langle \boldsymbol{v},\boldsymbol{z}\rangle^4] + 1)^{1/2} \\ &\le (d^4\sqrt{\mathbb{E}[\langle \boldsymbol{u},\boldsymbol{z}\rangle^8]\mathbb{E}[\langle \boldsymbol{v},\boldsymbol{z}\rangle^8]} + 1)^{1/2} = (d^4\cdot\mathbb{E}[z_1^8] + 1)^{1/2} \lesssim 1\,.\end{aligned}$$

In the above, we used the Cauchy-Schwarz inequality, rotational invariance of τ_d, and $\mathbb{E}[z_1^8] \lesssim 1/d^4$ respectively. Using hypercontractivity (Lemma 25), we have $\|g\|_{L^p(\tau_d)} \lesssim (p-1)^2$. We now use Lemma 27 to conclude that

$$\sigma_* \lesssim \sqrt{\frac{1}{m}\sup_{\boldsymbol{u},\boldsymbol{v}\in\mathbb{S}^{d-1}} \mathbb{E}[\mathcal{T}(y,r)^2(d\langle \boldsymbol{u},\boldsymbol{z}\rangle\langle \boldsymbol{v},\boldsymbol{z}\rangle - \langle \boldsymbol{u},\boldsymbol{v}\rangle)^2]} \lesssim \sqrt{\frac{\|\mathcal{T}\|_2^2\cdot\max\left(1,\log(\frac{\|\mathcal{T}\|_4}{\|\mathcal{T}\|_2})\right)}{m}} \lesssim \sqrt{\frac{\log(\kappa_\ell)}{m}}\,,$$

again we used that $\|\mathcal{T}\|_4 \lesssim \kappa_\ell$. Our next goal is to compute $\bar{R} = \mathbb{E}\left[\max_{i\in[n]}\|\boldsymbol{Y}_i\|_2^2\right]^{1/2}$, where $\boldsymbol{Y}_i = \frac{1}{m}(\mathcal{T}(y_i,r_i)d\boldsymbol{z}_i\boldsymbol{z}_i^\mathsf{T} - \boldsymbol{M}^*)$. And thus, $\|\boldsymbol{Y}_i\|_2 \le \frac{1}{m}(d|\mathcal{T}(y_i,r_i)| + \|\boldsymbol{M}^*\|_2) \lesssim \frac{1}{m}(d|\mathcal{T}(y_i,r_i)| + \beta_{d,2})$, using Lemma 5. For any $p \ge 3$, bounding the p^{th} moment

$$\mathbb{E}[\|\boldsymbol{Y}_i\|_2^p]^{1/p} \lesssim \frac{1}{m}\left(d\|\mathcal{T}\|_p + \beta_{d,2}\right) \lesssim \frac{d\,\kappa_\ell\,p^{k/2}}{m}\,,$$

Using Lemma 29, we have

$$\bar{R} = \mathbb{E}\left[\max_{i\in[m]}\|\boldsymbol{Y}_i\|_2^2\right]^{1/2} \lesssim \frac{d\cdot\kappa_\ell\,\log^{k/2} m}{m}\,.$$

The threshold for choosing R is:

$$\sigma^{1/2}\bar{R}^{1/2} + \sqrt{2}\bar{R} \lesssim \sqrt{\left(\frac{d\log\kappa_\ell}{m}\right)^{1/2}\cdot\frac{d\,\kappa_\ell\,\log^{k/2} m}{m}} + \frac{d\,\kappa_\ell\,\log^{k/2} m}{m} \lesssim \frac{d\,\kappa_\ell\,\log^{k/2} m}{m}\,,$$

where in the last step, we used the fact that we are in the regime $m \ge d$, only keeping the dominant term. Therefore, for any $\delta \ge 0$ we can choose some R that satisfies

$$R \lesssim \frac{d\,\kappa_\ell\log^{k/2}(m/\delta)}{m} \qquad\text{and}\qquad \mathbb{P}(\max_{i\in[m]}\|\boldsymbol{Y}_i\|_2 \ge R) \le \delta/2\,,$$

where in the last inequality, we used Lemma 29. Finally, we apply Lemma 28. With probability $1 - \delta/2 - de^{-t}$, we have

$$\|\boldsymbol{M}_m - \boldsymbol{M}^*\|_{\text{op}} \lesssim \sqrt{\frac{d\log(\kappa_\ell)}{m}} + t^{1/2}\sqrt{\frac{\log\kappa_\ell}{m}} + \left(\frac{d\,\kappa_\ell\log^{k/2}(m/\delta)}{m}\cdot\frac{d\log\kappa_\ell}{m}\right)^{1/3} t^{2/3} + \frac{d\kappa_\ell\log^{k/2}(m/\delta)}{m}t\,.$$

Choosing $t = \log(2d/\delta)$, we obtain with probability $1-\delta$,

$$\begin{aligned}\|\boldsymbol{M}_m - \boldsymbol{M}^*\|_{\text{op}} \lesssim{}& \sqrt{\frac{d\log\kappa_\ell}{m}} + \sqrt{\frac{\log(\kappa_\ell)\log(d/\delta)}{m}} + \left(\frac{d^2\kappa_\ell\log\kappa_\ell\,\log^{k/2}(m/\delta)\log^2(d/\delta)}{m^2}\right)^{1/3} \\ &+ \frac{d\kappa_\ell\log^{k/2}(m/\delta)\log(d/\delta)}{m}\,.\end{aligned}$$

Therefore, there exists a constant $C > 0$ such that, for

$$m_0 = \frac{C\,\kappa_\ell\, d}{\beta_{d,2}^2}\left(1 + \beta_{d,2}\log^{k/2+2}\left(\frac{d}{\delta\,\beta_{d,2}^2}\right)\right)$$

any $m \geq m_0$, with probability $1-\delta$,

$$\|\boldsymbol{M}_m - \boldsymbol{M}^*\|_{\mathrm{op}} \leq \frac{\beta_{d,2}}{16} \leq \frac{\lambda_1^*}{8}\,.$$

□

Proof of Theorem 2: Spectral algorithm, case $\ell = 2$: For any transformation $\mathcal{T}_\ell$ satisfying requirement Requirement 1, we have that choosing m sufficiently large that is

$$\mathsf{m} \leq \frac{C\,\kappa_\ell\, d}{\beta_{d,2}^2}\left(1 + \beta_{d,2}\log^{k/2+2}\left(\frac{d}{\delta\,\beta_{d,2}^2}\right)\right)$$

by Davis and Kahn's theorem, we have

$$\min_{s\in\{\pm 1\}}\|s\boldsymbol{v}_1(\boldsymbol{M}_{\mathsf{m}}) - \boldsymbol{v}_1(\boldsymbol{M}^*)\|_2 \leq \frac{\|\boldsymbol{M}_{\mathsf{m}} - \boldsymbol{M}^*\|_{\mathrm{op}}}{|\lambda_1^* - \lambda_2^*|}\,.$$

According to Lemma 5, this corresponds to

$$\min_{s\in\{\pm 1\}}\|s\hat{\boldsymbol{w}} - \boldsymbol{w}_*\|_2 \leq \frac{\lambda_1^*/8}{(1+o(1))\,\lambda_1^*} \leq \frac{1}{4}\,.$$

Rearranging terms, we obtain $|\langle \hat{\boldsymbol{w}}, \boldsymbol{w}_*\rangle| \geq 1/4$. Finally, the above sample complexity bound of m simplifies to the one provided in Theorem 2 under the stronger Assumption 2.

D.2 Analysis for $\ell = 1$

We now analyze the case $\ell = 1$ case, where the analysis is even simpler and follows by concentration of vector to its expected value. For simplicity, we will denote $\mathcal{T} = \mathcal{T}_1$. Let us first evaluate the expectation of the vector statistic $\boldsymbol{v}_m$ computed by the algorithm.

Lemma 7. *We have that* $\mathbb{E}[\hat{\boldsymbol{v}}_m] = \beta_{d,1}\cdot \boldsymbol{w}_*$.

Proof. For any $i \in [d]$,

$$\begin{aligned}
\mathbb{E}[\boldsymbol{v}_m]_i &= \mathbb{E}_{(y,r,\boldsymbol{z})\sim\mathbb{P}_{\boldsymbol{w}_*}}[\mathcal{T}(y,r)Q_1^{(d)}(z_i)] = \mathbb{E}_{(y,r,\boldsymbol{z})\sim\mathbb{P}_{\boldsymbol{w}_*}}[\mathcal{T}(y,r)Q_1^{(d)}(\langle \boldsymbol{z}, \boldsymbol{e}_i\rangle)]\\
&= \sum_{\ell=0}^{\infty}\beta_{d,\ell}\mathbb{E}_{(y,r,\boldsymbol{z})\sim\mathbb{P}_{\boldsymbol{w}_*}}[Q_\ell^{(d)}(\langle \boldsymbol{w}_*, \boldsymbol{z}\rangle)Q_1^{(d)}(\langle \boldsymbol{z}, \boldsymbol{e}_i\rangle)] = \beta_{d,1}Q_1(\langle \boldsymbol{w}_*, \boldsymbol{e}_i\rangle)/\sqrt{n_{d,1}}\\
&= \beta_{d,1}\cdot(\boldsymbol{w}_*)_i\,.
\end{aligned}$$

We conclude that $\mathbb{E}[\boldsymbol{v}_m] = \beta_{d,1}\boldsymbol{w}_*$. □

We now show that for sufficiently large sample size, our final estimator $\hat{\boldsymbol{w}}$ has the desired overlap with the ground-truth $\boldsymbol{w}_*$ via concentration arguments.

Lemma 8. *There exists a universal constant $C > 0$ such that for any $\delta > 0$ and any*

$$m \geq \frac{C\kappa_\ell\sqrt{d}}{\beta_{d,1}^2}\left(1 + \frac{1}{\sqrt{d}}\log^{(k+1)/2}\left(\frac{d\kappa_\ell}{\delta\beta_{d,1}}\right)\right) \quad \text{and} \quad m \geq \frac{C\kappa_\ell d}{\beta_{d,1}^2}\left(1 + \frac{1}{d}\log^{(k+1)/2}\left(\frac{d\kappa_\ell}{\delta\beta_{d,1}}\right)\right),$$

respectively, with probability $1-\delta$, we have

$$\frac{\langle \boldsymbol{v}_m, \boldsymbol{w}_*\rangle}{\|\boldsymbol{v}_m\|_2} \geq \frac{d^{-1/4}}{4} \quad \text{and} \quad \frac{\langle \boldsymbol{v}_m, \boldsymbol{w}_*\rangle}{\|\boldsymbol{v}_m\|_2} \geq \frac{1}{4}\,.$$

1142 *Proof.* Denote $\boldsymbol{v}^* = \mathbb{E}[\boldsymbol{v}_m]$ and define $X_i := \frac{1}{m}(\mathcal{T}(y,r)\sqrt{d}\langle \boldsymbol{z}_i, \boldsymbol{w}_*\rangle - \langle \boldsymbol{v}^*, \boldsymbol{w}_*\rangle)$. Calculating the
1143 variance

$$\sigma^2 = \sum_{i=1}^m \mathbb{E}[X_i^2] \lesssim \frac{1}{m}\mathbb{E}_{(y,r,\boldsymbol{z})\sim\mathbb{P}_{\boldsymbol{w}_*}}[\mathcal{T}(y,r)^2 d\langle \boldsymbol{z}, \boldsymbol{w}_*\rangle^2] \lesssim \frac{\log \kappa_\ell}{m},$$

where in the last inequality we used Lemma 27 and the fact that $\|\mathcal{T}\|_2 = 1$ and $g(\boldsymbol{z}) = d\langle \boldsymbol{w}_*, \boldsymbol{z}\rangle^2$ is a polynomial of degree two with $\|g\|_2 \lesssim 1$. Moreover, for any $p \geq 2$

$$\|X_i\|_p \leq \frac{1}{m}\left(\mathbb{E}[|\mathcal{T}(y_i,r_i)|^p |Q_1(\langle \boldsymbol{w}_*, \boldsymbol{z}_i\rangle)|^p]^{1/p} + \beta_{d,1}\right) \lesssim \frac{(\|\mathcal{T}\|_{2p}\|Q_1\|_{2p} + \beta_{d,1})}{m} \lesssim \frac{\kappa_\ell\, p^{(k+1)/2}}{m},$$

where in the last inequality, we used $\langle \boldsymbol{v}^*, \boldsymbol{w}^*\rangle = \beta_{d,1} \leq 1$ (cf. Lemma 7) and $\|\mathcal{T}\|_{2p} \leq \kappa_\ell\,(2p)^{k/2}$ and $\|Q_1\|_{2p} \leq \sqrt{2p}$ by hypercontractivity (Lemma 25). Applying Lemma 30, with probability $1-\delta$,

$$|\langle \boldsymbol{v}_m - \boldsymbol{v}^*, \boldsymbol{w}_*\rangle| \lesssim \sqrt{\frac{\log(\kappa_\ell)\log(1/\delta)}{m}} + \frac{\kappa_\ell \log(1/\delta)\log^{(k+1)/2}(m/\delta)}{m}.$$

1144 Therefore, there is a constant $C > 0$ such that with any $m \geq C\kappa_\ell\sqrt{d}/\beta_{d,1}^2$, with probability $1 - e^{-d^c}$
1145 for small enough $c > 0$

$$|\langle \boldsymbol{v}_m - \boldsymbol{v}^*, \boldsymbol{w}_*\rangle| \leq \frac{\beta_{d,1}}{4} \tag{55}$$

Now let $\boldsymbol{v}_m^\perp = \boldsymbol{v}_m - \langle \boldsymbol{v}_m, \boldsymbol{w}_*\rangle \boldsymbol{w}_*$ be the component of $\boldsymbol{v}_m$ orthogonal to $\boldsymbol{w}_*$. Our goal is to find a high probability bound on $\|\boldsymbol{v}_m^\perp\|_2$. Due to spherical symmetry, w.l.o.g., fix $\boldsymbol{w}_* = \boldsymbol{e}_1$ and so $S \sim \mathbb{P}_{\boldsymbol{e}_1}^m$, and the norm of the desired vector is given by

$$\|\boldsymbol{v}_m^\perp\|_2 = \sqrt{\sum_{j=2}^d (\boldsymbol{v}_m)_j^2} \text{ where } \boldsymbol{v}_m^\perp = \frac{1}{m}\sum_{i=1}^m \mathcal{T}(y_i,r_i)\sqrt{d}(\boldsymbol{z}_i)_{-1}.$$

Observe that $\boldsymbol{v}_m^\perp$ is a linear combination of m i.i.d. vectors, however, the coefficients of linear combinations $\mathcal{T}(y_i,r_i)$ are not independent from the vectors $(\boldsymbol{z}_i)_{-1}$ themselves, since it is coupled with $z_{i,1}$. We will decouple the laws and make coefficients independent of the vectors. To this end, consider $(y,r,\boldsymbol{z}) \sim \mathbb{P}_{\boldsymbol{e}_1}$ and $\tilde{\boldsymbol{z}} \sim \tau_{d-1}$ independent of $(y,r,\boldsymbol{z})$. Then the following two random variables have identical laws:

$$\mathcal{T}(y,r)\sqrt{d}(\boldsymbol{z})_{-1} \equiv \frac{\mathcal{T}(y,r)}{\sqrt{1-z_1^2}}\sqrt{d}\tilde{\boldsymbol{z}}$$

Using such argument for each of m samples, and $\boldsymbol{v}_m^\perp$ viewed as a random vector of variables $(\sqrt{d}\tilde{\boldsymbol{z}}_i)_{i\in[m]}$ is sub-gaussian with variance parameter $\sigma_*^2 \leq \frac{1}{m^2}\sum_{i=1}^m \frac{\mathcal{T}(y_i,r_i)^2}{(1-z_{i,1}^2)}$. Thus, with probability $1-2e^{-d}$, we have $\|\boldsymbol{v}_m^\perp\|_2 \lesssim \sigma_*\sqrt{d}$. Therefore, it suffices to bound σ_*. By Lemma 30, for any $\delta > 0$ such that $\log(1/\delta) < cd$ for some $c > 0$, we have with probability $1-\delta/2$,

$$\sigma_*^2 \lesssim \mathbb{E}[\sigma_*^2] + \frac{\sqrt{\kappa_\ell \log \kappa_\ell}}{m^{1.5}}\sqrt{\log(1/\delta)} + \frac{\kappa_\ell^2 \log(1/\delta)\log(m/\delta)^k}{m^2}.$$

Here, the condition $\log(1/\delta) < cd$ arises from the fact the function $g(z) = 1/(1-z_1^2)$ has $\|g\|_p \lesssim 1$ only for some $p < cd$ for some universal constant $c > 0$.

$$\sigma_*^2 \lesssim \frac{\log \kappa_\ell}{m} + \frac{\sqrt{\kappa_\ell \log \kappa_\ell}}{m^{1.5}}\sqrt{\log(1/\delta)} + \frac{\kappa_\ell^2 \log(1/\delta)\log(m/\delta)^k}{m^2}.$$

1146 Finally, for any $\delta < e^{-d^c}$, choosing sample size

$$m \geq \frac{C\kappa_\ell\sqrt{d}}{\beta_{d,1}^2}\left(1 + \frac{1}{\sqrt{d}}\log^{(k+1)/2}\left(\frac{d\kappa_\ell}{\delta\beta_{d,1}}\right)\right), \tag{56}$$

1147 with probability $1-\delta/2-e^{-2d}$

$$\sigma_*^2 \lesssim \frac{\beta_{d,1}^2}{C\sqrt{d}}, \quad \text{and thus,} \quad \|\boldsymbol{v}_m^\perp\| \lesssim \sigma_*\sqrt{d} \lesssim \frac{\beta_{d,1}\sqrt{d}}{\sqrt{C\sqrt{d}}} \leq \frac{\beta_{d,1}d^{1/4}}{\sqrt{C}}.$$

For $C > 1$ sufficiently large, we obtain with probability $1 - \delta/2 - e^{-2d}$,

$$\|\boldsymbol{v}_m^{\perp}\|_2 \leq \beta_{d,1} d^{1/4}$$

Combining this with (55), with probability $1 - \delta/2 - e^{-2d} - e^{-d^c}$

$$\|\boldsymbol{v}_m - \boldsymbol{v}^*\|_2 \leq 2\beta_{d,1} d^{1/4} \,.$$

Therefore, we finally analyze our overlap combining with (55). For $C > 0$ sufficiently large, for any $\delta > 0$, for any $m \geq \frac{C\kappa_\ell \sqrt{d}}{\beta_{d,1}^2}\left(1 + \frac{1}{\sqrt{d}} \log^{(k+1)/2}\left(\frac{\kappa_\ell\, d}{\beta_{d,1}\delta}\right)\right)$, with probability $1 - \delta$,

$$\frac{\langle \boldsymbol{v}_m, \boldsymbol{w}_* \rangle}{\|\boldsymbol{v}_m\|_2} \geq \frac{\langle \boldsymbol{v}^*, \boldsymbol{w}_* \rangle + \langle \boldsymbol{v}_m - \boldsymbol{v}^*, \boldsymbol{w}_* \rangle}{\|\boldsymbol{v}^*\|_2 + \|\boldsymbol{v}_m - \boldsymbol{v}^*\|_2} \geq \frac{\beta_{d,1} - \beta_{d,1}/4}{\beta_{d,1} + 2\beta_{d,1} d^{1/4}} \geq \frac{d^{-1/4}}{4} \,.$$

Similarly, if in Eq.(56), we instead chose a sample of size $m \geq \frac{C\kappa_\ell\, d}{\beta_{d,1}^2}\left(1 + \frac{1}{d} \log^{(k+1)/2}\left(\frac{\kappa_\ell\, d}{\delta\, \beta_{d,1}}\right)\right)$ for sufficiently large $C > 1$, then we obtain a tighter control on $\|\boldsymbol{v}_m - \boldsymbol{v}^*\|_2$, and directly achieve weak recovery. With probability $1 - \delta/2 - e^{-2d} - e^{-d^c}$

$$\|\boldsymbol{v}_m^{\perp}\| \lesssim \sigma_* \sqrt{d} \lesssim \frac{\beta_{d,1}\sqrt{d}}{\sqrt{Cd}} \leq \frac{\beta_{d,1}}{\sqrt{C}} \quad \text{and} \quad \|\boldsymbol{v}_m - \boldsymbol{v}^*\|_2 \leq 2\beta_{d,1} \,.$$

Combining this with (55), with probability $1 - \delta$

$$\frac{\langle \boldsymbol{v}_m, \boldsymbol{w}_* \rangle}{\|\boldsymbol{v}_m\|_2} \geq \frac{\langle \boldsymbol{v}^*, \boldsymbol{w}_* \rangle + \langle \boldsymbol{v}_m - \boldsymbol{v}^*, \boldsymbol{w}_* \rangle}{\|\boldsymbol{v}^*\|_2 + \|\boldsymbol{v}_m - \boldsymbol{v}^*\|_2} \geq \frac{\beta_{d,1} - \beta_{d,1}/4}{\beta_{d,1} + 2\beta_{d,1}} = \frac{1}{4} \,.$$

□

Note that the regime of interest where we can hope to succeed in polynomial sample and runtime is when $\|\xi_{d,1}\|_{L^2} \gg \mathrm{poly}(d)^{-1}$ i.e. which corresponds to $\beta_{d,1} \gg \mathrm{poly}(d)^{-1}$ for $\mathcal{T}$ from Requirement 1. In this regime, Lemma 8 establishes that with sample complexity

$$\mathsf{m} \leq \frac{C\kappa_\ell \sqrt{d}}{\beta_{d,1}^2}\sqrt{\log(1/\delta)} \quad \text{and} \quad \mathsf{m} \leq \frac{C\kappa_\ell d}{\beta_{d,1}^2}\sqrt{\log(1/\delta)}$$

one can achieve the overlaps

$$|\langle \hat{\boldsymbol{w}}, \boldsymbol{w}_* \rangle| \geq d^{-1/4}/4 \quad \text{and} \quad |\langle \hat{\boldsymbol{w}}, \boldsymbol{w}_* \rangle| \geq 1/4 \,.$$

This nearly finishes the proof of Theorem 2 for the case $\ell = 1$ (under stronger Assumption 2). Depending on the problem, the sample complexity bound either matches the one provided in Theorem 2, or it is suboptimal by a factor of $O(\sqrt{d})$. In the latter case, we can first get to $\Omega_d(d^{-1/4})$ overlap, followed by another boosting phase, as long as the following assumption holds.

Assumption 4. *There exists $\ell \geq 3$ and $c > 0$ such that* $\frac{\sqrt{d^\ell}}{\|\xi_{d,\ell}\|_{L^2}^2} \leq c\left(\frac{\sqrt{d}}{\|\xi_{d,1}\|_{L^2}^2} \vee d\right)$.

Note that this assumption holds for Gaussian SIMs according to the discussion in Section 4, i.e. all $\ell \equiv \mathsf{k}_\star \mod 2$ are all optimal for samples. The next section is dedicated to showing the guarantee for boosting procedure.

D.3 Boosting the overlap to achieve weak recovery

We now show how to boost the overlap from $\Omega_d(d^{-1/4})$ to $\Omega_d(1)$ using the following procedure.

Algorithm 3: A single step of the boosting algorithm on $\ell \geq 3$.

Input : An example set $S = \{(\boldsymbol{x}_i, y_i) : i \in [m]\} \sim_{iid} \mathbb{P}_{\boldsymbol{w}_*}$, the frequency $\ell \geq 3$, a transformation $\mathcal{T}_\ell$, and a vector $\boldsymbol{w}_0$ such that $\langle \boldsymbol{w}_0, \boldsymbol{w}_* \rangle \geq d^{-1/4}/4$.

Output : The vector $\hat{\boldsymbol{w}}$.

Let $\Upsilon = \lceil \log d \rceil$ be the total number of steps to be taken.

