Supplementary Material

Organization In Appendix A, we state some elementary probabilistic facts. The next two sections focus on proving our lemma on noisy location estimation. In Appendix B, we prove some critical lemmas used in the proof, and in Appendix C, we present the complete version of our location estimation algorithm, while addressing some typos in the main text. We mention some typos in footnotes and correct other minor typos without a mention.

Moving forward, in Appendix D, we introduce an algorithm and prove a hardness result for the specific version of list-decodable mean estimation we consider, which differs from prior work. Finally, in Appendix E, we state the final guarantees we can get for the problem of list-decodable stochastic optimization, incorporating our lemma from Appendix D.

A Elementary Probability Facts

In this section, we recall some elementary lemmas from probability theory.

Lemma A.1 (Hoeffding)**.** *Let $X_1, \dots X_n$ be independent random variables such that $X_i \in [a_i, b_i]$. Let $S_n := \frac{1}{n}\sum_{i=1}^n X_i$, then for all $t > 0$*

$$\Pr[|S_n - \mathbf{E}[S_n]| \geq t] \leq \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) .$$

Lemma A.2 (Multivariate Chebyshev)**.** *Let $X_1, \dots, X_m$ be independent random variables drawn from D where D is a distribution over $\mathbb{R}^d$ such that $\mathbf{E}_{X\sim D}[X] = 0$ and $\mathbf{E}_{X\sim D}[\|X\|^2] \leq \sigma^2$. Let $S_m := \frac{1}{m}\sum_{i=1}^m X_i$, then for all $t > 0$*

$$\Pr[\|S_m\| \geq t] \leq \sigma^2/mt^2 .$$

Proof. We first prove the following upper bound,

$$\begin{aligned}
\Pr\left[\forall v \ \|v\| = 1. \left|\frac{1}{m}\sum_{i\in[m]} X_i \cdot v\right| > t\right] &< \frac{\mathbf{E}[|\sum_{i\in[m]} X_i \cdot v|^2]}{m^2\, t^2} < \frac{\mathbf{E}[\sum_{i,j}(X_i\cdot v)(X_j \cdot v)]}{m^2\, t^2} \\
&< \frac{\sum_{i,j}\mathbf{E}[(X_i\cdot v)(X_j\cdot v)]}{m^2\, t^2} < \frac{\sum_i \mathbf{E}[(X_i\cdot v)^2]}{m^2\, t^2} < \frac{\sum_i \mathbf{E}[\|X_i\|^2]}{m^2\, t^2} \\
&< \frac{m\sigma^2}{m^2\, t^2} = \frac{\sigma^2}{m\, t^2} .
\end{aligned}$$

Since the inequality holds for all unit v, it also holds for the unit v in the direction of S_m, completing the proof. □

Fact A.3 (Inflation via conditional probability)**.** *Let y be a random variable with mean μ and variance σ^2 and let ξ be an arbitrary random variable independent of y, then*

$$\begin{aligned}
\Pr[\xi \in (a,b)] &< (1 + 2/A^2)\Pr[\xi + y \in (a + \mu - \sigma A, b + \mu + \sigma A) \\
&< (1 + 2/A^2)\Pr[\xi + y \in (a - |\mu| - |\sigma A|, b + |\mu| + |\sigma A|)] .
\end{aligned}$$

Proof. To do this, we inflate the intervals and use conditional probabilities.

$$\begin{aligned}
\Pr[\xi \in (a,b)] &= \Pr[\xi + y \in (a + \mu - \sigma A, b + \mu + \sigma A) \mid |y - \mu| < \sigma A] \\
&= \frac{\Pr[\xi + y \in (a + \mu - \sigma A, b + \mu + \sigma A) \text{ and } |y - \mu| < \sigma A]}{\Pr[|y - \mu| < \sigma A]} \\
&< (1 - 1/A^2)^{-1}\Pr[\xi + y \in (a + \mu - \sigma A, b + \mu + \sigma A) \text{ and } |y - \mu| < \sigma A] \\
&< (1 + 2/A^2)\Pr[\xi + y \in (a + \mu - \sigma A, b + \mu + \sigma A)] .
\end{aligned}$$

The second inequality above follows from observing that we are simply lengthening the interval.

We will often use the second version for ease of analysis. □

B Useful Lemmas

In this section, we present some helpful lemmas for the algorithm on noisy one-dimensional location estimation.

To recap the setting: We can access samples from distributions $\xi + y$ and $\xi + y' + t$. Here, $\Pr[\xi = 0] > \alpha$, y and y' are distributions with zero mean and bounded variance, and $t \in \mathbb{R}$ is an unknown translation. Our objective is to estimate the value of t.

B.1 Useful Lemma for Rough Estimation

Our algorithm for one-dimensional location estimation consists of two steps. In the first step, we obtain an initial estimate of the shift between the two distributions by computing pairwise differences of samples drawn from each distribution. This involves taking the median of the distribution of $x + y$, where x is symmetric and y has mean 0 and bounded variance.

The following lemma demonstrates that the median of this distribution is at most $O(\sigma\alpha^{-1/2})$, where σ is the standard deviation of y. Furthermore, this guarantee cannot be improved.

Fact B.1 (Median of Symmetric + Bounded-variance Distribution)**.** *Let x be a random variable symmetric around* 0 *such that $\alpha \in (0, 1)$, $\Pr[x = 0] \geq \alpha$. Let y be a random variable with mean* 0 *and variance σ^2. If S is a set of $O(1/\alpha^2 \log(1/\delta))$ samples drawn from the distribution of $x + y$,* $|\mathrm{median}(S)| \leq O(\sigma\alpha^{-1/2})$.

This guarantee is tight in the sense that there exist distributions for x and y satisfying the above constraints, such that $\mathrm{median}(x + y)$ *can be as large as* $\Omega(\sigma\alpha^{-1/2})$.

Proof. We show that $\Pr[x + y < -O(\sigma/\sqrt{\alpha})] < 0.5$ and $\Pr[x + y > O(\sigma/\sqrt{\alpha})] < 0.5$, as a result, $|\mathrm{median}(x + y)| < O(\sigma/\sqrt{\alpha})$. We will later transfer this guarantee to the uniform distribution over the samples.

Applying Fact A.3 to the first probability, we see that $\Pr[x + y < -O(\sigma/\sqrt{\alpha})] < (1 + \alpha)\Pr[x < -O(\sigma/\sqrt{\alpha})]$.

Since $\Pr[x = 0] \geq \alpha$, we see that $\Pr[x < 0] \leq 1/2 - \alpha$,

and so $\Pr[x+y < -O(\sigma/\sqrt{\alpha})] < (1+\alpha)(0.5-\alpha) = 0.5-\alpha+0.5\alpha-\alpha^2 = 0.5-0.5\alpha-\alpha^2 < 0.5$.

The upper bound follows similarly.

Since $\Pr[x = 0] \geq \alpha$ and $\Pr[|y| < O(\sigma/\sqrt{\alpha})] \geq 1 - \alpha$, we see $\Pr[|x + y| < O(\sigma/\sqrt{\alpha})] \geq \alpha/2$. Hoeffding's inequality (Lemma A.1) now implies that the empirical median also satisfies the above upper bound as long as the number of samples is greater than $O(1)/\alpha^2 \log(1/\delta)$.

To see that this is tight, consider the distribution centered at 0, whose density function is $2/(y + 2)^3$ in the range $[1, \infty)$, and is 0 otherwise.

