A Technical Lemmas

In this appendix, we introduce a series of technical lemmas that form the basis of our results. Our proofs rely on carefully analyzing how the conditional expectation of $\boldsymbol{\ell}^*$ changes when conditioning on different signals ψ. Lemmas 13-16 are the main tools we use to analyze the behavior of these conditional expectations.

Lemma 13 says that for random variable X, if we have some signal R that is equal to X plus noise with some small probability and is just pure noise otherwise, then the conditional expectation of X given the sign of R has magnitude that is $\Omega(\epsilon)$. In the "initial exploration" phase of our algorithm we explore a new (not previously explored) direction with very small probability. Lemma 13 implies that this exploration will lead to the conditional expectation of $\boldsymbol{\ell}^*$ in the newly-explored direction having magnitude proportional to the probability of exploration.

Lemma 13. *Suppose X is a real-valued K-sub-gaussian random variable such that $\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] = \sigma_X^2$. Let $R \sim X + N(0,1)$ with probability ϵ and $R \sim N(0,1)$ with probability $1-\epsilon$. Then there exists c_{L13} independent of X such that*

$$|\mathbb{E}[X \mid 1_{R>0}]| \geq c_{L13}\sigma_X^5\epsilon.$$

The proof of Lemma 13 can be found in Appendix H.1.

Lemmas 14 and 15 are the main technical tools that allow for us to exponentially grow the amount of exploration in any new direction. Informally, Lemma 14 says that even if X forms only an ϵ fraction of the signal r, conditioning on the sign of r will increase the conditional expectation of X by a multiplicative factor. This lemma will be applied to the expectation of $\boldsymbol{\ell}^*$ in the new direction we are trying to explore. Lemma 15 says that any random variable conditioned on the sign of r cannot have conditional expectation increase by more than $O(\epsilon)$. This will be applied to the expectation of $\boldsymbol{\ell}^*$ in all of the directions that we have already explored. These two lemmas combined allow our algorithm to multiplicatively increase the magnitude of the expectation of $\boldsymbol{\ell}^*$ in an unexplored direction *relative* to the already explored directions.

Lemma 14. *Let X be a K-sub-gaussian random variable satisfying $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = \sigma_X^2 \geq \mathrm{c_v}$. For $Z \sim N(0, \sigma^2)$ such that $Z \perp\!\!\!\perp X$ and $\epsilon > 0$, define $r = \epsilon X + Z$. Then there exists a constant δ_{L14} such that if $\epsilon/\sigma \leq \delta_{L14}$,*

$$|\mathbb{E}[X \mid 1_{r>0}]| \geq \frac{\epsilon\sigma_X^2}{2\sigma\sqrt{2\pi}} \geq \frac{\mathrm{c_v}\,\epsilon}{\sqrt{8\pi}\sigma} := \frac{c_{L14}\epsilon}{\sigma}.$$

The proof of Lemma 14 can be found in Appendix G.1.

Lemma 15. *For K-sub-gaussian random variables X, Y such that $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ and for $Z \sim N(0, \sigma^2)$ independent of X and Y, let $r \sim \epsilon X + Z$. Suppose $\frac{\epsilon}{\sigma} \leq \delta_{L15} := \min\left(1, \frac{1}{2K\sqrt{\log(2)}}\right)$. Then there exists a constant $c_{L15} > 0$ such that*

$$|\,\mathbb{E}[Y \mid 1_{r>0}]| \leq c_{L15}\epsilon/\sigma.$$

The proof of Lemma 15 can be found in Appendix G.2.

Lemma 16 is a more technical lemma that allows us to better understand the distribution of $\boldsymbol{\ell}^*$ when we condition on averages based on previous rewards. More specifically, this allows us to apply Lemma 7 to the random variable $\mathbf{z}$ as described in Section 2.2. The proof of Lemma 16 can be found in Appendix H.

Lemma 16. *For random variable $\mathbf{X}$ in $\mathbb{R}^d$, define $\mathbf{Y} = \mathbf{X} + \mathbf{W}$ where $\mathbf{W} \sim N(\mathbf{0}, \mathrm{Diag}(\mathbf{s}))$ and $\mathbf{W}$ is independent of $\mathbf{X}$. Define $\mathbf{Z}(\mathbf{Y}) = \mathbb{E}[\mathbf{X} \mid \mathbf{Y}]$. If $\min_{\|\mathbf{v}\|=1} \Pr(\langle \mathbf{X}, \mathbf{v}\rangle \geq c_d) \geq \epsilon_d$ and for all $i \in [1:d]$, $s_i \leq c_{L16} := \frac{c_d^2/32}{\log(4/\epsilon_d)}$ then*

$$\min_{\|\mathbf{v}\|=1} \mathbb{E}[(\langle \mathbf{Z}(\mathbf{Y}), \mathbf{v}\rangle)^+] \geq \frac{\epsilon_d c_d}{4}.$$

The final lemma for this section says that any vector in the span of the top eigenvectors of a positive semi-definite matrix can be represented as a linear combination of these top eigenvectors with coefficients that are not too large. The proof of Lemma 17 can be found in Appendix F.

Lemma 17. *Let $\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_j \in \mathbb{R}^d$ such that $\|\mathbf{v}_i\| = 1$. Define $\mathbf{M} = \sum_{i=1}^j \mathbf{v}_i^{\otimes 2}$. Suppose $\mathbf{w}_1, ..., \mathbf{w}_d$ are orthonormal eigenvectors of $\mathbf{M}$ with corresponding eigenvalues $\lambda_1 \geq ... \geq \lambda_d \geq 0$. Suppose $\lambda_\ell \geq \epsilon$. Then for any $\mathbf{u} \in \mathrm{Span}(\mathbf{w}_1, ..., \mathbf{w}_\ell)$ such that $\|\mathbf{u}\| \leq 1$, we have $\mathbf{u} = \sum_{i=1}^j c_i \mathbf{v}_i$, where $c_i := \sum_{k=1}^\ell \left(\frac{\langle \mathbf{u}, \mathbf{w}_k \rangle \langle \mathbf{v}_i, \mathbf{w}_k \rangle}{\lambda_k} \right)$. Furthermore, $\sum_{i=1}^j c_i^2 \leq \frac{1}{\epsilon}$.*

B Discussion on Assumptions

Here we illustrate why the first two conditions in Assumption 2 are important, by presenting an example where each is absent.

To see why Condition 1 (that $\boldsymbol{\ell}^*$ is not confined to any half-space) is necessary, suppose $d = 2$ and that ${\boldsymbol{\ell}^*}_1 \sim \text{Uniform}(0.5, 2)$ and ${\boldsymbol{\ell}^*}_2 \sim \text{Uniform}(-1, 1)$ and ${\boldsymbol{\ell}^*}_1 \perp\!\!\!\perp {\boldsymbol{\ell}^*}_2$. This example violates Condition 1, because $\Pr\left(\langle -\mathbf{e}_1, \boldsymbol{\ell}^* \rangle \geq 0\right) = 0$. However, the only BIC action will for every t be $\mathbf{A}^{(t)} = \mathbf{e}_1$. This is because $\mathbb{E}[{\boldsymbol{\ell}^*}_2] = 0$, and $\mathbb{E}[{\boldsymbol{\ell}^*}_1 \mid \Psi] > 0$ for any signal Ψ based on any historical actions and rewards. Condition 1 is also not overly restrictive so as to make λ-spectral exploration in poly(d) steps trivial. For example, suppose ${\boldsymbol{\ell}^*}_1 = \begin{cases} -1 & \text{w.p. } e^{-d} \\ 1 & \text{w.p. } 1 - e^{-d} \end{cases}$ and ${\boldsymbol{\ell}^*}_2 \sim \text{Uniform}(-1, 1)$. This prior distribution satisfies Condition 1, but there is still an exponentially small probability that the optimal action is not $\mathbf{A}^* = (1, 0)$. Therefore, in this example it is not trivial to guarantee λ-spectral exploration in a polynomial number of samples.

To see why Condition 2 (that $\boldsymbol{\ell}^*$ has non-degenerate covariance) is needed, consider the following example. Suppose $d = 2$ and that the coordinates of $\boldsymbol{\ell}^*$ are independent with:

$$
{\boldsymbol{\ell}^*}_1 = \begin{cases} -1 & \text{w.p. } e^{-d} \\ 1 & \text{w.p. } 1 - e^{-d} \end{cases}, \qquad {\boldsymbol{\ell}^*}_2 = \begin{cases} -2 & \text{w.p. } e^{-d} \\ 0 & \text{w.p. } 1 - 2e^{-d} \\ 2 & \text{w.p. } e^{-d} \end{cases}.
$$

Intuitively, this example once again requires exponentially many samples to λ-spectrally explore, because in order to explore ${\boldsymbol{\ell}^*}_2$ we need to find a signal based on the history where the expectation of ${\boldsymbol{\ell}^*}_2$ is not exponentially smaller than the expectation of ${\boldsymbol{\ell}^*}_1$. However, we cannot do exponential growth on the conditional expectation of ${\boldsymbol{\ell}^*}_2$, as we would need exponentially many samples to sufficiently increase the conditional expectation of ${\boldsymbol{\ell}^*}_2$. Similarly, we would need exponentially many samples to decrease the conditional expectation of ${\boldsymbol{\ell}^*}_1$.

We also note that any sub-gaussian random variable X satisfying $\mathbb{P}(|X| > t) \leq 2e^{-t^2/K^2}$ for all $t \geq 0$ (as in Condition 3) satisfies the following bounds on the moments of X:

$$
\mathbb{E}[X] \leq \mathbb{E}[|X|] = \int_0^\infty \mathbb{P}\left(|X| > t\right) dt \leq \int_0^\infty 2e^{-t^2/K^2} dt = K\sqrt{\pi}, \tag{1}
$$

$$
\mathbb{E}[X^2] = \int_0^\infty \mathbb{P}(X^2 > t) dt \leq \int_0^\infty 2e^{-t/K^2} dt = 2K^2. \tag{2}
$$

C Proof of Lemma 9

Proof of Lemma 9. Recall that the input parameters from Algorithm 5 come from their use in Algorithm 4. We compute as follows, with $\stackrel{d}{=}$ indicating equality in distribution.

$$
\begin{aligned}
\hat{y}_\ell &= \lambda c_{L16} \sum_{t'=0}^{\frac{1}{\lambda c_{L16}} - 1} \sum_{k=1}^j \frac{\langle \mathbf{v}_k, \mathbf{w}_\ell \rangle}{\lambda_\ell} q_k^{t'} \\
&\stackrel{d}{=} \lambda c_{L16} \sum_{t'=0}^{\frac{1}{\lambda c_{L16}} - 1} \sum_{k=1}^j \frac{\langle \mathbf{v}_k, \mathbf{w}_\ell \rangle}{\lambda_\ell} \langle \mathbf{v}_k, \boldsymbol{\ell}^* \rangle + \lambda c_{L16} \sum_{t'=0}^{\frac{1}{\lambda c_{L16}} - 1} \sum_{k=1}^j \frac{\langle \mathbf{v}_k, \mathbf{w}_\ell \rangle}{\lambda_\ell} N(0, 1) \\
&\stackrel{d}{=} \left\langle \boldsymbol{\ell}^*, \left(\lambda c_{L16} \sum_{t'=0}^{\frac{1}{\lambda c_{L16}} - 1} \sum_{k=1}^j \frac{\langle \mathbf{v}_k, \mathbf{w}_\ell \rangle}{\lambda_\ell} \mathbf{v}_k \right) \right\rangle + N\left(0, \lambda c_{L16} \sum_{k=1}^j \frac{\langle \mathbf{v}_k, \mathbf{w}_\ell \rangle^2}{\lambda_\ell^2} \right)
\end{aligned}
$$

$$\stackrel{d}{=} \left\langle \boldsymbol{\ell}^*, \left(\lambda c_{L16} \sum_{t'=0}^{\frac{1}{\lambda c_{L16}}-1} \mathbf{w}_\ell \right) \right\rangle + N\left(0, \frac{\lambda c_{L16}}{\lambda_\ell^2} \mathbf{w}_\ell^\top \mathbf{M} \mathbf{w}_\ell \right) \qquad [\text{Lemma 17 } (\mathbf{u} = \mathbf{w}_\ell)]$$

$$\stackrel{d}{=} x_\ell^* + N\left(0, c_{L16} \frac{\lambda}{\lambda_\ell}\right).$$