Divide the training set $S = \{S^{(t)}\}_{i \in [\Upsilon]}$ into disjoint collections of Υ steps, where $|S^{(t)}| = |S|/2^{t+2}$.

for $t = 1, \ldots, \Upsilon$ **do**

  $\boldsymbol{w}_t = \mathsf{boost\text{-}step}(\boldsymbol{w}_{t-1}, S^{(t)})$.

end

Rerurn $\hat{\boldsymbol{w}} = \boldsymbol{w}_\Upsilon$

Algorithm 4: boost-step

Input : An example set S of size m and a vector $\boldsymbol{v}$ such that $\langle \boldsymbol{v}, \boldsymbol{w}_* \rangle = \alpha \in [d^{-1/4}/4, 1/4]$.
Output : The new vector $\boldsymbol{v}_{\text{next}}$ with $\langle \boldsymbol{v}_{\text{next}}, \boldsymbol{w}_* \rangle = \alpha_{\text{next}}$.

Compute $\hat{\boldsymbol{v}} = \frac{1}{m} \sum_{i=1}^{m} \mathcal{T}_\ell(y_i, r_i) Q'_\ell(\langle \boldsymbol{v}, \boldsymbol{z}_i \rangle) \boldsymbol{z}_i$.
Return $\boldsymbol{v}_{\text{next}} = \frac{\hat{\boldsymbol{v}}}{\|\hat{\boldsymbol{v}}\|_2}$.

We have the following guarantee for one step of boosting algorithm.

Lemma 9. *There exists a constant $C = C(k, \ell)$ and $c = c(k, \ell)$ such that the following holds. For the input $\boldsymbol{v}$ such that $\langle \boldsymbol{v}, \boldsymbol{w}_* \rangle = \alpha \in [d^{-0.25}/4, 1/4]$ of the* boost-step *procedure (Algorithm 4), we have that for any*

$$m \geq C\kappa_\ell \frac{d}{\beta_{d,\ell}^2 \alpha^{2\ell-4}} \,,$$

with probability $1 - e^{-d^c}$, we have $\alpha_{\text{next}} = \langle \boldsymbol{v}_{\text{next}}, \boldsymbol{w}_* \rangle \geq 2\alpha$.

Using this lemma we can obtain the following theorem on the performance of the boosting algorithm.

Theorem 4. *There exists a constant $C = C(k, \ell) > 1$ and $c = c(k, \ell) > 0$ such that the following holds. On the initialization $|\langle \boldsymbol{w}_0, \boldsymbol{w}_* \rangle| \geq d^{-1/4}/4$, the boosting algorithm (Algorithm 3) on the training set S whose size is*

$$\mathsf{m} \leq C\kappa_\ell \frac{\sqrt{d^\ell}}{\beta_{d,\ell}^2} \,,$$

with probability $1 - e^{-d^c}$, we have $\langle \hat{\boldsymbol{w}}, \boldsymbol{w}_* \rangle \geq 1/4$.

Proof. According to Lemma 9, choosing $|S^{(1)}| \gtrsim \frac{\kappa_\ell d}{\beta_{d,\ell}^2 \alpha_0^{2\ell-4}} \asymp \frac{\kappa_\ell \sqrt{d^\ell}}{\beta_{d,\ell^2}}$ (hiding constants in k, ℓ), with probability $1 - e^{-d^c}$ we have $|\langle \boldsymbol{w}_1, \boldsymbol{w}_* \rangle| \geq 2\alpha_0$. The sample size threshold for subsequent iteration is strictly less than 1/2 of the previous one since $\alpha \in [d^{-1/4}, 1/4]$ and $\ell \geq 3$. Therefore, the overlap increases geometrically, and we have $\langle \boldsymbol{w}_\Upsilon, \boldsymbol{w}_* \rangle \geq 1/4$ in $\Upsilon = \log(d^{1/4}) \asymp \log d$ iterations. The probability of success is still $1 - e^{-d^c}$ by union bound over $\log d$ for smaller $c > 0$. For the total sample size it suffices to choose,

$$\mathsf{m} = |S| \leq \sum_{t=1}^{\Upsilon} |S^{(t)}| = |S^{(1)}| \sum_{t=1}^{\Upsilon} \frac{1}{2^{t-1}} \leq 2|S^{(1)}| \lesssim \kappa_\ell \frac{\sqrt{d^\ell}}{\beta_{d,\ell}^2} \,.$$

□

We now return to the deferred proof that shows the overlap increases geometrically in the boost-step procedure.

Proof of Lemma 9. Recall from Appendix B that we use $P_\ell(\cdot) = Q_\ell(\cdot)/\sqrt{n_{d,\ell}}$ to denote Gegenbauer polynomial that is normalized to have $P_\ell(1) = 1$. We first note that

$$\begin{aligned}
\boldsymbol{v}^* &= \mathbb{E}[\hat{\boldsymbol{v}}] = \mathbb{E}_{\mathbb{P}_{\boldsymbol{w}_*}}[\mathcal{T}_\ell(y, r) Q'_\ell(\langle \boldsymbol{v}, \boldsymbol{z} \rangle) \boldsymbol{z}] = \nabla_{\boldsymbol{v}} \mathbb{E}_{\mathbb{P}_{\boldsymbol{w}_*}}[\mathcal{T}_\ell(y, r) Q_\ell(\langle \boldsymbol{v}, \boldsymbol{z} \rangle)] \\
&= \beta_{d,\ell} \nabla_{\boldsymbol{v}} \mathbb{E}[Q_\ell(\langle \boldsymbol{w}_*, \boldsymbol{z} \rangle) Q_\ell(\langle \boldsymbol{v}, \boldsymbol{z} \rangle)] \\
&= \beta_{d,\ell} \nabla_{\boldsymbol{v}} P_\ell(\langle \boldsymbol{w}_*, \boldsymbol{v} \rangle) = \beta_{d,\ell} P'_\ell(\langle \boldsymbol{v}, \boldsymbol{w}_* \rangle) \boldsymbol{w}_* \\
&= \beta_{d,\ell} P'_\ell(\alpha) \boldsymbol{w}_* \,.
\end{aligned}$$

Consider any fixed $\boldsymbol{w} \in \mathbb{S}^{d-1}$ and consider the following analysis. Define $X_i = \frac{1}{m}(\mathcal{T}_\ell(y_i, r_i) Q'_\ell(\langle \boldsymbol{v}, \boldsymbol{z}_i \rangle) \langle \boldsymbol{z}_i, \boldsymbol{w} \rangle - \langle \boldsymbol{v}_*, \boldsymbol{w} \rangle)$. Using Lemma 27, we can say that

$$\mathbb{E}[X_i^2] \leq \frac{2}{m^2} \mathbb{E}_{\mathbb{P}_{\boldsymbol{w}}}[\mathcal{T}_\ell(y, r)^2 \, Q'_\ell(\langle \boldsymbol{v}, \boldsymbol{z} \rangle)^2 \langle \boldsymbol{z}, \boldsymbol{w} \rangle^2] \lesssim \frac{1}{m^2} \log^\ell(\kappa_\ell) \,,$$

where we used the fact that $g(\boldsymbol{z}) = Q'_\ell(\langle \boldsymbol{v}, \boldsymbol{z} \rangle)^2 \langle \boldsymbol{z}, \boldsymbol{w} \rangle^2$ is a polynomial of degree 2ℓ with $\|g\|_2 \lesssim 1$. Thus, by Lemma 25, we have $\|g\|_p \lesssim p^\ell$, which allows us to use Lemma 27. Similarly, computing

$$\|X_i\|_p \lesssim \frac{1}{m} \mathbb{E}[|\mathcal{T}_\ell(y, r) \, \underbrace{Q'_\ell(\langle \boldsymbol{v}, \boldsymbol{z} \rangle) \langle \boldsymbol{z}, \boldsymbol{w} \rangle}_{:= g(\boldsymbol{z})} |^p]^{1/p} \leq \frac{1}{m} \|\mathcal{T}_\ell\|_{2p} \|g\|_{2p} \lesssim \frac{\kappa_\ell \, p^{(k+\ell)/2}}{m} \,,$$

where we used the facts that $\|\mathcal{T}_\ell\|_{2p} \lesssim \kappa_\ell p^{k/2}$ and $g(\boldsymbol{z})$ is a polynomial of degree ℓ with $\|g\|_2 \lesssim 1$, and thus, by hypercontractivity, we have $\|g\|_{2p} \lesssim p^{\ell/2}$. Using Lemma 30, with probability $1-\delta$

$$|\langle \hat{\boldsymbol{v}} - \boldsymbol{v}^*, \boldsymbol{w} \rangle| \lesssim \sqrt{\frac{\log^\ell \kappa_\ell \log(1/\delta)}{m}} + \frac{\kappa_\ell \, \log(1/\delta) \, \log^{(k+\ell)/2}(m/\delta)}{m} \,.$$

Therefore, invoking this guarantee for $\boldsymbol{w} \in \{\boldsymbol{w}_*, \boldsymbol{v}\}$, we obtain that with probability $1 - e^{-d^c}$,

$$|\langle \hat{\boldsymbol{v}} - \boldsymbol{v}^*, \boldsymbol{w}_* \rangle| + |\langle \hat{\boldsymbol{v}} - \boldsymbol{v}^*, \boldsymbol{v} \rangle| \lesssim \sqrt{\frac{\log^\ell \kappa_\ell \log(1/\delta)}{m}} + \frac{\kappa_\ell \, \log(1/\delta) \, \log^{(k+\ell)/2}(m/\delta)}{m} \,.$$

Our next goal is to bound $\|\hat{\boldsymbol{v}}^\perp\|_2$, where $\hat{\boldsymbol{v}}^\perp$ is the component of $\hat{\boldsymbol{v}}$ orthogonal to $\mathrm{span}\{\boldsymbol{w}_*, \boldsymbol{v}\}$. Due to rotational symmetry, w.l.o.g., let us say $\boldsymbol{w}_* = \boldsymbol{e}_1$ and $\boldsymbol{v} = \alpha \boldsymbol{e}_1 + \sqrt{1-\alpha^2}\boldsymbol{e}_2$. Then $\hat{\boldsymbol{v}}^\perp = \hat{\boldsymbol{v}} - (\hat{\boldsymbol{v}})_1 \boldsymbol{e}_1 - (\hat{\boldsymbol{v}})_2 \boldsymbol{e}_2$. So our goal is to bound

$$\|\hat{\boldsymbol{v}}^\perp\|_2 \quad \text{where} \quad \hat{\boldsymbol{v}} = \frac{1}{m} \sum_{i=1}^m \mathcal{T}_\ell(y_i, r_i) Q_\ell'(\alpha z_{i,1} + \sqrt{1-\alpha^2} z_{i,2}) (\boldsymbol{z}_i)_{3:d} \,.$$

Consider the following analysis similar to the proof of Lemma 8. For a single sample $(y, r, \boldsymbol{z}) \sim \mathbb{P}_{\boldsymbol{e}_1}$, one can define $\tilde{\boldsymbol{z}} \sim S^{d-3}$ independent of $(y, r, \boldsymbol{z})$. The following two random variables have identical distribution.

$$\mathcal{T}_\ell(y, r) Q_\ell'(\alpha z_1 + \sqrt{1-\alpha^2} z_2) (\boldsymbol{z}_i)_{3:d} \equiv \mathcal{T}_\ell(y, r) \frac{Q_\ell'(\alpha z_1 + \sqrt{1-\alpha^2} z_2)}{\sqrt{1 - z_1^2 - z_2^2}} \tilde{\boldsymbol{z}} \,.$$

Now let us define $\sqrt{d}\tilde{\boldsymbol{z}} \sim$ consider $\boldsymbol{z}_i^\perp$, the component of $\boldsymbol{z}_i$ that is orthogonal to . Using the identical argument from Lemma 7 that $\hat{\boldsymbol{v}}^\perp$ is sub-Gaussian in random variables $(\tilde{\boldsymbol{z}}_i)_{i \in [m]}$, with probability $1 - e^{-2d}$, we have $\|\hat{\boldsymbol{v}}^\perp\| \lesssim \sigma_* \sqrt{d}$, where the parameter

$$\sigma_*^2 = \frac{1}{m^2} \sum_{i=1}^m \mathcal{T}_\ell(y_i, r_i)^2 \frac{Q_\ell'(\alpha z_{i,1} + \sqrt{1-\alpha^2}\, z_{i,2})^2}{d\,(1 - z_{i,1}^2 - z_{i,2}^2)} \,.$$

Using exactly the same bounding strategy used in Lemma 8, for any $\delta \geq d^{-d^c}$, we have with probability $1 - e^{-d^c}$,

$$\sigma_*^2 \lesssim \frac{\log \kappa_\ell}{m} + \frac{\sqrt{\kappa_\ell \log \kappa_\ell}}{m^{1.5}} \sqrt{\log(1/\delta)} + \frac{\kappa_\ell^2 \log(1/\delta) \log^{k+\ell}(m/\delta)}{m^2} \,.$$

Overall, we can conclude that there exists some constant $C(k, \ell) > 1$ such that choosing any

$$m \geq C \kappa_\ell \, \frac{d}{\beta_{d,\ell}^2 \alpha^{2\ell - 4}} ,$$

with probability $1 - e^{-d^c}$,

$$\|\hat{\boldsymbol{v}}^\perp\| \lesssim \sigma_* \sqrt{d} \lesssim \frac{\beta_{d,\ell} \alpha^{\ell-2} \sqrt{d}}{\sqrt{Cd}} \leq \frac{\alpha^{\ell-2} \beta_{d,\ell}}{16} \quad \text{and} \quad \langle \hat{\boldsymbol{v}} - \boldsymbol{v}^*, \boldsymbol{w}_* \rangle \leq \frac{\alpha^{\ell-1} \beta_{d,\ell}}{4} \,.$$

Combining we have

$$\|\hat{\boldsymbol{v}} - \boldsymbol{v}^*\|_2 \leq (1 + o(1)) \frac{\alpha^{\ell-2}\, \beta_{d,\ell}}{8} \,.$$

Finally, analyzing the quantity of desired interest under this high probability event that happens with probability $1 - e^{-d^c}$:

$$\begin{aligned}
\alpha_{\text{next}} = \langle \boldsymbol{v}_{\text{next}}, \boldsymbol{w}_* \rangle = \frac{\langle \hat{\boldsymbol{v}}, \boldsymbol{w}_* \rangle}{\|\hat{\boldsymbol{v}}\|_2} &\geq \frac{\langle \boldsymbol{v}^*, \boldsymbol{w}_* \rangle + \langle \hat{\boldsymbol{v}} - \boldsymbol{v}^*, \boldsymbol{w}_* \rangle}{\|\boldsymbol{v}^*\|_2 + \|\hat{\boldsymbol{v}} - \boldsymbol{v}^*\|_2} \geq \frac{\beta_{d,\ell} P_\ell'(\alpha) - {\beta_{d,\ell}\alpha^{\ell-1}}/{100}}{\beta_{d,\ell} P_\ell'(\alpha) + {\beta_{d,\ell}\alpha^{\ell-2}}/{50}} \\
&\geq (1 + o(1)) \left(\frac{\alpha^{\ell-1} - \frac{\alpha^{\ell-1}}{100}}{\alpha^{\ell-1} + \frac{\alpha^{\ell-2}}{50}} \right) \geq \frac{98\alpha}{50\alpha + 1} \geq 2\alpha \,.
\end{aligned}$$

$\square$

E Online SGD estimator

In this section, we present the analysis of the online SGD on the harmonic loss which corresponds to theorem 2Part 3. As in Section section 3, we will implement the algorithm on $\mathcal{T}_\ell$ for $\ell > 2$. We work under the following assumption

Assumption 5. *For each $\ell \geq 1$, we assume that there exists $\mathcal{T}_\ell$ such that $\|\mathcal{T}_\ell\|_{L^2} = 1$, $\|\mathcal{T}_\ell\|_\infty \leq \kappa_\ell$, consider $\mathcal{T}_\ell := \xi_{d,\ell}/\|\xi_{d,\ell}\|_{L^2}$, and we have the following inequality*

$$\|\xi_{d,\ell}\|_{L^4} \leq C\|\xi_{d,\ell}\|_{L^2} \quad , \tag{57}$$

and also we gave the following

$$\mathbb{E}[\mathcal{T}_\ell(y,r)Q_\ell(\langle \boldsymbol{w}, \boldsymbol{z}\rangle)] = \beta_{d,\ell} > 0.$$

In the following we introduce the following loss

$$L(\boldsymbol{w}; \boldsymbol{z}_i, y_i, r_i) = \left[(\mathcal{T}_\ell(y,r) - Q_\ell(\langle \boldsymbol{w}, \boldsymbol{z}\rangle))^2\right]. \tag{58}$$

We first present the online SGD algorithm

Algorithm 5: Online SGD algorithm on the frequency ℓ.

Input : An example set $S = \{(\boldsymbol{x}_i, y_i) : i \in [m]\} \sim_{iid} \mathbb{P}_{\boldsymbol{w}_*}$, the frequency $\ell > 2$, and a transformation $\mathcal{T}_\ell$, a step size η and a number of step-size.

Output : An estimator $\hat{\boldsymbol{w}} \in \mathbb{R}^d$.

Decompose $\boldsymbol{x}_i = (r_i, \boldsymbol{z}_i)$.
Sample $\boldsymbol{w}_0 \in \mathbb{S}^{d-1}$.
for $i = 1, \ldots, N$ **do**
    Compute $L_i := L(\boldsymbol{w}_{i-1}; \boldsymbol{z}_i, y_i, r_i)$
    Let $\boldsymbol{w}_i := \frac{\boldsymbol{w}_{i-1} - \eta\nabla^{\mathbb{S}^{d-1}}_{\boldsymbol{w}_{i-1}} L_i}{\|\boldsymbol{w}_{i-1} - \nabla^{\mathbb{S}^{d-1}}_{\boldsymbol{w}_{i-1}} L_i}\|$
end
Return $\hat{\boldsymbol{w}_N}$.

We state the formal statement for weak recovery of online SGD. We focus on weak recovery, and refer to the section for explanations on how to boost the weak to strong recovery. We also only focus on $\ell > 3$, however notice that the proof can be adapted to the cases $\ell \in \{1, 2\}$.

Theorem 5 (Online SGD for learning ν_d)**.** *Let $(\boldsymbol{w}_t)_{t\geq 0}$ the iterates of the SGD dynamics with the loss given by Eq.(58), with $b > s_\ell$ (where s_ℓ only depends on ℓ) then conditionally on $m_0 \geq \frac{b}{\sqrt{d}}$, we then have*

$$\tau^+_{1/2} \leq \frac{C_\ell d^{\ell-1}}{\beta^2_{d,\ell}},$$

with probability at least $c > 0$ for some constant $c > 0$.

The good initialization probability is at least constant [49]. Thus, the above online SGD algorithm succeeds in total with probability at least constant bounded away from zero. We can then boost the confidence to $1 - \delta$ by just choosing the best estimator over multiple starts, and $O(\log(1/\delta))$ trails suffice.

Let introduce some notations for the following, denote $m_t = \langle \boldsymbol{w}_t, \boldsymbol{w}_*\rangle$, we define $\tau^-_c = \inf\{t \geq 0 \colon m_t \leq c\}$, and $\tau_c = \inf\{t \geq 0 \colon m_t \geq c\}$.

Proof of theorem 5. Let $\ell \geq 3$. Consider a transformation $\mathcal{T}_\ell$ given by assumption 5. We then have the following

$$\mathbb{E}[\mathcal{T}(y,r)|\boldsymbol{z}] = \sum_{\ell=0}^{\infty} \beta_{d,\ell} Q_\ell(\langle \boldsymbol{z}, \boldsymbol{w}_*\rangle).$$

We first state a lemma on the population loss defined as

$$\mathcal{L}(\boldsymbol{w}) = \mathbb{E}[L(\boldsymbol{w}; \boldsymbol{z}, y, r)]. \tag{59}$$

Lemma 10 (Population loss)**.** *Let $\ell \geq 1$, consider the normalized tranformation $\mathcal{T}_\ell$ given by Assumption 5, we then have the following inequality*

$$\forall m \geq 2\sqrt{\frac{s^*}{d}}, \quad \langle \nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}} \mathcal{L}(\boldsymbol{w}), \boldsymbol{w}_* \rangle \leq -2(1-m^2)\beta_{d,\ell} \frac{\ell(\ell+d-2)}{d-1} \langle \boldsymbol{w}, \boldsymbol{w}_* \rangle^{\ell-1}, \tag{60}$$

where $s^* = \sqrt{\frac{(\ell-2)(\ell+d-3)}{(\ell-d/2-3)(\ell+d/2-2)}} \cos(\pi/\ell)$.

Discretization bounds. In this part, we give bounds on the discretization error from the online SGD and the population loss gradient flow. Consider the SGD iterations

$$\boldsymbol{w}_{t+1} = \frac{\boldsymbol{w}_t - \eta \nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}} L(\boldsymbol{w}_t; \boldsymbol{z}_t, r_t, y_t)}{\| \boldsymbol{w}_t - \eta \nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}} L(\boldsymbol{w}_t; \boldsymbol{z}_t, r_t, y_t) \|},$$

initialized at $\boldsymbol{w}_0$ uniformly on the sphere $\mathbb{S}^{d-1}$.