Call this $D_{y^{-3}}$. Observe that $\Pr_{D_{y^{-3}}}[z > t] < O(1) \int_t^\infty y^{-3} dy = C/t^2$.

Let x be a symmetric distribution whose distribution takes the value 0 with probability α and takes the values $\pm\alpha^{-1/2}100C^{1/2} + 10$ with probability $0.5(1 - \alpha)$.

We show that the median of the distribution of $x + y$ where y is drawn from $D_{y^{-3}}$, is larger than $\Omega(\alpha^{-1/2})$.

To see this, we show that the probability that $x + y$ takes a value smaller than $100\alpha^{-1/2}C^{1/2}$ is less than half, implying that the median has to be larger than this quantity.

x takes three values. Note that $y + 100\alpha^{-1/2}C^{1/2} + 12$ places no mass in the region $(-\infty, 100\alpha^{-1/2}C^{1/2} + 10]$. So to estimate the probability that $x + y$ takes a value smaller than $100\alpha^{-1/2}C^{1/2} + 10$, we only need to consider contributions from the other two possible values. By choosing α small enough, so that $100\alpha^{-1/2}C^{1/2} > 10$, we see

$$\begin{aligned}
&\Pr[x + y < 100\alpha^{-1/2}C^{1/2} + 10] \\
&< 0.5(1-\alpha)\Pr[y < 200\alpha^{-1/2}C^{1/2} + 20] + \alpha \Pr[y < 100\alpha^{-1/2}C^{1/2} + 10] \\
&< 0.5(1-\alpha)(1 - C/(200\alpha^{-1/2}C^{1/2} + 20)^2) + \alpha(1 - C/(100\alpha^{-1/2}C^{1/2} + 10)^2) \\
&< 0.5(1-\alpha)(1 - \alpha^{1/2}/(400)^2) + \alpha \\
&< 0.5 + 0.5\alpha - \alpha^{1/2}/8 \cdot (400^2)\ .
\end{aligned}$$

We are done when $0.5\alpha - \alpha^{1/2}/8 \cdot (400^2) < 0$, this happens for $\alpha^{1/2} < 2/(8 \cdot (400^2))$. □

B.2 Useful Lemma for Finer Estimation

In the second step of our location-estimation lemma, we refine the estimate of t. To do this, we first re-center the distributions based on our rough estimate, so that the shift after re-centering is bounded. Then, we identify an interval I centered around 0 such that, when conditioning on $\xi + z$ falling within this interval, the expected value of $\xi + z$ remains the same as when conditioning on ξ falling within the same interval. This expectation will help us get an improved estimate, which we use to get an improved re-centering of our original distributions, and repeat the process.

To identify such an interval, we search for a pair of bounded-length intervals equidistant from the origin (for e.g. $(-10\sigma, -5\sigma)$ and $(5\sigma, 10\sigma)$) that contain very little probability mass. By doing so, when z is added to ξ, the amount of probability mass shifted into the interval $(-5\sigma, 5\sigma)$ z remains small.

In this subsection, we prove Lemma B.2, which states that any positive sequence which has a finite sum must eventually have one small element. The lemma also gives a concrete upper bound on which element of the sequence satisfies this property.

Lemma B.2. *$a_i \geq 0$ for all i and $\sum_{i=1}^{\infty} a_i < C$ for some constant C. Also, suppose we have $\eta \in [0, 1]$. Then there is an i such that $1 \leq L < i < (C/a_0 + L)^{1/\eta}$ such that $ia_i < \eta \sum_{j=1}^{i} a_j$.*

Consider a partition of the reals into length L intervals. In our proof, we will use Lemma B.2 on the sequence a_i, where a_i corresponds to an upper bound on the mass of ξ contained in the i-th intervals equidistant from the origin on either side, and the mass that crosses them (i.e., the mass of ξ that is moved either inside or out of the interval when z is added to it).

We need the following calculation to prove Lemma B.2.

Notation: For integer $i \geq 1$ and $\eta \in (0, 1)$, define $(i - \eta)! := \Pi_{j=1}^{i}(j - \eta)$.

Fact B.3. *Let $A_k := 1 + \sum_{t=1}^{k-1} \frac{\eta(t-1)!}{(t-\eta)!}$. Then, for $k \geq 2$, $A_k = (k-1)!/(k-1-\eta)!$.*

Proof. We prove this by induction. By definition, our hypothesis holds for A_2 because $A_2 = 1 + \eta/(1-\eta) = 1/(1-\eta) = (2-1)!/(2-1-\eta)!$. Suppose it holds for all $2 \leq t \leq k$. We then show that it holds for $t = k + 1$.

$$\begin{aligned}
A_{k+1} = 1 + \sum_{t=1}^{k} \frac{\eta(t-1)!}{(t-\eta)!} &= A_k + \frac{\eta\,(k-1)!}{(k-\eta)!} \\
&= \frac{(k-1)!}{(k-1-\eta)!} + \frac{\eta\,(k-1)!}{(k-\eta)!} = \frac{(k-1)!}{(k-1-\eta)!}\left(1 + \frac{\eta}{k-\eta}\right) \\
&= \frac{(k-1)!}{(k-1-\eta)!}\,\frac{k}{k-\eta} = \frac{k!}{(k-\eta)!}.
\end{aligned}$$

□

Proof of Lemma B.2 Let $U = (C/a_0 + L)^{1/\eta}$ and suppose towards a contradiction that there is no such i that satisfies the lemma. Specifically, all integers $i \in [1, U]$, we will assume that for $ia_i \geq \eta \sum_{j=1}^{i} a_j$. We then show that this implies $i^{1-\eta} a_i \geq \eta a_0$ for all i in the range.

Consider the inductive hypothesis on t given by $a_t \geq \eta \frac{(t-1)!}{(t-\eta)!} \cdot a_0$. The base case when $t = 1$ is true since $a_1 \geq \eta a_0/(1-\eta)$ by our assumption. Suppose the inductive hypothesis holds for integers $t \in [1, k-1]$. We show this for $t = k$ below.

$$\begin{aligned} a_k &\geq \frac{\eta}{k-\eta} \sum_{t=0}^{k-1} a_t \\ &\geq \frac{a_0 \eta}{k-\eta} \left(1 + \sum_{t=1}^{k-1} \frac{\eta(t-1)!}{(t-\eta)!} \right) \\ &= a_0 \eta \, \frac{(k-1)!}{(k-\eta)!} \,. \end{aligned}$$

The final equality follows from Fact B.3 which states that $(k-1)!/(k-1-\eta)! = 1 + \sum_{t=1}^{k-1} \frac{\eta(t-1)!}{(t-\eta)!}$. Simplifying this further, we see that since $(i - \eta) \geq i \exp(-\eta/i)$ for all $i \in [1, k]$,

$$\begin{aligned} a_k &\geq a_0 \eta \, \frac{(k-1)!}{(k-\eta)!} \\ &\geq a_0 \eta \, \frac{(k-1)!}{k! \exp(-\eta/k)} \\ &\geq a_0 \eta \, (1/k) \, (1/\exp(-\eta(\sum_{i=1}^{k} 1/i))) \\ &\geq a_0 \eta \, (1/k) \, (1/\exp(-\eta \log(k)/20)) \\ &\geq (a_0/2) \eta \, (1/k^{1-\eta/20}) \,. \end{aligned}$$

Finally, observe that

$$\begin{aligned} C = \sum_{i=L}^{U} a_i > a_0 \eta \sum_{i=L}^{U} (1/i^{1-\eta/20}) \\ > a_0 \eta \int_L^U (1/x^{1-\eta}) \, dx \\ = a_0 (U^\eta - L^\eta). \end{aligned}$$

If $a_0(U^\eta - L^\eta) > C$, we have a contradiction, since $\sum_{i=L}^{\infty} a_i < C$. This follows when $U > (C/a_0 + L)^{1/\eta}$. □

C Noisy Location Estimation

In this section, we state and prove the guarantees of our algorithms for noisy location estimation (Lemma 3.1 and Lemma 3.5).