We will apply Lemma 16 with $\mathbf{X} = \mathbf{x}^*$, $\mathbf{Y} = \hat{\mathbf{y}}$, and $\mathbf{Z}(\mathbf{Y}) = \mathbf{z}(\hat{\mathbf{y}})$. As shown above, (and using that $\frac{\lambda}{\lambda_\ell} \leq 1$ for all $\ell \leq \ell_\lambda$) we have that $\mathbf{Y} - \mathbf{X}$ has the appropriate distribution. The last thing we need to show is that

$$\begin{aligned}
\min_{\|\mathbf{q}\|=1} \Pr\left(\langle \mathbf{X}, \mathbf{q} \rangle \geq c_d\right) &= \min_{\|\mathbf{q}\|=1} \Pr\left(\sum_{\ell=1}^{\ell_\lambda} \langle \boldsymbol{\ell}^*, \mathbf{w}_\ell \rangle q_\ell \geq c_d \right) \\
&= \min_{\|\mathbf{q}\|=1} \Pr\left(\left\langle \boldsymbol{\ell}^*, \left(\sum_{\ell=1}^{\ell_\lambda} q_\ell \mathbf{w}_\ell \right) \right\rangle \geq c_d \right) \\
&\geq \min_{\|\mathbf{v}\|=1} \Pr\left(\langle \boldsymbol{\ell}^*, \mathbf{v} \rangle \geq c_d\right) \\
&\geq \epsilon_d. \qquad [\text{Assumption 1}].
\end{aligned}$$

This means we can apply Lemma 16 to get that

$$\min_{\|\mathbf{v}\|=1} \mathbb{E}[\langle \mathbf{z}(\hat{\mathbf{y}}), \mathbf{v} \rangle^+] \geq \frac{\epsilon_d c_d}{4}.$$

We have therefore shown that $\mathbf{z}(\hat{\mathbf{y}})$ satisfies the assumptions of Lemma 7 with $\epsilon = \frac{\epsilon_d c_d}{4}$. ☐

D Proof of Lemma 6

Proof of Lemma 6. The BIC optimal action given ψ is

$$\begin{aligned}
\mathbf{A}^* &= \arg\max_{\mathbf{A} \in S^{d-1}} \mathbb{E}[\langle \mathbf{A}, \boldsymbol{\ell}^* \rangle \mid \psi] \\
&= \arg\max_{\mathbf{A} \in S^{d-1}} \langle \mathbf{A}, \mathbb{E}[\boldsymbol{\ell}^* \mid \psi] \rangle.
\end{aligned}$$

Therefore, the BIC action is $\mathbf{A}^* = \frac{\mathbb{E}[\boldsymbol{\ell}^* \mid \psi]}{\|\mathbb{E}[\boldsymbol{\ell}^* \mid \psi]\|}$ if $\|\mathbb{E}[\boldsymbol{\ell}^* \mid \psi]\| \neq 0$. If $\|\mathbb{E}[\boldsymbol{\ell}^* \mid \psi]\| = 0$, then any action is BIC including $\mathbf{v}$. ☐

E Proof of Lemma 8

We will prove the following equivalent lemma.

Lemma 18. *In the setting of Lemma 8,*

$$\sum_{i=1}^{\ell} \lambda_i' \leq (1 - \epsilon/2) + \sum_{i=1}^{\ell} \lambda_i$$

We first observe that Lemma 18 implies the desired Lemma 8.

Proof of Lemma 8. By linearity of trace,

$$\sum_{i=1}^{d} \lambda_i' = 1 + \sum_{i=1}^{d} \lambda_i.$$

Therefore, Lemma 18 implies the desired result that

$$\sum_{i=\ell+1}^{d} \lambda_i' \geq \epsilon/2 + \sum_{i=\ell+1}^{d} \lambda_i. \qquad ☐$$

Proof of Lemma 18. First, note the following, where the max is over $\mathbf{x}_1, ..., \mathbf{x}_\ell$ that are orthonormal.

$$\sum_{i=1}^{\ell} \lambda_i' = \max_{\mathbf{x}_1,\ldots,\mathbf{x}_\ell} \sum_{i=1}^{\ell} \mathbf{x}_i^T \mathbf{M}' \mathbf{x}_i = \max_{\mathbf{x}_1,\ldots,\mathbf{x}_\ell} \left(\sum_{i=1}^{\ell} \mathbf{x}_i^\top \mathbf{M} \mathbf{x}_i + \sum_{i=1}^{\ell} \langle \mathbf{x}_i, \mathbf{u} \rangle^2 \right).$$

Define

$$\mathbf{x}_1^*, ..., \mathbf{x}_\ell^* = \arg\max_{\mathbf{x}_1,\ldots,\mathbf{x}_\ell} \sum_{i=1}^{\ell} \mathbf{x}_i^\top \mathbf{M}' \mathbf{x}_i. \tag{4}$$

We will now prove (by contradiction) that for all $i \leq \ell$, $\|\mathcal{P}_{S^\perp}(\mathbf{x}_i^*)\|_2^2 \leq \frac{\epsilon^2}{100d^2}$. Suppose that there exists some $i' \in 1, ..., \ell$ such that $\|\mathcal{P}_{S^\perp}(\mathbf{x}_{i'}^*)\|_2^2 > \frac{\epsilon^2}{100d^2}$. Then we find

$$\begin{aligned}
&\sum_{i=1}^{\ell} (\mathbf{x}_i^*)^\top \mathbf{M}' \mathbf{x}_i^* = \sum_{i=1}^{\ell} (\mathbf{x}_i^*)^\top \mathbf{M} \mathbf{x}_i^* + \sum_{i=1}^{\ell} \langle \mathbf{x}_i^*, \mathbf{u} \rangle^2 \\
&\leq 1 + \sum_{i=1}^{\ell} (\mathbf{x}_i^*)^\top \mathbf{M} \mathbf{x}_i^* && [\mathbf{x}_i^* \text{ orthonormal so } \sum_{i=1}^{\ell} \langle \mathbf{x}_i^*, \mathbf{u} \rangle^2 \leq \|\mathbf{u}\|^2 \leq 1] \\
&= 1 + \sum_{i=1}^{\ell} \left(\mathcal{P}_{S^\perp}(\mathbf{x}_i^*)^\top \mathbf{M} \mathcal{P}_{S^\perp}(\mathbf{x}_i^*) + \mathcal{P}_S(\mathbf{x}_i^*)^\top \mathbf{M} \mathcal{P}_S(\mathbf{x}_i^*) \right) \\
&= 1 + \sum_{i=1}^{\ell} \mathcal{P}_{S^\perp}(\mathbf{x}_i^*)^\top \mathbf{M} \mathcal{P}_{S^\perp}(\mathbf{x}_i^*) + \sum_{i=1}^{\ell} \mathcal{P}_S(\mathbf{x}_i^*)^\top \mathbf{M} \mathcal{P}_S(\mathbf{x}_i^*) \\
&= 1 + d\epsilon + \sum_{i=1}^{\ell} \mathcal{P}_S(\mathbf{x}_i^*)^\top \mathbf{M} \mathcal{P}_S(\mathbf{x}_i^*) && [S^\perp = \text{span of evectors with evalues} \leq \epsilon] \\
&= 1 + d\epsilon + \sum_{i=1}^{\ell} \left(\sum_{k=1}^{\ell_\epsilon} \langle \mathbf{x}_i^*, \mathbf{w}_k \rangle \mathbf{w}_k \right)^\top \mathbf{M} \left(\sum_{k=1}^{\ell_\epsilon} \langle \mathbf{x}_i^*, \mathbf{w}_k \rangle \mathbf{w}_k \right) \\
&= 1 + d\epsilon + \sum_{i=1}^{\ell} \sum_{k=1}^{\ell_\epsilon} \langle \mathbf{x}_i^*, \mathbf{w}_k \rangle^2 \mathbf{w}_k^\top \mathbf{M} \mathbf{w}_k \\
&= 1 + d\epsilon + \sum_{i=1}^{\ell} \sum_{k=1}^{\ell_\epsilon} \langle \mathbf{x}_i^*, \mathbf{w}_k \rangle^2 \lambda_k.
\end{aligned} \tag{5}$$

Because $\mathbf{x}_i^*$ are orthonormal, we know that $\sum_{i=1}^{\ell} \langle \mathbf{x}_i^*, \mathbf{w}_k \rangle^2 \leq \|\mathbf{w}_k\|_2^2 = 1$. Because we assumed that $\|\mathcal{P}_{S^\perp}(\mathbf{x}_{i'}^*)\|_2^2 > \frac{\epsilon^2}{100d^2}$, we know that $\sum_{i=1}^{\ell} \sum_{k=1}^{\ell_\epsilon} \langle \mathbf{x}_i^*, \mathbf{w}_k \rangle^2 = \sum_{i=1}^{\ell} \|\mathcal{P}_S(\mathbf{x}_i^*)\|^2 \leq \ell - \frac{\epsilon^2}{100d^2}$. Combining these two statements with the fact that λ_k is decreasing in k,

$$\sum_{i=1}^{\ell} \sum_{k=1}^{\ell_\epsilon} \langle \mathbf{x}_i^*, \mathbf{w}_k \rangle^2 \lambda_k = \sum_{k=1}^{\ell_\epsilon} \sum_{i=1}^{\ell} \langle \mathbf{x}_i^*, \mathbf{w}_k \rangle^2 \lambda_k \leq \left(1 - \frac{\epsilon^2}{100d^2}\right) \lambda_\ell + \sum_{i=1}^{\ell-1} \lambda_i.$$

Continuing where we left off with Equation (5), we have that

$$\begin{aligned}
&= 1 + d\epsilon + \sum_{i=1}^{\ell} \sum_{k=1}^{\ell_\epsilon} \langle \mathbf{x}_i^*, \mathbf{w}_k \rangle^2 \lambda_k \leq 1 + d\epsilon + \left(1 - \frac{\epsilon^2}{100d^2}\right) \lambda_\ell + \sum_{i=1}^{\ell-1} \lambda_i \\
&\leq 1 + d\epsilon - \frac{\epsilon^2}{100d^2} \lambda_\ell + \sum_{i=1}^{\ell} \lambda_i \\
&\leq 1 + d\epsilon - 2d + \sum_{i=1}^{\ell} \lambda_i && [\lambda_\ell \geq \frac{200d^3}{\epsilon^2}] \\
&< \sum_{i=1}^{\ell} \lambda_i && [\epsilon < 1]
\end{aligned}$$

$$= \sum_{i=1}^{\ell} \mathbf{w}_i^\top \mathbf{M} \mathbf{w}_i \leq \sum_{i=1}^{\ell} \mathbf{w}_i^\top \mathbf{M}' \mathbf{w}_i.$$

Therefore, we have a contradiction, as $\mathbf{x}_1^*, ..., \mathbf{x}_\ell^*$ cannot be a solution to Equation (4) because these vectors are strictly beaten by $\mathbf{w}_1, ..., \mathbf{w}_\ell$.

Therefore, we have shown that $\|\mathcal{P}_{S^\perp}(\mathbf{x}_i^*)\|_2^2 \leq \frac{\epsilon^2}{100d^2}$ for all i.