Proposition 4 (Discretization bound)**.** *With probability at least $1 - p_{\eta, \mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3}$, where*

$$p_{\eta,\mathcal{K}_1,\mathcal{K}_2,\mathcal{K}_3} = \frac{\mathcal{K}_2 d T \eta^2}{b} + \exp\left(- \frac{b^2}{2(\beta_{d,\ell}^2 + \mathcal{K}_2 d^2 \eta^2) d \eta^2 T + 2 b d^{1/2} \eta (\beta_{d,\ell} + \eta d^{3/2})} \right) + \mathcal{K}_3 T d^{1/2} \eta^3 + \sqrt{\mathcal{K}_1 \mathcal{K}_2} T \eta^2 d^{-1},$$

where we have

$$\begin{aligned}
\mathcal{K}_1 &= K\ell \left(\frac{4e}{\ell} \right)^{\ell/2} \log(\|\mathcal{T}_\ell\|_4^2)^{\ell/2}, \\
\mathcal{K}_2 &= 4K\ell \left(\frac{4e}{4\ell} \right)^{4\ell/2} \|\mathcal{T}\|_4^4 \log\left(\frac{\|\mathcal{T}_\ell\|_8^2}{\|\mathcal{T}\|_4^4} \right)^{4\ell/2}, \\
\mathcal{K}_3 &= 2K \left(\frac{4e}{\ell} \right)^{\ell/2} \log(\|\mathcal{T}_\ell\|_4^2)^{\ell/2}.
\end{aligned}$$

We then have

$$m_T \geq \frac{m_0}{2} + \eta \beta_{\ell,d} \frac{\ell \cdot (\ell+d-2)}{d-1} \sum_{t=0}^{T-1} (1-m_t^2) m_t^{\ell-1}.$$

Conditioned on the event $\{T \leq \tau_{1/2}^+ \wedge \tau_{2s^*/\sqrt{d}}^-\}$, we have the following inequality

$$m_T \geq \frac{s^*}{\sqrt{d}} + \eta \beta_{\ell,d} \frac{\ell \cdot (\ell+d-2)}{(d-1)2^{\ell+1}} \sum_{t=0}^{T-1} m_t^{\ell-1}.$$

Sample complexity for weak recovery We are interested in the weak recovery setting i.e we want to bound $\tau_{1/2}$. We work under the event $\{T \leq \tau_{1/2}^+ \wedge \tau_{2s^*/\sqrt{d}}^-\}$, we then can apply the same analysis as in [49]. We choose $\eta = c_\ell \beta_{\ell,d} d^{-\ell/2}$, we then have

$$\eta \tau_{1/2}^+ \leq \frac{2d^{\ell/2-1}}{\beta_{\ell,d} \frac{\ell \cdot (\ell+d-2)}{(d-1)2^{\ell+1}}}.$$

Rearranging this, it gives us

$$\tau_{1/2}^+ \leq \frac{2d^{\ell-1}}{\beta_{\ell,d}^2 c_\ell \frac{\ell \cdot (\ell+d-2)}{(d-1)2^{\ell+1}}} \leq \frac{2d^{\ell-1}}{\beta_{\ell,d}^2} \cdot \left(c_\ell \frac{\ell \cdot (\ell+d-2)}{(d-1)2^{\ell+1}} \right)^{-1}.$$

□

E.1 Proof of Lemma 10

Proof. We remind that we have $\mathcal{T}_\ell$ which is normalized and we also have the expansion

$$\mathbb{E}[\mathcal{T}_\ell(y,r)|\boldsymbol{z}] = \sum_{i=0}^{+\infty} \beta_{d,i} Q_i(\langle \boldsymbol{w}_*, \boldsymbol{z}\rangle). \tag{61}$$

Consider the population mean-squared loss (we can also directly use the correlation loss)

$$\mathcal{L}(\boldsymbol{w}) = \mathbb{E}_{(y,r,\boldsymbol{z})}\left[\left(\mathcal{T}_\ell(y,r) - Q_\ell^{(d)}(\langle \boldsymbol{w}, \boldsymbol{z}\rangle)\right)^2\right] = 2 - 2\beta_{d,\ell}\mathbb{E}\left[\mathcal{T}_\ell(y,r) Q_\ell^{(d)}(\langle \boldsymbol{w}, \boldsymbol{z}\rangle)\right].$$

Using the above decomposition (61), and orthogonality of spherical harmonics

$$\mathcal{L}(\boldsymbol{w}) = 2 - 2\beta_{d,\ell}\mathbb{E}[Q_\ell^{(d)}(\langle \boldsymbol{w}_*, \boldsymbol{z}\rangle) Q_\ell^{(d)}(\langle \boldsymbol{w}, \boldsymbol{z}\rangle)].$$

Using the identity (33), we then have

$$\mathcal{L}(\boldsymbol{w}) = 2 - 2\frac{\beta_{d,\ell}}{\sqrt{n_{\ell,d}}} Q_\ell^{(d)}(\langle \boldsymbol{w}_*, \boldsymbol{w}\rangle).$$

Let denote $m := \langle \boldsymbol{w}_*, \boldsymbol{w}\rangle$, we can rewrite the loss in term of the overlap parameter. We now want to compute the spherical gradient of the population loss

$$\langle \nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}} \mathcal{L}(\boldsymbol{w}), \boldsymbol{w}^*\rangle = -(1-m^2)\ell'(m) = -2(1-m^2)\frac{\beta_{d,\ell}}{\sqrt{n_{d,\ell}}} Q_\ell^{(d)}(\langle \boldsymbol{w}_*, \boldsymbol{w}\rangle)'.$$

We can use the representation of the derivative of Gegenbauer i.e $Q_\ell^{(d)}(\boldsymbol{z})' = \frac{\ell(\ell+d-2)\sqrt{n_{d,\ell}}}{(d-1)\sqrt{n_{d+2,\ell-1}}} Q_{\ell-1}^{d+2}(\boldsymbol{z})$ ([49] Fact C.3). So, the loss can be written as

$$\langle \nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}} \mathcal{L}(\boldsymbol{w}), \boldsymbol{w}^*\rangle = -(1-m^2)\ell'(m) = -2(1-m^2)\frac{\beta_{d,\ell}\ell(\ell+d-2)}{(d-1)\sqrt{n_{d+2,\ell-1}}} Q_{\ell-1}^{(d+2)}(\langle \boldsymbol{w}_*, \boldsymbol{w}\rangle).$$

We can use the facts C.4 and C.5 from [49] (note that the Gegenbauer polynomials in [49] is normalized such that $P_\ell^{(d)}(1) = 1$, meanwhile we consider $Q_\ell^{(d)}(1) = \sqrt{B(d,\ell)}$) to state that

$$\forall m \geq 2\sqrt{\frac{s^*}{d}}, \quad \langle \nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}} \mathcal{L}(\boldsymbol{w}), \boldsymbol{w}_*\rangle \leq -2(1-m^2)\beta_{d,\ell}\frac{\ell^*(\ell^*+d-2)}{d-1}\langle \boldsymbol{w}, \boldsymbol{w}_*\rangle^{\ell^*-1}, \tag{62}$$

where $s^* = \sqrt{\frac{(\ell-2)(\ell+d-3)}{(\ell-d/2-3)(\ell+d/2-2)}}\cos(\pi/\ell)$. □

E.2 Proof of Proposition 4

Proof. In the following, we denote $r_t = \|\boldsymbol{w}_t - \eta\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}} L(\boldsymbol{w}_t; \boldsymbol{z}_t, y_t)\|$ and the martingale part $\boldsymbol{M}_t = L(\boldsymbol{w}_t; \boldsymbol{z}_t, y_t) - \mathbb{E}[L(\boldsymbol{w}_t; \boldsymbol{z}_t, y_t)]$. We have the recursion

$$m_{t+1} = \frac{1}{r_t}\left(m_t - \eta\langle \nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}} \mathcal{L}(\boldsymbol{w}_t), \boldsymbol{w}_*\rangle - \eta\langle \nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}} \boldsymbol{M}_t, \boldsymbol{w}^*\rangle\right). \tag{63}$$

The strategy of the proof is to use the results from [49]. The proofs of these lemmas are classical bounds of martingale relying on some assumptions on the moments of the gradients. Notice here that C is no longer a constant and extra care of the analysis is necessary for the proof of Lemma [49, Lemma B.4]. To prove the lemmas [49, Lemmas B.2,B.3,B.4], we need to prove the bounds on the growth of gradients norms of [49, Lemma B.8]. We check this

$$\begin{aligned}
\mathbb{E}_{(y,r,\boldsymbol{z})}[\|\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}}(L(\boldsymbol{w}_t; \boldsymbol{z}_t, r_t, y_t)\|^2] &= \mathbb{E}_{(y,r,\boldsymbol{z})}[\|\mathsf{P}_{\boldsymbol{w}_t}(\boldsymbol{z}_t)(Q_{\ell^*}^{(d)})'(\langle \boldsymbol{w}, \boldsymbol{z}_t\rangle)\mathcal{T}(y,r)\|^2] \\
&\leq \mathbb{E}_{(y,r,\boldsymbol{z})}\left[\left|(Q_{\ell^*}^{(d)})'(\langle \boldsymbol{w}, \boldsymbol{z}_t\rangle)\mathcal{T}(y,r)\right|^2\right] \\
&\leq C(d)^2\mathbb{E}\left[\left|Q_{\ell^*-1}^{(d-2)}(\langle \boldsymbol{w}, \boldsymbol{z}\rangle)\right|^2 \mathcal{T}(y,r)^2\right] \\
&\leq d\ell\left(\frac{4e}{\ell}\right)^{\ell/2}\mathbb{E}\left[\mathcal{T}(y,r)^2\right]\log(1/\mathbb{E}\left[\mathcal{T}(y,r)\right])^{\ell/2} \\
&\leq Kd\ell\left(\frac{4e}{\ell}\right)^{\ell/2}\log(\|\mathcal{T}\|_4^2)^{\ell/2} \\
&\leq d\mathcal{K}_1.
\end{aligned}$$

where we have used Lemma 27, the identity and the hypercontractivity and Jensen inequality in the last line.

$$\begin{aligned}
\mathbb{E}_{(y,r,z)}[\|\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}}(L(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t)\|^4] &= \mathbb{E}_{(y,r,z)}[\|\mathsf{P}_{\boldsymbol{w}_t}(\boldsymbol{z}_t)(Q_{\ell^*}^{(d)})'(\langle \boldsymbol{w},\boldsymbol{z}_t\rangle)\mathcal{T}_\ell(y,r)\|^4] \\
&\le \mathbb{E}_{(y,r,\boldsymbol{z})}\left[\left|(Q_{\ell^*}^{(d)})'(\langle \boldsymbol{w},\boldsymbol{z}_t\rangle)\mathcal{T}_\ell(y,r)\right|^4\right] \\
&\le C(d)^4\mathbb{E}\left[\left|Q_{\ell^*-1}^{(d-2)}(\langle \boldsymbol{w},\boldsymbol{z}\rangle)\right|^4|\mathcal{T}_\ell(y,r)|^4\right] \\
&\le 4C(d)^4\ell\left(\frac{4e}{4\ell}\right)^{4\ell/2}\|\mathcal{T}_\ell\|_4^4\log\left(\frac{\|\mathcal{T}_\ell\|_8^2}{\|\mathcal{T}\|_4^4}\right)^{4\ell/2}. \\
&\le d^2 4K\ell\left(\frac{4e}{4\ell}\right)^{4\ell/2}\|\mathcal{T}_\ell\|_4^4\log\left(\frac{\|\mathcal{T}_\ell\|_8^2}{\|\mathcal{T}_\ell\|_4^4}\right)^{4\ell/2} \\
&\le d^2\mathcal{K}_2.
\end{aligned}$$

We have $\boldsymbol{M}(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t) = \ell(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t) - \mathbb{E}[\ell(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t)]$, hence

$$\begin{aligned}
\langle\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}}\boldsymbol{M}(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t),\boldsymbol{w}_*\rangle &= \langle\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}}(L(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t),\boldsymbol{w}_*\rangle - \mathbb{E}[\langle\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}}(L(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t),\boldsymbol{w}^*\rangle] \\
&\quad - \langle\boldsymbol{w},\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}}(L(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t)\rangle m + \mathbb{E}[m\langle\boldsymbol{w},\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}}(L(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t)\rangle].
\end{aligned}$$

We then have

$$\begin{aligned}
&\mathbb{E}_{(y,r,\boldsymbol{z})}[\langle\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}}(\boldsymbol{M}(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t),\boldsymbol{w}_*\rangle^2] \\
&\le 2\mathrm{Var}_{(y,r,\boldsymbol{z})}(\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}}(L(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t),\boldsymbol{w}_*\rangle) + 2\mathrm{Var}_{(y,r,\boldsymbol{z})}(\langle\boldsymbol{w},\nabla_{\boldsymbol{w}}^{\mathbb{S}^{d-1}}(L(\boldsymbol{w}_t;\boldsymbol{z}_t,r_t,y_t)\rangle m) \\
&\le 2\mathbb{E}[\langle\boldsymbol{z}_t,\boldsymbol{w}_*\rangle^2(Q_{\ell^*}^{(d)})'(\langle\boldsymbol{w},\boldsymbol{z}_t\rangle)^2\mathcal{T}(y,r)^2] + 2\mathbb{E}[\langle\boldsymbol{z}_t,\boldsymbol{w}_*\rangle^2(Q_{\ell^*}^{(d)})'(\langle\boldsymbol{w},\boldsymbol{z}_t\rangle)^2\mathcal{T}_\ell(y,r)^2] \\
&\le 2K\left(\frac{4e}{\ell}\right)^{\ell/2}\log(\|\mathcal{T}_\ell\|_4^2)^{\ell/2} := \mathcal{K}_3,
\end{aligned}$$

where we have used the inequality $(a+b)^2 \le 2(a^2+b^2)$, hypercontractivity of Gegenbauer polynomials.

Conditioned on the event $\{T \le \tau^-_{2s^*/\sqrt{d}}\}$, and using the inequality (62), we have

$$m_{t+1} \ge \frac{1}{r_t}\left(m_t + 2\eta(1-m_t^2)\beta_{\ell,d}\frac{\ell\cdot(\ell^*+d-2)}{d-1}m_t^{\ell-1} - \eta\langle\nabla^{\mathbb{S}^{d-1}}\boldsymbol{M}_t,\boldsymbol{w}^*\rangle\right).$$

Using the following bound on r_t, which is for all $t \in \mathbb{N}$, we have

$$1/r_t \ge 1 - \eta^2\|\nabla_{\boldsymbol{w}_t}L(\boldsymbol{w}_t;y_t,r_t,\boldsymbol{z}_t)\|^2,$$

and plugging this into previous inequality, we have

$$m_{t+1} \ge m_t + 2\eta(1-m_t^2)\beta_{\ell^*,d}\frac{\ell^*\cdot(\ell^*+d-2)}{d-1}m_t^{\ell^*-1} - \eta\langle\nabla^{\mathbb{S}^{d-1}}\boldsymbol{M}_t,\boldsymbol{w}^*\rangle - \eta^2|m_t|\,\|\nabla_{\boldsymbol{w}_t}L(\boldsymbol{w}_t;y_t,r_t,\boldsymbol{z}_t)\|^2 - \eta^3\xi_T, \tag{64}$$

where $\xi_T = \|\nabla_{\boldsymbol{w}}L(\boldsymbol{w}_T;y_T,r_T,\boldsymbol{z}_T)\|^2\cdot|\langle\nabla_{\boldsymbol{w}}L(\boldsymbol{w}_T;y_T,r_T,\boldsymbol{z}_T),\boldsymbol{w}^*\rangle|^2$.

We use [49, Lemma B.3] to state that with probability at least $1-\frac{\mathcal{K}_3 t\eta^2}{\lambda^2}$, we have for all $\lambda > 0$,

$$\sup_{t\le T}\eta\left|\sum_{k=0}^{t-1}\left\langle\nabla^{\mathbb{S}^{d-1}}\boldsymbol{M}_t,\boldsymbol{w}^*\right\rangle\right| \le \lambda. \tag{65}$$

We employ [49, Lemma B.6] to state that for all $\lambda > 0$, with probability at least $1-\frac{\sqrt{\mathcal{K}_1\mathcal{K}_2}Td\eta^3}{\lambda}$, we have

$$\sup_{t\le T}\eta^3\sum_{k=0}^{t-1}\xi_k \le \lambda. \tag{66}$$

We sum the equation 64 to obtain

$$m_T \geq m_0 + 2\eta \frac{\beta_{\ell,d}\ell \cdot (\ell+d-2)}{d-1} \sum_{t=0}^{T-1} (1-m_t^2) m_t^{\ell-1} - \eta \sum_{t=0}^{T-1} \langle \nabla^{\mathbb{S}^{d-1}} \boldsymbol{M}_t, \boldsymbol{w}^* \rangle$$
$$- \eta^2 \sum_{t=0}^{T-1} |m_t| \, \|\nabla_{\boldsymbol{w}_t} L(\boldsymbol{w}_t; y_t, r_t, \boldsymbol{z}_t)\|^2 - \sum_{t=0}^{T-1} \eta^3 \xi_T,$$

We then use (65),(66) and plug it into (64), and use $\lambda = b/\sqrt{d}$, to state that with probability at least $1 - \mathcal{K}_3 T d^{1/2} \eta^3 - \sqrt{\mathcal{K}_1 \mathcal{K}_2} T \eta^2 d^{-1}$, we have

$$m_T \geq \frac{7m_0}{10} + 2\eta \frac{\beta_{\ell,d}\ell \cdot (\ell+d-2)}{d-1} \sum_{t=0}^{T-1} (1-m_t^2) m_t^{\ell-1} - \eta^2 \sum_{t=0}^{T-1} |m_t| \, \|\nabla_{\boldsymbol{w}_t} L(\boldsymbol{w}_t; y_t, r_t, \boldsymbol{z}_t)\|^2. \tag{67}$$

We now bound the term coming from the projection step in inequality (67). We adapt the proof to take account the dependency on $\|\mathcal{T}_\ell\|_2$.

Lemma 11. *For all $\lambda > 0$, if for all $t \leq T, m_t \in [2b/\sqrt{d}, 1/2]$, and $\eta \leq \frac{c\|\xi_{d,\ell^*}\|_2}{d}$, with probability at least $1 - \frac{\mathcal{K}_2 T d^{1/2} \eta^2}{\lambda} - \exp\left(-\frac{\lambda^2}{2(\beta_{\ell,d}^2 + \mathcal{K}_2 d^2 \eta^2)\eta^2 T + 2\lambda\eta(\beta_{\ell,d} + \eta d^{3/2})}\right)$, we have*

$$\eta^2 \sum_{t=0}^{T-1} |m_t| \, \|\nabla_{\boldsymbol{w}_t} L(\boldsymbol{w}_t; y_t, r_t, \boldsymbol{z}_t)\|^2 + \eta \sum_{t=0}^{T-1} (1-m_t^2) \beta_{d,\ell} \frac{\ell \cdot (\ell+d-2)}{d-1} m_t^{\ell-1} \leq 2\lambda.$$

Proof. The proof is a slight adaptation of [49, Lemma B.4,B.5]. An adaptation of the proof of Lemma B.4 gives us the following. For all $\lambda > 0$, if for all $t \leq T, m_t \in [2b/\sqrt{d}, 1/2]$, and $\eta > 0$, we have

$$\mathbb{P}(\eta \sum_{t=0}^{T-1} D_t \leq -\lambda) \leq \exp\left(-\frac{\lambda^2}{2(\beta_{\ell,d}^2 + \mathcal{K}_2 d^2 \eta^2)\eta^2 T + 2\lambda\eta(\beta_{\ell,d} + \eta d^{3/2})}\right).$$

Besides, the adaptation of Lemma B.5 gives us

$$\mathbb{P}\left(\sup_{t \leq T} \eta^2 \sum_{t=0}^{T-1} |m_t| \|\nabla_{\boldsymbol{w}} L\|^2 1_{\|\nabla_{\boldsymbol{w}} L\| > d^{3/2}} \geq \lambda\right) \leq \frac{\mathcal{K}_2 T d^{1/2} \eta^2}{\lambda}.$$

Combining the two inequalities, we end up the desired claim. □

We then use $\lambda = b/\sqrt{d}$, and we obtain that with probability at least $1 - p_{\eta,\mathcal{K}_1,\mathcal{K}_2,\mathcal{K}_3}$ where

$$p_{\eta,\mathcal{K}_1,\mathcal{K}_2,\mathcal{K}_3} = \frac{\mathcal{K}_2 d T \eta^2}{b} + \exp\left(-\frac{b^2}{2(\beta_{k,d}^2 + \mathcal{K}_2 d^2 \eta^2)\eta^2 d T + 2 b d^{1/2} \eta (\beta_{k,d} + \eta d^{3/2})}\right) + \mathcal{K}_3 T d^{1/2} \eta^3 + \sqrt{\mathcal{K}_1 \mathcal{K}_2} T \eta^2 d^{-1},$$

we have

$$m_T \geq \frac{m_0}{2} + \beta_{\ell,d} \frac{\ell \cdot (\ell+d-2)}{d-1} \eta \sum_{t=0}^{T-1} (1-m_t^2) m_t^{\ell-1}.$$

Conditioned on the event $\{T \leq \tau_{1/2}^{+} \wedge \tau_{2s^*/\sqrt{d}}^{-}\}$, we have the following inequality

$$m_T \geq \frac{s^*}{\sqrt{d}} + \eta \beta_{\ell,d} \frac{\ell \cdot (\ell+d-2)}{(d-1)2^{\ell+1}} \sum_{t=0}^{T-1} m_t^{\ell-1}.$$

□

F Harmonic tensor unfolding

In this section, we present the analysis of the harmonic tensor unfolding which corresponds to theorem 2[Part 3]. As in the two previous sections, we will implement the algorithm on $\mathcal{T}_\ell$ for $\ell > 2$.

Assumption 6. *For $\nu_d \in \mathfrak{L}_d$ and $\ell \geq 1$, there exist $\mathcal{T}_\ell : \mathcal{Y} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ and $\kappa_\ell > 1$ such that $\|\mathcal{T}_\ell\|_{L^2} = 1$, $\|\mathcal{T}_\ell\|_\infty \leq \kappa_\ell$ and $\mathbb{E}[\mathcal{T}_\ell(y, r) Q_\ell(\langle \boldsymbol{w}, \boldsymbol{z}\rangle)] = \beta_{d,\ell} > 0$.*

We remind some notations for tensors. Let $\boldsymbol{A}, \boldsymbol{B} \in (\mathbb{R}^d)^{\otimes \ell}$, we define the inner-product

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{i_1,\dots,i_\ell} \boldsymbol{A}_{i_1,\dots,i_\ell} \boldsymbol{B}_{i_1,\dots,i_\ell}.$$

We call the tensor $\boldsymbol{X}$ symmetric if, for any permutation $\pi \in \mathfrak{S}_\ell$, $\boldsymbol{X}^\pi = \boldsymbol{X}$. We also denote the unfolded matrix of the tensor $\boldsymbol{T} \in (\mathbb{R}^d)^{\otimes 2\ell}$ as $\textbf{Mat}_q(\boldsymbol{T}) \in \mathbb{R}^{d^q} \times \mathbb{R}^{d^{\ell-q}}$ with coefficients given by

$$(\textbf{Mat}(\boldsymbol{T})_q)_{(i_1,\dots,i_q),(j_1,\dots,j_{\ell-q})} = \boldsymbol{T}_{i_1,\dots,i_\ell,j_1,\dots,j_\ell},$$

where we identify $(i_1, \dots, i_q)$ with $1 + \sum_{k=1}^q (i_k - 1) d^{k-1}$, and $(j_1, \dots, j_{\ell-q})$ with $1 + \sum_{k=1}^{\ell-q} (j_k - 1) d^{k-1}$. In the following, we will drop the dependancy on q since the unfolding parameter will be made clear. We now present the Harmonic tensor unfolding. The main difference with the spherical tensor unfolding is that the Harmonic tensor is the projection of the tensor $\boldsymbol{z}^{\otimes \ell}$ into the space of traceless symmetric tensor, and it allows to decouple the different harmonic subspaces and it yields a simpler analysis. In the following two subsections, we will work under the following assumption

We define the Harmonic tensors.

Definition 2 (Harmonic Tensors)**.** *For every $k > 0$, we define $\mathcal{H}_k \in \mathrm{Sym}((\mathbb{R}^d)^{\otimes k})$ the unique symmetric tensor such that for all $\boldsymbol{w} \in \mathbb{S}^{d-1}$, and for all $\boldsymbol{z} \in \mathbb{S}^{d-1}$, we have*

$$Q_k^{(d)}(\langle \boldsymbol{z}, \boldsymbol{w} \rangle) = \langle \mathcal{H}_k(\boldsymbol{z}), \boldsymbol{w}^{\otimes k} \rangle. \tag{68}$$

Proof of uniqueness of Zonal tensor. Let T, T' two symmetric tensors that satisfy the identity 68.Consider the Warring decomposition of $T - T'$ i.e there exists $K > 0$, and $\boldsymbol{v}_1, \dots, \boldsymbol{v}_K \in \mathbb{S}^{d-1}$ such that

$$T - T' = \sum_{i=1}^K \alpha_i \boldsymbol{v}_i^{\otimes k}.$$

Using that for all $\boldsymbol{w} \in \mathbb{S}^{d-1}$, $\langle \boldsymbol{w}^{\otimes k}, T - T' \rangle = 0$, we have

$$\begin{aligned} \|T - T'\|^2 &= \sum_i^K \alpha_i \langle T - T', \boldsymbol{v}_i^{\otimes k} \rangle \\ &= 0. \end{aligned}$$

We then conclude that $T = T'$, and it proves the uniqueness. $\square$

The uniqueness property is due to the fact that the tensor is defined to be symmetric.

Remark F.1. These tensors seem to appear in physics litterature as Harmonic tensors. Besides another interpretation of these tensors, is that it can be seen as the projection into the space of symmetric, trace-less tensors of the tensor $\boldsymbol{z}^{\otimes \ell}$.

These tensor satisfy a Reproducing property Eq.72. We state here some properties of the Harmonic Tensors. The algorithm of harmonic tensor unfolding uses the harmonic tensors defined above, and unfold it. Given enough samples, the unfold matrix has a left-singular value that has good alignment with the feature vector $\boldsymbol{w}^{\otimes \ell}$. We then compute the vectorization. We show that the vectorization does not hurt the alignment of the signal.

We can now state our result on tensor unfolding recovery.