C.1 One-dimensional Noisy Location Estimation

Throughout the technical summary and some parts of the proof, we make the assumption that the variables y and y' were bounded. Extending this assumption to bounded-variance distributions requires significant effort.

Our algorithm for one-dimensional noisy location estimation (Algorithm 4) can be thought of as a two-step process. The first step involves a rough initial estimation algorithm, while the second step employs an iterative algorithm that progressively refines the estimate by a factor of η in each iteration.

Due to space limitations and for ease of exposition, the algorithm we present in the main body is a sketch of the refinement procedure.

In this Algorithm 4 , we introduce the definition of $\hat{P}$ (the empirical estimate of $\tilde{P}(\cdot)$), which is an upper bound on the probability mentioned earlier. This probability can be calculated using samples from $\xi + z$ and $\xi + z'$. Additionally, we incorporate the iterative refinement process within the algorithm.

Algorithm 4 One-dimensional Location Estimation: Shift1D$(S_1, S_2, \eta, \sigma, \alpha)$

Input: Sample sets $S_1, S_2 \subset \mathbb{R}^d$ of size m, $\alpha, \eta \in (0,1)$, $\sigma > 0$

1. Let $T = O(\log_{1/\eta}(1/\alpha))$. For $j \in \{1,2\}$, partition S_j into T equal pieces, $S_j^{(i)}$ for $i \in [T]$.
2. $D = \{a - b \mid a \in S_1^{(1)}, b \in S_2^{(1)}\}$.
3. $t'(1) := \mathrm{median}(D)$.
4. Set $A = O(1/\sqrt{\alpha})$.
5. Repeat steps 6 to 12, for i going from 2 to T:
6. $S_1^{(i)} := S_1^{(i)} - t'_r(i-1)$.
7. For $j \in \{1,2\}$

$$\hat{P}_j(i) := O(1) \Pr_{x \sim S_j^{(i)}}[|x| \in A\sigma(i-5, i+5)]$$
$$+ O(1) \sum_{j=1}^{i-1} (1/(i-j)^2) \Pr_{x \sim S_j^{(i)}}[|x| \in Aj\sigma + A\sigma[-4,5).]$$

8. Let $\hat{P}(i) = \hat{P}_1(i) + \hat{P}_2(i)$.
9. Identify an integer $k \in [1/(\alpha\eta), (O(1)/\alpha\eta)^{1/\eta}]$ such that

$$\hat{P}(k) \leq \eta \sum_{j \in \{1,2\}} \Pr_{x \sim S_j^{(i)}}[|x| \in A\sigma k] \pm O(\eta/i).$$

10. $t'(i) := t'(i-1) + \mathbf{E}_{z \sim S_1^{(i)}}[z \mid |z| \leq A\sigma k] - \mathbf{E}_{z \sim S_2^{(i)}}[z \mid |z| \leq A\sigma k]$.
12. $A := \eta A$.
11. Return $t'(T)$

Lemma C.1 (One-dimensional location-estimation)**.** *There is an algorithm (Algorithm 4) which, given* $\mathrm{poly}((O(1)/\eta\alpha)^{1/\eta}, \log(1/\delta\eta\alpha))$ *samples of the form $\xi + y + t$ and $\xi + y'$, where $t \in \mathbb{R}$ is an unknown translation, runs in time* $\tilde{\mathrm{poly}}((O(1)/\eta\alpha)^{1/\eta}, \log(1/\delta\eta\alpha))$ *and recovers t' such that $|t - t'| \leq O(\eta\sigma)$.*

Proof. Our proof is based on the following claims:

Claim C.2 (Rough Estimate)**.** *There is an algorithm which, given $m = O((1/\alpha^4)\ \log(1/\delta))$ samples of the kind $\xi + y + t$ and $\xi + y'$, where $t \in \mathbb{R}$ is an unknown translation, returns t'_r satisfying $|t'_r - t| < O(\sigma\alpha^{-1/2})$.* [2]

Claim C.3 (Fine Estimate)**.** *Suppose z, z' have means bounded from above by $A\sigma$ and variances at most σ^2 and suppose $\alpha \in (0,1)$ and $\eta \in (0, 1/2)$. Then in* $\mathrm{poly}((O(1)/\alpha\eta)^{1/\eta}, \log(1/\delta\eta\alpha))$ *samples and* $\mathrm{poly}((O(1)/\alpha\eta)^{1/\eta}, \log(1/\delta\eta\alpha))$ *time, it is possible to recover $k \in [1/\eta\alpha, (O(1)/\eta\alpha)^{O(1/\eta)}]$ such that*

$$\widehat{\mathbf{E}}[\xi + z \mid |\xi + z| \leq A\sigma k] - \widehat{\mathbf{E}}[\xi + z' \mid |\xi + z'| \leq A\sigma k] = \mathbf{E}[z] - \mathbf{E}[z'] \pm \eta\,(A\sigma).$$

Using Claim 3.2, we first identify a rough estimate t'_r satisfying $|t'_r - t| < O(\sigma\alpha^{-1/2})$. This allows us to re-center y'. Let the re-centered distribution be denoted by $z' = y'$ and $z = y + t - t'_r$. Then z and z' are such that $\mathbf{E}[z]$ and $\mathbf{E}[z']$ are both at most $O(\sigma\alpha^{-1/2})$ in magnitude, and have variance at most σ^2.

Claim 3.4 then allows us to estimate t'_f such that $|(\mathbf{E}[z] - \mathbf{E}[z']) - t'_f| = |t - t'_r - t'_f| \leq \eta\, O(\sigma\alpha^{-1/2})$.

Setting $t' = t'_r + t'_f$, we see that our estimate t' is now η times closer to t compared to t'_r.

[2] Typo in main body: Missing $\alpha^{-1/2}$ term.

To refine this estimate further, we can obtain fresh samples and re-center using t' instead of t'_r. Repeating this process $O(\log_{1/\eta}(1/\alpha)) = O(\log_\eta(\alpha))$ times is sufficient to obtain an estimate that incurs an error of $\eta \cdot \eta^{\log_\eta(\alpha^{1/2})} \cdot O(\sigma\alpha^{-1/2}) \le O(\eta\sigma)$.

This results in a runtime and sample complexity that is only $O(\log_{1/\eta}(1/\alpha))$ times the runtime and sample complexity required by Claim 3.4. This amounts to the final runtime and sample complexity being $\mathrm{poly}((O(1)/\alpha\eta)^{1/\eta}, \log(1/\delta\eta\alpha))$.

We now prove Claim 3.2 and Claim 3.4.

Claim 3.2 shows that the median of the distribution of pairwise differences of $\xi + y + t$ and $\xi + y'$ estimates the mean up to an error of $\sigma\alpha^{-1/2}$.