In the following equation, we define P_S as the projection matrix for projecting a vector onto S. Now, we have that

$$\begin{aligned}
\sum_{i=1}^{\ell} \lambda_i' &= \sum_{i=1}^{\ell} (\mathbf{x}_i^*)^\top \mathbf{M} \mathbf{x}_i^* + \sum_{i=1}^{\ell} \langle \mathbf{x}_i^*, \mathbf{u} \rangle^2 \\
&\leq \sum_{i=1}^{\ell} \lambda_i + \sum_{i=1}^{\ell} \langle \mathbf{x}_i^*, \mathbf{u} \rangle^2 \\
&\leq \sum_{i=1}^{\ell} \lambda_i + \sum_{i=1}^{\ell} \Big(\langle \mathcal{P}_S(\mathbf{x}_i^*), \mathcal{P}_S(\mathbf{u}) \rangle + \langle \mathcal{P}_{S^\perp}(\mathbf{x}_i^*), \mathcal{P}_{S^\perp}(\mathbf{u}) \rangle \Big)^2 \\
&\leq \sum_{i=1}^{\ell} \lambda_i + \sum_{i=1}^{\ell} (|\langle \mathcal{P}_S(\mathbf{x}_i^*), \mathcal{P}_S(\mathbf{u}) \rangle| + \frac{\epsilon}{10d})^2 && [\|\mathcal{P}_{S^\perp}(\mathbf{x}_i^*)\|_2^2 \leq \frac{\epsilon^2}{100d^2}] \\
&= \sum_{i=1}^{\ell} \lambda_i + \sum_{i=1}^{\ell} \left(\langle \mathcal{P}_S(\mathbf{x}_i^*), \mathcal{P}_S(\mathbf{u}) \rangle^2 + \frac{\epsilon}{5d} |\langle \mathcal{P}_S(\mathbf{x}_i^*), \mathcal{P}_S(\mathbf{u}) \rangle| + \frac{\epsilon^2}{100d^2} \right) \\
&\leq \sum_{i=1}^{\ell} \lambda_i + \sum_{i=1}^{\ell} \left(\langle \mathcal{P}_S(\mathbf{x}_i^*), \mathcal{P}_S(\mathbf{u}) \rangle^2 + \frac{\epsilon}{5d} + \frac{\epsilon^2}{100d^2} \right) \\
&\leq \sum_{i=1}^{\ell} \lambda_i + \frac{\epsilon}{5} + \frac{\epsilon^2}{100d} + \sum_{i=1}^{\ell} \langle \mathcal{P}_S(\mathbf{x}_i^*), \mathcal{P}_S(\mathbf{u}) \rangle^2 \\
&= \sum_{i=1}^{\ell} \lambda_i + \frac{\epsilon}{5} + \frac{\epsilon^2}{100d} + \sum_{i=1}^{\ell} \big((\mathbf{x}_i^*)^\top P_S^\top P_S \mathbf{u} \big)^2 \\
&= \sum_{i=1}^{\ell} \lambda_i + \frac{\epsilon}{5} + \frac{\epsilon^2}{100d} + \sum_{i=1}^{\ell} \big((\mathbf{x}_i^*)^\top P_S \mathbf{u} \big)^2 \\
&= \sum_{i=1}^{\ell} \lambda_i + \frac{\epsilon}{5} + \frac{\epsilon^2}{100d} + \sum_{i=1}^{\ell} \langle \mathbf{x}_i^*, \mathcal{P}_S(\mathbf{u}) \rangle^2 \\
&\leq \sum_{i=1}^{\ell} \lambda_i + \frac{\epsilon}{5} + \frac{\epsilon^2}{100d} + \|\mathcal{P}_S(\mathbf{u})\|_2^2 \\
&\leq \sum_{i=1}^{\ell} \lambda_i + \frac{\epsilon}{5} + \frac{\epsilon^2}{100d} + 1 - \epsilon \\
&\leq \sum_{i=1}^{\ell} \lambda_i + (1 - \epsilon/2).
\end{aligned}$$

This completes the proof of Lemma 18. □

F Proof of Lemma 17

Proof of Lemma 17. Define $\mathbf{A} \in \mathbb{R}^{j \times d}$ with rows corresponding to $\mathbf{v}_1, ..., \mathbf{v}_j$. Then $\mathbf{M} = \mathbf{A}^\top \mathbf{A}$. We want to write $\mathbf{u} = \mathbf{A}^\top \mathbf{c}$ for $\mathbf{c} \in \mathbb{R}^j$.

Because $\mathbf{w}_1, ..., \mathbf{w}_\ell$ are orthonormal and $\mathbf{u} \in \mathrm{Span}(\mathbf{w}_1, ..., \mathbf{w}_\ell)$, we have that

$$\mathbf{u} = \sum_{i=1}^{\ell} \langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{w}_i$$

Because $\mathbf{w}_i$ is an eigenvector of $\mathbf{M}$ with eigenvalue λ_i, we know that for any $i \leq \ell$

$$\lambda_i \mathbf{w}_i = \mathbf{M}\mathbf{w}_i = \mathbf{A}^\top \mathbf{A}\mathbf{w}_i$$

Rearranging terms and multiplying both sides by $\langle \mathbf{u}, \mathbf{w}_i \rangle$, we have that for any $i \leq \ell$

$$\langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{w}_i = \mathbf{A}^\top \left(\frac{\langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{A}\mathbf{w}_i}{\lambda_i} \right).$$

Therefore, we have that

$$\mathbf{u} = \sum_{i=1}^{\ell} \langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{w}_i = \sum_{i=1}^{\ell} \mathbf{A}^\top \left(\frac{\langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{A}\mathbf{w}_i}{\lambda_i} \right) = \mathbf{A}^\top \sum_{i=1}^{\ell} \left(\frac{\langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{A}\mathbf{w}_i}{\lambda_i} \right).$$

Now, we can define

$$\mathbf{c} = \sum_{i=1}^{\ell} \left(\frac{\langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{A}\mathbf{w}_i}{\lambda_i} \right) = \mathbf{A} \sum_{i=1}^{\ell} \frac{\langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{w}_i}{\lambda_i}.$$

This implies that

$$\begin{aligned}
\|\mathbf{c}\|_2^2 &= \left\| \mathbf{A} \sum_{i=1}^{\ell} \frac{\langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{w}_i}{\lambda_i} \right\|_2^2 \\
&= \left(\sum_{i=1}^{\ell} \frac{\langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{w}_i}{\lambda_i} \right)^\top \mathbf{M} \left(\sum_{i=1}^{\ell} \frac{\langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{w}_i}{\lambda_i} \right) \\
&= \sum_{i=1}^{\ell} \lambda_i \left(\frac{\langle \mathbf{u}, \mathbf{w}_i \rangle}{\lambda_i} \right)^2 && [\mathbf{w}_i^\top \mathbf{M}\mathbf{w}_i = \lambda_i] \\
&= \sum_{i=1}^{\ell} \frac{\langle \mathbf{u}, \mathbf{w}_i \rangle^2}{\lambda_i} \\
&\leq \frac{\|\mathbf{u}\|_2^2}{\epsilon} \\
&\leq \frac{1}{\epsilon}.
\end{aligned}$$

This vector $\mathbf{c}$ therefore satisfies the desired properties. ☐

G Proof of Lemma 7

We begin by proving the following lemma.

Lemma 19. *Let μ be a probability distribution on $\mathbb{R}^d$ with finite first moment and suppose*

$$\min_{\|\mathbf{v}\|=1} \mathbb{E}^{\mathbf{x}\sim\mu}[\langle \mathbf{v}, \mathbf{x} \rangle_+] \geq \epsilon.$$

Then for any $\mathbf{w}$ with $\|\mathbf{w}\| < \epsilon$ there is a $[0,1]$-valued measurable function f such that $\mathbb{E}[\mathbf{x}f(\mathbf{x})] = \mathbf{w}$.

Proof of Lemma 19. Let K be the set of possible vectors $\mathbb{E}[\mathbf{x}f(\mathbf{x})]$ where f is a a $[0,1]$-valued measurable function. Then for any $\mathbf{a}, \mathbf{b} \in K$, there exist corresponding $[0,1]$-valued functions f^a and f^b such that $\mathbb{E}[\mathbf{x}f^a(\mathbf{x})] = \mathbf{a}$ and $\mathbb{E}[\mathbf{x}f^b(\mathbf{x})] = \mathbf{b}$. Therefore for any $t \in [0,1]$, $t\mathbf{a} + (1-t)\mathbf{b} \in K$ because $\mathbb{E}[\mathbf{x}f^t(\mathbf{x})] = t\mathbf{a} + (1-t)\mathbf{b}$ for $f^t(\mathbf{x}) = tf^a(\mathbf{x}) + (1-t)f^b(\mathbf{x})$. This implies that K is convex.

We will now prove the desired result by contradiction. Suppose $\mathbf{w} \notin K$. Because K is convex, if $\mathbf{w} \notin K$ then there is a "separating hyperplane" unit vector $\mathbf{v}$ such that

$$\sup_{\mathbf{u} \in K} \langle \mathbf{v}, \mathbf{u} \rangle \leq \langle \mathbf{v}, \mathbf{w} \rangle.$$

(Note that we do not argue here that K is closed.) Because by assumption we have that $\|\mathbf{w}\| < \epsilon$, this implies that

$$\sup_{\mathbf{u} \in K} \langle \mathbf{v}, \mathbf{u} \rangle < \epsilon.$$

For any $\mathbf{v}$, by definition of K and linearity of expectation we have that

$$\begin{aligned}
\sup_{\mathbf{u} \in K} \langle \mathbf{v}, \mathbf{u} \rangle &= \sup_{f:\mathbb{R}^d \to [0,1]} \langle \mathbf{v}, \mathbb{E}^{x \sim \mu}[\mathbf{x} f(\mathbf{x})] \rangle \\
&= \sup_{f:\mathbb{R}^d \to [0,1]} \mathbb{E}^{\mathbf{x} \sim \mu}[f(\mathbf{x}) \langle \mathbf{v}, \mathbf{x} \rangle] \\
&= \mathbb{E}^{x \sim \mu}[\langle \mathbf{v}, \mathbf{x} \rangle_+] \\
&\geq \epsilon.
\end{aligned}$$

where the last line followed from the assumption of the lemma. This gives a contradiction, and therefore $\mathbf{w} \in K$ must be true. $\square$

Proof of Lemma 7. Applying the above lemma with $\mathbf{w} = \mathbf{0}$, there exists $f^0 : \mathbb{R}^d \to [0,1]$ such that $\mathbb{E}[\mathbf{x} f^0(\mathbf{x})] = \mathbf{0}$. Define the function f' as $f'(\mathbf{x}) = \frac{f^0(\mathbf{x}) + 2\epsilon}{4\max(\|\mathbb{E}[\mathbf{x}]\|, 1)}$. By this construction,

$$\|\mathbb{E}[\mathbf{x} f'(\mathbf{x})]\| = \left\| \frac{\mathbb{E}[\mathbf{x} f^0(\mathbf{x})] + 2\epsilon \mathbb{E}[\mathbf{x}]}{4\max(\|\mathbb{E}[\mathbf{x}]\|, 1)} \right\| = \frac{2\epsilon \|\mathbb{E}[\mathbf{x}]\|}{4\max(\|\mathbb{E}[\mathbf{x}]\|, 1)} \leq \frac{\epsilon}{2}.$$