Theorem 6 (Tensor unfolding)**.** *Let $\hat{\boldsymbol{v}}_1$ denote the the top right singular vector of $\hat{\boldsymbol{M}}_n$. Under Assumption 6, there exists a constant $\mathcal{C}_\ell$, such that for for any $\delta > 0$ with probability at least $1-\delta$, and for*

$$n \geq \frac{\kappa_\ell d^{\ell/2}\ell \log(2d/\delta)}{\|\xi_{d,\ell}\|_2 \varepsilon} \vee \frac{\mathcal{C}_\ell \kappa_\ell^2 d^{\ell/2} k \log(2d/\delta)}{\|\xi_{d,\ell}\|_2^2 \varepsilon^2},$$

we have

$$\langle \boldsymbol{w}, \hat{\boldsymbol{v}} \rangle \geq 1 - \varepsilon. \tag{69}$$

In particular, in view of the boosting lemma, it is sufficient to consider $\varepsilon = 1/2$, and we end up with the following sample complexity

$$n \geq \frac{\kappa_\ell d^{\ell/2}\ell \log(2d/\delta)}{\|\xi_{d,\ell}\|_2} \vee \frac{\mathcal{C}_\ell \kappa_\ell^2 d^{\ell/2} k \log(2d/\delta)}{\|\xi_{d,\ell}\|_2^2}.$$

It proves the part about harmonic tensor in the theorem.

The rest of section is devoted to the proof of this result. We first prove some results on Harmonic tensors, then we focus on even and odd cases.

F.1 Preliminaries on Harmonic tensors.

Lemma 12. *We have the following properties:*

- *$\mathcal{H}_k$ has the following expression*

$$\mathcal{H}_k(\boldsymbol{z}) = \sum_{j=0}^{\lfloor k/2 \rfloor} c_{k,j} \mathrm{Sym}\left(\boldsymbol{z}^{\otimes k-2j} \otimes \mathbf{I}_d^{\otimes j}\right), \tag{70}$$

  where $c_{k,j} = (-1)^k \frac{\Gamma(k-j+d/2-1)2^{k-2j}}{\Gamma(d/2-1)k!(n-2k)!} \frac{\sqrt{B(d,k)}}{Q(1)} \simeq d^{k/2-j}$.

- *Let $O \in \mathcal{O}(d)$, we then have*

$$\mathcal{H}_k(O\boldsymbol{z}) = O^{\otimes k}\mathcal{H}_k(\boldsymbol{z}).$$

Proof. We decompose the Gegenbauer polynomials into monomials

$$Q_k^{(d)}(\langle \boldsymbol{z}, \boldsymbol{w}\rangle) = \sum_{j=0}^{\lfloor k/2 \rfloor} c_{k,j} \langle \boldsymbol{w}, \boldsymbol{z}\rangle^{k-2j},$$

where we have that

$$c_{k,j} = (-1)^k \frac{\Gamma(k-j+d/2-1)2^{k-2j}}{\Gamma(d/2-1)k!(n-2k)!} \frac{\sqrt{B(d,k)}}{Q(1)} \simeq d^{k/2-j}.$$

Besides, we have

$$\begin{aligned}
\left\langle \boldsymbol{w}^{\otimes k}, \left(\boldsymbol{z}^{\otimes(k-2j)} \otimes \mathbf{I}_d^{\otimes j}\right)\right\rangle &= \sum_{i_1,\dots,i_k} \boldsymbol{w}_{i_1} \dots \boldsymbol{w}_{i_k} \boldsymbol{z}_{i_1} \dots \boldsymbol{z}_{i_{k-2j}} \delta_{i_{k-2j+1,k-2j+2}} \dots \delta_{i_{k-1,k}} \\
&= \prod_{j=1}^{k-2j} \sum_j \boldsymbol{w}_{i_j} \boldsymbol{z}_{i_j} \prod_{j=1}^{k-2j} \sum_j \boldsymbol{w}_{i_j} \boldsymbol{w}_{i_j} \\
&= \langle \boldsymbol{w}, \boldsymbol{z}\rangle^{k-2j} \|\boldsymbol{w}\|^j.
\end{aligned}$$

So, we end up with the claimed identity. For the second identity, we use the fact that

$$Q_k(\langle O\boldsymbol{w}, \boldsymbol{z}\rangle) = \langle \mathcal{H}(\boldsymbol{z}), (O\boldsymbol{w})^{\otimes k}\rangle = \langle O^{\otimes k}\mathcal{H}(\boldsymbol{z}), \boldsymbol{w}^{\otimes k}\rangle,$$

and

$$Q_k(\langle \boldsymbol{w}, O\boldsymbol{z}\rangle) = \langle \mathcal{H}(O\boldsymbol{z}), \boldsymbol{w}^{\otimes k}\rangle,$$

we can identify the two elements. $\square$

We remind the zonal property of Gegenbauer polynomials

Lemma 13 (Zonal property of Gegenbauer ([12]))**.** *Let $f \in L^2(\mathbb{S}^{d-1})$. Consider the projection P_{V_k} of f into the spherical harmonics of degree k denoted V_k, this projection can be written as*

$$\mathsf{P}_{V_k} f(\boldsymbol{x}) = \sqrt{n_{d,k}} \cdot \mathbb{E}_{\boldsymbol{z}}[f(\boldsymbol{z}) Q_k^{(d)}(\langle \boldsymbol{z}, \boldsymbol{x} \rangle)]. \tag{71}$$

Proof. We have the following identity 32 for Gegenbauer polynomials, for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{S}^{d-1}$,

$$Q_k(\langle \boldsymbol{z}, \boldsymbol{x}) = \frac{1}{\sqrt{n_{d,k}}} \sum_{i=1}^{n_{d,k}} Y_{k,i}(\boldsymbol{z}) Y_{k,i}(\boldsymbol{x}).$$

We then have that

$$\mathbb{E}_{\boldsymbol{z}}[Q_k(\langle \boldsymbol{z}, \boldsymbol{x} \rangle) f(\boldsymbol{z})] = \frac{1}{\sqrt{n_{d,k}}} \sum_{i=1}^{n_{d,k}} \langle Y_{k,i}, f \rangle Y_{k,i}(\boldsymbol{x}) = \frac{\mathsf{P}_{V_k} f}{\sqrt{n_{d,k}}}.$$

□

Definition 3 (Reproducing Property of Tensor)**.** *Consider the following operator $\boldsymbol{M}_f : L^2(\mathbb{S}^{d-1}) \to$ $\mathrm{Sym}(\mathbb{R}^d)^{\otimes k}$ which is defined as the unique element such that*

$$\mathsf{P}_{V_k} f(\boldsymbol{x}) = \langle \boldsymbol{x}^{\otimes k}, \boldsymbol{M}_k(f) \rangle. \tag{72}$$

The unicity comes from the same reasoning as Definition 1.

Lemma 14. *Let $\ell \in \mathbb{N}$, and $k \in \mathbb{N}$, we have the following identity*

$$\mathbb{E}_{\boldsymbol{z} \in \mathbb{S}^{d-1}} \left[Q_\ell^{(d)}(\langle \boldsymbol{z}, \boldsymbol{w} \rangle) \mathcal{H}_k(\boldsymbol{z}) \right] = 1_{\ell = k} \frac{\mathcal{H}_k(\boldsymbol{w})}{\sqrt{n_{d,k}}}. \tag{73}$$

Proof. We have the following form for the Reproducing property

$$\boldsymbol{M}_k(f) = \sqrt{n_{d,k}} \cdot \mathbb{E}_{\boldsymbol{z}} \left[f(\boldsymbol{z}) \mathcal{H}_k(\boldsymbol{z}) \right].$$

Using Zonal property of Gegenbauer polynomials (71), we have

$$\begin{aligned} \mathsf{P}_{V_k} f(\boldsymbol{x}) &= \sqrt{n_{d,k}} \int_{\mathbb{S}^{d-1}} f(\boldsymbol{z}) Q_k^{(d)}(\langle \boldsymbol{z}, \boldsymbol{x} \rangle) \mathrm{d}\sigma_{d-1}(\boldsymbol{z}) \\ &= \sqrt{n_{d,k}} \int_{\mathbb{S}^{d-1}} f(\boldsymbol{z}) \langle \boldsymbol{x}^{\otimes k}, \mathcal{H}_k(\boldsymbol{z}) \rangle \mathrm{d}\sigma_{d-1}(\boldsymbol{z}) \\ &= \langle \boldsymbol{x}^{\otimes k}, \sqrt{n_{d,k}} \int_{\mathbb{S}^{d-1}} f(\boldsymbol{z}) \mathcal{H}_k(\boldsymbol{z}) \mathrm{d}\sigma_{d-1}(\boldsymbol{z}) \rangle. \end{aligned}$$

We then use the fact that the uniqueness to identify the two tensors, and so we have that

$$\boldsymbol{M}_k(f) = \sqrt{n_{d,k}} \int_{\mathbb{S}^{d-1}} f(\boldsymbol{z}) \mathcal{H}_k(\boldsymbol{z}) \mathrm{d}\sigma_{d-1}.$$

We have

$$1_{\ell = k} Q_k(\langle \boldsymbol{w}^*, \boldsymbol{z} \rangle) = \mathsf{P}_{V_\ell} Q_k(\langle \boldsymbol{w}^*, \boldsymbol{z} \rangle) = \langle \boldsymbol{x}^{\otimes \ell}, \boldsymbol{M}_\ell(Q_k(\langle \boldsymbol{w}^*, \cdot \rangle)) \rangle.$$

Using uniqueness of the linear form, we have if $\ell \neq k$,

$$\boldsymbol{M}_\ell(Q_k(\langle \boldsymbol{w}^*, \cdot \rangle)) = 0.$$

If $\ell = k$, we have

$$\boldsymbol{M}_k(Q_k(\langle \boldsymbol{w}^*, \cdot \rangle)) = \mathcal{H}_k(\boldsymbol{w}^*).$$

Using the above equality and the fact above, we have that

$$\begin{aligned} \sqrt{n_{d,k}} \mathbb{E}_{\boldsymbol{z}}[Q_k^{(d)}(\langle \boldsymbol{z}, \boldsymbol{w} \rangle) \mathcal{H}_k(\boldsymbol{z})] &= \boldsymbol{M}_k(\boldsymbol{z} \to Q_k(\langle \boldsymbol{w}^*, \boldsymbol{z} \rangle)) \\ &= \mathcal{H}_k(\boldsymbol{w}). \end{aligned}$$

We then conclude the following

$$\mathbb{E}_{\boldsymbol{z}}[Q_k^{(d)}(\langle \boldsymbol{z}, \boldsymbol{w} \rangle) \mathcal{H}_k(\boldsymbol{z})] = \frac{\mathcal{H}_k(\boldsymbol{w})}{\sqrt{n_{d,k}}}.$$

□

Denote $\boldsymbol{M}_n$ the unfolded matrix $\textbf{Mat}\left(\frac{1}{n}\sum_{i=1}^n \mathcal{T}_k(y_i,r_i)\mathcal{H}_k(\boldsymbol{w})\right)$ obtained from tensor $\mathcal{H}_k$, and we denote $\boldsymbol{M} := \mathbb{E}[\boldsymbol{M}_n]$.

Proposition 5. *Let $k \geq 2$. There exists a constant K only depending on $\mathcal{C}$ such that*

$$\|\boldsymbol{M} - c_{k,0}\boldsymbol{w}^{\otimes m}(\boldsymbol{w}^{\otimes m})^{\mathsf{T}}\|_{\text{op}} \leq \mathcal{C}d^{k/2-1/2}. \tag{74}$$

Proof. We use the following Lemma

Lemma 15 (Lemma 4.13 from [46]). *Let $\mathcal{T} \in (\mathbb{R}^d)^{\otimes k}$ be an k-order tensor with same dimension, we then have*

$$\|\textbf{\textit{Mat}}(\mathcal{T})\|_{op} \leq \|\mathcal{T}\|_F. \tag{75}$$

We then compute the Frobenius norm of the tensor $\mathcal{H}_k(\boldsymbol{w}) - c_{k,0}\boldsymbol{w}^{\otimes k}$

$$\begin{aligned}
\|\mathcal{H}_k(\boldsymbol{w}) - c_{k,0}\boldsymbol{w}^{\otimes k}\|_F &= \left\|\sum_{\ell=0}^{\lfloor m/2\rfloor} c_{k,\ell}\text{Sym}(\boldsymbol{w}^{\otimes(k-2\ell)}\otimes \mathbf{I}_d^{\otimes \ell}) - c_{k,0}\boldsymbol{w}^{\otimes k}\right\|_F \\
&= \left\|\sum_{\ell=1}^{\lfloor m/2\rfloor} c_{k,\ell}\text{Sym}(\boldsymbol{w}^{\otimes(k-2\ell)}\otimes \mathbf{I}_d^{\otimes \ell})\right\|_F \\
&\leq \sum_{\ell=1}^{\lfloor m/2\rfloor} c_{k,\ell}\left\|\text{Sym}(\boldsymbol{w}^{\otimes(k-2\ell)}\otimes \mathbf{I}_d^{\otimes \ell})\right\|_F \\
&\leq \sum_{\ell=1}^{\lfloor m/2\rfloor} c_{k,\ell}\|\boldsymbol{w}\|^{k-2l}\|\mathbf{I}_d\|_F^{\ell} \\
&\leq \sum_{\ell=1}^{\lfloor m/2\rfloor} c_{k,\ell}d^{\ell/2} \leq \mathcal{C}d^{k/2-1/2},
\end{aligned}$$

where we have used that $\|\boldsymbol{AB}\|_F \leq \|\boldsymbol{A}\|_F\cdot\|\boldsymbol{B}\|_F$, $\|\text{Sym}\boldsymbol{A}\|_F \leq \|\boldsymbol{A}\|_F$. Using the fact that **Mat** is linear, and $\textbf{Mat}(\boldsymbol{w}^{\otimes k}) = \boldsymbol{w}^{\otimes k/2}(\boldsymbol{w}^{\otimes k/2})^{\mathsf{T}}$, we end up with

$$\textbf{Mat}\left(\mathcal{H}_k(\boldsymbol{w}) - c_{k,0}\boldsymbol{w}^{\otimes k}\right) = \boldsymbol{M} - c_{k,0}\boldsymbol{w}^{\otimes k/2}(\boldsymbol{w}^{\otimes k/2})^{\mathsf{T}}.$$

Using inequality 75 from Lemma 15, we have

$$\|\boldsymbol{M} - c_{k,0}\boldsymbol{w}^{\otimes k/2}(\boldsymbol{w}^{\otimes k/2})^{\mathsf{T}}\|_{\text{op}} \leq \mathcal{C}d^{k/2-1/2}.$$

□

F.2 Proof of Theorem 6 for even case.

This subsection is devoted to the study of the matrix unfolding algorithm in the even case. In all the section, we consider $\ell = 2m$. In the following, we will denote $\hat{T}_\ell \in (\mathbb{R}^d)^{\otimes \ell}$ the tensor of the algorithm, and we denote $M_n \in \mathbb{R}^{d^m\times d^m}$ the unfolded matrix. We begin with the following lemma about the expectation of the unfold matrix.

Lemma 16. *We have the identity*

$$\mathbb{E}[\mathcal{T}_\ell(y,r)\mathcal{H}_\ell(\boldsymbol{z})] = \frac{\beta_{d,\ell}\mathcal{H}_\ell(\boldsymbol{w})}{\sqrt{n_{d,\ell}}}. \tag{76}$$

Proof of 16. We first decompose $\mathbb{E}[\mathcal{T}_\ell(y,r)|\boldsymbol{z}]$ into Gegenbauer basis i.e

$$\mathbb{E}[\mathcal{T}_\ell(y,r)|\boldsymbol{z}] = \sum_{i=0}^{+\infty}\beta_{d,i}Q_i(\langle\boldsymbol{w},\boldsymbol{z}\rangle).$$

We then have the following identity

$$\mathbb{E}[\mathcal{T}_\ell(y,r)\mathcal{H}_\ell(\boldsymbol{z})] = \sum_{i=0}^{+\infty}\beta_{d,i}\mathbb{E}[Q_i(\langle\boldsymbol{w},\boldsymbol{z}\rangle)\mathcal{H}_\ell(\boldsymbol{z})] = \frac{\beta_{d,\ell}\mathcal{H}_\ell(\boldsymbol{w})}{\sqrt{n_{d,\ell}}}.$$

Using the proposition we have

$$\mathbb{E}[\mathcal{T}_\ell(y,r)\mathcal{H}_\ell(\boldsymbol{z})] = \mathcal{C}_\ell \beta_{d,\ell} \boldsymbol{w}^{\otimes \ell} + \boldsymbol{W},$$

where $\|\boldsymbol{W}\|_F = o(d^{-1/2})$. $\square$

Combining the equation (76) and proposition 5, we have that the largest eigenvalue is given by $c_{\ell,0}$ and associated to the eigenvector $\boldsymbol{w}_*^{\otimes m}$. We now control the fluctuations of the unfold matrix. Define $\boldsymbol{W}_i = \mathbf{Mat}(\mathcal{T}_\ell(y_i, r_i)\mathcal{H}_\ell(\boldsymbol{z}_i))$.

Lemma 17. *Let $\delta > 0$. With probability at least $1-\delta$, the following inequality holds*

$$\left\| \frac{1}{n}\sum_{i=1}^n \boldsymbol{W}_i - \mathbb{E}[\boldsymbol{W}] \right\|_{\text{op}} \le \frac{\kappa_\ell d^m}{3n} \ell \log\left(\frac{2d}{\delta}\right) + \sqrt{\frac{4\mathcal{C}_\ell \kappa_\ell^2 d^m \ell \log(\frac{2d}{\delta})}{n}}. \tag{77}$$

With these two results, we can prove the theorem 6.

Proof of theorem 6 for even case. Combining the lemma 16 and lemma 17, and the Davis-Kahan theorem, we have with probability at least $1-\delta$, we have

$$\langle \hat{\boldsymbol{w}}, \boldsymbol{w}_* \rangle^2 \ge 1 - \left(\frac{\frac{\kappa_\ell d^m}{3n}\ell \log\left(\frac{2d}{\delta}\right) + \sqrt{\frac{4\mathcal{C}_\ell \kappa_\ell^2 d^m \ell \log(\frac{2d}{\delta})}{n}}}{\beta_{d,\ell}} \right)^2.$$

Taking $n \ge \frac{\kappa_\ell d^m \ell \log(2d/\delta)}{\|\xi_{d,\ell}\|_2 \varepsilon} \vee 4\frac{\mathcal{C}_\ell \kappa_\ell^2 d^m \ell \log(2d/\delta)}{\|\xi_{d,\ell}\|_2^2 \varepsilon^2}$, we have the desired result. $\square$

We now prove the Lemma 17

Proof of Lemma 17. Denote $\boldsymbol{Z}_i = \mathcal{T}_\ell(y_i, r_i)\mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}_i))$. We bound the variance of $\boldsymbol{M}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i$

$$\begin{aligned}
\sigma^2 &= \left\| \mathbb{E}\left[\boldsymbol{M}_n \boldsymbol{M}_n^\top \right] \right\|_{\text{op}} \\
&= \frac{1}{n^2} \left\| \mathbb{E}\left[\sum_{k=1}^n (\boldsymbol{Z}_k - \mathbb{E}[\boldsymbol{Z}_k]) \sum_{k=1}^n (\boldsymbol{Z}_\ell - \mathbb{E}[\boldsymbol{Z}_\ell])^\top \right] \right\|_{\text{op}} \\
&\le \frac{1}{n} \|\mathbb{E}[(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])^\top]\|_{\text{op}} \\
&\le \frac{1}{n} \|\mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}^\top]\|_{\text{op}} \\
&\le \frac{1}{n} \sup_{\boldsymbol{x} \in \mathbb{S}^{d^m - 1}} \mathbb{E}[\mathcal{T}_\ell(\boldsymbol{y}, \boldsymbol{r})^2 \boldsymbol{x} \mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z})) \mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}))^\top \boldsymbol{x}^\top]. \\
&\le \frac{\|\mathcal{T}_\ell\|_\infty^2}{n} \|\mathbb{E}[\mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z})) \mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}))^\top]\|_{\text{op}}.
\end{aligned}$$

We bound the last term in the following. We then have

$$\mathbf{Mat}(\mathcal{H}_\ell)\mathbf{Mat}(\mathcal{H}_\ell)^\top = \mathrm{Cont}(\mathcal{H}_\ell(\boldsymbol{z}) \otimes \mathcal{H}_\ell(\boldsymbol{z})),$$

where the contraction operation is defined as

$$\mathrm{Cont}(\boldsymbol{T}_1 \otimes \boldsymbol{T}_2)_{(i_1,\ldots,i_m),(j_1,\ldots,j_m)} = \sum_{\ell_1,\ldots,\ell_m} (\boldsymbol{T}_1)_{i_1,\ldots,i_m,\ell_1,\ldots,\ell_m} (\boldsymbol{T}_2)_{\ell_1,\ldots,\ell_m,j_1,\ldots,j_m}.$$

Lemma 18. *There exists an universal constant $\mathcal{C}_\ell$ that only depend on ℓ such that we have*

$$\|\mathbb{E}\left[\mathrm{Cont}(\mathcal{H}_\ell(\boldsymbol{z}) \otimes \mathcal{H}_\ell(\boldsymbol{z}))\right]\|_{\text{op}} \le \mathcal{C}_\ell d^m.$$

Using the lemma 18, we bound σ^2

$$\sigma^2 \leq \frac{\mathcal{C}_\ell \kappa_\ell^2 d^m}{n}.$$

We bound the operator norm almsot surely for all $i \in \{1, \dots, n\}$ by the following

$$\|\mathcal{T}_\ell(y, r)\mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z})) - \mathbb{E}[\mathcal{T}_\ell(y, r)\mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}))]\|_{\text{op}} \leq \frac{\kappa_\ell d^m}{n}.$$

Using Berstein inequality for matrices, with probability at least $1 - \delta$, we have

$$\left\| \frac{1}{n} \sum_{k=1}^n \boldsymbol{X}_\ell - \mathbb{E}[\boldsymbol{X}] \right\|_{\text{op}} \leq \frac{\kappa_\ell d^m}{3n} \ell \log\left(\frac{2d}{\delta}\right) + \sqrt{\frac{4\mathcal{C}_\ell \kappa_k^2 d^m \ell \log(\frac{2d}{\delta})}{n}}.$$

This is the desired result. $\square$

Proof of Claim 18. We have

$$\begin{aligned}
\mathbb{E}[\mathcal{H}_\ell(\boldsymbol{z}) \otimes \mathcal{H}_\ell(\boldsymbol{z})] &= \sum_{q=0}^{\lfloor \ell/2 \rfloor} c_{\ell,q} \mathbb{E}\left[\mathcal{H}_\ell(\boldsymbol{z}) \otimes \text{Sym}(\boldsymbol{z}^{\otimes(\ell-2q')} \otimes \mathbf{I}_d^{\otimes q'})\right] \\
&= c_{\ell,0} \mathbb{E}\left[\mathcal{H}_\ell(\boldsymbol{z}) \otimes \boldsymbol{z}^\ell\right],
\end{aligned}$$

where we used that for all $(i_1, \dots, i_\ell)$, we have $\text{Sym}(\boldsymbol{z}^{\otimes(\ell-2i)} \otimes \mathbf{I}_d^{\otimes i})_{i_1,\dots,i_\ell} \in V_{\ell-2\ell}$, and so we deduce $\mathbb{E}[\mathcal{H}_\ell(\boldsymbol{z})\text{Sym}(\boldsymbol{z}^{\otimes(\ell-2i)} \otimes \mathbf{I}_d^{\otimes i})_{i_1,\dots,i_\ell}] = 0$, which implies that for all $i < \ell$, we have

$$\mathbb{E}[\mathcal{H}_\ell(\boldsymbol{z}) \otimes \text{Sym}(\boldsymbol{z}^{\otimes(\ell-2i)} \otimes \mathbf{I}_d^{\otimes i})] = 0.$$

By definition of the Harmonic tensor,

$$\mathbb{E}[\mathcal{H}_\ell(\boldsymbol{z}) \otimes \mathcal{H}_\ell(\boldsymbol{z})] = c_{\ell,0} \sum_{q=0}^{\lfloor \ell/2 \rfloor} c_{\ell,q} \mathbb{E}\left[\text{Sym}(\boldsymbol{z}^{\otimes(\ell-2q)} \otimes \mathbf{I}_d^{\otimes q}) \otimes \boldsymbol{z}^\ell\right].$$

We state the Wick formula for spherical measure.