Proof of Claim 3.2 Let $\tilde{\xi}$ be a random variable with the same distribution as ξ and independently drawn. We have independent samples from the distributions of $\xi + y + t$ and $\xi + y'$. Applying Fact B.1 to these distributions, we see that if we have at least $O(1/\alpha^4)\log(1/\delta)$ samples from the distribution of $(\xi - \tilde{\xi}) + (y - y') + t$, these samples will have a median of $t \pm O(\sigma/\sqrt{\alpha})$. □

Proof of Claim 3.4 To identify such a k, the idea is to ensure that $\mathbf{E}[\xi + z \mid |\xi + z| \le A\sigma k] = \mathbf{E}[\xi + z \mid |\xi| \le A\sigma k] \pm O(A\eta\sigma) = \mathbf{E}[\xi \mid |\xi| \le A\sigma k] + \mathbf{E}[z] \pm O(A\eta\sigma)$, and similarly for z'. The theorem follows by taking the difference of these equations.

Before we proceed, we will need the following definitions: let $P(i, z)$ be defined as follows:

$$\begin{aligned} P(i,z) := &\Pr[|\xi| \in A\sigma(i-1, i+1)] \\ &+ \Pr[|\xi| < Ai\sigma, |\xi + z| > Ai\sigma] + \Pr[|\xi| > Ai\sigma, |\xi + z| < Ai\sigma] \,. \end{aligned}$$

This will help us bound the final error terms that arise in the calculation. We will need the following upper bound on $P(i, z) + P(i, z')$.

Claim C.4. *There exists a function $\tilde{P} : \mathbb{N} \to \mathbb{R}^+$ satisfying:*

1. *For all $i \in \mathbb{N}$, $\tilde{P}(i) \ge P(i, z) + P(i, z')$ which can be computed using samples from $\xi + z$ and $\xi + z'$.*
2. *There is a $k \in [(1/\alpha\eta), (C/\alpha + 1/\alpha\eta)^{1/\eta}]$ such that $k\tilde{P}(k) < \eta \sum_{j=1}^{k} \tilde{P}(k)$.*
3. *$\sum_{j=1}^{k} \tilde{P}(k) = O(\Pr[|\xi + z| \le A\sigma k] + \Pr[|\xi + z'| \le A\sigma k])$.*
4. *With probability $1 - \delta$, for all $i < (O(1)/\eta\alpha)^{O(1)/\eta}$, $\tilde{P}(i)$ can be estimated to an accuracy of less than $O(\eta/i)$ by using $\mathrm{poly}((O(1)/\eta\alpha)^{1/\eta}, \log(1/\delta\alpha\eta))$ samples from $\xi + z$ and $\xi + z'$.*

We defer the proof of Claim C.4 to Appendix C.2, and continue with our proof showing that $\mathbf{E}[\xi + z \mid |\xi + z| \le A\sigma k] \approx \mathbf{E}[\xi \mid |\xi| \le A\sigma k]$ for k satisfying the conclusions of Claim C.4. To this end, observe the following for $f(\xi, z)$ being either 1 or $\xi + z$.

$$\begin{aligned} &|\mathbf{E}[f(\xi,z)\,\mathbf{1}(|\xi| \le \sigma i)] - \mathbf{E}[f(\xi,z)\,\mathbf{1}(|\xi + z| \le \sigma i)]| \\ &\le |\mathbf{E}[f(\xi,z)\,\mathbf{1}(|\xi + z| > \sigma i)\,\mathbf{1}(|\xi| \le \sigma i)]| + |\mathbf{E}[f(\xi,z)\,\mathbf{1}(|\xi + z| \le \sigma i)\,\mathbf{1}(|\xi| > \sigma i)]|. \end{aligned}$$

By setting $f(\xi, z) := 1$ and considering the case where $i = k$ satisfies the conclusions of Claim C.4, we can bound the "error terms"

$\Pr[|\xi + z| \le A\sigma k$ and $|\xi| > A\sigma k]$ and $\Pr[|\xi + z| > A\sigma k$ and $|\xi| \le A\sigma k]$ in terms of $\tilde{P}(k)$.

Furthermore, $\tilde{P}(k)$ itself is upper bounded by $O(\eta/k)(\Pr[|\xi + z| \le A\sigma k] + \Pr[|\xi + z'| \le A\sigma k])$ as per Item 2 and Item 3. Putting these facts together, we have that

$$\begin{aligned} &|\Pr[|\xi| \le A\sigma k] - \Pr[|\xi + z| \le A\sigma k]| \\ &= O(\eta/k)(\Pr[|\xi + z| \le A\sigma k] + \Pr[|\xi + z'| \le A\sigma k]). \end{aligned}$$

A similar claim holds for the distribution over z'. An application of the triangle inequality now implies

$$\begin{aligned}&|\Pr[|\xi + z'| \le A\sigma k] - \Pr[|\xi + z| \le A\sigma k]| \\ &= O(\eta/k)\,(\Pr[|\xi + z| \le A\sigma k] + \Pr[|\xi + z'| \le A\sigma k]).\end{aligned}$$

If $|A - B| < \tau(A + B)$ it follows that $(1 - \tau)/(1 + \tau) < A/B < (1 + \tau)/(1 - \tau)$. For $\tau \in (0, 1/2]$, this means $A = \Theta(B)$. Applying this to our case, we can conclude that $\Pr[|\xi + z'| \le A\sigma k] = \Theta(\Pr[|\xi + z| \le A\sigma k])$. Substituting this equivalence back into the previous expression, we obtain:

$$|\Pr[|\xi| \le A\sigma k] - \Pr[|\xi + z| \le A\sigma k]| = O(\eta/k)\,(\Pr[|\xi + z| \le A\sigma k])\,. \tag{1}$$

Similarly, when $f(\xi, z) := \xi + z$, we need to control the error terms: $\mathbf{E}[(\xi + z)\,\mathbf{1}(|\xi + z| \le A\sigma k)\,\mathbf{1}(|\xi| > A\sigma k)]$ and $\mathbf{E}[(\xi + z)\,\mathbf{1}(|\xi + z| > A\sigma k)\,\mathbf{1}(|\xi| \le A\sigma k)]$.

Observe that $(\xi + z)\,\mathbf{1}(|\xi + z| \le A\sigma k)\,\mathbf{1}(|\xi| > A\sigma k)$ has a nonzero value with probability at most $\mathbf{E}[\mathbf{1}(|\xi + z| \le A\sigma k)\,\mathbf{1}(|\xi| > A\sigma k)] < \tilde{P}(k)$. Also, the magnitude of $(\xi + z)$ in this event is at most $A\sigma k$. Putting these together, we get that

$$|\mathbf{E}[(\xi + z)\,\mathbf{1}(|\xi + z| \le A\sigma k)\,\mathbf{1}(|\xi| > A\sigma k)]| < A\sigma k\tilde{P}(k) < O(A\sigma\eta)\,\sum_{j=1}^{k} \tilde{P}(j).$$

Unfortunately, we cannot use the same argument to bound $\mathbf{E}[(\xi+z)\,\mathbf{1}(|\xi+z| > A\sigma k)\,\mathbf{1}(|\xi| \le A\sigma k)]$, since $|\xi + z|$ is no longer bounded by $A\sigma k$ in this event. However, we can break the sum $\xi + z$ as follows: $\xi + z = \xi + z\mathbf{1}(|z| > A\sigma k) + z\mathbf{1}(|z| \le A\sigma k)$. This allows us to get the following bound:

$$\begin{aligned}&\mathbf{E}[(\xi + z)\,\mathbf{1}(|\xi + z| > A\sigma k)\,\mathbf{1}(|\xi| \le A\sigma k)] \\ &< 2A\sigma k\tilde{P}(k) + \mathbf{E}[z\,\mathbf{1}(|z| > A\sigma k)\,\mathbf{1}(|\xi + z| > A\sigma k)\,\mathbf{1}(|\xi| \le A\sigma k)] \\ &< 2A\sigma k\tilde{P}(k) + \mathbf{E}[z\,\mathbf{1}(|z| > A\sigma k)] \\ &< 2A\sigma k\tilde{P}(k) + A\sigma/k \\ &< O(A\eta\sigma\sum_{j=1}^{k}\tilde{P}(j)) + O(A\eta\sigma\alpha),\end{aligned}$$

where the third inequality follows by an application of Chebyshev's inequality, and the final inequality follows by choosing $k \ge 1/(\eta\alpha)$.