Therefore, $\mathbf{w} := -\mathbb{E}[\mathbf{x} f'(\mathbf{x})]$ satisfies $\|\mathbf{w}\| < \epsilon$. Applying the above lemma again, there exists $f^w : \mathbb{R}^d \to [0,1]$ such that $\mathbb{E}[\mathbf{x} f^w(\mathbf{x})] = -\mathbb{E}[\mathbf{x} f'(\mathbf{x})]$. Now define $f(\mathbf{x}) = \frac{f'(\mathbf{x}) + f^w(\mathbf{x})}{2}$. By construction, we have that $1 \geq f(\mathbf{x}) \geq \frac{f'(\mathbf{x})}{2} \geq \frac{\epsilon}{4\max(\|\mathbb{E}[\mathbf{x}]\|, 1)}$. Furthermore, by linearity of expectation and the construction of f^w we have that

$$\mathbb{E}[\mathbf{x} f(\mathbf{x})] = \mathbb{E}\left[\mathbf{x} \frac{f'(\mathbf{x}) + f^w(\mathbf{x})}{2}\right] = \frac{1}{2}\left(\mathbb{E}[\mathbf{x} f'(\mathbf{x})] + \mathbb{E}[\mathbf{x} f^w(\mathbf{x})]\right) = \frac{1}{2}\left(\mathbb{E}[\mathbf{x} f'(\mathbf{x})] - \mathbb{E}[\mathbf{x} f'(\mathbf{x})]\right) = 0.$$

Therefore, $f(\mathbf{x})$ is a $\left[\frac{\epsilon}{4\max(\|\mathbb{E}[\mathbf{x}]\|, 1)}, 1\right]$-valued function that satisfies $\mathbb{E}[\mathbf{x} f(\mathbf{x})] = 0$ as desired. $\square$

G.1 Proof of Lemma 14

Lemma 20. *Let $\Phi^C(x) = \mathbb{P}(Z > x)$ for $Z \sim N(0,1)$. Then for $|x| \leq 1$,*

$$\left| \Phi^C(x) - \left(\frac{1}{2} - \frac{1}{\sqrt{2\pi}} x \right) \right| \leq |x^3|/15.$$

Proof of Lemma 20. A third order Taylor expansion shows the error is at most

$$|x^3| \cdot \frac{\sup_{|y| \leq 1} |(\Phi^C)'''(y)|}{6}.$$

For $|y| \leq 1$ we easily compute

$$|(\Phi^C)'''(y)| = \frac{|y^2 - 1| \, e^{-y^2/2}}{\sqrt{2\pi}} \leq 1/\sqrt{2\pi} \leq 2/5. \qquad \square$$

Lemma 21. *If X is K-sub-gaussian, then for any event E and any $\mathbb{P}(E) \geq a > 0$, we have*

$$\begin{aligned}
\mathbb{E}[X^2 1_E] &\leq K^2 \Pr(E) \log(2/a) + K^2 a, \\
\mathbb{E}[X 1_E] &\leq O(\Pr(E) \log(1/a)). \qquad (6)
\end{aligned}$$

Proof of Lemma 21. We prove both estimates using the tail-sum formula. For the truncated second moment,

$$\begin{aligned}
\mathbb{E}[X^2 1_E] &= \int_0^\infty \Pr(|X^2 1_E| \geq t)dt \\
&\leq \int_0^\infty \min(\Pr(E), \Pr(X^2 \geq t))dt \\
&\leq \int_0^\infty \min(\Pr(E), 2e^{-t/K^2})dt \\
&\leq \Pr(E)\log(2/a)K^2 + \int_{\log(2/a)K^2}^\infty 2e^{-t/K^2}dt \\
&= K^2 \Pr(E)\log(2/a) + K^2 a.
\end{aligned}$$

Similarly for the truncated first moment,

$$\begin{aligned}
\mathbb{E}[X 1_E] &= \int_0^\infty \Pr(|X 1_E| \geq t)dt \\
&\leq \int_0^\infty \min(\Pr(E), \Pr(X \geq t))dt \\
&\leq \int_0^\infty \min(\Pr(E), 2e^{-t^2/K^2})dt \\
&\leq \Pr(E)\sqrt{\log(3/a)}K + \int_{\sqrt{\log(3/a)}K}^\infty 2e^{-t^2/K^2}dt \\
&= O(\Pr(E)\log(1/a)).
\end{aligned}$$

□

Proof of Lemma 14. Let $d\mu_X$ be the law of X and $d\mu_{X|r>0}$ be the conditional law of X given the event $r > 0$. Then by Bayes rule,

$$\begin{aligned}
d\mu_{X|r>0}(x) &= \frac{\Pr(r > 0 \mid X = x)d\mu_X(x)}{\Pr(r > 0)} \\
&= \frac{\Pr\left(N(\epsilon x, \sigma^2) > 0\right)d\mu_X(x)}{\Pr(r > 0)} \\
&= \frac{\Phi^C(-\epsilon x/\sigma)d\mu_X(x)}{\Pr(r > 0)}.
\end{aligned}$$

Note that

$$\mathbb{E}[X \mid r > 0] = \int_{-\infty}^\infty x d\mu_{X|r>0}(x) = \frac{1}{\Pr(r > 0)}\int_{-\infty}^\infty x\Phi^C(-\epsilon x/\sigma)d\mu_X(x).$$

Since $\Pr(r > 0) \leq 1$ it suffices to lower-bound the latter integral by a suitable positive value. As long as $\epsilon/\sigma \leq \delta_{L14} \leq \sqrt{\frac{1}{4\sigma_X^2\sqrt{2\pi}}}$, we have

$$\begin{aligned}
&\int_{-\infty}^\infty x\Phi^C(-\epsilon x/\sigma)d\mu_X(x) \\
&= \int_{-\infty}^\infty x\left(\frac{1}{2} + \frac{1}{\sqrt{2\pi}}\frac{\epsilon x}{\sigma}\right)d\mu_X(x) \\
&\qquad + \int_{-\infty}^\infty x\left(\Phi^C(\frac{-\epsilon x}{\sigma}) - \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\frac{\epsilon x}{\sigma}\right)d\mu_X(x) \\
&= \left(\frac{\epsilon\,\mathbb{E}[X^2]}{\sigma\sqrt{2\pi}} + \int_{-\infty}^\infty x\left(\Phi^C(\frac{-\epsilon x}{\sigma}) - \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\frac{\epsilon x}{\sigma}\right)d\mu_X(x)\right) && [\mathbb{E}[X] = 0] \\
&\geq \left(\frac{\epsilon\sigma_X^2}{\sigma\sqrt{2\pi}} - \frac{2\,\mathbb{E}[X^4]\epsilon^3}{\sigma^3}\right) && \text{[Inequality (8) Below]} \\
&\geq \frac{\epsilon\sigma_X^2}{2\sigma\sqrt{2\pi}}. && [\frac{\epsilon^2}{\sigma^2} \leq \frac{\sigma_X^2}{4\,\mathbb{E}[X^4]\sqrt{2\pi}}]
\end{aligned}$$

Above we used the following estimate (8). For sufficiently small δ_{L14},

$$\begin{aligned}
&\int_{-\infty}^{\infty} x\left(\Phi^C(\frac{-\epsilon x}{\sigma}) - \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\frac{\epsilon x}{\sigma}\right) d\mu_X(x) \\
&= \int_{|x|\le\frac{\sigma}{10\epsilon}} x\left(\Phi^C(\frac{-\epsilon x}{\sigma}) - \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\frac{\epsilon x}{\sigma}\right) d\mu_X(x) + \int_{|x|>\frac{\sigma}{10\epsilon}} x\left(\Phi^C(\frac{-\epsilon x}{\sigma}) - \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\frac{\epsilon x}{\sigma}\right) d\mu_X(x) \\
&\ge \int_{|x|\le\frac{\sigma}{10\epsilon}} x\left(\Phi^C(\frac{-\epsilon x}{\sigma}) - \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\frac{\epsilon x}{\sigma}\right) d\mu_X(x) - \int_{|x|>\frac{\sigma}{10\epsilon}} \left(\frac{|x|}{2} + \frac{\epsilon x^2}{\sigma\sqrt{2\pi}}\right) d\mu_X(x) \\
&\ge -\int_{|x|\le\frac{\sigma}{10\epsilon}} |x|\left|\frac{\epsilon x}{\sigma}\right|^3 d\mu_X(x) - \int_{|x|>\frac{\sigma}{10\epsilon}} \left(\frac{|x|}{2} + \frac{\epsilon x^2}{\sigma\sqrt{2\pi}}\right) d\mu_X(x) && \text{[Lemma 20]} \\
&= -\frac{\mathbb{E}[X^4]\epsilon^3}{\sigma^3} - \int_{|x|>\frac{\sigma}{10\epsilon}} \left(\frac{|x|}{2} + \frac{\epsilon x^2}{\sigma\sqrt{2\pi}}\right) d\mu_X(x) \\
&\ge -\frac{\mathbb{E}[X^4]\epsilon^3}{\sigma^3} - \int_{|x|>\frac{\sigma}{10\epsilon}} x^2 d\mu_X(x) && [\text{if } \epsilon/\sigma \le \delta_{L14} \le 1/10] \\
&\ge -\frac{\mathbb{E}[X^4]\epsilon^3}{\sigma^3} - \left(K^2\Pr(E)\log(2/a) + K^2 a\right) && [\text{Lemma 21}, E := \{|x| > \frac{\sigma}{10\epsilon}\}, a = 2e^{-\frac{\sigma^2}{100\epsilon^2K^2}}] \\
&\ge -\frac{\mathbb{E}[X^4]\epsilon^3}{\sigma^3} - 2K^2 e^{-\frac{\sigma^2}{100K^2\epsilon^2}}\left(\frac{\sigma^2}{100\epsilon^2K^2} + 1\right) && [\Pr(E) \le 2e^{-\sigma^2/(100\epsilon^2K^2)}] \\
&\ge -\frac{2\,\mathbb{E}[X^4]\epsilon^3}{\sigma^3} && [2K^2 e^{-\frac{\sigma^2}{100K^2\epsilon^2}}\left(\frac{\sigma^2}{100\epsilon^2K^2} + 1\right) \le \frac{(\sigma_X^2)^2\epsilon^3}{\sigma^3} \le \frac{\mathbb{E}[X^4]\epsilon^3}{\sigma^3}]
\end{aligned} \tag{8}$$

where the second to last line holds for sufficiently small ϵ/σ. $\square$

G.2 Proof of Lemma 15

Proof of Lemma 15. First, we observe that $\mathbb{E}[Y \mid r > 0] = \frac{\mathbb{E}[Y\mathbf{1}\{r>0\}]}{\mathbb{P}(r>0)}$. If $\frac{\epsilon}{\sigma}K\sqrt{\log(4)} \le 1/2$, we also have that

$$\begin{aligned}
&\Pr(r > 0) \\
&\ge \Pr(r > 0 \mid X \ge -K\sqrt{\log(4)})\Pr(X \ge -K\sqrt{\log(4)}) \\
&\ge \Phi^C(\frac{\epsilon}{\sigma}K\sqrt{\log(4)})\Pr(X \ge -K\sqrt{\log(4)}) \\
&\ge \Phi^C(\frac{\epsilon}{\sigma}K\sqrt{\log(4)})\left(1 - 2e^{-K^2\log(4)/K^2}\right) \\
&\ge \Phi^C(1/2)1/2 && [\frac{\epsilon}{\sigma}K\sqrt{\log(4)} \le 1/2] \\
&\ge 1/8.
\end{aligned}$$

Therefore, it is sufficient to upper bound $|\mathbb{E}[Y\mathbf{1}\{r > 0\}]|$. By law of total expectation,

$$\mathbb{E}[Y\mathbf{1}\{r > 0\}] = \mathbb{E}[\mathbb{E}[Y\mathbf{1}\{r > 0\} \mid X]] = \mathbb{E}\left[\mathbb{E}[Y \mid X]\mathbb{P}(Z > -\frac{\epsilon}{\sigma}X)\right].$$