Lemma 19 (Wick formula)**.** *Let z be a uniform vector on $\mathbb{S}^{d-1}$, we then have*

$$\mathbb{E}_{\boldsymbol{z}}\left[\prod_{j=1}^{\ell} \boldsymbol{z}_{k_j}\right] = \frac{1_{\ell \equiv 0[2]}}{d(d+2)\cdots(d+2\ell-2)} \sum_{P \in P_\ell^2(\{k_1,\dots,k_\ell\})} \prod_{\{r,s\}} \delta_{k_r,k_s}, \tag{78}$$

where $P_\ell^2(\{k_1, \dots, k_\ell\})$ denotes the set of pairings of the set $\{k_1, \dots, k_\ell\}$.

By Wick formula (19),

$$\begin{aligned}
&\mathbb{E}\left[\text{Sym}(\boldsymbol{z}^{\otimes(\ell-2q)} \otimes \mathbf{I}_d^{\otimes q}) \otimes \boldsymbol{z}^{\otimes \ell}\right]_{(i_1,\dots,i_\ell),(j_1,\dots,j_\ell)} \\
&= \sum_{\sigma \in \mathfrak{S}_\ell} \frac{1}{\ell!} \prod_{k=1}^q \delta_{i_{\sigma(\ell-2q+2k-1)}, i_{\sigma(\ell-2q+2k)}} \mathbb{E}\left[\prod_{k=1}^{\ell-2q} \boldsymbol{z}_{i_\sigma(k)} \prod_{k=1}^{\ell} \boldsymbol{z}_{j_k}\right] \\
&= \frac{1}{\ell!^2 d \cdots (d + 2(\ell - 1 - q))} \sum_{\sigma \in \mathfrak{S}_\ell} \prod_{k=1}^q \delta_{i_{\sigma(\ell-2q+2k-1)}, i_{\sigma(\ell-2q+2k)}} \sum_{P \in P_{2(\ell-q)}^2 (i_{\sigma(\leq \ell-2q)}, j)} \prod_{\{r,s\}} \delta_{\gamma_r,\gamma_s} \\
&\simeq \frac{\mathcal{C}_\ell d^{-\ell+q}}{\ell!} \sum_{\sigma \in \mathfrak{S}_\ell} \prod_{k=1}^q \delta_{i_{\sigma(\ell-2q+2k-1)}, i_{\sigma(\ell-2q+2k)}} \sum_{P \in P_{2(\ell-q)}^2 (i_{\sigma(\leq \ell-2q)}, j)} \prod_{\{r,s\}} \delta_{\gamma_r,\gamma_s},
\end{aligned}$$

where γ is an enumeration of the set $\{i_{\sigma(1)}, \dots, i_{\sigma(\ell-2q)}, j_1, \dots, j_\ell\}$. Let focus on $\mathbb{E}\left[\text{Cont}(\mathcal{H}_\ell(\boldsymbol{z}) \otimes \mathcal{H}_\ell(\boldsymbol{z}))\right]_{(i_1,\dots,i_m),(j_1,\dots,j_m)}$. By linearity, we only need to compute it for

$\mathrm{Sym}(\boldsymbol{z}^{\otimes(\ell-2q)} \otimes \mathbf{I}_d^{\otimes q}) \otimes \boldsymbol{z}^{\ell}$,

$$
\begin{aligned}
&\mathbb{E}\left[\mathrm{Cont}(\mathrm{Sym}(\boldsymbol{z}^{\otimes(\ell-2q)} \otimes \mathbf{I}_d^{\otimes q}) \otimes \boldsymbol{z}^{\ell})\right]_{(i_1,\dots,i_m),(j_1,\dots,j_m)} \\
&= \sum_{k_1,\dots,k_m} \mathbb{E}[\mathrm{Sym}(\boldsymbol{z}^{\otimes(\ell-2q)} \otimes \mathbf{I}_d^{\otimes q}) \otimes \boldsymbol{z}^{\ell}]_{(i_1,\dots,i_m,k_1,\dots,k_m),(j_1,\dots,j_m,k_1,\dots,k_m)} \\
&= \mathcal{C}_\ell d^{-\ell+q} \sum_{k_1,\dots,k_m} \sum_{\sigma\in\mathfrak{S}_\ell} \prod_{k=1}^{q} \delta_{i_{\sigma(\ell-2q+2k-1)}, i_{\sigma(\ell-2q+2k)}} \sum_{P\in P^2_{2(\ell-q)}(i_{\sigma(\leq \ell-2q)}, j)} \prod_{\{r,s\}} \delta_{k_r,k_s} \\
&= \mathcal{C}_\ell d^{-\ell+q} \sum_{k_1,\dots,k_m} \sum_{\sigma\in\mathfrak{S}_\ell} \sum_{P\in P^2_{2(\ell-q)}(i_{\sigma(\leq \ell-2q)}, j)} \prod_{\{r,s\}} \delta_{\gamma_r,\gamma_s} \prod_{k=1}^{q} \delta_{i_{\sigma(\ell-2q+2k-1)}, i_{\sigma(\ell-2q+2k)}} \\
&= \mathcal{C}_\ell d^{-\ell+q} \sum_{\sigma\in\mathfrak{S}_\ell} \sum_{P\in P^2_{2(\ell-q)}(i_{\sigma(\leq \ell-2q)}, j)} \sum_{k_1,\dots,k_m} \prod_{\{r,s\}} \delta_{\gamma_r,\gamma_s} \prod_{k=1}^{q} \delta_{i_{\sigma(\ell-2q+2k-1)}, i_{\sigma(\ell-2q+2k)}}.
\end{aligned}
$$

Consider the matrix $\boldsymbol{M}_{P,\sigma} \in \mathbb{R}^{d^m \times d^m}$ with coefficients given by

$$
(\boldsymbol{M}_{P,\sigma})_{(i_1,\dots,i_m),(j_1,\dots,j_m)} = \sum_{k_1,\dots,k_m} \prod_{k=1}^{q} \delta_{i_{\sigma(\ell-2q+k)}, i_{\sigma(\ell-2q+k+1)}} \prod_{\{r,s\}} \delta_{\gamma_r,\gamma_s}.
$$

The rest of the analysis is devoted to bound the operator norm of these matrices, In the following, we denote $\boldsymbol{M}_P := \boldsymbol{M}_{P,id}$. Fix $P \in P^2_{2(\ell-q)}$, and let assume $\sigma = i_d$ without losing any generality. Denote $\mathcal{L} := \{i \in \{i_1,\dots,i_m,j_1,\dots,j_m\} \colon \exists k_q \text{ paired with } i\}$ i.e the set of indices which are paired with a term $\{k_i\}_{i=1,\dots,m}$, and also denote $n_P = |\{k_i \colon \exists k \text{ such that } k_i \text{ is paired with } k_k\}|$. We then have

$$
\begin{aligned}
(\boldsymbol{M}_P)_{(i_1,\dots,i_m),(j_1,\dots,j_m)} &= \prod_{i\in n_p} (\sum_{k_1,k_2} \delta_{k_1,k_2}) \prod_{i\in\mathcal{L}} (\sum_{k_i} \delta_{k_i,i}) \prod_{k=1\colon \ell\notin\mathcal{L}}^{q} \delta_{i_{k-2q+\ell}, i_{k-2q+k+1}} \prod_{\{r,s\}\in P\colon r,s\notin\mathcal{L}} \delta_{\gamma_r,\gamma_s} \\
&= d^{|n_P|} \prod_{k=1\colon \ell\notin\mathcal{L}}^{q} \delta_{i_{k-2q+k}, i_{k-2q+k+1}} \prod_{\{r,s\}\in P\colon r,s\notin\mathcal{L}} \delta_{\gamma_r,\gamma_s}.
\end{aligned}
$$

We now prove the following inequality. Suppose we have a matrix $\boldsymbol{M} \in \mathbb{R}^{d^m \times d^m}$ with coefficients given by

$$
\boldsymbol{M}_{(i_1,\dots,i_m),(j_1,\dots,j_m)} = \prod_{i=1}^{r} \delta_{c_{2i}, c_{2i+1}},
$$

where c_i is a set of $2r$ different indices of $\{i_1,\dots,i_m,j_1,\dots,j_m\}$. We then have the following

$$
\begin{aligned}
\|\boldsymbol{M}\|_{\mathrm{op}} &\leq \|\boldsymbol{M}\|_F \\
&\leq \sqrt{\sum_{k_1,k_2} \boldsymbol{M}^2_{k_1,k_2}} \\
&\leq d^{m-r/2}.
\end{aligned}
$$

We then have the following inequality

$$
\|\boldsymbol{M}\|_{\mathrm{op}} \leq d^{m-r/2}.
$$

Applying this fact to $\boldsymbol{M}_P$, which has exactly $2\ell - 2n_P - |\mathcal{L}|$ kronecker, we then have

$$
\|\boldsymbol{M}_P\|_{\mathrm{op}} \leq d^{m+2n_P-\ell+|\mathcal{L}|/2}.
$$

Using the fact that $2m = 2n_P + |\mathcal{L}|$, we then deduce that

$$
\|\boldsymbol{M}_P\|_{\mathrm{op}} \leq d^{2m+n_P-\ell} \leq d^{n_P} \leq d^m.
$$

This reasonning hold true for all the matrices $\boldsymbol{M}_P$, we then have by triangular inequality

$$
c_{\ell,0} c_{\ell,q} \left\| \mathbb{E}\left[\mathrm{Cont}(\mathrm{Sym}(\boldsymbol{z}^{\otimes(\ell-2q)} \otimes \mathbf{I}_d^{\otimes q}) \otimes \boldsymbol{z}^{\ell})\right] \right\|_{\mathrm{op}} \leq \mathcal{C}_\ell d^m,
$$

with $\mathcal{C}_\ell$ which is still a universal constant. We now bound the operator norm of $\mathrm{Cont}(\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z}))$, using the triangular inequality, we have the following

$$\begin{aligned}
&\|\mathbb{E}\left[\mathrm{Cont}(\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z}))\right]\|_{\mathrm{op}}\\
&\leq c_{\ell,0}\sum_{q}^{\lfloor k/2\rfloor} c_{k,q}\left\|\mathbb{E}\left[\mathrm{Cont}(\mathrm{Sym}(\boldsymbol{z}^{\otimes(\ell-2q)}\otimes\mathbf{I}_d^{\otimes q})\otimes\boldsymbol{z}^{\ell})\right]\right\|_{\mathrm{op}}\\
&\leq \mathcal{C}_\ell d^m.
\end{aligned}$$

□

We can adapt the proof to the case where $\mathcal{T}_\ell$ satisfies the weaker assumption using similar ideas of the analysis of Section . We are not giving details of this adaptation in this appendix.

F.3 Proof of Proposition 6 for odd case.

Here we focus on $\ell = 2k+1$. In the following, we denote $\boldsymbol{M}_n = \frac{1}{n}\sum_{i=1}^n \mathcal{T}_{2k+1}(y_i,r_i)\mathbf{Mat}\left(\mathcal{H}_{2k+1}(\boldsymbol{z}_i)\right)\in\mathbb{R}^{d^k\times d^{k+1}}$, and consider the matrix $\boldsymbol{T}_n = \boldsymbol{M}_n\boldsymbol{M}_n^\top\in\mathbb{R}^{d^k\times d^k}$, and the matrix $\boldsymbol{M} = \mathbb{E}[\boldsymbol{M}_n]$,

Proof of Proposition 6 for odd case. We first state some lemmas needed for the proof. We follow the idea of [26] and we decompose the matrix of interest into different parts

$$\boldsymbol{T}_n = \boldsymbol{M}\boldsymbol{M}^\top + \boldsymbol{M}(\boldsymbol{M}_n-\boldsymbol{M})^\top + (\boldsymbol{M}_n-\boldsymbol{M})\boldsymbol{M}^\top + (\boldsymbol{M}_n-\boldsymbol{M})(\boldsymbol{M}_n-\boldsymbol{M})^\top. \tag{79}$$

The next lemmas bound the norm of matrices in the above equation.

Lemma 20. *Let $n\geq 0$, with probability at least $1-\delta$, we have*

$$\left\|(\boldsymbol{M}_n-\boldsymbol{M})(\boldsymbol{M}_n-\boldsymbol{M})^\top - \frac{1}{n}\mathbb{E}\left[\mathcal{T}_k(y,r)^2\boldsymbol{Mat}(\mathcal{H}_k(\boldsymbol{z}))\boldsymbol{Mat}(\mathcal{H}_k(\boldsymbol{z}))^\top\right]\right\|_{\mathrm{op}} \leq \sigma_1\vee\sigma_2, \tag{80}$$

where we introduced σ_1 and σ_2 defined as

$$\sigma_1 \leq \frac{\kappa_k^2 d^{2k+1}\log(d/\delta)}{n^2} + \sqrt{\frac{\mathcal{C}_k\kappa_k^4 d^{2k+1}}{n^2} + \frac{\mathcal{C}_k\kappa_k^4 d^{3k+1}\log(d/\delta)}{n^3}}, \tag{81}$$

and

$$\sigma_2 \leq \frac{\kappa_k^2 d^{2k+1}}{3n^2}k\log\left(\frac{2d}{\delta}\right) + \sqrt{\frac{4\mathcal{C}_k\kappa_k^2 d^{2k+1}k\log(\frac{2d}{\delta})}{n^2}}. \tag{82}$$

Lemma 21. *Let $n\geq 0$, with probability at least $1-\delta$, we have*

$$\|(\boldsymbol{M}_n-\boldsymbol{M})\boldsymbol{M}^\top\|_{\mathrm{op}} \leq \frac{\mathcal{C}_{2k+1}d^k\|\mathcal{T}_{2k+1}\|_\infty^2}{n} + \mathcal{C}_{2k+1}\sqrt{\frac{d^{2k}}{n^2}}\log(C/\delta)^k. \tag{83}$$

We also compute the expectation of the unfolded matrix.

Lemma 22. *Denote $\boldsymbol{v}_1\in\mathbb{R}^d$ the singular values associated to the largest left singular values λ_1, we then have*

$$\boldsymbol{v}_1 = \boldsymbol{w}_*^{\otimes 2k+1}, \qquad \lambda_1 \simeq \mathcal{C}_k\beta_{d,k}^2.$$

With the lemmas in hands, we can prove the desired result. Let apply the argument by [26]. Denote $\boldsymbol{u}\in\mathbb{R}^{d^k}$ the top eigenvector of $\boldsymbol{T}_n$, which is also the top eigenvector of $\boldsymbol{M}_n := \boldsymbol{T}_n\boldsymbol{T}_n^\top - cI_{d^k}$. We have with probability at least $1-4\delta$,

$$\begin{aligned}
\boldsymbol{u}^\top\boldsymbol{T}_n\boldsymbol{u} &= \boldsymbol{u}^\top\left(\boldsymbol{M}\boldsymbol{M}^\top + \boldsymbol{M}(\boldsymbol{M}_n-\boldsymbol{M})^\top + (\boldsymbol{M}_n-\boldsymbol{M})\boldsymbol{M}^\top + (\boldsymbol{M}_n-\boldsymbol{M})(\boldsymbol{M}_n-\boldsymbol{M})^\top)\right)\boldsymbol{u}\\
&= \langle\boldsymbol{u},\boldsymbol{M}\boldsymbol{M}^\top\boldsymbol{u}\rangle + \boldsymbol{u}^\top\left(\boldsymbol{M}(\boldsymbol{M}_n-\boldsymbol{M})^\top + (\boldsymbol{M}_n-\boldsymbol{M})\boldsymbol{M}^\top + (\boldsymbol{M}_n-\boldsymbol{M})(\boldsymbol{M}_n-\boldsymbol{M})^\top)\right)\boldsymbol{u}\\
&\leq \beta_{d,k}^2\langle\boldsymbol{w}^{\otimes k},\boldsymbol{u}\rangle + 2\sigma_3 + \sigma_1\vee\sigma_2 + \boldsymbol{u}^\top\frac{1}{n}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top] - \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{X}]^\top\right)\boldsymbol{u}
\end{aligned}$$

We then notice that for $n \geq \frac{Cd^{k/2}\log^2(d)}{\delta\|\xi_{d,l}\|_2^2\sqrt{\varepsilon}}$, we have that

$$\sigma_1 \leq \frac{\kappa_k^2 d^{2k+1}\log(d/\delta)}{n^2} + \sqrt{\frac{\mathcal{C}_k\kappa_k^4 d^{2k+1}}{n^2} + \frac{\mathcal{C}_k\kappa_k^4 d^{3k+1}\log(d/\delta)}{n^3}} \leq \mathcal{C}_k\beta_{d,k}^2\varepsilon,$$

and we also have

$$\sigma_2 \leq \frac{\kappa_k^2 d^{2k+1}}{3n^2}k\log\left(\frac{2d}{\delta}\right) + \sqrt{\frac{4\mathcal{C}_k\kappa_k^2 d^{2k+1}k\log(\frac{2d}{\delta})}{n^2}} \leq \mathcal{C}_k\beta_{d,k}^2\varepsilon.$$

So, we deduce that

$$\sigma_2 \vee \sigma_1 \leq \mathcal{C}_k\beta_{d,k}^2\varepsilon.$$

We also have the following bound

$$\sigma_3 \leq \frac{\mathcal{C}_{2k+1}d^k\|\mathcal{T}_{2k+1}\|_\infty^2}{n} + \mathcal{C}_{2k+1}\sqrt{\frac{d^{2k}}{n^2}}\log(C/\delta)^k \leq \mathcal{C}_{2k+1}\beta_{d,k}^2\varepsilon.$$

Combining the bounds, we then have

$$\boldsymbol{u}^\mathsf{T}\boldsymbol{T}_n\boldsymbol{u} \leq \beta_{d,k}^2\langle\boldsymbol{w}^{\otimes k}, \boldsymbol{u}\rangle + \mathcal{C}_{2k+1}\beta_{d,k}^2\varepsilon$$

We also use that $\frac{1}{n}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\mathsf{T}] - \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{X}]^\mathsf{T}\right)$ can be decomposed into three different parts; one part aligned with $\boldsymbol{w}^{\otimes k}$, a diagonal part $c\mathbf{I}_{d^k}$ and $\boldsymbol{W} \in \mathbb{R}^{d^k \times d^k}$, and we have that $\|\boldsymbol{W}\|_{\mathrm{op}} \leq \frac{d^{k+1/2}}{n}$. Using Courant-Fisher theorem, we have

$$\boldsymbol{u}^\mathsf{T}\boldsymbol{T}_n\boldsymbol{u} \geq (\boldsymbol{w}^{\otimes k})^\mathsf{T}\boldsymbol{T}_n(\boldsymbol{w}^{\otimes k}) \geq \lambda^2 - O(\varepsilon\lambda^2).$$

So we end up with

$$\langle\boldsymbol{w}^{\otimes k}, \boldsymbol{u}\rangle \geq 1 - O(\varepsilon).$$

□

F.4 Proof of the lemmas.

We now provide proofs of the lemmas (20).

Proof of Lemma.20. For the sake of clarity, in the following we denote the matrix $\boldsymbol{X}_i := \mathcal{T}(\boldsymbol{y}_i, \boldsymbol{r}_i)\mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}_i))$, and denote $\boldsymbol{Y}_i = \boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i]$. We notice that the matrix $(\boldsymbol{M}_n - \boldsymbol{M})(\boldsymbol{M}_n - \boldsymbol{M})^\mathsf{T}$ is a U-statistic

$$(\boldsymbol{M}_n - \boldsymbol{M})(\boldsymbol{M}_n - \boldsymbol{M})^\mathsf{T} = \frac{1}{n^2}\sum_{1\leq i,j\leq n}(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])(\boldsymbol{X}_j - \mathbb{E}[\boldsymbol{X}_j])^\mathsf{T}. \tag{84}$$

We decompose this sum as

$$\frac{1}{n^2}\sum_{i=1}^n(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])^\mathsf{T} + \frac{1}{n^2}\sum_{1\leq i\neq j\leq n}(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])(\boldsymbol{X}_j - \mathbb{E}[\boldsymbol{X}_j])^\mathsf{T}. \tag{85}$$

We now have the two lemmas.

Lemma 23. *With probability at least $1 - 2\delta$, we have*

$$\left\|\frac{1}{n^2}\sum_{i=1}^n\boldsymbol{Y}_i\boldsymbol{Y}_i^\mathsf{T} - \frac{1}{n}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\mathsf{T}] - \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{X}]^\mathsf{T}\right)\right\|_{\mathrm{op}} \leq \sigma_1, \tag{86}$$

where $\sigma_1 = \frac{\kappa_k^2 d^{2k+1}\log(d/\delta)}{n^2} + \sqrt{\frac{\mathcal{C}_k\kappa_k^4 d^{2k+1}}{n^2} + \frac{\mathcal{C}_k\kappa_k^4 d^{3k+1}\log(d/\delta)}{n^3}}$, and

$$\left\|\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])(\boldsymbol{X}_j - \mathbb{E}[\boldsymbol{X}_j])^\mathsf{T}\right\|_{\mathrm{op}} \leq \sigma_2, \tag{87}$$

where $\sigma_2 = \frac{\kappa_k^2 d^{2k+1}}{3n^2}k\log\left(\frac{2d}{\delta}\right) + \sqrt{\frac{4\mathcal{C}_k\kappa_k^2 d^{2k+1}k\log(\frac{2d}{\delta})}{n^2}}$.