Putting everything together, we see

$$\mathbf{E}[(\xi + z)\,\mathbf{1}(|\xi + z| \le A\sigma k)] = \mathbf{E}[(\xi + z)\,\mathbf{1}(|\xi| \le A\sigma k)] \pm O(A\sigma\eta\alpha + A\sigma\eta\sum_{j=1}^{k}\tilde{P}(j)). \tag{2}$$

To finally compute the conditional probability, we use Equation (1) and Equation (2) to get

$$\begin{aligned}\mathbf{E}[(\xi + z) \mid |\xi| \le A\sigma k] &= \frac{\mathbf{E}[(\xi + z)\,\mathbf{1}(|\xi + z| \le A\sigma k)] \pm O(A\sigma\eta\alpha + A\sigma\eta\sum_{j=1}^{k}\tilde{P}(j))}{\Pr[|\xi| \le A\sigma k]} \\ &= \frac{\mathbf{E}[(\xi + z)\,\mathbf{1}(|\xi + z| \le A\sigma k)] \pm O(A\sigma\eta\alpha + A\sigma\eta\sum_{j=1}^{k}\tilde{P}(j))}{(1 + \Theta(\eta/k))\;\Pr[|\xi + z| \le A\sigma k]} \\ &= (1 - \Theta(\eta/k))\;\mathbf{E}[(\xi + z) \mid |\xi + z| \le A\sigma k] \\ &\qquad \pm O(1)\,\frac{A\sigma\eta\alpha + A\sigma\eta\Pr[|\xi + z| \le A\sigma k]}{\Pr[|\xi + z| \le A\sigma k]} \\ &= \mathbf{E}[(\xi + z) \mid |\xi + z| \le A\sigma k] \pm O(A\eta\sigma)\,,\end{aligned}$$

where the second inequality is a consequence of Item 3, and the last is due to the fact that $\Pr[|\xi + z| \le A\sigma k] \ge \alpha/2$ whenever $k > 2$, which follows from an application of Fact A.3 while noting the fact that $\Pr[\xi = 0] \ge \alpha$.

Taking a difference for the above calculations for z and z', we see that,

$$\mathbf{E}[(\xi + z) \mid |\xi + z| \le A\sigma k] - \mathbf{E}[(\xi + z') \mid |\xi + z'| \le A\sigma k] = \mathbf{E}[z] - \mathbf{E}[z'] \pm O(A\eta\sigma).$$

Consider this final error, and let $O(A\eta\sigma) < CA\eta\sigma$ for some constant C. Repeating the above argument initially setting $\eta = \eta'/C$, where C is the constant gives us the guarantee we need.

Finally, we estimate the runtime and sample complexity of our algorithm. The main bottleneck in our algorithm is the repeated estimation of $\tilde{P}(i)$ and estimation of $\mathbf{E}[(\xi + z) \mid |\xi + z| \le A\sigma k]$.

According to Item 4, each time we estimate $\tilde{P}(i)$ to the desired accuracy, we draw $\mathrm{poly}((O(1)/\alpha\eta)^{1/\eta}, \log(1/\delta\eta\alpha))$ samples.

An application of Hoeffding's inequality (Lemma A.1) then allows us to estimate the conditional expectation $\mathbf{E}[(\xi + z) \mid |\xi + z| \le A\sigma k]$ to an accuracy of $\eta A\sigma$ by drawing $\mathrm{poly}((O(1)/\alpha\eta)^{1/\eta}, \log(1/\delta\eta\alpha))$ samples as well. The exponential dependence here comes from the exponential upper bound on k.

□

C.2 Proof of Claim C.4

In this section, we prove the existence of $\tilde{P}(\cdot)$ which is an upper bound on $P(i, z) + P(i, z')$, which we can estimate using samples from $\xi + z$ and $\xi + z'$.

Proof of Claim C.4

Proof of Item 1:

Recall the definition of $P(i, z)$.

$$P(i, z) := \Pr[|\xi| \in A\sigma(i-1, i+1)] + \Pr[|\xi| < Ai\sigma, |\xi + z| > Ai\sigma] + \Pr[|\xi| > Ai\sigma, |\xi + z| < Ai\sigma] \,.$$

For Item 1 to hold, we need to define $\tilde{P}(i)$ to be an upper bound on $P(i, z) + P(i, z')$ which can be computed using samples from $\xi + z$ and $\xi + z'$. To this end, we bound $P(i, z)$ as follows. First, note that we can adjust the endpoints of the intervals to get

$$P(i, z) < 3\Pr[|\xi| \in A\sigma(i-1, i+1)] + \Pr[|\xi| < A(i-1)\sigma, |\xi + z| > Ai\sigma] + \Pr[|\xi| > A(i+1)\sigma, |\xi + z| < Ai\sigma] \,.$$

Then, we partition the ranges in the definition above into intervals of length $A\sigma$ to get:

$$\begin{aligned} P(i, z) < 3\Pr[|\xi| \in A\sigma(i-1, i+1)] &\\ + \sum_{j=1}^{i-2} \Pr[|\xi| \in Aj\sigma + [0, A\sigma), |\xi + z| > Ai\sigma] &\\ + \sum_{j=1}^{i-1} \Pr[|\xi| > A(i+1)\sigma, |\xi + z| \in Aj\sigma + [0, A\sigma)]. \end{aligned}$$

Next, an application of the triangle inequality to $|\xi| \in Aj\sigma + [0, A\sigma)$ and $|\xi + z| > Ai\sigma$ implies that $|z| \ge A(i - j - 1)\sigma$. Similarly, the same kind of argument when $|\xi| > A(i+1)\sigma$ and $|\xi + z| \in Aj\sigma + [0, A\sigma)$ demonstrates that $|-z| = |\xi + z - \xi| \ge A(i - j)\sigma$. We then use Fact A.3 to move from $|\xi + z|$ to $|\xi|$ in the third term.

$$\begin{aligned} P(i, z) < 3\Pr[|\xi| \in A\sigma(i-1, i+1)] &\\ + \sum_{j=1}^{i-2} \Pr[|\xi| \in Aj\sigma + A\sigma[0, 1), |z| \ge (i - j - 1)A\sigma] &\\ + O(1) \sum_{j=1}^{i-1} \Pr[|\xi| \in Aj\sigma + A\sigma[-2, 3), |z| \ge (i - j)A\sigma] \,. \end{aligned}$$

An application of Chebyshev's inequality to z, using the independence of z and ξ, gives that

$$\begin{aligned}P(i,z) &< 3\Pr[|\xi| \in A\sigma(i-1,i+1)] \\ &\quad + O(1)\sum_{j=1}^{i-2}(1/(i-j-1)^2)\Pr[|\xi| \in Aj\sigma + A\sigma[0,1)] \\ &\quad + O(1)\sum_{j=1}^{i-1}(1/(i-j)^2)\Pr[|\xi| \in Aj\sigma + A\sigma[-2,3)]\ .\end{aligned}$$