Define $Q(X) = \mathbb{E}[Y \mid X]$. Because Y is sub-gaussian, the random variable $Q(X)$ must also be sub-gaussian with parameter $\sqrt{18}K$ (see [Van Handel, 2014, Exercise 3.1]). Furthermore, $\mathbb{E}[Q(X)] = \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Y] = 0$. Now, we have the following

$$\begin{aligned}
&|\mathbb{E}[Y\mathbf{1}\{r > 0\}]| \\
&= \left|\mathbb{E}\left[\mathbb{E}[Y \mid X]\mathbb{P}(Z > -\frac{\epsilon}{\sigma}X)\right]\right| \\
&= \left|\int_{-\infty}^{\infty} Q(x)\Phi^C(-\frac{\epsilon x}{\sigma})d\mu_X(x)\right| \\
&= \left|\int_{-\infty}^{\infty} \frac{1}{2}Q(x)d\mu_X(x) + \int_{-\infty}^{\infty} Q(x)\left(\Phi^C(-\frac{\epsilon x}{\sigma}) - \frac{1}{2}\right) d\mu_X(x)\right|
\end{aligned}$$

$$
\begin{aligned}
&= \left| \frac{1}{2}\,\mathbb{E}[Q(X)] + \int_{-\infty}^{\infty} Q(x)\left(\Phi^C(-\frac{\epsilon x}{\sigma}) - \frac{1}{2}\right) d\mu_X(x) \right| \\
&= \left| \int_{-\infty}^{\infty} Q(x)\left(\Phi^C(-\frac{\epsilon x}{\sigma}) - \frac{1}{2}\right) d\mu_X(x) \right| \\
&\le \int_{|x|\le \frac{\sigma}{10\epsilon}} |Q(x)| \left(\frac{|\epsilon x|}{\sigma\sqrt{2\pi}} + \left|\frac{\epsilon x}{\sigma}\right|^3 \right) d\mu_X(x) \\
&\quad + \int_{|x|> \frac{\sigma}{10\epsilon}} Q(x)\left(\Phi^C(-\frac{\epsilon x}{\sigma}) - \frac{1}{2}\right) d\mu_X(x) && \text{[Lemma 20]} \\
&\le \frac{\epsilon}{\sigma}\int_{-\infty}^{\infty} |Q(x)| \left(\frac{|x|}{\sqrt{2\pi}} + |x^3| \right) d\mu_X(x) + \frac{1}{2}\int_{|x|> \frac{\sigma}{10\epsilon}} |Q(x)| d\mu_X(x) && [\epsilon/\sigma \le 1] \\
&\le \frac{\epsilon}{\sigma}\sqrt{\mathbb{E}[Q(X)^2]\,\mathbb{E}\left[\left(\frac{|x|}{\sqrt{2\pi}} + |x^3|\right)^2\right]} + \frac{1}{2}O\left(\frac{\sigma}{10\epsilon K}(2e^{-\frac{\sigma^2}{100\epsilon^2 K^2}})\right) && \text{[Lemma 21 Equation (6)]} \\
&\le O(\epsilon/\sigma).
\end{aligned}
$$

where in the second to last line we used that $\Pr(|X| > \frac{\sigma}{10\epsilon}) \le 2e^{-\frac{\sigma^2}{100\epsilon^2 K^2}}$ and that $Q(X)$ is $K\sqrt{18}$-sub-gaussian. The last line again uses that both X and $Q(X)$ are sub-gaussian.

Therefore, we have shown that

$$
|\,\mathbb{E}[Y \mid r > 0]| = \left| \frac{\mathbb{E}[Y\mathbf{1}\{r > 0\}]}{\mathbb{P}(r > 0)} \right| \le 8\,|\mathbb{E}[Y\mathbf{1}\{r > 0\}]| = O(\epsilon/\sigma).
$$

By symmetric arguments to the ones above, we have the same bound on $|\,\mathbb{E}[Y \mid r \le 0]|$. ☐

H Proof of Lemma 16

Proof of Lemma 16. Fix any $\mathbf{v}$ such that $\|\mathbf{v}\| = 1$. First, we will show that

$$
\Pr(|\mathbf{v}\cdot(\mathbf{Z}(\mathbf{Y}) - \mathbf{X})| \ge c_d/2) \le \epsilon_d/2.
$$

To do this, we will show that $\langle \mathbf{v}, (\mathbf{Z}(\mathbf{Y}) - \mathbf{X})\rangle$ is a sub-gaussian random variable. Let $\hat{\mathbf{X}}$ be a draw from the distribution of $\mathbf{X} \mid \mathbf{Y}$. Then we have that

$$
\langle \mathbf{v}, \hat{\mathbf{X}} - \mathbf{X}\rangle = \langle \mathbf{v}, \hat{\mathbf{X}} - \mathbf{Y}\rangle + \langle \mathbf{v}, \mathbf{Y} - \mathbf{X}\rangle.
$$

By standard properties of posterior samples, $\langle \mathbf{v}, \hat{\mathbf{X}} - \mathbf{Y}\rangle$ and $\langle \mathbf{v}, \mathbf{Y} - \mathbf{X}\rangle$ are identically distributed with distribution $N(0, \sigma^2)$ for $\sigma^2 = \sum_{i=1}^d \mathbf{v}_i^2 s_i$ (here one averages over all randomness). Therefore, we have that

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(t\langle \mathbf{v}, \mathbf{Z}(\mathbf{Y}) - \mathbf{X}\rangle\right)\right] \\
&= \mathbb{E}\left[\exp\left(t\langle \mathbf{v}, \mathbb{E}[\hat{\mathbf{X}}] - \mathbf{X}\rangle\right)\right] \\
&\le \mathbb{E}\left[\exp\left(t\langle \mathbf{v}, \hat{\mathbf{X}} - \mathbf{X}\rangle\right)\right] && \text{[Jensen]} \\
&= \mathbb{E}\left[\exp\left(t\langle \mathbf{v}, \hat{\mathbf{X}} - \mathbf{Y} + \mathbf{Y} - \mathbf{X}\rangle\right)\right] \\
&\le \sqrt{\mathbb{E}\left[\exp\left(2t\langle \mathbf{v}, \hat{\mathbf{X}} - \mathbf{Y}\rangle\right)\right]\mathbb{E}\left[\exp\left(2t\langle \mathbf{v}, \mathbf{Y} - \mathbf{X}\rangle\right)\right]} && \text{[Cauchy-Schwarz]} \\
&\le e^{2t^2\sigma^2}.
\end{aligned}
$$

Therefore, $\langle \mathbf{v}, (\mathbf{Z}(\mathbf{Y}) - \mathbf{X})\rangle$ is sub-gaussian and satisfies the tail bound

$$
\Pr(|\langle \mathbf{v}, (\mathbf{Z}(\mathbf{Y}) - \mathbf{X})\rangle| > t) \le 2e^{-t^2/(8\sigma^2)}.
$$

Taking $t = c_d/2$, because $\sigma^2 = \sum_{i=0}^d \mathbf{v}_i^2 s_i \le \max_i s_i \le \frac{c_d^2/32}{\log(4/\epsilon_d)}$, we have that

$$
\Pr(|\langle \mathbf{v}, (\mathbf{Z}(\mathbf{Y}) - \mathbf{X})\rangle| > \frac{c_d}{2}) \le 2e^{-c_d^2/(32\sigma^2)} = \epsilon_d/2. \tag{9}
$$

Now, we can prove the desired result that

$$
\begin{aligned}
&\mathbb{E}[(\langle \mathbf{Z}(\mathbf{Y}), \mathbf{v}\rangle)^+] \\
&\geq \mathbb{E}\left[(\langle \mathbf{Z}(\mathbf{Y}), \mathbf{v}\rangle)^+ \middle| |\langle \mathbf{v}, (\mathbf{Z}(\mathbf{Y}) - \mathbf{X})\rangle| \leq \frac{c_d}{2}, \langle \mathbf{X}, \mathbf{v}\rangle \geq c_d\right] \Pr\left(|\langle \mathbf{v}, (\mathbf{Z}(\mathbf{Y}) - \mathbf{X})\rangle| \leq \frac{c_d}{2}, \langle \mathbf{X}, \mathbf{v}\rangle \geq c_d\right) \\
&\geq \mathbb{E}\left[\frac{c_d}{2} \middle| |\langle \mathbf{v}, (\mathbf{Z}(\mathbf{Y}) - \mathbf{X})\rangle| \leq \frac{c_d}{2}, \langle \mathbf{X}, \mathbf{v}\rangle \geq c_d\right] \Pr\left(|\langle \mathbf{v}, (\mathbf{Z}(\mathbf{Y}) - \mathbf{X})\rangle| \leq \frac{c_d}{2}, \langle \mathbf{X}, \mathbf{v}\rangle \geq c_d\right) \\
&= \frac{c_d}{2} \Pr\left(|\langle \mathbf{v}, (\mathbf{Z}(\mathbf{Y}) - \mathbf{X})\rangle| \leq \frac{c_d}{2}, \langle \mathbf{X}, \mathbf{v}\rangle \geq c_d\right) \\
&\geq \frac{c_d}{2}\left(\Pr\left(\langle \mathbf{X}, \mathbf{v}\rangle \geq c_d\right) - \Pr\left(\langle \mathbf{v}, (\mathbf{Z}(\mathbf{Y}) - \mathbf{X})\rangle| > \frac{c_d}{2}\right)\right) \\
&\geq \frac{c_d}{2}\left(\epsilon_d - \frac{\epsilon_d}{2}\right) && \text{[Eq (9) and lemma assum]} \\
&= \frac{c_d \epsilon_d}{4}.
\end{aligned}
$$

□

H.1 Proof of Lemma 13

Proof of Lemma 13. Let B be a Bernoulli random variable such that $\Pr(B = 1) = \epsilon$ and let $Z \sim N(0, 1)$ be independent of B and X. Then we can write $R \sim X \cdot B + Z$.

Then we have that

$$
\begin{aligned}
&\mathbb{E}[X \mid R > 0] \\
&= \mathbb{E}[X \mid R > 0, B = 1] \Pr(B = 1 \mid R > 0) + E[X \mid R > 0, B = 0] \Pr(B = 0 \mid R > 0) \\
&= \mathbb{E}[X \mid X + Z > 0] \Pr(B = 1 \mid R > 0) \\
&= \mathbb{E}[X \mid X + Z > 0] \frac{\Pr(R > 0 \mid B = 1) \Pr(B = 1)}{\Pr(R > 0)} \\
&\geq \mathbb{E}[X \mid X + Z > 0] \Pr(R > 0 \mid B = 1) \Pr(B = 1) \\
&= \mathbb{E}[X \mid X + Z > 0] \Pr(X + Z > 0)\epsilon. && (10)
\end{aligned}
$$

Next, we need to lower bound $\mathbb{E}[X \mid X + Z > 0] \Pr(X + Z > 0)$. Applying Baye's rule gives

$$
\begin{aligned}
&\mathbb{E}[X \mid X + Z > 0] \Pr(X + Z > 0) \\
&= \Pr(X + Z > 0) \int_{-\infty}^{\infty} x d\mu_{X|X+Z>0}(x) \\
&= \int_{-\infty}^{\infty} x \left(\Phi^C(-x)\right) d\mu_X(x) \\
&= \int_{-\infty}^{\infty} x \left(\Phi^C(-x) - 1/2\right) d\mu_X(x) && [\mathbb{E}[X] = 0]
\end{aligned}
$$

Note that $\left(\Phi^C(-x) - 1/2\right)$ has the same sign as x and has magnitude increasing in $|x|$. Therefore,

$$
\begin{aligned}
&\geq \int_{x \geq \frac{\sigma_X}{\sqrt{10}}} \frac{\sigma_X}{\sqrt{10}} \mathbb{P}\left(0 \leq Z \leq \frac{\sigma_X}{\sqrt{10}}\right) d\mu_X(x) + \int_{x \leq -\frac{\sigma_X}{\sqrt{10}}} \frac{\sigma_X}{\sqrt{10}} \mathbb{P}\left(0 \leq Z \leq \frac{\sigma_X}{\sqrt{10}}\right) d\mu_X(x) \\
&= \frac{\sigma_X \mathbb{P}(0 \leq Z \leq \frac{\sigma_X}{\sqrt{10}})}{\sqrt{10}} \mathbb{P}(|X| > \frac{\sigma_X}{\sqrt{10}}) \\
&\geq \frac{\sigma_X \mathbb{P}(0 \leq Z \leq \frac{\sigma_X}{\sqrt{10}})}{\sqrt{10}} \left(\frac{4\sigma_X^2}{5K^2 \log\left(\frac{20K^2}{\sigma_X^2}\right)}\right). && \text{[Equation (12) below]} \\
&\geq \Omega(\sigma_X^5). && (11)
\end{aligned}
$$

Combining Equations (10) and (11) gives the desired result of the lemma.