Using the two bounds (86) and (87), with probability at least $1-2\delta$, we have

$$\left\|(\boldsymbol{M}_n-\boldsymbol{M})(\boldsymbol{M}_n-\boldsymbol{M})^\mathsf{T}-\frac{1}{n}\mathbb{E}\left[\mathcal{T}(y,r)^2\mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}))\mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}))^\mathsf{T}\right]\right\|_{\mathrm{op}}\leq\sigma_1\vee\sigma_2. \quad (88)$$

□

Proof of Lemma 23. Let bound the matrix variance statistics of the sum

$$\begin{aligned}\sigma^2&=\sup_{\boldsymbol{x}}\frac{1}{n^4}\mathbb{E}\left[\boldsymbol{x}^\mathsf{T}\sum_{i,j}^n\boldsymbol{Y}_i\boldsymbol{Y}_i^\mathsf{T}\boldsymbol{Y}_j\boldsymbol{Y}_j^\mathsf{T}\right]\boldsymbol{x}\\&=\sup_{\boldsymbol{x}}\frac{1}{n^3}\mathbb{E}\left[\boldsymbol{x}^\mathsf{T}\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\right]\boldsymbol{x}+\frac{1}{n^2}\mathbb{E}\left[\boldsymbol{x}^\mathsf{T}\boldsymbol{Y}_1\boldsymbol{Y}_1^\mathsf{T}\boldsymbol{Y}_2\boldsymbol{Y}_2^\mathsf{T}\right]\boldsymbol{x}\\&=\sup_{\boldsymbol{x}}\frac{1}{n^3}\mathbb{E}\left[\boldsymbol{x}^\mathsf{T}\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\right]\boldsymbol{x}+\frac{1}{n^2}\boldsymbol{x}^\mathsf{T}\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\right]\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\right]\boldsymbol{x}\end{aligned}$$

We expand this and we end up bounding the following quantities

$$\begin{aligned}&\mathbb{E}[\mathcal{T}(\boldsymbol{y},\boldsymbol{r})^4\boldsymbol{x}^\mathsf{T}\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))^\mathsf{T}\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))^\mathsf{T}\boldsymbol{x}]\\&\leq\|\mathcal{T}\|_\infty^4\sup_{\boldsymbol{x}}\mathbb{E}[\boldsymbol{x}^\mathsf{T}\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))^\mathsf{T}\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))^\mathsf{T}\boldsymbol{x}].\end{aligned}$$

Lemma 24. *We have the following bound*

$$\|\mathbb{E}[\boldsymbol{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))\boldsymbol{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))^\mathsf{T}\boldsymbol{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))\boldsymbol{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))^\mathsf{T}]\|_{\mathrm{op}}\leq\mathcal{C}_k d^{3k+3/2}.$$

So, we obtain the following bound

$$\frac{1}{n^3}\mathbb{E}\left[\boldsymbol{x}^\mathsf{T}\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\right]\boldsymbol{x}\leq\frac{\mathcal{C}_k\|\mathcal{T}_\ell\|_\infty^4 d^{3k+3/2}}{n^3}.$$

For the other term, using the same reasonning we have

$$\frac{1}{n^2}\boldsymbol{x}^\mathsf{T}\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\right]\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\right]\boldsymbol{x}\leq\frac{\mathcal{C}_k\|\mathcal{T}_\ell\|_\infty^4 d^{2k+1}}{n^2}$$

We then have the following bound on the matrix variance statistic

$$\sigma^2\leq\frac{\mathcal{C}_k\|\mathcal{T}_\ell\|_\infty^4 d^{2k+1}}{n^2}+\frac{\mathcal{C}_k\|\mathcal{T}_\ell\|_\infty^4 d^{3k+3/2}}{n^3}.$$

Let now bound the operator norm almost surely for all $i\in\{1,\dots,n\}$

$$\left\|\frac{1}{n^2}\boldsymbol{Y}_i\boldsymbol{Y}_i^\mathsf{T}\right\|_{\mathrm{op}}\leq\frac{\|\mathcal{T}_\ell\|_\infty^2 d^{2k+1}}{n^2}.$$

Using Berstein's inequality, with probability at least $1-\delta$, we have

$$\begin{aligned}&\left\|\frac{1}{n^2}\sum_{i=1}^n\boldsymbol{Y}_i\boldsymbol{Y}_i^\mathsf{T}-\frac{1}{n}\left(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\mathsf{T}]-\mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{X}]^\mathsf{T}\right)\right\|_{\mathrm{op}}\\&\leq\frac{\|\mathcal{T}_k\|_\infty^2 d^{2k+1}\log(d/\delta)}{n^2}+\sqrt{\frac{\mathcal{C}_k\|\mathcal{T}_\ell\|_\infty^4 d^{2k+1}}{n^2}+\frac{\mathcal{C}_k\|\mathcal{T}_\ell\|_\infty^4 d^{3k+1}\log(d/\delta)}{n^3}}\\&\leq\frac{\kappa_k^2 d^{2k+1}\log(d/\delta)}{n^2}+\sqrt{\frac{\mathcal{C}_k\kappa_k^4 d^{2k+1}}{n^2}+\frac{\mathcal{C}_k\kappa_k^4 d^{3k+1}\log(d/\delta)}{n^3}}.\end{aligned}$$

This proves the desired claim. We prove the second identity. We proceed by first decoupling the matrices (see [45, 19] Chapter 6.1), and we then focus on the $\sum_{1=i\neq j\leq n}\boldsymbol{X}_i\boldsymbol{X}_j'$, where $\boldsymbol{X}_j$ are i.i.d copies of $\boldsymbol{X}_i$, and rewriting the sum like $\frac{1}{n^2}\sum_{i=1}^n\left(\boldsymbol{X}_i\sum_{j\neq i}\boldsymbol{X}_j'\right)$,

$$\boldsymbol{P}\left(\left\|\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\boldsymbol{Y}_i\boldsymbol{Y}_j^\mathsf{T}\right\|_{\mathrm{op}}\geq t\right)\leq K\boldsymbol{P}\left(\left\|\frac{1}{n^2}\sum_{i=1}^n\left(\boldsymbol{Y}_i\sum_{j\neq i}\boldsymbol{Y}_j'^\mathsf{T}\right)\right\|_{\mathrm{op}}\geq t\right),$$

We use the following observation

$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\boldsymbol{Y}_i\sum_{j\neq i}\boldsymbol{Y}_j'^{\mathsf{T}}\right)=\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Y}_i\right)\cdot\left(\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{Y}_j'^{\mathsf{T}}\right)-\frac{1}{n^2}\sum_{i=1}^{n}\boldsymbol{Y}_i\boldsymbol{Y}_i'^{\mathsf{T}}.$$

We now use Bernstein's inequality for each of the terms, and we get that with probability with at least $1-2K\delta$, we have

$$\left\|\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\boldsymbol{Y}_i\boldsymbol{Y}_j^{\mathsf{T}}\right\|_{\mathrm{op}}\leq\sigma_2, \tag{89}$$

where σ_2 is given by

$$\sigma_2=\frac{\kappa_k^2 d^{2k+1}}{3n^2}k\log\left(\frac{2d}{\delta}\right)+\sqrt{\frac{4\mathcal{C}_k\kappa_k^2 d^{2k+1}k\log(\frac{2d}{\delta})}{n^2}}.$$

$\square$

Proof of Lemma 21. We want to bound the following quantity

$$\begin{aligned}
&\|(\boldsymbol{M}_n-\boldsymbol{M})\boldsymbol{M}^{\mathsf{T}}\|_{\mathrm{op}}\\
&=\left\|(\boldsymbol{M}_n-\boldsymbol{M})\frac{\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{w}))^{\mathsf{T}}}{\sqrt{n_{d,k}}}\right\|_{\mathrm{op}}\\
&=\mathcal{C}_k\beta_{d,k}\|(\boldsymbol{M}_n-\boldsymbol{M})^{\mathsf{T}}\boldsymbol{w}^{\otimes k+1}(\boldsymbol{w}^{\otimes k})^{\mathsf{T}}\|_{\mathrm{op}}+\left\|(\boldsymbol{M}_n-\boldsymbol{M})\left(\frac{\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{w}))^{\mathsf{T}}}{\sqrt{n_{d,k}}}-\mathcal{C}_k\beta_{d,k}\boldsymbol{w}^{\otimes k+1}(\boldsymbol{w}^{\otimes k})^{\mathsf{T}}\right)\right\|_{\mathrm{op}}\\
&=\mathcal{C}_k\beta_{d,k}\|(\boldsymbol{M}_n-\boldsymbol{M})\boldsymbol{w}^{\otimes k+1}\|_2+\|(\boldsymbol{M}_n-\boldsymbol{M})\|_{\mathrm{op}}\left\|\left(\frac{\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{w}))^{\mathsf{T}}}{\sqrt{n_{d,k}}}-\mathcal{C}_k\beta_{d,k}\boldsymbol{w}^{\otimes k+1}(\boldsymbol{w}^{\otimes k})^{\mathsf{T}}\right)\right\|_{\mathrm{op}}\\
&\leq\mathcal{C}_k\beta_{d,k}\|(\boldsymbol{M}_n-\boldsymbol{M})\boldsymbol{w}^{\otimes k+1}\|_2+\mathcal{C}_k d^{-1}\left\|\boldsymbol{M}_n-\boldsymbol{M}\right\|_{\mathrm{op}}.
\end{aligned}$$

We then have with probability at least $1-\delta$,

$$\|\boldsymbol{M}_n-\boldsymbol{M}\|_{\mathrm{op}}\leq\frac{\kappa_k d^{k+1}}{3n}k\log\left(\frac{2d}{\delta}\right)+\sqrt{\frac{4\mathcal{C}_k\kappa_k^2 d^{k+1}k\log(\frac{2d}{\delta})}{n}},$$

Now, we focus on the first term, without losing any generality, we assume $\boldsymbol{w}^*=e_1$. We define the function $f(\boldsymbol{z})=\|(\boldsymbol{M}_n(\boldsymbol{z})-\boldsymbol{M}))e_1^{\otimes k+1}\|^2$, we notice that f is a degree k polynomial of the sphere, we then use Hypercontractivity lemma 25 to state the following tail bound

$$\mathbb{P}\left(|f(\boldsymbol{z})-\mathbb{E}[f(\boldsymbol{z})]|\geq t\right)\leq ce^{-\left(\frac{t^2}{C\mathrm{Var}(f(\boldsymbol{z}))}\right)^{1/k}}, \tag{90}$$

where c, C are universal constants, and $\mathrm{Var}(f(\boldsymbol{z})$ is the variance of $f(\boldsymbol{z})$. We first compute the expectation of f,

$$\begin{aligned}
\mathbb{E}[f(\boldsymbol{z})]&=\mathbb{E}[\|(\boldsymbol{M}_n-\boldsymbol{M})e_1^{\otimes k+1}\|^2]\\
&=\frac{1}{n}\mathbb{E}\left[\mathcal{T}_{2k+1}(y,r)^2\|\left(\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))-\mathbb{E}[\boldsymbol{Y}(\boldsymbol{z})]\right)e_1^{\otimes k+1}\|^2\right]\\
&\leq\frac{\|\mathcal{T}_{2k+1}\|_\infty^2}{n}\sum_{i_1,\ldots,i_k}\mathbb{E}[\mathbf{Mat}(\mathcal{H}_{2k+1}(\boldsymbol{z}))^2_{i_1,\ldots,i_k,1,\ldots,1}]\\
&\leq\frac{\mathcal{C}_k d^k\|\mathcal{T}_{2k+1}\|_\infty^2}{n},
\end{aligned}$$

where we used the same computations that in proof of the even case to state that $\mathbb{E}[\mathbf{Mat}(\mathcal{H}_{2k+1})(\boldsymbol{z})^2_{i_1,\ldots,i_k,1,\ldots,1}] \leq \mathcal{C}_k$. We now compute the variance parameter of f,

$$\begin{aligned}
\mathrm{Var}(f(\boldsymbol{z})) &\leq \mathbb{E}[\|(\boldsymbol{M}_n - \boldsymbol{M})e_1^{\otimes k+1}\|^4] \\
&= \mathbb{E}\left[\left(\frac{1}{n^2}\sum_{i\neq j}\langle \boldsymbol{Y}_i e_1^{\otimes k+1}, \boldsymbol{Y}_j e_1^{\otimes k+1}\rangle\right)^2\right] \\
&= \frac{1}{n^4}\mathbb{E}\left[\sum_{i_1\neq j_1, i_2\neq j_2}\left(\langle \boldsymbol{Y}_{i_1} e_1^{\otimes k+1}, \boldsymbol{Y}_{j_1} e_1^{\otimes k+1}\rangle\right)\left(\langle \boldsymbol{Y}_{i_2} e_1^{\otimes k+1}, \boldsymbol{Y}_{j_2} e_1^{\otimes k+1}\rangle\right)\right] \\
&= \frac{1}{n^2}\mathbb{E}\left[\langle \boldsymbol{Y}_1 e_1^{\otimes k+1}, \boldsymbol{Y}_2 e_1^{\otimes k+1}\rangle^2\right] \\
&\leq \mathcal{C}_k \frac{d^{2k}}{n^2}.
\end{aligned}$$

Combining 90 with the two above bounds, we then have with probability at least $1-\delta$,

$$\|(\boldsymbol{M}_n - \boldsymbol{M})\boldsymbol{M}^\top\|_{\mathrm{op}} \leq \frac{\mathcal{C}_{2k+1} d^k \kappa_k^2}{n} + \mathcal{C}_{2k+1}\frac{d^k}{n}\kappa_k^2 \log(C/\delta)^k.$$

This is the desired result. $\square$

Proof of Lemma 22. The proof is very similar to the proof of lemma 16. $\square$

Proof of Lemma 24. The proof is similar in spirit to lemma 18, however we need to control an order 4 tensor. In this proof, we use the following shortcuts $\mathcal{S}_q(\boldsymbol{z}) := \mathrm{Sym}(\boldsymbol{z}^{\otimes \ell - 2q} \otimes \mathbf{I}_d^{\otimes q})$. We then bound this in the following way

$$\mathbb{E}[\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})] = \sum_{q_1,q_2,q_3,q_4} c_{\ell,q_1}c_{\ell,q_2}c_{\ell,q_3}c_{\ell,q_4}\mathbb{E}[\mathcal{S}_{q_1}(\boldsymbol{z})\otimes\mathcal{S}_{q_2}(\boldsymbol{z})\otimes\mathcal{S}_{q_3}(\boldsymbol{z})\otimes\mathcal{S}_{q_4}(\boldsymbol{z})].$$

We denote $\boldsymbol{i}^1 = (i_1, \ldots, i_\ell)$, and the same for $\boldsymbol{i}^2, \boldsymbol{i}^3$, and $\boldsymbol{i}^4$. We can focus on the term inside expectation, and we get by expanding this term as in proof of lemma

$$\begin{aligned}
&\mathbb{E}[\mathcal{S}_{q_1}(\boldsymbol{z})\otimes\mathcal{S}_{q_2}(\boldsymbol{z})\otimes\mathcal{S}_{q_3}(\boldsymbol{z})\otimes\mathcal{S}_{q_4}(\boldsymbol{z})]_{\boldsymbol{i}^1,\boldsymbol{i}^2,\boldsymbol{i}^3,\boldsymbol{i}^4} \\
&= \sum_{\sigma_1,\sigma_2,\sigma_3,\sigma_4\in\mathfrak{S}_k}\frac{1}{\ell!^4}\mathbb{E}\left[\prod_{j=1}^{\ell-2q_1}\boldsymbol{z}_{i^1_{\sigma_1(j)}}\prod_{j=1}^{\ell-2q_2}\boldsymbol{z}_{i^2_{\sigma_2(j)}}\prod_{j=1}^{\ell-2q_3}\boldsymbol{z}_{i^3_{\sigma_3(j)}}\prod_{j=1}^{\ell-2q_4}\boldsymbol{z}_{i^4_{\sigma_4(j)}}\right]\prod_{l=1}^{q_1}\delta_{i^1_{\sigma_1(\ell-2q_1+2l-1)},i^1_{\sigma_1(\ell-2q_1+2l)}} \\
&\times\prod_{l=1}^{q_2}\delta_{i^2_{\sigma_2(\ell-2q_2+2l-1)},i^2_{\sigma_2(\ell-2q_2+2l)}}\prod_{l=1}^{q_3}\delta_{i^3_{\sigma_3(\ell-2q_3+2l-1)},i^3_{\sigma_3(\ell-2q_3+2l)}}\prod_{l=1}^{q_4}\delta_{i^4_{\sigma_4(\ell-2q_4+2l-1)},i^4_{\sigma_4(\ell-2q_4+2l)}} \\
&\simeq \mathcal{C}_\ell d^{2\ell-q_1-q_2-q_3-q_4}\sum_{\sigma_1,\sigma_2,\sigma_3,\sigma_4\in\mathfrak{S}_\ell}\sum_{P\in P^2_{4\ell-2q_1-2q_2-2q_3-2q_4}}\prod_{\{r,s\}}\delta_{\gamma_r,\gamma_s}\prod_{l=1}^{q_1}\delta_{i^1_{\sigma_1(\ell-2q_1+2l-1)},i^1_{\sigma_1(\ell-2q_1+2l)}} \\
&\times\prod_{l=1}^{q_2}\delta_{i^2_{\sigma_2(\ell-2q_2+2l-1)},i^2_{\sigma_2(\ell-2q_2+2l)}}\prod_{l=1}^{q_3}\delta_{i^3_{\sigma_3(\ell-2q_3+2l-1)},i^3_{\sigma_3(\ell-2q_3+2l)}}\prod_{l=1}^{q_4}\delta_{i^4_{\sigma_4(\ell-2q_4+2l-1)},i^4_{\sigma_4(\ell-2q_4+2l)}},
\end{aligned}$$

where we introduced γ is an enumeration of the set over the partition. We notice that we have

$$\mathrm{Cont}\left(\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})\right) = \mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}))\mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}))^\top\mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}))\mathbf{Mat}(\mathcal{H}_\ell(\boldsymbol{z}))^\top.$$

We compute $\mathbb{E}[\mathrm{Cont}\,(\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z}))]_{(i_1,\dots,i_k),(j_1,\dots,j_k)}$,

$$
\begin{aligned}
&\mathbb{E}\left[\mathcal{S}_{q_1}(\boldsymbol{z})\otimes\mathcal{S}_{q_2}(\boldsymbol{z})\otimes\mathcal{S}_{q_3}(\boldsymbol{z})\otimes\mathcal{S}_{q_4}(\boldsymbol{z})\right]_{i_1,\dots,i_k),(j_1,\dots,j_k)}\\
&=\sum_{\boldsymbol{u}^1,\boldsymbol{u}^2,\boldsymbol{u}^3}\mathbb{E}\left[\mathcal{S}_{q_1}(\boldsymbol{z})\otimes\mathcal{S}_{q_2}(\boldsymbol{z})\otimes\mathcal{S}_{q_3}(\boldsymbol{z})\otimes\mathcal{S}_{q_4}(\boldsymbol{z})\right]_{(\boldsymbol{i},\boldsymbol{u}^1),(\boldsymbol{u}^1,\boldsymbol{u}^2),(\boldsymbol{u}^2,\boldsymbol{u}^3),(\boldsymbol{u}^3,\boldsymbol{j})}\\
&\simeq\mathcal{C}_\ell d^{2\ell-q_1-q_2-q_3-q_4}\sum_{\sigma_1,\sigma_2,\sigma_3,\sigma_4\in\mathfrak{S}_\ell}\;\sum_{P\in P^2_{4\ell-2q_1-2q_2-2q_3-2q_4}}\;\sum_{\boldsymbol{u}^1,\boldsymbol{u}^2,\boldsymbol{u}^3}\prod_{\{r,s\}}\delta_{\gamma_r,\gamma_s}\prod_{l=1}^{q_1}\delta_{i^1_{\sigma_1(\ell-2q_1+2l-1)},i^1_{\sigma_1(\ell-2q_1+2l)}}\\
&\times\prod_{l=1}^{q_2}\delta_{i^2_{\sigma_2(\ell-2q_2+2l-1)},i^2_{\sigma_2(\ell-2q_2+2l)}}\prod_{l=1}^{q_3}\delta_{i^3_{\sigma_3(\ell-2q_3+2l-1)},i^3_{\sigma_3(\ell-2q_3+2l)}}\prod_{l=1}^{q_4}\delta_{i^4_{\sigma_4(\ell-2q_4+2l-1)},i^4_{\sigma_4(\ell-2q_4+2l)}}.
\end{aligned}
$$

Let define the matrix $\boldsymbol{M}_{P,\sigma}$ with coefficients given by

$$
\begin{aligned}
(\boldsymbol{M}_{P,\sigma})_{(i_1,\dots,i_k),(j_1,\dots,j_k)}=&\sum_{\boldsymbol{u}^1,\boldsymbol{u}^2,\boldsymbol{u}^3}\prod_{\{r,s\}}\delta_{\gamma_r,\gamma_s}\prod_{l=1}^{q_1}\delta_{i^1_{\sigma_1(\ell-2q_1+2l-1)},i^1_{\sigma_1(\ell-2q_1+2l)}}\prod_{l=1}^{q_2}\delta_{i^2_{\sigma_2(\ell-2q_2+2l-1)},i^2_{\sigma_2(\ell-2q_2+2l)}}\\
&\times\prod_{l=1}^{q_3}\delta_{i^3_{\sigma_3(\ell-2q_3+2l-1)},i^3_{\sigma_3(\ell-2q_3+2l)}}\prod_{l=1}^{q_4}\delta_{i^4_{\sigma_4(\ell-2q_4+2l-1)},i^4_{\sigma_4(\ell-2q_4+2l)}}.
\end{aligned}
$$

Using the same reasoning as in the proof of lemma 18, we are bounding the operator norm of this matrix. We denote $\boldsymbol{M}_P:=\boldsymbol{M}_{P,id}$. Fix $P\in P\in P^2_{4\ell-2q_1-2q_2-2q_3-2q_4}$, and let assume $\sigma=i_d$ without losing any generality. Denote $\mathcal{L}:=\{i\in\{i_1,\dots,i_k,j_1,\dots,j_k\}\colon\exists p\in\{\boldsymbol{u}^1,\boldsymbol{u}^2,\boldsymbol{u}^3\}\text{ paired with }i\}$ i.e the set of indices which are paired with a term $\{\boldsymbol{u}^1,\boldsymbol{u}^2,\boldsymbol{u}^3\}$, and also denote $n_P=\{p\in\{\boldsymbol{u}^1,\boldsymbol{u}^2,\boldsymbol{u}^3\}\colon\exists u\in\{\boldsymbol{u}^1,\boldsymbol{u}^2,\boldsymbol{u}^3\}\text{ such that }p\text{ is paired with }u\}$. In the following, we make some abuse of notations and use i to denote an enumeration of $\{i_1,\dots,i_m,j_1,\dots,j_m\}$ for the sake of clarity, we then have

$$
\begin{aligned}
&(\boldsymbol{M}_P)_{(i_1,\dots,i_k),(j_1,\dots,j_k)}\\
&=\sum_{\boldsymbol{u}^1,\boldsymbol{u}^2,\boldsymbol{u}^3}\prod_{\{r,s\}}\delta_{\gamma_r,\gamma_s}\prod_{l=1}^{q_1}\delta_{i^1_{\sigma_1(\ell-2q_1+2l-1)},i^1_{\sigma_1(\ell-2q_1+2l)}}\prod_{l=1}^{q_2}\delta_{i^2_{\sigma_2(\ell-2q_2+2l-1)},i^2_{\sigma_2(\ell-2q_2+2l)}}\\
&\times\prod_{l=1}^{q_3}\delta_{i^3_{\sigma_3(\ell-2q_3+2l-1)},i^3_{\sigma_3(\ell-2q_3+2l)}}\prod_{l=1}^{q_4}\delta_{i^4_{\sigma_4(\ell-2q_4+2l-1)},i^4_{\sigma_4(\ell-2q_4+2l)}}.\\
&=\prod_{i\in n_p}\Big(\sum_{q_1,q_2}\delta_{q_1,q_2}\Big)\prod_{i\in\mathcal{L}}\Big(\sum_{q_i}\delta_{\ell_i,i}\Big)\prod_{\{r,s\}\in P\colon r,s\notin\mathcal{L}}\delta_{\gamma_r,\gamma_s}\prod_{n=1\colon n\notin\mathcal{L}}^{q_1}\delta_{i_{\sigma_1(\ell-2q_1+2n-1)},i_{\sigma_1(\ell-2q_1+2n)}}\\
&\times\prod_{n=1\colon n\notin\mathcal{L}}^{q_2}\delta_{i_{\sigma_2(\ell-2q_2+2n-1)},i_{\sigma_2(\ell-2q_2+2n)}}\prod_{n=1\colon n\notin\mathcal{L}}^{q_3}\delta_{i_{\sigma_3(\ell-2q+2n)},i_{\sigma_3(\ell-2q+2n+1)}}\prod_{n=1\colon n\notin\mathcal{L}}^{q_4}\delta_{i_{\sigma_4(\ell-2q_4+2n-1)},i_{\sigma_4(\ell-2q_4+2n)}}\\
&=d^{|n_P|}\prod_{\{r,s\}\in P\colon r,s\notin\mathcal{L}}\delta_{\gamma_r,\gamma_s}\prod_{n=1\colon n\notin\mathcal{L}}^{q_1}\delta_{i_{\sigma_1(\ell-2q_1+2n-1)},i_{\sigma_1(\ell-2q_1+2n)}}\prod_{n=1\colon n\notin\mathcal{L}}^{q_2}\delta_{i_{\sigma_2(\ell-2q_2+2n-1)},i_{\sigma_2(\ell-2q_2+2n)}}\\
&\times\prod_{n=1\colon n\notin\mathcal{L}}^{q_3}\delta_{i_{\sigma_3(\ell-2q+2n)},i_{\sigma_3(\ell-2q+2n+1)}}\prod_{n=1\colon n\notin\mathcal{L}}^{q_4}\delta_{i_{\sigma_4(\ell-2q_4+2n-1)},i_{\sigma_4(\ell-2q_4+2n)}}
\end{aligned}
$$

We proceed as in the proof of lemma 18, and we overall get the following inequality

$$\|\boldsymbol{M}_P\|_{\mathrm{op}}\leq\mathcal{C}_\ell d^{3k+3/2}.$$

Using triangular inequality, we end up with the following bound

$$\left\|\mathbb{E}\left[\mathrm{Cont}(\mathcal{S}_{q_1}(\boldsymbol{w})\otimes\mathcal{S}_{q_2}(\boldsymbol{w})\otimes\mathcal{S}_{q_3}(\boldsymbol{w})\otimes\mathcal{S}_{q_4}(\boldsymbol{w})\right]\right\|_{\mathrm{op}}\leq\mathcal{C}_\ell d^{3k+3/2}.$$

Combining all the bounds, we then have the desired statement

$$\left\|\mathbb{E}\left[\mathrm{Cont}\left(\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})\otimes\mathcal{H}_\ell(\boldsymbol{z})\right)\right]\right\|_{\mathrm{op}}\leq\mathcal{C}_\ell d^{3k+3/2}.$$

$\square$

G Proofs for Gaussian SIMs

In this, we shall prove the results from Section 4. We first start by showing the rates on L^2 norm of coefficients $\xi_{d,\ell}$ for a Gaussian SIM of generative exponent $\mathsf{k}_\star$ (cf. Lemma 1). We now prove our lemma on the rates of $\xi_{d,\ell}$ when one is allowed to exploit the norm in the query.