Another application of Fact A.3 applied to $(\xi+z)-z$ then gives us

$$\begin{aligned}P(i,z) &< 3\Pr[|\xi| \in A\sigma(i-5,i+5)] \\ &\quad + O(1)\sum_{j=1}^{i-2}(1/(i-j-1)^2)\Pr[|\xi+z| \in Aj\sigma + A\sigma[-2,3)] \\ &\quad + O(1)\sum_{j=1}^{i-1}(1/(i-j)^2)\Pr[|\xi+z| \in Aj\sigma + A\sigma[-4,5)].\end{aligned}$$

Finally, extending all intervals so that they match, and observing that $\sum_{j=1}^{i-2}(1/(i-j-1)^2)\Pr[|\xi+z| \in Aj\sigma + A\sigma[-2,3)] \le \sum_{j=1}^{i-1}(1/(i-j)^2)\Pr[|\xi+z| < Aj\sigma + A\sigma[-4,5)]$, we get

$$\begin{aligned}P(i,z) &< O(1)\Pr[|\xi| \in A\sigma(i-5,i+5)] \\ &\quad + O(1)\sum_{j=1}^{i-1}(1/(i-j)^2)\Pr[|\xi+z| \in Aj\sigma + A\sigma[-4,5)].\end{aligned}$$

We now let $\tilde{P}(i,z)$ denote the final upper bound on $P(i,z)$. The value of having $\tilde{P}(i,z)$ is that it can be computed using samples from $\xi+z$.

$$\begin{aligned}\tilde{P}(i,z) &:= O(1)\Pr[|\xi| \in A\sigma(i-5,i+5)] \\ &\quad + O(1)\sum_{j=1}^{i-1}(1/(i-j)^2)\Pr[|\xi+z| \in Aj\sigma + A\sigma[-4,5)]\ .\end{aligned}$$

We defined $\tilde{P}(i) = \tilde{P}(i,z) + \tilde{P}(i,z')$.

Proof of Item 2:

First observe that $\sum_{i=1}^{\infty}\tilde{P}(i) < C$ for some constant C. It is clear that this is true of the first term, since every interval will get over-counted at most 10 times. To see that the second term can be bounded, observe that

$$\begin{aligned}&\sum_{i=1}^{\infty}\sum_{j=1}^{i-1}(1/(i-j)^2)\Pr[|\xi+z| \in Aj\sigma + A\sigma[-4,5)] \\ &< \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}(1/(i-j)^2)\Pr[|\xi+z| \in Aj\sigma + A\sigma[-4,5)] \\ &< \sum_{j=1}^{\infty}\Pr[|\xi+z| \in Aj\sigma + A\sigma[-4,5)]\sum_{i=1}^{\infty}(1/(i-j)^2) = O(1)\ .\end{aligned}$$

The first inequality follows by extending the limits of summation.

The final inequality follows from the fact that the total probability is at most 1, every interval of size σ gets over-counted at most finitely many times, and the fact that $\sum_1^{\infty} 1/k^2 = O(1)$.

Item 2 now follows from the fact that $\tilde{P}(i)$, $i \ge 1$ is a positive sequence that sums to a finite quantity, and $\tilde{P}(1) \ge \alpha/2$, since the interval $\tilde{P}(1)$ upper bounds is contains at least a constant fraction of the mass of ξ at 0 that is moved by z, z', and $\Pr[\xi = 0] \ge \alpha$.

Applying Lemma B.2, we get our result.

Proof of Item 3:

Let k be such that Item 2 holds, i.e. $k\tilde{P}(k) < \eta \sum_{j=1}^{k} \tilde{P}(k)$, then the goal is to show $\sum_{j=1}^{k} \tilde{P}(k) = O(\Pr[|\xi + z| < A\sigma k] + \Pr[|\xi + z'| < A\sigma k])$.

We first consider the sum over i, of $\tilde{P}(i, z)$. It is easy to see that this is

$$\sum_{i=1}^{k} \tilde{P}(i, z) = O(1) \Pr[|\xi + z| \le A\sigma(k + 5)] \\ + O(1) \sum_{i=1}^{k} \sum_{j=1}^{i-1} (1/(i - j)^2) \Pr[|\xi + z| \in Aj\sigma + A\sigma[-4, 5)] \ .$$

The first term on the RHS is almost what we want. We now show how to bound the second term,

$$\begin{aligned}
&\sum_{i=1}^{k} \sum_{j=1}^{i-1} (1/(i - j)^2) \Pr[|\xi + z| \in Aj\sigma + A\sigma[-4, 5)] \\
&< \sum_{j=1}^{k-1} \sum_{i=0; i\neq j}^{k} (1/(i - j)^2) \Pr[|\xi + z| \in Aj\sigma + A\sigma[-4, 5)] \\
&= \sum_{j=1}^{k-1} \Pr[|\xi + z| \in Aj\sigma + A\sigma[-4, 5)] \sum_{i=0; i\neq j}^{k} (1/(i - j)^2) \\
&< O(1) \sum_{j=1}^{k-1} \Pr[|\xi + z| \in Aj\sigma + A\sigma[-4, 5)] \\
&< O(1) \Pr[|\xi + z| < A(k + 5)\sigma] \ .
\end{aligned}$$

The first inequality holds since any pair of (i, j) that has a nonzero term in the first sum will also occur in the second sum, and all terms are non-negative.

The second equality is just pulling the common j term out.

The third inequality follows from the fact that $\sum_{i=1}^{\infty} 1/i^2 = O(1)$.

The fourth inequality follows from the fact that each σ-length interval is overcounted at most a constant number of times.

This allows us to bound $\sum_{i=1}^{k} \tilde{P}(i, z)$ by $O(\Pr[|\xi + z| < A\sigma(k + 5)])$ overall. Similarly for $\sum_{i=1}^{k} \tilde{P}(i, z')$, we can obtain a bound of $O(\Pr[|\xi + z'| < A\sigma(k + 5)])$. Putting these together, we see $\sum_{i=1}^{k} \tilde{P}(i) \le O(\Pr[|\xi + z| < A\sigma(k + 5)] + \Pr[|\xi + z'| < A\sigma(k + 5)])$. Finally, to get the upper bound claimed in Item 3, observe that

$$\begin{aligned}
&\Pr[|\xi + z| < A\sigma(k + 5)] + \Pr[|\xi + z'| < A\sigma(k + 5)] \\
&= \Pr[|\xi + z| < A\sigma(k - 4)] + \Pr[|\xi + z'| < A\sigma(k - 4)] \\
&\quad + \Pr[|\xi + z| \in A\sigma(k - 4, k + 5)] + \Pr[|\xi + z'| \in A\sigma(k - 4, k + 5)] \\
&\le \Pr[|\xi + z| < A\sigma(k - 4)] + \Pr[|\xi + z'| < A\sigma(k - 4)] \\
&\quad + \tilde{P}(k) \\
&\le \Pr[|\xi + z| < A\sigma(k - 4)] + \Pr[|\xi + z'| < A\sigma(k - 4)] \\
&\quad + O(\eta/k)(\Pr[|\xi + z| < A\sigma(k + 5)] + \Pr[|\xi + z'| < A\sigma(k + 5)]) \ .
\end{aligned}$$

Rearranging the inequality and by scaling η such that that $O(\eta/k) \le 1/2$, we see that $\Pr[|\xi + z| < A\sigma(k+5)] + \Pr[|\xi + z'| < A\sigma(k+5)] = O(\Pr[|\xi + z| < A\sigma(k-4)] + \Pr[|\xi + z'| < A\sigma(k-4)]) = O(\Pr[|\xi + z| < A\sigma k] + \Pr[|\xi + z'| < A\sigma k])$, completing our proof of Item 3.