It remains to show the lower bound on $\mathbb{P}(|X| > \frac{\sigma_X}{\sqrt{10}})$ used in the penultimate line above.

Define $a = K^2 \log\left(\frac{20K^2}{\sigma_X^2}\right) \geq \sigma_X^2 \log(10)/2 > \sigma_X^2/10$ (using Equation (2)). Next, we observe that

$$\begin{aligned}
&\mathbb{E}[X^2] \\
&= \int_0^\infty \mathbb{P}(X^2 > t)dt \\
&= \int_0^\infty \mathbb{P}(X > \sqrt{t})dt \\
&= \int_0^{\sigma_X^2/10} \mathbb{P}(X > \sqrt{t})dt + \int_{\sigma_X^2/10}^a \mathbb{P}(X > \sqrt{t})dt + \int_a^\infty \mathbb{P}(X > \sqrt{t})dt \\
&\leq \sigma_X^2/10 + \int_{\sigma_X^2/10}^a \mathbb{P}(X > \sqrt{t})dt + \int_a^\infty 2e^{-t/K^2}dt \\
&= \sigma_X^2/10 + \int_{\sigma_X^2/10}^a \mathbb{P}(X > \sqrt{t})dt + 2K^2 e^{-a/K^2} \\
&= \sigma_X^2/5 + \int_{\sigma_X^2/10}^a \mathbb{P}(X > \sqrt{t})dt && \text{[Def of } a\text{]} \\
&\leq \sigma_X^2/5 + \left(a - \frac{\sigma_X^2}{10}\right)\mathbb{P}\left(X > \frac{\sigma_X}{\sqrt{10}}\right). && [\mathbb{P}(X > \sqrt{t}) \text{ monotone decr.}]
\end{aligned}$$

Since $\mathbb{E}[X^2] = \sigma_X^2$, this implies that

$$\left(a - \frac{\sigma_X^2}{10}\right)\mathbb{P}\left(X > \frac{\sigma_X}{\sqrt{10}}\right) \geq \frac{4\sigma_X^2}{5}.$$

Therefore, we can conclude that

$$\mathbb{P}\left(X > \frac{\sigma_X}{\sqrt{10}}\right) \geq \frac{4\sigma_X^2/5}{a - \sigma_X^2/10} \geq \frac{4\sigma_X^2/5}{a} = \frac{4\sigma_X^2}{5K^2 \log\left(\frac{20K^2}{\sigma_X^2}\right)}. \tag{12}$$

By symmetry, identical logic as above gives the desired upper bound on $\mathbb{E}[X \mid R \leq 0]$. □

I Proof of Proposition 10

Proof of Proposition 10. We first show that $\mathbf{A}^{(t)}$ on Line 6 satisfies $\mathbf{A}^{(t)} \in S^\perp$ when $\Psi = 1$. Recall $\mathbf{x}^*$ defined as $x_\ell^* = \langle \boldsymbol{\ell}^*, \mathbf{w}_\ell \rangle$ and recall that $\mathbf{z}(\mathbf{y}) = \mathbb{E}[\mathbf{x}^* \mid \hat{\mathbf{y}} = \mathbf{y}]$. By construction,

$$\begin{aligned}
&\mathbb{E}[\mathbf{x}^* \mid \Psi = 1] \\
&= \int \mathbb{E}[\mathbf{x}^* \mid \Psi = 1, \hat{\mathbf{y}} = \mathbf{y}] d\mu_{\hat{\mathbf{y}} \mid \Psi = 1}(\mathbf{y}) \\
&= \int \mathbb{E}[\mathbf{x}^* \mid \hat{\mathbf{y}} = \mathbf{y}] \frac{\Pr(\Psi = 1 \mid \hat{\mathbf{y}} = \mathbf{y})}{\Pr(\Psi = 1)} d\mu_{\hat{\mathbf{y}}}(\mathbf{y}) && [\Psi = 1 \text{ is a function of } \hat{\mathbf{y}}] \\
&= \frac{1}{\Pr(\Psi = 1)} \int \mathbb{E}[\mathbf{x}^* \mid \hat{\mathbf{y}} = \mathbf{y}] f\left(\mathbf{z}(\mathbf{y})\right) d\mu_{\hat{\mathbf{y}}}(\mathbf{y}) \\
&= \frac{1}{\Pr(\Psi = 1)} \int \mathbf{z}(\mathbf{y}) f(\mathbf{z}(\mathbf{y})) d\mu_{\hat{\mathbf{y}}}(\mathbf{y}) \\
&= \frac{1}{\Pr(\Psi = 1)} \mathbb{E}\left[\mathbf{z}(\hat{\mathbf{y}}) f(\mathbf{z}(\hat{\mathbf{y}}))\right] \\
&= 0. && [\text{Definition of } f]
\end{aligned}$$

Because $x_\ell^* = \langle \boldsymbol{\ell}^*, \mathbf{w}_\ell \rangle$ for $\ell \leq \ell_\lambda$, this implies that $\mathbb{E}[\langle \boldsymbol{\ell}^*, \mathbf{w}_\ell \rangle \mid \Psi = 1] = 0$ for $\ell \leq \ell_\lambda$. Therefore, we must have that $\mathbb{E}[\boldsymbol{\ell}^* \mid \Psi = 1] \in S^\perp$. By construction of $\mathbf{A}^{(t)}$ in Line 6, this implies that $\mathbf{A}^{(t)} \in S^\perp$ when $\Psi = 1$.

Define

$$\mathbf{A} = \begin{cases} \frac{\mathbb{E}[\boldsymbol{\ell}^* \mid \Psi=1]}{\|\mathbb{E}[\boldsymbol{\ell}^* \mid \Psi=1]\|_2} & \text{if } \mathbb{E}[\boldsymbol{\ell}^* \mid \Psi = 1] \neq 0 \\ \mathbf{w}_{\ell_\lambda+1} & \text{otherwise,} \end{cases}$$

in other words $\mathbf{A}$ is equal to $\mathbf{A}^{(t)}$ when $\Psi = 1$.

By the choice of f, we have that $\mathbb{P}(\Psi = 1) \geq \frac{\epsilon_d c_d}{16 \max(\|\mathbb{E}[\mathbf{z}(\hat{\mathbf{y}})]\|, 1)} \geq \frac{\epsilon_d c_d}{16\left(K\sqrt{\pi}+1\right)}$, where in the last line we used that Equation (B) implies

$$\max(\|\mathbb{E}[\mathbf{z}(\hat{\mathbf{y}})]\|, 1) = \max(\|\mathbb{E}[\mathbf{x}^*]\|, 1) \leq \max\left(\|\mathbb{E}[\boldsymbol{\ell}^*]\|, 1\right) \leq \max\left(K\sqrt{\pi}, 1\right) \leq K\sqrt{\pi} + 1.$$

By construction, we therefore have that for any realization of $\mathbf{z}(\hat{\mathbf{y}})$, the probability that $R = r^{(t)} = \langle \boldsymbol{\ell}^*, \mathbf{A}^{(t)} \rangle + w_t = \langle \boldsymbol{\ell}^*, \mathbf{A} \rangle + w_t$ is exactly $\frac{\epsilon_d c_d}{16\left(K\sqrt{\pi}+1\right)}$ and otherwise $R \sim N(0, 1)$.

We can now apply Lemma 13 with $X = \langle \boldsymbol{\ell}^*, \mathbf{A} \rangle$, and $\epsilon = \frac{\epsilon_d c_d}{16\left(K\sqrt{\pi}+1\right)}$ to get that either $\mathbf{a} = \mathbf{w}_{\ell_\lambda+1}$ or

$$|\langle \mathbf{A}, \mathbf{a} \rangle| = |\langle \mathbf{A}, \mathbb{E}[\boldsymbol{\ell}^* \mid 1_{R>0}] \rangle| = |\mathbb{E}[\langle \boldsymbol{\ell}^*, \mathbf{A} \rangle \mid 1_{R>0}]| \geq \frac{c_{L13} \epsilon_d c_d \mathrm{Var}(\langle \boldsymbol{\ell}^*, \mathbf{A} \rangle)^{2.5}}{16\left(K\sqrt{\pi}+1\right)} \geq \frac{c_{L13} \epsilon_d c_d \, \mathrm{c_v}^{2.5}}{16\left(K\sqrt{\pi}+1\right)},$$

where in the last inequality we used Assumption 2.

Because $\mathbf{A} \in S^\perp$ and $\|\mathbf{A}\| = 1$, the previous equation implies the desired result that $\|\mathcal{P}_{S^\perp}(\mathbf{a})\| \geq \frac{c_{L13} \epsilon_d c_d \, \mathrm{c_v}^{2.5}}{16\left(K\sqrt{\pi}+1\right)}$. $\square$

J Proof of Proposition 11

Proof of Proposition 11. The first step is to rewrite R from Algorithm 6 Line 7 so that we can apply Lemmas 14 and 15.

Define

$$W := \sum_{t'=t}^{t+L-1} \left(w_{t'} - \sum_{k=1}^{j} (c_k q_k^{t'} - c_k \langle \mathbf{v}_k, \boldsymbol{\ell}^* \rangle) \right) = \sum_{t'=t}^{t+L-1} \left(w_{t'} - \sum_{k=1}^{j} c_k q_k^{t'} \right) + L \sum_{k=1}^{j} c_k \langle \mathbf{v}_k, \boldsymbol{\ell}^* \rangle.$$

Note that W is normally distributed with mean 0 and variance $\sigma_W^2 := L(1 + \sum_{k=1}^{j} c_k^2)$. By construction, we can rewrite R as

$$\begin{aligned} R &= \sum_{t'=t}^{t+L-1} \left(r^{(t')} - \sum_{k=1}^{j} c_k q_k^{t'} \right) \\ &= \sum_{t'=t}^{t+L-1} \left((\langle \mathbf{a}, \boldsymbol{\ell}^* \rangle + w_{t'}) - \sum_{k=1}^{j} c_k q_k^{t'} \right) \\ &= \sum_{t'=t}^{t+L-1} \langle \mathbf{a}, \boldsymbol{\ell}^* \rangle - L \sum_{k=1}^{j} c_k \langle \mathbf{v}_k, \boldsymbol{\ell}^* \rangle + W \\ &= L \langle \mathbf{a}, \boldsymbol{\ell}^* \rangle - L \langle \mathcal{P}_S(\mathbf{a}), \boldsymbol{\ell}^* \rangle + W && \left[\text{Lemma 17 implies } \mathcal{P}_S(\mathbf{a}) = \sum_{k=1}^{j} c_k \mathbf{v}_k\right] \\ &= L \langle (\mathbf{a} - \mathcal{P}_S(\mathbf{a})), \boldsymbol{\ell}^* \rangle + W \\ &= L \langle \mathcal{P}_{S^\perp}(\mathbf{a}), \boldsymbol{\ell}^* \rangle + W \\ &= L \|\mathcal{P}_{S^\perp}(\mathbf{a})\| \langle \mathbf{x}, \boldsymbol{\ell}^* \rangle + W. \end{aligned} \tag{13}$$

Therefore, R is exactly in the form necessary to apply Lemmas 14 and 15. In order to apply these lemmas, we need that $\frac{L\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\|}{\sigma_W} \leq \delta_{L14}$ and $\frac{L\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\|}{\sigma_W} \leq \delta_{L15}$ respectively.