Proof of Lemma 1. We start by recalling that, from [15], Generative exponent $\mathsf{k}_\star(\rho)$, is only defined for ρ whose ν_d is such that $\nu_d \ll \bar{\nu}_{d,0}$, where $\bar{\nu}_{d,0} := \nu_{d,Y} \otimes \chi_d \otimes \tau_{d,1}$ is completely decoupled null, and $\|\frac{\mathrm{d}\nu_d}{\mathrm{d}\bar{\nu}_{d,0}}\|_{L^2(\bar{\nu}_{d,0})}$ is bounded by a constant independent of d. Let $\{\nu_d\}_d$ be a sequence of problems for Gaussian SIM ρ, i.e. $\nu_R = \chi_d$ and $\nu_d(Y \mid Z, R) = \rho(Y \mid Z \cdot R) = \rho(Y \mid X)$. Let $\bar{\nu}_{d,0} = \nu_Y \otimes \nu_R \otimes \nu_Z$. Let us consider the log-likelihood ratio decomposition in $L^2(\bar{\nu}_{d,0})$ identical to the one in [15, Lemma D.1]

$$\frac{\mathrm{d}\nu_d}{\mathrm{d}\bar{\nu}_{d,0}}(y,r,z) - 1 \overset{L^2(\bar{\nu}_{d,0})}{=} \sum_{k \geq \mathsf{k}_\star} \zeta_k(y)\mathrm{He}_k(r \cdot z)\,, \quad \zeta_k(y) = \mathbb{E}_{(Y,R,Z)\sim\nu_d}[\mathrm{He}_k(R \cdot Z) \mid Y = y].$$

Denote $\lambda_k = \|\zeta_k\|_{\nu_Y}$, which are completely determined the model ρ independent of d, and by definition of generative exponent (2) we have $\lambda_{\mathsf{k}_\star}^2 > 0$. We now use the decomposition of Hermite into Gegenbauer polynomials from Proposition 2 to rewrite the above as

$$\frac{\mathrm{d}\nu_d}{\mathrm{d}\bar{\nu}_{d,0}}(y,r,z) - 1 \overset{L^2(\bar{\nu}_{d,0})}{=} \sum_{k \geq \mathsf{k}_\star} \zeta_k(y) \sum_{\ell=0}^{k} \beta_{k,\ell}(r) Q_\ell^{(d)}(z) = \sum_{\ell=0}^{\infty} Q_\ell^{(d)}(z) \sum_{k \geq \mathsf{k}_\star} \beta_{k,\ell}(r)\zeta_k(y)\,.$$

We can now expand the same directly in the Gegenbauer basis

$$\frac{\mathrm{d}\nu_d}{\mathrm{d}\bar{\nu}_{d,0}}(y,r,z) \overset{L^2(\bar{\nu}_{d,0})}{=} \bar{\xi}_{d,0}(y,r) + \sum_{\ell \geq 1} \bar{\xi}_{d,0}(y,r)\xi_{d,\ell}(y,r) Q_\ell^{(d)}(z)\,,$$

where

$$\bar{\xi}_{d,0}(y,r) = \frac{\mathrm{d}\nu_{d,Y,R}}{\mathrm{d}\nu_{d,Y} \otimes \nu_{d,R}}(y,r) \quad \text{and} \quad \xi_{d,\ell}(y,r) = \mathbb{E}_{(Y,R,Z)\sim\nu_d}[Q_\ell^{(d)}(Z) \mid Y = y, R = r]$$

Equating both sides, we find that

$$\bar{\xi}_{d,0}(y,r) = \frac{\mathrm{d}\nu_{d,Y,R}}{\mathrm{d}\nu_{d,Y} \otimes \nu_{d,R}}(y,r) = 1 + \underbrace{\sum_{k \geq \mathsf{k}_\star} \zeta_k(y)\beta_{k,0}(r)}_{:=\psi(y,r)} = 1 + \psi(y,r)\,,$$

and for $\ell \geq 1$

$$\bar{\xi}_{d,\ell}(y,r)\xi_{d,\ell}(y,r) \overset{L^2(\bar{\nu}_{d,0})}{=} \sum_{k \geq \mathsf{k}_\star} \zeta_k(y)\beta_{k,\ell}(r) = \sum_{k \in \mathcal{I}_\ell} \zeta_k(y)\beta_{k,\ell}(r) \text{ where } \mathcal{I}_\ell := \{k \geq \mathsf{k}_\star : k \equiv \ell \mod 2\}\,.$$

In the last equality, we used the fact that $\beta_{k,\ell}(r) = 0$ for $\ell \not\equiv k \mod 2$. Our goal is to bound

$$\begin{aligned} \mathbb{E}_{\nu_d}[\xi_{d,\ell}(y,r)^2] &= \mathbb{E}_{\bar{\nu}_{d,0}}[\bar{\xi}_{d,0}(y,r)\xi_{d,\ell}(y,r)^2] = \mathbb{E}_{\bar{\nu}_{d,0}}[\bar{\xi}_{d,0}(y,r)^2\xi_{d,\ell}(y,r)^2\bar{\xi}_{d,0}(y,r)^{-1}] \\ &= (1 + o_d(1))\mathbb{E}_{\bar{\nu}_{d,0}}[\bar{\xi}_{d,0}(y,r)^2\xi_{d,\ell}(y,r)^2]. \end{aligned}$$

Thus, the problem has reduced to computing the expecation of the following term under $\bar{\nu}_{d,0}$

$$\bar{\xi}_{d,0}(y,r)^2\xi_{d,\ell}(y,r)^2 = \sum_{k \in \mathcal{I}_\ell} \beta_{k,\ell}(r)^2\zeta_k^2(y) + 2\sum_{(k_2 > k_1) \in \mathcal{I}_\ell} \zeta_{k_1}(y)\zeta_{k_1}(y)c_{k_1,\ell}(r)c_{k_2,\ell}(r)\,.$$

$$\mathbb{E}_{\bar{\nu}_{d,0}}[\bar{\xi}_{d,0}(y,r)^2\xi_{d,\ell}(y,r)^2] = \sum_{k \in \mathcal{I}_\ell} \mathbb{E}[\beta_{k,\ell}(r)^2]\,\lambda_k^2 + 2\sum_{(k_2 > k_1) \in \mathcal{I}_\ell} \mathbb{E}[\zeta_{k_1}(y)\zeta_{k_1}(y)]\mathbb{E}[c_{k_1,\ell}(r)c_{k_2,\ell}(r)]\,. \tag{91}$$

We now use the bound from Lemma 3 and find the rates for each term in the above. For any $\ell < \mathsf{k}_\star$ with $\ell \equiv \mathsf{k}_\star \mod 2$ then

$$\sum_{k \geq \mathcal{I}_\ell} \mathbb{E}[\beta_{k,\ell}(r)^2]\lambda_k^2 \asymp \lambda_{\mathsf{k}_\star}^2\, d^{-(\mathsf{k}_\star - \ell)/2} + \sum_{\mathsf{k}_\star < k \in \mathcal{I}_\ell} \lambda_k^2\, d^{-(k-\ell)/2} \asymp \lambda_{\mathsf{k}_\star} d^{-(\mathsf{k}_\star - \ell)/2}\,,$$

where in the last step we noticed that the latter term forms a geometric series with decaying ratio whose leading term is of smaller order than the former term. We now show that the cross terms from (91) are of smaller order in the absolute value.

$$
\begin{aligned}
|\sum_{(k_2>k_1)\in\mathcal{I}_\ell} \mathbb{E}[\zeta_{k_1}(y)\zeta_{k_1}(y)]\mathbb{E}[c_{k_1,\ell}(r)c_{k_2,\ell}(r)]| &\le \sum_{(k_2>k_1)\in\mathcal{I}_\ell} |\mathbb{E}[\zeta_{k_1}(y)\zeta_{k_1}(y)]| \cdot |\mathbb{E}[c_{k_1,\ell}(r)c_{k_2,\ell}(r)]| \\
&\le \sum_{(k_2>k_1)\in\mathcal{I}_\ell} |\lambda_{k_1}\lambda_{k_2}| \cdot \sqrt{\mathbb{E}[c_{k_1,\ell}(r)^2]\cdot\mathbb{E}[c_{k_2,\ell}(r)^2]} \\
&\lesssim \sum_{(k_2>k_1)\in\mathcal{I}_\ell} d^{-\left(\frac{k_1+k_2}{4}-\frac{\ell}{2}\right)} \lesssim \sum_{k_1\in\mathcal{I}_\ell} d^{-\left(\frac{k_1-\ell}{2}+1\right)} \\
&\lesssim d^{-\left(\frac{\mathsf{k}_\star-\ell}{2}+1\right)}.
\end{aligned}
$$

Substituting this bound in (91), we conclude that for any $\ell < \mathsf{k}_\star$ with $\ell \equiv \mathsf{k}_\star \mod 2$, we have

$$
\mathbb{E}_{\bar{\nu}_{d,0}}[\bar{\xi}_{d,0}(y,r)^2\xi_{d,\ell}(y,r)^2] \asymp d^{-(\mathsf{k}_\star-\ell)/2}.
$$

We now do similar simplifications for the case $\ell < \mathsf{k}_\star$ such that $\ell \not\equiv \mathsf{k}_\star \mod 2$.

$$
\sum_{k\ge\mathcal{I}_\ell} \mathbb{E}[\beta_{k,\ell}(r)^2]\lambda_k^2 \asymp \min_{\substack{k>\mathsf{k}_\star\\ k\in\mathcal{I}_\ell}} \lambda_{\mathsf{k}}^2\, d^{-(k-\ell)/2} \lesssim d^{-(\mathsf{k}_\star+1-\ell)/2}.
$$

Bounding the cross terms

$$
\begin{aligned}
|\sum_{(k_2>k_1)\in\mathcal{I}_\ell} \mathbb{E}[\zeta_{k_1}(y)\zeta_{k_1}(y)]\mathbb{E}[c_{k_1,\ell}(r)c_{k_2,\ell}(r)]| &\le \sum_{(k_2>k_1)\in\mathcal{I}_\ell} |\lambda_{k_1}\lambda_{k_2}| \cdot \sqrt{\mathbb{E}[c_{k_1,\ell}(r)^2]\cdot\mathbb{E}[c_{k_2,\ell}(r)^2]} \\
&\lesssim \sum_{(k_2>k_1)\in\mathcal{I}_\ell} d^{-\left(\frac{k_1+k_2}{4}-\frac{\ell}{2}\right)} \lesssim \sum_{k_1\in\mathcal{I}_\ell} d^{-\left(\frac{k_1-\ell}{2}+1\right)} \\
&\lesssim d^{-\left(\frac{\mathsf{k}_\star+1-\ell}{2}+1\right)}.
\end{aligned}
$$

Putting the bounds in (91), for any $\ell < \mathsf{k}_\star$ such that $\ell \not\equiv \mathsf{k}_\star \mod 2$, we have

$$
\mathbb{E}_{\bar{\nu}_{d,0}}[\bar{\xi}_{d,0}(y,r)\xi_{d,\ell}(y,r)^2] \lesssim d^{-(\mathsf{k}_\star+1-\ell)/2}.
$$

□

We next show the proof of Lemma 2 which helps us characterize the complexity when one is only allowed to use the directional component $\boldsymbol{z}$.

Proof of Lemma 2. The proof is very similar to that of Lemma 1, but with a slightly more care in the decomposition to account for the fact that we only observe $(y, \boldsymbol{z})$. For clarity, we will continue to denote the original problem $(y, \boldsymbol{x})$ with ν_d, and the new spherical single index model where one only observes $(y, \boldsymbol{z})$ by υ_d.

Again, we have $\{\nu_d\}_d$ with $\nu_R = \chi_d$ and $\rho(Y \mid X) = \rho(Y \mid Z \cdot R) = \nu_d(Y \mid Z, R)$. Let $\bar{\nu}_{d,0} = \nu_Y \otimes \nu_R \otimes \nu_Z$ be the completely decoupled null. We will also let $\{\upsilon_d\}_d$ be the sequence of problem associated with $\{\nu_d\}_d$ where we only observe $(y, \boldsymbol{z})$. We have

$$
\begin{aligned}
\frac{\mathrm{d}\nu_d}{\mathrm{d}\bar{\nu}_{d,0}}(y,r,z) - 1 &\overset{L^2(\bar{\nu}_{d,0})}{=} \sum_{k\ge\mathsf{k}_\star} \zeta_k(y)\mathrm{He}_k(r\cdot z), \text{ where } \zeta_k(y) = \mathbb{E}_{(Y,R,Z)\sim\nu_d}[\mathrm{He}_k(R\cdot Z) \mid Y = y] \\
&\overset{L^2(\bar{\nu}_{d,0})}{=} \sum_{k\ge\mathsf{k}_\star} \zeta_k(y) \sum_{\ell=0}^{k} \beta_{k,\ell}(r) Q_\ell^{(d)}(z) = \sum_{\ell=0}^{\infty} Q_\ell^{(d)}(z) \sum_{k\ge\mathsf{k}_\star} \beta_{k,\ell}(r)\zeta_k(y),
\end{aligned}
$$

where in the second line, we used the harmonic decomposition of Hermite from Proposition 2. We marginalize the radius to explicitly write the likelihood ratio of only (y, z) part under ν_d and $\bar{\nu}_{d,0}$, is identical to that of υ_d and $\upsilon_{d,0} = \nu_Y \otimes \tau_{d,1}$.

$$
\frac{\mathrm{d}\upsilon_d}{\mathrm{d}\upsilon_{d,0}}(y,z) - 1 \overset{L^2(\upsilon_{d,0})}{=} \sum_{\ell=0}^{\infty} Q_\ell^{(d)}(z) \sum_{k\ge\mathsf{k}_\star} \mathbb{E}[\beta_{k,\ell}(r)]\zeta_k(y).
$$

We can also expand the log likelihood ratio of (y, z) directly in the Gegenbauer basis

$$\frac{\mathrm{d}\upsilon_d}{\mathrm{d}\upsilon_{d,0}}(y,z) - 1 \overset{L^2(\upsilon_{d,0})}{=} \sum_{\ell \geq 1} \xi_{d,\ell}(y) Q_\ell^{(d)}(z)\,, \text{ where } \xi_{d,\ell}(y) = \mathbb{E}_{(Y,Z)\sim\upsilon_d}[Q_\ell^{(d)}(Z) \mid Y = y].$$

Equating both, we have for any $\ell \geq 1$

$$\xi_{d,\ell}(y) \overset{L^2(\upsilon_{d,0})}{=} \sum_{k \geq \mathsf{k}_\star} \zeta_k(y)\mathbb{E}[\beta_{k,\ell}(r)] = \sum_{k \in \mathcal{I}_\ell} \zeta_k(y)\mathbb{E}[\beta_{k,\ell}(r)] \text{ where } \mathcal{I}_\ell := \{k \geq \mathsf{k}_\star : k \equiv \ell \mod 2\}\,.$$

Squaring both sides

$$\xi_{d,\ell}(y)^2 = \sum_{k \geq \mathcal{I}_\ell} \mathbb{E}[\beta_{k,\ell}(r)]^2 \zeta_k^2(y) + 2 \sum_{(k_2 > k_1) \in \mathcal{I}_\ell} \zeta_{k_1}(y)\zeta_{k_1}(y)\mathbb{E}[c_{k_1,\ell}(r)]\mathbb{E}[c_{k_2,\ell}(r)]\,.$$

$$\|\xi_{d,\ell}\|^2_{L^2(\upsilon_Y)} = \sum_{k \in \mathcal{I}_\ell} \mathbb{E}[\beta_{k,\ell}(r)]^2\,\lambda_k^2 + 2 \sum_{(k_2 > k_1) \in \mathcal{I}_\ell} \mathbb{E}[\zeta_{k_1}(y)\zeta_{k_1}(y)]\mathbb{E}[c_{k_1,\ell}(r)]\mathbb{E}[c_{k_2,\ell}(r)]\,. \tag{92}$$

We now use the rates on $\mathbb{E}_{r\sim\chi_d}[\beta_{k,\ell}(r)]^2$ from Lemma 3 to carry out the simplification similar to the one done in the proof of Lemma 1. For any $\ell < \mathsf{k}_\star$ with $\ell \equiv \mathsf{k}_\star \mod 2$

$$\sum_{k \in \mathcal{I}_\ell} \mathbb{E}[\beta_{k,\ell}(r)]^2\,\lambda_k^2 \asymp \lambda_{\mathsf{k}_\star}^2\, d^{-(\mathsf{k}_\star - \ell)} + \sum_{\mathsf{k}_\star < k \in \mathcal{I}_\ell} \lambda_k^2\, d^{-(k-\ell)} \asymp \lambda_{\mathsf{k}_\star} d^{-(\mathsf{k}_\star - \ell)}\,,$$

where the step followed by observing that it is a sum of geometric series whose rate is dominated by the first term. We now show the bound on the magnitude of the cross terms

$$\begin{aligned}
|\sum_{(k_2>k_1)\in\mathcal{I}_\ell} \mathbb{E}[\zeta_{k_1}(y)\zeta_{k_1}(y)]\mathbb{E}[c_{k_1,\ell}(r)]\mathbb{E}[c_{k_2,\ell}(r)]| &\leq \sum_{(k_2>k_1)\in\mathcal{I}_\ell} |\mathbb{E}[\zeta_{k_1}(y)\zeta_{k_1}(y)]| \cdot |\mathbb{E}[c_{k_1,\ell}(r)]\mathbb{E}[c_{k_2,\ell}(r)]| \\
&\leq \sum_{(k_2>k_1)\in\mathcal{I}_\ell} |\lambda_{k_1}\lambda_{k_2}| \cdot \sqrt{\mathbb{E}[c_{k_1,\ell}(r)]^2 \cdot \mathbb{E}[c_{k_2,\ell}(r)]^2} \\
&\lesssim \sum_{(k_2>k_1)\in\mathcal{I}_\ell} d^{-\left(\frac{k_1+k_2}{2} - \ell\right)} \lesssim \sum_{k_1 \in \mathcal{I}_\ell} d^{-(k_1 - \ell + 1)} \\
&\lesssim d^{-(\mathsf{k}_\star - \ell + 1)}\,.
\end{aligned}$$

Combining these rates with (92), we obtain for any $\ell < \mathsf{k}_\star$ with $\ell \equiv \mathsf{k}_\star \mod 2$,

$$\|\xi_{d,\ell}\|^2_{L^2(\upsilon_Y)} = \mathbb{E}_\upsilon[\xi_{d,\ell}(y)^2] \asymp d^{-(\mathsf{k}_\star - \ell)}\,.$$

We do similar calculation now for $\ell < \mathsf{k}_\star$ such that $\ell \not\equiv \mathsf{k}_\star \mod 2$.

$$\sum_{k \geq \mathcal{I}_\ell} \mathbb{E}[\beta_{k,\ell}(r)]^2 \lambda_k^2 \asymp \min_{\substack{k > \mathsf{k}_\star \\ k \in \mathcal{I}_\ell}} \lambda_{\mathsf{k}}^2\, d^{-(k-\ell)} \lesssim d^{-(\mathsf{k}_\star - \ell + 1)}\,.$$

Bounding the cross terms

$$\begin{aligned}
|\sum_{(k_2>k_1)\in\mathcal{I}_\ell} \mathbb{E}[\zeta_{k_1}(y)\zeta_{k_1}(y)]\mathbb{E}[c_{k_1,\ell}(r)]\mathbb{E}[c_{k_2,\ell}(r)]| &\leq \sum_{(k_2>k_1)\in\mathcal{I}_\ell} |\lambda_{k_1}\lambda_{k_2}| \cdot |\mathbb{E}[c_{k_1,\ell}(r)] \cdot \mathbb{E}[c_{k_2,\ell}(r)]| \\
\lesssim \sum_{(k_2>k_1)\in\mathcal{I}_\ell} d^{-\left(\frac{k_1+k_2}{2} - \ell\right)} \lesssim \sum_{k_1 \in \mathcal{I}_\ell} d^{-(k_1 - \ell + 1)} \lesssim d^{-(\mathsf{k}_\star - \ell + 2)}\,.
\end{aligned}$$

Substituting these bounds in (92), for any $\ell < \mathsf{k}_\star$ such that $\ell \not\equiv \mathsf{k}_\star \mod 2$, we have

$$\mathbb{E}_{\bar{\upsilon}_{d,0}}[\xi_{d,\ell}(y,r)^2] \lesssim d^{-(\mathsf{k}_\star - \ell + 1)}\,.$$

□

H Information-theoretic sample complexity

In this section, we prove an information-theoretic sample complexity upper bound. We exhibit an estimator (which is not computable in polynomial time) that achieves strong recovery with $O(d)$ samples for general spherically symmetric measure under a mild assumption on the sequence $\{\nu_d\}$ that there is $\ell \geq 1$ such that $\|\xi_{d,\ell}\|_{L^2} = \Omega_d(1)$ (cf. Appendix A). The proof is directly adapted from [15, Theorem 6.1].