Proof of Item 4:

Finally, to see Item 4 holds, observe that $0 < \tilde{P}(i) < O(1)$. Let $B = (O(1)/\eta\alpha)^{1/\eta}$ denote the maximum index before which we can find a k such that $k\tilde{P}(k) \leq \eta \sum_{i=1}^{k} \tilde{P}(i)$. Now, To estimate $\tilde{P}(i)$ empirically, we partition the interval $(-BA\sigma, BA\sigma)$ into B intervals of length $A\sigma$ each, and estimate the probability of $\xi + z$ falling in each interval. If we estimate each of these probabilities to an accuracy of $\eta/(100\,B)$, we can estimate $\tilde{P}(i)$ to an accuracy of $O(\eta/i)$.

An application of Hoeffding's inequality (Lemma A.1) tells us that each estimate will require $O(B^2/\eta^2 \log(1/\delta))$ samples. Taking a union bound over all these intervals, we see that we will require $O(B^2/\eta^2 \log(B/\delta))$ samples.

Finally, another union bound over each $i \in [0, B]$ implies that we will need $O(B^2/\eta^2 \log(B^2/\delta))$ samples. Substituting the value of B back in, we see that this amounts to requiring $(O(1)/\eta\alpha)^{2/\eta} \log(1/\eta\alpha\delta)$ samples.

Estimating $\tilde{P}(i)$ will take time polynomial in the number of samples, and so we take time $\tilde{O}(O(1)/\eta\alpha)^{O(1)/\eta})$. □

C.3 High-dimensional Noisy Location Estimation

In this section, we explain how to use our one-dimensional location estimation algorithm to get an algorithm for noisy location estimation in d dimensions.

The algorithm performs one-dimensional location estimation coordinate-wise, after a random rotation.

We need to perform such a rotation to ensure that every coordinate has a known variance bound of $\sigma/\sqrt{d}$.

Algorithm 5 High-dimensional Location Estimation: ShiftHighD$(S_1, S_2, \eta, \sigma, \alpha)$

input: Sample sets $S_1, S_2 \subset \mathbb{R}^d$ of size m, $\eta \in (0, 1)$, $\sigma > 0$, α

1. Sample $R_{i,j}$ i.i.d. from the uniform distribution over $\{\pm 1/\sqrt{d}\}$ for $i, j \in [d]$
2. Represent S_1 and S_2 in the basis given by the rows of R: $r_1, \ldots, r_d$.
3. **for** $i \in [d]$ **do**
    $v'_i :=$ Shift1D$(S_1 \cdot e_i, S_2 \cdot e_i, \eta, O(\sigma/\sqrt{d}), \alpha)$
end
4. Change the representation of v' back to the standard basis.
5. *Probability Amplification:* Repeat steps 1-4, $T := O(\log(1/\delta))$ times to get $C := \{v'_1, \ldots, v'_T\}$
6. Find a ball of radius $O(\eta\sigma)$ centered at one of the v'_i containing $> 90\%$ of C. If such a vector exists, set v' to be this vector. Otherwise set v' to be an arbitrary element of C.
5. Return v'.

Lemma C.5 (Location Estimation). *Let $y_i := \xi + z_i$ for $i \in \{1, 2\}$ where $\Pr[\xi = 0] \geq \alpha$ and $z_i \sim D_i$ are distributions over $\mathbb{R}^d$ satisfying $\mathbf{E}_{D_i}[x] = 0$ and $\mathbf{E}_{D_i}[\|x\|^2] \leq \sigma^2$. Let $v \in \mathbb{R}^d$ be an unknown shift. There is an algorithm (Algorithm 5), which draws $m = \mathrm{poly}((O(1)/\eta\alpha)^{1/\eta}, \log(1/\delta\epsilon\alpha))$ samples each from y_1 and $y_2 + v$, runs in time $\mathrm{poly}(d, (O(1)/\eta\alpha)^{1/\eta}, \log(1/\delta\epsilon\alpha))$ and returns v' satisfying $\|v' - v\| \leq O(\eta\sigma)$ with probability $1 - \delta$.*

Proof. Consider a matrix R whose entries $R_{i,j}$ are independently drawn from the uniform distribution over $\pm 1/\sqrt{d}$. and whose diagonals are $1/\sqrt{d}$.

Our goal is to show that with probability at least 99%, the standard deviation of each coordinate of Rz is bounded by $O(\sigma/\sqrt{d})$, i.e., the standard deviation of $Rz \cdot e_i$ is at most $O(\sigma/\sqrt{d})$ for all integer i in $[d]$.

We can then amplify this probability to ensure that the algorithm fails with a probability that is exponentially small.

To see this, observe that $Rz \cdot e_i = r_i \cdot z$, and so $\mathbf{E}_z[r_i \cdot z] = 0$.

$$\begin{aligned}\mathop{\mathbf{E}}_z[(r_i \cdot z)^2] &= \sum_{p\in[d],q\in[d]} R_{i,p}R_{i,q}\,\mathbf{E}[z_p z_q] \\ &= \sum_{i=1}^{d} \mathbf{E}[z_i^2]/d + 2\sum_{p,q\in[d],p<q} R_{i,p}R_{i,q}\,\mathbf{E}[z_p z_q] \\ &\le (\sigma^2/d) + 2\sum_{p,q\in[d],p<q} R_{i,p}R_{i,q}\,\mathbf{E}[z_p z_q]\ .\end{aligned}$$

We now bound the second term with probability 99% via applying Chebyshev's inequality. Observe that $\mathbf{E}[z_p z_q] \le \sqrt{\mathbf{E}[z_p^2]\,\mathbf{E}[z_q^2]} \le \sigma^2$. Since $R_{i,p}$ and $R_{i,q}$ are drawn independently and $p \neq q$, we see that the variables $R_{i,p}R_{i,q}$ and $R_{i,l}R_{i,m}$ pairwise independent for pairs $(p,q) \neq (l,m)$, this implies $\Pr[|\sum_{p,q\in[d],p<q} R_{i,p}R_{i,q}\,\mathbf{E}[z_p z_q]| > T] \le \frac{O(\sigma^4)}{d\ d^2T^2}$. By choosing $T = O(\sigma^2/d)$, we see that the right-hand side above is at most $0.001/d$.

A union bound over all the coordinates then tells us that with probability 99%, the variance of each coordinate is at most $O(\sigma^2/d)$.

Then, for each coordinate i, we can identify $v_i' = v_i \pm O(\eta\sigma/\sqrt{d})$ through an application of Lemma 3.1. Putting these together with probability at least 99%, we find v' satisfying $\|v' - v\|^2 \le O(\eta^2\sigma^2)$.

Changing between these basis representations maintains the quality of our estimate since the new basis contains unit vectors nearly orthogonal to each other. With high probability, the inner products between these are around $O(1/\sqrt{d})$ for every pairwise comparison, so R approximates a random rotation.