To see this, note that $\mathcal{P}_{S^\perp}(\mathbf{a}) \le \sqrt{\lambda}$ (as otherwise ExponentialGrowth would not have been called), and therefore

$$\begin{aligned}
\frac{L\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\|}{\sigma_W} &\le \frac{L\sqrt{\lambda}}{\sqrt{L(1+\sum_{k=1}^{j} c_k^2)}} \\
&= \sqrt{\frac{4\lambda d(\mathbb{E}[\boldsymbol{\ell}^*{}_1]+1)^2}{c_{L14}^2}} \\
&\le \sqrt{\frac{4\lambda d(K\sqrt{\pi}+1)^2}{(\mathrm{c_v}/\sqrt{8\pi})^2}} && \text{[Equation (B), Assum 2]} \\
&\le \min(\delta_{L14}, \delta_{L15}, 1/c_{L15}), && (14)
\end{aligned}$$

where in the last line we used $\lambda \le \min(\delta_{L14}, \delta_{L15}, 1/c_{L15})^2 \frac{(\mathrm{c_v}/\sqrt{8\pi})^2}{4d(K\sqrt{\pi}+1)^2}$ by our assumption on λ.

Applying Lemmas 14 and 15 gives the following two bounds. Define $\mathbf{y} = \mathbb{E}[\boldsymbol{\ell}^* \mid 1_{R>0}]$. The first is a lower bound on $|\langle \mathbf{x}, \mathbf{y}\rangle|$. Importantly, we can apply Lemma 14 for $X = \langle \mathbf{x}, \boldsymbol{\ell}^*\rangle$ because of Equations (13) and (14).

$$\begin{aligned}
|\langle \mathbf{x}, \mathbf{y}\rangle| &= |\langle \mathbf{x}, \mathbb{E}[\boldsymbol{\ell}^* \mid 1_{R>0}]\rangle| \\
&= |\,\mathbb{E}[\langle \mathbf{x}, \boldsymbol{\ell}^*\rangle \mid 1_{R>0}]| \\
&\ge \frac{c_{L14} L\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\|}{\sigma_W} && \text{[Lemma 14]} \\
&= \frac{c_{L14}\sqrt{L}\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\|}{\sqrt{1+\sum_{k=1}^{j} c_k^2}} \\
&= c_{L14}\sqrt{\frac{4d(\mathbb{E}[\boldsymbol{\ell}^*{}_1]+1)^2}{c_{L14}^2}}\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\| \\
&= \sqrt{4d(\mathbb{E}[\boldsymbol{\ell}^*{}_1]+1)^2}\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\| \\
&= 2\sqrt{d}(\mathbb{E}[\boldsymbol{\ell}^*{}_1]+1)\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\|. && (15)
\end{aligned}$$

The next equation is an upper bound on $\|\mathbf{y}_i\|$ for all $i \in [d]$:

$$\begin{aligned}
|\mathbf{y}_i| &= \mathbb{E}[\boldsymbol{\ell}^*{}_i \mid 1_{R>0}] \\
&\le \mathbb{E}[\boldsymbol{\ell}^*{}_i] + c_{L15} L\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\|/\sigma_W && \text{[Lemma 15]} \\
&\le \mathbb{E}[\boldsymbol{\ell}^*{}_i] + 1. && \text{[Equation (14)]}
\end{aligned}$$

Using the above equation, we can bound $\|\mathbf{y}\|_2$ as follows. Because $\mathbb{E}[\boldsymbol{\ell}^*{}_1] \ge \mathbb{E}[\boldsymbol{\ell}^*{}_i]$ for all i,

$$\|\mathbf{y}\|_2 \le \sqrt{\sum_{i=1}^{d}\left(\mathbb{E}[\boldsymbol{\ell}^*{}_i]+1\right)^2} \le \sqrt{d}(\mathbb{E}[\boldsymbol{\ell}^*{}_1]+1).$$

Equation (15) implies that $\mathbf{y} \neq \mathbf{0}$. This implies by construction that $\mathbf{b} = \mathrm{Exploit}(1_{R>0}, \mathbf{w}_{\ell_\lambda+1}) = \frac{\mathbf{y}}{\|\mathbf{y}\|}$. Putting everything together, we have that

$$|\langle \mathbf{x}, \mathbf{b}\rangle| = \frac{|\langle \mathbf{x}, \mathbf{y}\rangle|}{\|\mathbf{y}\|_2} \ge \frac{2\sqrt{d}(\mathbb{E}[\boldsymbol{\ell}^*{}_1]+1))\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\|}{\sqrt{d}(\mathbb{E}[\boldsymbol{\ell}^*{}_1]+1)} \ge 2\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\|.$$

Finally, because $\mathbf{x} \in S^\perp$, this implies the desired result that

$$\left\|\mathcal{P}_{S^\perp}(\mathbf{b})\right\| \ge |\langle \mathbf{x}, \mathbf{b}\rangle| \ge 2\left\|\mathcal{P}_{S^\perp}(\mathbf{a})\right\|. \qquad \square$$

K Proof of Theorem 12

Proof of Theorem 12. We begin by proving that Algorithm 4 is BIC. There are four places where we set $\mathbf{A}^{(t)}$. The first is in the Line 4 of Algorithm 4, where we set $\mathbf{A}^{(t)} = \mathbf{e}_1$. This is BIC because we assumed (without loss of generality) that $\mathbb{E}[\boldsymbol{\ell}^*{}_i] = 0$ for all $i > 1$ and $\mathbb{E}[\boldsymbol{\ell}^*{}_1] \geq 0$.

The second place we set $\mathbf{A}^{(t)}$ is in Line 6 of Algorithm 5. This choice of $\mathbf{A}^{(t)}$ is BIC with the signal Ψ by construction and Lemma 6.

The third place we set $\mathbf{A}^{(t)}$ is in Line 5 of Algorithm 6. In order for this to be BIC, we must show that every input $\mathbf{a}$ to Algorithm 3 is BIC. The first time Algorithm 6 is used for any fixed value of j, the input action $\mathbf{a}$ is the action returned by Algorithm 5. This is BIC for signal R defined on Line 7 of Algorithm 5 by construction. Each subsequent call to Algorithm 6 for a fixed value of j uses an action $\mathbf{a}$ that is returned by the previous call to Algorithm 6. This is BIC for signal R defined on Line 7 of Algorithm 6.

The final time we set an action is on Line 17 of Algorithm 4 This action is again an action returned by the last call to Algorithm 6, which as argued above is BIC for signal R.

The rest of the proof will focus on bounding the sample complexity of Algorithm 4.

First, we will bound the number of times the inner while loop (Line 13) calls Algorithm 6 for each value of j. By Proposition 10, the action returned by Algorithm 5 satisfies $\|\mathcal{P}_{S^\perp}(\mathbf{a})\| \geq c_{P10}\, \mathrm{c_v}^{2.5}\, \epsilon_d c_d$. Furthermore, by Proposition 11, $\|\mathcal{P}_{S^\perp}(\mathbf{a})\|$ doubles with each call to Algorithm 6. Therefore, $\|\mathcal{P}_{S^\perp}(\mathbf{a})\| \geq \sqrt{\lambda}$ will be satisfied after at most $\log_2\left(\frac{\sqrt{\lambda}}{c_{P10}\, \mathrm{c_v}^{2.5}\, \epsilon_d c_d}\right) = O\left(\log(\frac{1}{\mathrm{c_v}\, \epsilon_d c_d})\right)$ calls to Algorithm 6.

Next we will bound the number of steps in each call to Algorithm 6, which is equivalent to bounding the L defined on Line 4 of Algorithm 6. To do this, we note that the c_i in Algorithm 6 are the same as the c_i in Lemma 17 with $\epsilon = \lambda$, $\ell = \ell_\lambda$, $\mathbf{u} = \mathcal{P}_{S^\perp}(\mathbf{a})$, and $\mathbf{v}_1, ..., \mathbf{v}_j$. This implies that

$$\sum_{k=1}^{j} c_k^2 \leq \frac{1}{\lambda}. \qquad \text{[Lemma 17]}$$

Therefore, we can bound L as follows:

$$\begin{aligned}
L &= \frac{4d(\mathbb{E}[\boldsymbol{\ell}^*{}_1] + 1)(1 + \sum_{k=1}^{j} c_k^2)}{c_{L14}^2} \\
&\leq \frac{4d(\mathbb{E}[\boldsymbol{\ell}^*{}_1] + 1)(1 + \frac{1}{\lambda})}{c_{L14}^2} \\
&\leq \frac{4d(K\sqrt{\pi} + 1)(1 + \frac{1}{\lambda})}{\mathrm{c_v}^2/(8\pi)} && \text{[Assum 2, Eq (B)]} \\
&= O\left(\frac{d}{\lambda\, \mathrm{c_v}^2}\right).
\end{aligned}$$

For each loop of the while loop on Line 8, we also have $\kappa = O(\frac{\log(1/\epsilon_d)}{\lambda c_d^2} + \frac{d}{\lambda\, \mathrm{c_v}^2})$ steps in the loop on Line 16. All together, this gives that each iteration of the loop on Line 8 takes at most

$$O\left(\frac{d \log(\frac{1}{\mathrm{c_v}\, \epsilon_d c_d})}{\lambda\, \mathrm{c_v}^2} + \frac{\log(1/\epsilon_d)}{\lambda c_d^2}\right) = O\left(\log\left(\frac{1}{\mathrm{c_v}\, \epsilon_d c_d}\right)\left(\frac{d}{\lambda\, \mathrm{c_v}^2} + \frac{1}{\lambda c_d^2}\right)\right)$$

steps. Next, we will bound the number of iterations of the while loop on Line 8.