Proposition 6. *Under the assumption that there exists $\ell \geq 1$ such that $\|\xi_{d,\ell}\|_{L^2} = \Theta_d(1)$, there exists an estimator (non polynomially computable) that returns $\hat{\boldsymbol{w}} \in \mathbb{S}^{d-1}$ that satisfies $|\langle \hat{\boldsymbol{w}}, \boldsymbol{w}_* \rangle| \geq 1 - \varepsilon$, with probability at least $1 - 2e^{-d}$ with information theoretic sample complexity $\mathsf{m} = O\left(\frac{d}{\varepsilon^2}\log(1/\varepsilon)\right)$, hiding constants in ℓ and $\|\xi_{d,\ell}\|_{L^2}$.*

Proof. For any $\delta > 0$, let $\mathcal{N}_\delta$ be a δ-net of $\mathbb{S}^{d-1}$, and we can choose $\mathcal{N}_\delta$ such that $|\mathcal{N}_\delta| \leq \left(\frac{3}{\delta}\right)^d$. Consider the following $g(y, r, z) = \xi_{d,\ell}(y, r) Q_\ell(z)$. For simplicity, let us denote $\beta_{d,\ell} = \|\xi_{d,\ell}\|_{L^2}$ Fix a truncation $R > 0$, and denote $L_n(\boldsymbol{w})$ defined as

$$L_n(\boldsymbol{w}) := \frac{1}{n}\sum_{i=1}^{n} g(\langle \boldsymbol{w}, \boldsymbol{z}_i \rangle, y_i, r_i) \mathbf{1}_{|g(\langle \boldsymbol{w}, \boldsymbol{z}_i\rangle, y_i, r_i)| \leq R}.$$

We consider the min-max estimator

$$\hat{\boldsymbol{w}} \in \arg\min_{\hat{\boldsymbol{w}} \in \mathbb{S}^{d-1}} \max_{\boldsymbol{w} \in \mathcal{N}_\delta} \left| L_n(\boldsymbol{w}) - \frac{\beta_{d,\ell}^2 Q_\ell(\langle \boldsymbol{w}, \hat{\boldsymbol{w}} \rangle)}{\sqrt{n_{d,\ell}}} \right|.$$

Using Lemma 27, we have

$$\mathbb{E}[g(\langle \boldsymbol{w}, \boldsymbol{z} \rangle, y, r)^2] = \mathbb{E}[\xi_{d,\ell}(\langle \boldsymbol{w}, \boldsymbol{z} \rangle)^2 Q_\ell(\langle \boldsymbol{w}, \boldsymbol{z} \rangle^2)] \leq \beta_{d,\ell}^2 \log(3/\beta_{d,\ell}^2)^{k/2}.$$

Using Bernstein's lemma, we have for any $\boldsymbol{w} \in \mathbb{S}^{d-1}$, with probability at least $1 - 2e^{-t}$

$$|L_n(\boldsymbol{w}) - \mathbb{E}[L_n(\boldsymbol{w})]| \leq \sqrt{\frac{\beta_{d,\ell}^2 \log\left(\frac{3}{\beta_{d,\ell}^2}\right)^{\ell/2} t}{n}} + \frac{Rt}{n}.$$

By union bound and setting $t = d\left(\log\left(\frac{3}{\delta}\right) + 1\right)$, we then have with probability at least $1 - 2e^{-d}$,

$$\sup_{\boldsymbol{w} \in \mathcal{N}_\delta} |L_n(\boldsymbol{w}) - \mathbb{E}[L_n(\boldsymbol{w})]| \lesssim \sqrt{\frac{\beta_{d,\ell}^2 \log\left(\frac{3}{\beta_{d,\ell}^2}\right)^{k/2} d\log\left(\frac{3}{\delta}\right)}{n}} + \frac{Rd\left(\log\left(\frac{1}{\delta}\right)\right)}{n}. \tag{93}$$

We bound the effect of the truncation

$$\begin{aligned}
|\mathbb{E}[L_n(\boldsymbol{w}) - \mathbb{E}[g(\langle \boldsymbol{w}, \boldsymbol{z} \rangle, y, r)]]| &= |\mathbb{E}[g(\langle \boldsymbol{w}, \boldsymbol{z} \rangle, y, r) 1_{|g(\langle \boldsymbol{w}, \boldsymbol{z}\rangle, y, r)| \geq R}]| \\
&\leq \sqrt{\mathbb{E}[g(\langle \boldsymbol{w}, \boldsymbol{z} \rangle, y, r)^2] \mathbb{P}(|g(\langle \boldsymbol{w}, \boldsymbol{z} \rangle, y, r)| > R)} \\
&\leq \beta_{d,\ell} \sqrt{\mathbb{P}(|g(\langle \boldsymbol{w}, \boldsymbol{z} \rangle, y, r)| > R)}\,.
\end{aligned}$$

We then have the following control on the moments of $|g(\langle \boldsymbol{w}, \boldsymbol{z} \rangle, y, r))|$

$$\mathbb{E}[|g(\langle \boldsymbol{w}, \boldsymbol{z} \rangle, y, r)|^p]^{1/p} \leq \mathbb{E}[|Q_\ell(\langle \boldsymbol{w}, \boldsymbol{z} \rangle)|^p |\xi_{d,\ell}(y, r)|^p]^{1/p} \leq (2p)^\ell,$$

by using Jensen inequality and spherical hypercontractivity. By taking $R \geq (2e)^\ell$, and $\delta = R^{1/\ell}/2e$

$$\mathbb{P}(|g(\langle \boldsymbol{w}, \boldsymbol{z} \rangle, y, r)| > R) \leq \frac{2p^{\ell p}}{R^p} \leq \exp\left(-\frac{\ell}{2e} R^{1/\ell}\right)$$

Combining the two inequalities gives

$$|\mathbb{E}[L_n(\boldsymbol{w})] - \mathbb{E}[g(\langle \boldsymbol{w}, \boldsymbol{z} \rangle, y, r)]| \leq \beta_{d,\ell} \exp\left(-\frac{\ell}{4e} R^{1/\ell}\right).$$

1545 Combining the above inequalities, we then have

$$\begin{aligned}
&\max_{\boldsymbol{w}\in\mathcal{N}_\delta}\left|\frac{\beta_{d,\ell}^2 Q_\ell(\langle \boldsymbol{w},\boldsymbol{w}_*\rangle)}{\sqrt{n_{d,\ell}}} - \frac{\beta_{d,\ell}^2 Q_\ell(\langle \boldsymbol{w},\hat{\boldsymbol{w}}\rangle)}{\sqrt{n_{d,\ell}}}\right| \\
&\le \max_{\boldsymbol{w}\in\mathcal{N}_\delta}\left|L_n(\boldsymbol{w}) - \frac{\beta_{d,\ell}^2 Q_\ell(\langle \boldsymbol{w},\boldsymbol{w}_*\rangle)}{\sqrt{n_{d,\ell}}}\right| + \max_{\boldsymbol{w}\in\mathcal{N}_\delta}\left|L_n(\boldsymbol{w}) - \frac{\beta_{d,\ell}^2 Q_\ell(\langle \boldsymbol{w},\hat{\boldsymbol{w}}\rangle)}{\sqrt{n_{d,\ell}}}\right| \\
&\le 2\max_{\boldsymbol{w}\in\mathcal{N}_\delta}\left|L_n(\boldsymbol{w}) - \frac{\beta_{d,\ell}^2 Q_\ell(\langle \boldsymbol{w},\boldsymbol{w}_*\rangle)}{\sqrt{n_{d,\ell}}}\right| \\
&\le 2\max_{\boldsymbol{w}\in\mathcal{N}_\delta}\left|L_n(\boldsymbol{w}) - \mathbb{E}\left[g(\langle \boldsymbol{w},\boldsymbol{z}\rangle, y, r)\right]\right| \\
&\le \max_{\boldsymbol{w}\in\mathcal{N}_\delta}\left|L_n(\boldsymbol{w}) - \mathbb{E}\left[L_n(\boldsymbol{w})\right]\right| + 3^\ell \exp\left(-\frac{\ell}{4e}R^{1/\ell}\right) \\
&\le \sqrt{\frac{\beta_{d,\ell}^2 \log\left(\frac{3}{\beta_{d,\ell}^2}\right)^{k/2} t}{n}} + \frac{Rt}{n} + 3^\ell \exp\left(-\frac{\ell}{4e}R^{1/\ell}\right).
\end{aligned}$$

1546 We have the following

$$\begin{aligned}
\left|\frac{Q_\ell(\langle \boldsymbol{w},\boldsymbol{w}_*\rangle) - Q_\ell(\langle \boldsymbol{w},\hat{\boldsymbol{w}}\rangle)}{\sqrt{n_{d,\ell}}}\right| &= \frac{1}{\sqrt{n_{d,\ell}}}\left|\sum_{q=0}^{\lfloor \ell/2\rfloor} c_{\ell,q}\left(\langle \boldsymbol{w},\boldsymbol{w}_*\rangle^{\ell-2q} - \langle \boldsymbol{w},\hat{\boldsymbol{w}}\rangle^{\ell-2q}\right)\right| \\
&= \frac{c_{\ell,0}}{\sqrt{n_{d,\ell}}}\left|\langle \boldsymbol{w},\boldsymbol{w}_*\rangle^{\ell} - \langle \boldsymbol{w},\hat{\boldsymbol{w}}\rangle^{\ell}\right| + O(d^{-1}).
\end{aligned}$$

1547 We then deduce that

$$\max_{\boldsymbol{w}\in\mathcal{N}_\delta}\left|\frac{\beta_{d,\ell}^2 Q_\ell(\langle \boldsymbol{w},\boldsymbol{w}_*\rangle)}{\sqrt{n_{d,\ell}}} - \frac{\beta_{d,\ell}^2 Q_\ell(\langle \boldsymbol{w},\hat{\boldsymbol{w}}\rangle)}{\sqrt{n_{d,\ell}}}\right| = \frac{\beta_{d,\ell}^2 c_{\ell,0}}{\sqrt{n_{d,\ell}}}\max_{\boldsymbol{w}\in\mathcal{N}_\delta}\left|\langle \boldsymbol{w},\boldsymbol{w}_*\rangle^{\ell} - \langle \boldsymbol{w},\hat{\boldsymbol{w}}\rangle^{\ell}\right| + O(d^{-1}).$$

1548 Using [22, Lemma 25], we then have

$$\frac{\beta_{d,\ell}^2 c_{\ell,0}}{\sqrt{n_{d,\ell}}}\min_{s\in\{\pm 1\}}\|s\hat{\boldsymbol{w}} - \boldsymbol{w}_*\| \lesssim \beta_{d,\ell}^2\left(\max_{\boldsymbol{w}\in\mathcal{N}_\delta}\left|\frac{Q_\ell(\langle \boldsymbol{w},\boldsymbol{w}_*\rangle)}{\sqrt{n_{d,\ell}}} - \frac{Q_\ell(\langle \boldsymbol{w},\hat{\boldsymbol{w}}\rangle)}{\sqrt{n_{d,\ell}}}\right| + \delta + O(d^{-1})\right)$$

1549 Using this inequality above, and plugging the inequality, we have

$$\min_{s\in\{\pm 1\}}\|s\hat{\boldsymbol{w}} - \boldsymbol{w}_*\|_2 \lesssim \sqrt{\frac{\beta_{d,\ell}^2 \log(\frac{3}{\beta_{d,\ell}^2})^{\ell/2} d\log\left(\frac{3}{\delta}\right)}{n}} + \delta + \frac{Rd\log\left(\frac{3}{\delta}\right)}{n} + 3^\ell\exp\left(-\frac{\ell}{4e}R^{1/\ell}\right)$$

1550 Choosing $R = (4e\log(3/\delta))^\ell$, it yields

$$\frac{\beta_{d,\ell}^2 c_{\ell,0}}{\sqrt{n_{d,\ell}}}\min_{s\in\{\pm 1\}}\|s\hat{\boldsymbol{w}} - \boldsymbol{w}_*\| \lesssim \delta + \sqrt{\frac{\beta_{d,\ell}^2 \log(\frac{3}{\beta_{d,\ell}^2})^{\ell/2} d\log\left(\frac{3}{\delta}\right)}{n}} + \frac{Rd\log\left(\frac{3}{\delta}\right)}{n} + 3^\ell\exp\left(-\frac{\ell}{4e}R^{1/\ell}\right).$$

1551 Taking $\delta = O(\varepsilon\beta_{d,\ell}^2)$ concludes the proof. □

1552 I Additional technical results

1553 The distribution $\tau_d = \mathrm{Unif}(\mathbb{S}^{d-1})$ has a well-celebrated hypercontractivity property.

Lemma 25 (Spherical Hypercontractivity [3]). *For any $\ell \in \mathbb{N}$ and $f \in L^2(\tau_d)$ which is a degree ℓ polynomial, for any $p \ge 2$, we have*

$$\|f\|_{L^p(\tau_d)} \le (p-1)^{\ell/2}\|f\|_{L^2(\tau_d)}.$$

1554 We will often couple this with the following standard tail-bound.

Lemma 26 (Lemma 24 in [14]). *Let $\delta \geq 0$ and X be a mean zero random variable satisfying*

$$\mathbb{E}[|X|^p]^{1/p} \leq B\, p^{k/2} \;\; \text{for } p = \frac{2\log(1/\delta)}{k}\,,$$

for some k. Then with probability $1-\delta$, we have $|X| \leq B\, p^{k/2}$.

Similar to [14], we will use the following lemma to bound $\mathbb{E}[XY]$ instead of standard Cauchy-Schwarz, when we have a tight bound $\|X\|_1$ and all moments $\|Y\|_p$ but a very loose bound on $\|X\|_2$.

Lemma 27 (Lemma 23 in [14]). *Let X, Y be random variables with $\|Y\|_p \leq B\, p^{k/2}$. Then*

$$\mathbb{E}[XY] \leq \|X\|_1 \cdot B \cdot (2e)^{k/2} \cdot \max\left(1, \frac{2}{k}\log\left(\frac{\|X\|_2}{\|X\|_1}\right)\right)^{k/2}$$

.

Lemma 28 (Lemma I.5 in [15]). *Let $\boldsymbol{Y} = \sum_{i=1}^n \boldsymbol{Z}_i$, where $\boldsymbol{Z}_i$ are self-adjoint, mean zero, i.i.d. Define*

$$\sigma = \|\mathbb{E}[\boldsymbol{Y}^2]\|_2^{1/2}, \;\; \sigma^* = \sup_{\boldsymbol{v},\boldsymbol{w}\in\mathbb{S}^{d-1}} \mathbb{E}[(\langle \boldsymbol{v}, \boldsymbol{Y}\boldsymbol{w}\rangle)^2]^{1/2}, \;\; \bar{R} = \mathbb{E}[\max_{i\in[n]} \|\boldsymbol{Z}_i\|^2]^{1/2}.$$

Then for

$$R \geq \bar{R}^{1/2}\sigma^{1/2} + \sqrt{2}\bar{R}$$

and $t \geq 0$, if $\delta = \mathbb{P}(\max\|\boldsymbol{Z}_i\| \geq R)$ with probability $1-\delta-de^{-t}$

$$\|\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}]\| \leq 2\sigma + \sigma_* t^{1/2} + R^{1/3}\sigma^{2/3}t^{2/3} + Rt.$$

Lemma 29. *Let $\{Z_i\}_{i\in[n]}$ be a sequence of independent random variables with polynomial tails, i.e. there exists B, k such that $\mathbb{E}[|Z_i|^p]^{1/p} \leq Bp^{k/2}$. Define $R = \max_{i\in[n]} Z_i$. Then for any $p \leq \log n/k$, we have $\mathbb{E}[|R|^p]^{1/p} \leq B\log^{k/2}(n)$ and for any $\delta \geq 0$, with probability at least $1-\delta$, $R \leq B\log^{k/2}(n/\delta)$.*

Lemma 30 (Lemma I.3 from [15]). *Let $X_1, \ldots, X_n$ be independent mean zero random variables such that for all $p \geq 2$, we have $\|X_i\|_p \leq Bp^{k/2}$ for some k and let $\sigma^2 = \sum_{i=1}^n \mathbb{E}[X_i^2]$. Let $Y = \sum_{i=1}^n X_i$. Then with pribability at least $1-\delta$,*

$$|Y| \lesssim_k \sigma\sqrt{\log(1/\delta)} + B\log(1/\delta)\log(n/\delta)^{k/2}\,.$$

Lemma 31 ([44]). *Let $\boldsymbol{X}_k$ be i.i.d random matrices of dimensions $d_1 \times d_2$. Assume that each matrix is bounded by*

$$\forall k, \|\boldsymbol{X}_k - \mathbb{E}[\boldsymbol{X}_k]\|_{\text{op}} \leq L.$$

Consider $v(\boldsymbol{Z}) = \max\{\|\sum_{k=1}^n \mathbb{E}[(\boldsymbol{X}_k - \mathbb{E}[\boldsymbol{X}_k])(\boldsymbol{X}_k - \mathbb{E}[\boldsymbol{X}_k])^\mathsf{T}]\|, \|\sum_{k=1}^n \mathbb{E}[(\boldsymbol{X}_k - \mathbb{E}[\boldsymbol{X}_k])^\mathsf{T}(\boldsymbol{X}_k - \mathbb{E}[\boldsymbol{X}_k])]\|\}$, then with probability at least $1-\delta$,

$$\left\|\sum_{k=1}^n (\boldsymbol{X}_k - \mathbb{E}[\boldsymbol{X}_k])\right\|_{\text{op}} \leq \frac{L}{3}\log\left(\frac{d_1+d_2}{\delta}\right) + \sqrt{4v\log\left(\frac{d_1+d_2}{\delta}\right)}.$$

References

[1] Emmanuel Abbe, Enric Boix Adsera, and Theodor Misiakiewicz. Sgd learning on neural networks: leap complexity and saddle-to-saddle dynamics. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 2552–2623. PMLR, 2023.

[2] Gerard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Online stochastic gradient descent on non-convex losses from high-dimensional inference. *Journal of Machine Learning Research*, 22(106):1–51, 2021.

[3] William Beckner. Sobolev inequalities, the poisson semigroup, and analysis on the sphere sn. *Proceedings of the National Academy of Sciences*, 89(11):4816–4819, 1992.

[4] Raphaël Berthier, Andrea Montanari, and Kangjie Zhou. Learning time-scales in two-layers neural networks. *Foundations of Computational Mathematics*, pages 1–84, 2024.

[5] Alberto Bietti, Joan Bruna, and Loucas Pillaud-Vivien. On learning gaussian multi-index models with gradient flow. *arXiv preprint arXiv:2310.19793*, 2023.

[6] Alberto Bietti, Joan Bruna, Clayton Sanford, and Min Jae Song. Learning single-index models with shallow neural networks. *Advances in Neural Information Processing Systems*, 35:9768–9783, 2022.

[7] Giulio Biroli, Chiara Cammarota, and Federico Ricci-Tersenghi. How to iron out rough landscapes and get optimal performances: averaged gradient descent and its application to tensor pca. *Journal of Physics A: Mathematical and Theoretical*, 53(17):174003, 2020.

[8] George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 26(2):211–243, 1964.

[9] Joan Bruna and Daniel Hsu. Survey on algorithms for multi-index models. *arXiv preprint arXiv:2504.05426*, 2025.

[10] Nader H Bshouty and Vitaly Feldman. On using extended statistical queries to avoid membership queries. *Journal of Machine Learning Research*, 2(Feb):359–395, 2002.

[11] Theodore S Chihara. *An introduction to orthogonal polynomials*. Courier Corporation, 2011.

[12] Feng Dai. *Approximation theory and harmonic analysis on spheres and balls*. Springer, 2013.

[13] Arnak S Dalalyan, Anatoly Juditsky, and Vladimir Spokoiny. A new algorithm for estimating the effective dimension-reduction subspace. *The Journal of Machine Learning Research*, 9:1647–1678, 2008.

[14] Alex Damian, Eshaan Nichani, Rong Ge, and Jason D Lee. Smoothing the landscape boosts the signal for sgd: Optimal sample complexity for learning single index models. *Advances in Neural Information Processing Systems*, 36, 2024.

[15] Alex Damian, Loucas Pillaud-Vivien, Jason Lee, and Joan Bruna. Computational-statistical gaps in gaussian single-index models. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 1262–1262. PMLR, 2024.

[16] Alexandru Damian, Jason Lee, and Mahdi Soltanolkotabi. Neural networks can learn representations with gradient descent. In *Conference on Learning Theory*, pages 5413–5452. PMLR, 2022.

[17] Yatin Dandi, Florent Krzakala, Bruno Loureiro, Luca Pesce, and Ludovic Stephan. How two-layer neural networks learn, one (giant) step at a time. *arXiv preprint arXiv:2305.18270*, 2023.

[18] Yatin Dandi, Emanuele Troiani, Luca Arnaboldi, Luca Pesce, Lenka Zdeborová, and Florent Krzakala. The benefits of reusing batches for gradient descent in two-layer networks: Breaking the curse of information and leap exponents. *arXiv preprint arXiv:2402.03220*, 2024.

[19] Victor H. de la Peña and Stephen J. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate u-statistics, 1999.

[20] *NIST Digital Library of Mathematical Functions*. https://dlmf.nist.gov/, Release 1.2.3 of 2024-12-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[21] Rishabh Dudeja and Daniel Hsu. Learning single-index models in gaussian space. In *Conference On Learning Theory*, pages 1887–1930. PMLR, 2018.

[22] Rishabh Dudeja and Daniel Hsu. Statistical-computational trade-offs in tensor pca and related problems via communication complexity. *The Annals of Statistics*, 52(1):131–156, 2024.

[23] Wolfgang Härdle and Thomas M Stoker. Investigating smooth multiple regression by the method of average derivatives. *Journal of the American statistical Association*, 84(408):986–995, 1989.

[24] Samuel Hopkins. *Statistical inference and the sum of squares method*. PhD thesis, Cornell University, 2018.

[25] Samuel B Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 178–191, 2016.

[26] Samuel B. Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In Peter Grünwald, Elad Hazan, and Satyen Kale, editors, *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 956–1006, Paris, France, 03–06 Jul 2015. PMLR.

[27] Hidehiko Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of econometrics*, 58(1-2):71–120, 1993.

[28] Nirmit Joshi, Theodor Misiakiewicz, and Nathan Srebro. On the complexity of learning sparse functions with statistical and gradient queries. *arXiv preprint arXiv:2407.05622*, 2024.

[29] Sham M Kakade, Varun Kanade, Ohad Shamir, and Adam Kalai. Efficient learning of generalized linear and single index models with isotonic regression. *Advances in Neural Information Processing Systems*, 24, 2011.

[30] Adam Tauman Kalai and Ravi Sastry. The isotron algorithm: High-dimensional isotonic regression. In *COLT*, volume 1, page 9, 2009.

[31] Michael Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45(6):983–1006, 1998.

[32] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2019.

[33] Ker-Chau Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991.

[34] Juan Carlos López Carreño, Rosalba Mendoza Suárez, and Jairo Alonso Mendoza S. Connection between the hermite and gegenbauer polynomials using the lewanowicz method. *International Journal of Mathematics & Computer Science*, 18(4), 2023.

[35] Peter McCullagh. Generalized linear models. *European Journal of Operational Research*, 16(3):285–292, 1984.

[36] Song Mei, Yu Bai, and Andrea Montanari. The landscape of empirical risk for nonconvex losses. *The Annals of Statistics*, 46(6A):2747–2774, 2018.

[37] Marco Mondelli and Andrea Montanari. Fundamental limits of weak recovery with applications to phase retrieval. In *Conference On Learning Theory*, pages 1445–1450. PMLR, 2018.

[38] Andrea Montanari and Emile Richard. A statistical model for tensor pca. *Advances in neural information processing systems*, 27, 2014.

[39] Lev Reyzin. Statistical queries and statistical algorithms: Foundations and applications. *arXiv preprint arXiv:2004.00557*, 2020.

[40] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022.

[41] Min Jae Song, Ilias Zadik, and Joan Bruna. On the cryptographic hardness of learning single periodic neurons. *Advances in neural information processing systems*, 34:29602–29615, 2021.

[42] Gabor Szeg. *Orthogonal polynomials*, volume 23. American Mathematical Soc., 1939.

[43] Yan Shuo Tan and Roman Vershynin. Online stochastic gradient descent with arbitrary initialization solves non-smooth, non-convex phase retrieval. *Journal of Machine Learning Research*, 24(58):1–47, 2023.

[44] Joel A Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[45] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[46] Miaoyan Wang, Khanh Dao Duc, Jonathan Fischer, and Yun S. Song. Operator norm inequalities between tensor unfoldings on the partition lattice. *Linear Algebra and its Applications*, 520:44–66, May 2017.

[47] Alexander S Wein, Ahmed El Alaoui, and Cristopher Moore. The kikuchi hierarchy and tensor pca. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1446–1468. IEEE, 2019.

[48] Wikipedia contributors. Gegenbauer polynomials — Wikipedia, the free encyclopedia, 2024. [Online; accessed 28-December-2024].

[49] Aaron Zweig, Loucas Pillaud-Vivien, and Joan Bruna. On single-index models beyond gaussian data. *Advances in Neural Information Processing Systems*, 36, 2024.