For each j, we will apply Lemma 8 with $\epsilon = \lambda$, $\mathbf{u} = \mathbf{v}_{j+1}$, and the vectors $\mathbf{v}_1, ..., \mathbf{v}_j$. By construction of the algorithm, $S^\perp$ is non-empty because the algorithm has not yet terminated, and $\|\mathcal{P}_{S^\perp}(\mathbf{v}_{j+1})\|^2 \geq \lambda$ by the termination condition of the while loop on Line 13 of Algorithm 4. Therefore, this satisfies the assumption of Lemma 8. Define $\lambda_1^j, ..., \lambda_d^j$ as the eigenvalues of

$\mathbf{M}^j := \sum_{i=1}^{j} \mathbf{v}_i^{\otimes 2}$ and define ℓ^j as the largest index such that $\lambda_{\ell^j}^j \geq 200d^3/\lambda^2$ (and $\ell^j = 0$ if all eigenvalues of $\mathbf{M}^j$ are less than $200d^3/\lambda^2$). Now define

$$\Delta^j = \sum_{i=\ell^j+1}^{d} \left(\frac{200d^3}{\lambda^2} - \lambda_i \right).$$

Note that for any fixed i, the ith eigenvalue does not decrease between $\mathbf{M}^j$ and $\mathbf{M}^{j+1}$. Because of this monotonicity, Lemma 8 implies that for every round j, either

$$\ell^{j+1} \geq \ell^j + 1 \quad \text{or} \quad \Delta^{j+1} \leq \Delta^j - \frac{\lambda}{2}.$$

Because $\ell^1 \geq 0$ and $\Delta^1 \leq \frac{200d^3}{\lambda^2} \cdot d = \frac{200d^4}{\lambda^2}$, this implies that after $\frac{\frac{200d^4}{\lambda^2}}{\lambda/2} + d$ applications of Lemma 8, either $\ell^j = d$ or $\Delta^j = 0$. This means that after $\frac{400d^4}{\lambda^2} + d$ applications of Lemma 8, the smallest eigenvalue of $\mathbf{M}^j$ must be at least $200d^3/\lambda^2 \geq \lambda$. However, this means that the algorithm must terminate before round $400d^4/\lambda^3 + d$. Therefore, the number of iterations of the while loop on Line 8 is less than $O(d^4/\lambda^3)$. Putting everything together, the total number of steps needed for λ-exploration is upper bounded by

$$O\left(\log\left(\frac{1}{\mathrm{c_v}\, \epsilon_d c_d}\right)\left(\frac{d}{\lambda\, \mathrm{c_v}^2} + \frac{1}{\lambda c_d^2}\right)\right) \cdot O\left(\frac{d^4}{\lambda^3}\right) = O\left(\log\left(\frac{1}{\mathrm{c_v}\, \epsilon_d c_d}\right)\left(\frac{d^5}{\lambda^4\, \mathrm{c_v}^2} + \frac{d^4}{c_d^2 \lambda^4}\right)\right). \quad \square$$

L Proof of Proposition 4

Proof of Proposition 4. First, for any unit vector $\mathbf{v}$, we have $B_{r/3}(2r\mathbf{v}/3) \subseteq \mathcal{K} \subseteq B_1(0)$. Therefore

$$\mu(B_{r/3}(2r\mathbf{v}/3)) = \mathrm{Vol}(B_{r/3}(2r\mathbf{v}/3))/\mathrm{Vol}(\mathcal{K}) \geq \mathrm{Vol}(B_{r/3}(2r\mathbf{v}/3))/\mathrm{Vol}(B_1(0)) = (r/3)^d.$$

Since $\langle \mathbf{x}, \mathbf{v} \rangle \geq r/3$ for all $\mathbf{x} \in B_{r/3}(2r\mathbf{v}/3)$, this confirms the values $(c_d, \epsilon_d) = (r/3, (r/3)^d)$.

The bound on $\mathrm{c_v}$ follows by [Sellke, 2023, Lemma 3.2] and Jensen's inequality since $\mathcal{K}$ has width at least $2r$ in any direction. The bound on K is trivial since $2e^{-(t/1.25)^2} \geq 1$ for $|t| \leq 1$. $\square$

M Proof of Proposition 5

We first recall several useful facts on log-concave distributions. Throughout we take μ to be α-log-concave and β-log-smooth with mode $\mathbf{x}^*$ and mean $\bar{\mathbf{x}}$, possibly in dimension 1. (The proof will use 1-dimensional projections of the original measure μ.) We will write $\mathbf{x} \sim \mu$ instead of $\boldsymbol{\ell}^* \sim \mu$.

Fact 22 ([Dwivedi et al., 2019, Lemma 5], [Durmus and Moulines, 2019, Theorem 1])**.** *For $x \sim \mu$, we have $\mathbb{E}[\|x - x^*\|^2] \leq 1/\alpha$ and with probability $1 - \delta$:*

$$\|x - x^*\|_2 \leq 2\alpha^{-1/2}\left(1 + \sqrt{\frac{\log(1/\delta)}{d}} + \sqrt[4]{\frac{\log(1/\delta)}{d}}\right).$$

Fact 23 ([Chewi and Pooladian, 2023, Lemma 2])**.** *We have the covariance bounds*

$$\frac{I_d}{\alpha d} \succeq \mathit{Cov}(\mu) \succeq \frac{I_d}{\beta d}. \tag{17}$$

Fact 24. *Any 1-dimensional projection of μ is also αd-log-concave and βd-log-smooth.*

Proof. Preservation of strong log-concavity under projection is well known, see e.g. [Saumard and Wellner, 2014, Theorem 3.8]. For log-smoothness, supposing for convenience that the projection is onto the first coordinate axis, the claim is proved by the following standard computation. With $e^{-f(\mathbf{x})}$ the density of μ and $e^{-g(x)}$ the density of the projection of μ to the first coordinate axis, one may compute as in [Saumard and Wellner, 2014, Proof of Proposition 7.1] that

$$g''(x) = \mathbb{E}^{\mu}[\partial_{1,1} f(\mathbf{x}) | x_1 = x] - \mathrm{Var}^{\mu}[\partial_1 f(\mathbf{x}) | x_1 = x] \leq \mathbb{E}^{\mu}[\partial_{1,1} f(\mathbf{x}) | x_1 = x] \leq \beta d.$$

This completes the proof. $\square$

Proof of Proposition 5. We have $c_v \ge \frac{1}{\beta d}$ directly from (17).

For ϵ_d, let $\bar{\mathbf{x}}$ be the mean under μ and note that from (17), we find

$$\begin{aligned}
\|\bar{\mathbf{x}} - \mathbf{x}^*\| &= \sup_{\|\mathbf{w}\|=1} \langle \bar{\mathbf{x}} - \mathbf{x}^*, \mathbf{w} \rangle \\
&= \sup_{\|\mathbf{w}\|=1} \mathbb{E}^{\mathbf{x}\sim\mu}[\langle \mathbf{x} - \mathbf{x}^*, \mathbf{w} \rangle] \\
&\le \sqrt{\sup_{\|\mathbf{w}\|=1} \mathbb{E}^{\mathbf{x}\sim\mu}[\langle \mathbf{x} - \mathbf{x}^*, \mathbf{w} \rangle^2]} \\
&\le \sqrt{\langle \mathrm{Cov}(\mu), \mathbf{w}^{\otimes 2} \rangle} \\
&\le 1/\sqrt{\alpha d}.
\end{aligned}$$

Fixing a unit vector $\mathbf{v}$ as in Assumption 2, we consider the projection P onto the 1-dimensional subspace spanned by $\mathbf{v}$, and let $P(\mu)$ be the pushforward of μ under the projection (to which Fact 24 applies). Identifying $P(\mathbb{R}^d)$ isometrically with $\mathbb{R}$, let $\hat{x}$ be the mode of $P(\mu)$. Then the same argument as above applies to $P(\mu)$ shows $\|P(\bar{\mathbf{x}}) - \hat{x}\| \le 1/\sqrt{\alpha d}$, and so

$$\|\hat{x}\| \le \|\mathbf{x}^*\| + \frac{2}{\sqrt{\alpha d}} \le \gamma + \frac{2}{\sqrt{\alpha d}}.$$

(Note that if $\bar{\mathbf{x}} = 0$ then this shows $\|\hat{x}\| \le 1/\sqrt{\alpha d}$, which following the arguments below leads to $\epsilon_d \ge \Omega(1)$ as mentioned below Proposition 5.)

Write $f : \mathbb{R} \to \mathbb{R}_+$ for the density of $P(\mu)$, and $g(x) = \log f(x)$. We have $f'(\hat{x}) = 0$ and so $g'(\hat{x}) = 0$ also. By Fact 24, we have $g''(x) \in [-\beta d, -\alpha d]$ for all x, so for $x \ge \hat{x}$ we have:

$$g'(x) = g'(x) - g'(\hat{x}) = \int_{\hat{x}}^{x} g''(y)dy \in [-\beta d(x - \hat{x}), -\alpha d(x - \hat{x})].$$

Integrating again, we find

$$g(x) - g(\hat{x}) = \int_{\hat{x}}^{x} g'(y)dy \in [-\beta d(x - \hat{x})^2/2, -\alpha d(x - \hat{x})^2/2].$$

Identical reasoning gives the same conclusion for $x \le \hat{x}$. Translating back to $f = e^g$, we conclude that for each $x \in \mathbb{R}$:

$$e^{-\beta d|x-\hat{x}|^2/2} \le \frac{f(x)}{f(\hat{x})} \le e^{-\alpha d|x-\hat{x}|^2/2}.$$

It follows that for $\mathbf{x} = \boldsymbol{\ell}^* \sim \mu$ and $x \sim P(\mu)$:

$$\Pr[\langle \mathbf{v}, \mathbf{x} \rangle \ge c_d] \ge \Pr[x \ge \gamma + \frac{2}{\sqrt{\alpha d}} + c_d].$$

Letting $J = \gamma + \frac{2}{\sqrt{\alpha d}} + c_d$, the latter probability is at least

$$\frac{\int_J^\infty e^{-\beta d z^2/2} dz}{\int_{\mathbb{R}} e^{-\alpha d z^2/2} dz} = \sqrt{\alpha/\beta} \cdot \Phi^C(J\sqrt{\beta d}) \ge \frac{J e^{-J^2 \beta d/2} \sqrt{\alpha d}}{(1 + J^2 \beta d)\sqrt{2\pi}}.$$

The last inequality follows from the classical bound $\Phi^C(\kappa) \ge \frac{\varphi(\kappa)\kappa}{1+\kappa^2}$ where φ is the standard Gaussian density Gordon [1941]. This confirms the value of ϵ_d.

For K we consider a similar projection, and note that by Fact 24 and the Bakry-Emery theory for strongly log-concave measures (see e.g. [Anderson et al., 2010, Lemma 2.3.3], we have

$$\mathbb{E}[e^{\lambda \langle \mathbf{v}, \mathbf{x} - \bar{\mathbf{x}} \rangle}] \le e^{\lambda^2 \alpha d/2}, \quad \forall \lambda \in \mathbb{R}.$$

Thus with $J_0 = \gamma + \frac{1}{\sqrt{\alpha d}} \ge \|\langle \bar{\mathbf{x}}, \mathbf{v} \rangle\|$, we have (using $\lambda J_0 \le \frac{1+\lambda^2 J_0^2}{2}$):

$$\mathbb{E}[e^{\lambda \langle \mathbf{v}, \mathbf{x} \rangle}] \le e^{(\lambda^2 \alpha d/2) + \lambda J_0} \le e^{0.5} \cdot e^{\lambda^2 (J_0^2 + \alpha d)/2} \le 2 e^{\lambda^2 (J_0^2 + \alpha d)/2}.$$

It follows by the usual Markov inequality arguments that

$$\Pr[|\langle \mathbf{v}, \mathbf{x} \rangle| \geq t] \leq 4e^{-\frac{t^2}{2(J_0^2+\alpha d)}}.$$

Since probabilities are at most 1 and $a \leq \sqrt{a}$ for $a \leq 1$ we find

$$\Pr[|\langle \mathbf{v}, \mathbf{x} \rangle| \geq t] \leq 2e^{-\frac{t^2}{4(J_0^2+\alpha d)}}$$

which completes the verification of K since $(J_0^2 + \alpha d)^{1/2} \leq J_0 + \sqrt{\alpha d}$.

For the counterexample, we may take $(\alpha, \beta) = (1, 2)$ and let ν be the distribution on $\mathbb{R}$ with density proportional to $e^{-dx^2 \cdot (1+1_{x \geq 0})/2}$. Then let $\mu = \nu^{\otimes d}$, so that $x \sim \mu$ has IID coordinates with law ν. Then the mode $\mathbf{x}^*$ is indeed zero but the mean of ν is non-zero, so taking $\mathbf{v} = (1, 1, \ldots, 1)/\sqrt{d}$, a Chernoff estimate shows $\Pr^{\mathbf{x} \sim \mu}[\langle \mathbf{x}, \mathbf{v} \rangle \geq 0] \leq e^{-\Omega(d)}$. $\square$