Probability Amplification: The current guarantee ensures that we obtain a good candidate with a constant probability of success. However, for the final algorithmic guarantee, we need a higher probability of success. To achieve this, we modify the algorithm as follows:

1. Run the algorithm T times, each time returning a candidate v_i' that is, with probability 99%, within $O(\eta\sigma)$ distance from the true solution.
2. Construct a list of candidates $C = \{v_1', \ldots, v_T'\}$.
3. Identify a ball of radius $O(\sigma\eta)$ centered at one of the v_i' that contains at least 90% of the remaining points.
4. Return the corresponding v_i' as the final output.
5. If no such v_i' exists, return any vector from C.

Let E denote the event that a point is within $O(\eta\sigma)$ to the true solution.

This will succeed with probability $1 - \exp(-T)$. To see why, observe the chance that we recover $(2/3)T$ vectors outside the event E is less than $(0.01)^{2/3\,T}\binom{T}{2/3T} < (0.047)^T\binom{T}{T/2} < (0.047)^T(2^T/\sqrt{T}) < (0.095)^T$. □

D List-Decodable Mean Estimation

This section presents an algorithm for list-decodable mean estimation when the inlier distribution follows $\mathcal{D}_\sigma$. Here, $\mathcal{D}_\sigma$ represents a set of distributions over $\mathbb{R}^d$ defined as $\mathcal{D}_\sigma := \{D \mid \mathbf{E}_D[|x - \mathbf{E}_D[x]|^2] \le \sigma^2\}$. In our setting, we receive samples from $\xi + z$, where $\Pr[\xi = 0] > \alpha$, where α can be close to 0. Our objective is to estimate the mean with a high degree of precision.

Note that the guarantees provided by prior work do not directly apply to our setting. Prior work examines a more aggressive setting where arbitrary outliers are drawn with a probability of $1 - \alpha$. These outliers might not have the additive structure we have.

Recall the definition of an (α, β, s)-LDME algorithm:

Definition D.1 (Algorithm for List-Decodable Mean Estimation)**.** *Algorithm $\mathcal{A}$ is an (α, β, s)-LDME algorithm for $\mathcal{D}$ (a set of candidate inlier distributions) if with probability $1 - \delta_{\mathcal{A}}$, it returns a list $\mathcal{L}$ of size s such that $\min_{\hat{\mu} \in \mathcal{L}} \|\hat{\mu} - \mathbf{E}_{x \sim D}[x]\| \leq \beta$ for $D \in \mathcal{D}$ when given $m_{\mathcal{A}}$ samples of the kind $z + \xi$ for $z \sim D$ and $\Pr[\xi = 0] \geq \alpha$. If $1 - \alpha$ is a sufficiently small constant less than $1/2$, then $s = 1$.*

We now prove Fact 2.1 which we restate below for convenience.

Fact D.2 (List-decoding algorithm)**.** *There is an $(\alpha, \eta\sigma, \tilde{O}((1/\alpha)^{2/\eta^2}))$-LDME [3] algorithm for the inlier distribution belonging to $\mathcal{D}_\sigma$ which runs in time $\tilde{O}(d(1/\alpha)^{2/\eta^2})$ and succeeds with probability $1 - \delta$. Conversely, any algorithm which returns a list, one of which makes an error of at most $O(\eta\sigma)$ in ℓ_2 norm to the true mean, must have a list whose size grows exponentially in $1/\eta$.*

If $1 - \alpha$ is a sufficiently small constant less than half, then the list size is 1 to get an error of $O(\sqrt{1 - \alpha}\,\sigma)$.

Proof. **Algorithm:** Consider the following algorithm:

1. If $\alpha < c$ and $\eta > \sqrt{\alpha}$: Run any stability-based robust mean estimation algorithm from [10] and return a singleton list containing the output of the algorithm.
2. Otherwise, for integer each $i \in [1, 100(1/\alpha)^{2/\eta^2} \log(1/\delta)^2]$ sample $1/\eta^2$ samples and let their mean be μ_i.
3. Return the list $\mathcal{L} = \{\mu_i \mid i \in [1, 100(1/\alpha)^{2/\eta^2} \log(1/\delta)^2]\}$.

If the algorithm returns in the first step, then the guarantees follow from the guarantees of the algorithm for robust mean estimation from [10] (Proposition 1.5 on page 4).

Otherwise, observe that the probability that every one of $1/\eta^2$ samples drawn is an inlier, is α^{1/η^2}.

Hence, with probability $1 - \delta$ we see that if we draw $1/\eta^2$ samples $O((1/\alpha)^{2/\eta^2} \log(1/\delta)^2)$ times, there are at least $O(\log(1/\delta))$ sets of samples containing only inliers. Then, the mean of one of these concentrates to an error of $O(\eta\sigma)$ by an application of Lemma A.2. More precisely, Lemma A.2 ensures that with probability 99%, the mean of a set of $1/\eta^2$ inliers concentrates up to an error of $O(\eta\sigma)$. Repeating this $\log(1/\delta)$ times, and hence get our result.

Hardness: To see that the list size must be at least $\exp(1/\eta)$, consider the set of inlier distributions given by $\{D_s \mid s \in \{\pm 1\}^d\}$ where each D_s is defined as follows: D_s is a distribution over $\mathbb{R}^d$ such that each coordinate independently takes the value s_i with probability $1/d$, and 0 otherwise.

Each D_s defined above belongs to $\mathcal{D}_{\sqrt{1-1/d}}$ since $\mathbf{E}_{x \sim D_s}[x] = s/d$ and

$$\sigma^2 := \mathop{\mathbf{E}}_{x \sim D_s}[\|x - s/d\|^2] = \sum_{i=1}^{d} \mathop{\mathbf{E}}_{x_i \sim (D_s)_i}[(x_i - s_i/d)^2]$$
$$= \sum_{i=1}^{d} (1 - 1/d)(1/d)^2 + (1/d)(1 - 1/d)^2 = (1 - 1/d).$$

We will set the oblivious noise distribution for each D_s to be $-D_s$. Our objective is to demonstrate that the distribution of $D_s - D_s$ is the same for all s and is independent of s. This means that we cannot identify s by seeing samples from $D_s - D_s$.

Then, since the means of D_s and D_s' for any distinct pair $s, s' \in \{\pm 1\}^d$ differ by at least $1/d$, if we set $d = 1/\eta$ we see that there are $2^{1/\eta}$ possible different values of the original mean, each pair being at least η far apart, which is larger than $\eta\sigma^2 = \eta(1 - \eta)$.

We can assume, without loss of generality, that $s = \mathbf{1}$, where $\mathbf{1}$ represents the all-ones vector. Each coordinate of D_s can be viewed as a random coin flip, taking the value 0 with probability $1 - 1/d$ and 1 with probability $1/d$.

[3] Typo in main body: η instead of the correct $\eta\sigma$.

The probability of obtaining the all-zeros vector is given by $(1 - 1/d)^d$, which approaches a constant value for sufficiently large d, and so, $\Pr_{x \sim D_s}[x = 0] \geq 0.001$, i.e., the α for the oblivious noise is at least a constant. In fact, it can be as large as $1/e > 0.35$ for large enough d.

Let the oblivious noise be $-D$. Now, consider the distribution of $x + y$, where x follows the distribution D and y follows the distribution $-D$. If we focus on the first coordinate, $x_1 + y_1$, we observe that it follows a symmetric distribution over $\{-1, 0, 1\}$ which does not depend on s_1. Also, each coordinate exhibits the same distribution, and they are drawn independently of one another. Hence, the final distribution is independent of s, so we get our result. ☐

E Proof of Corollary 4.2

Below, we restate Corollary 4.2 for convenience.

Corollary E.1. *Given access to oblivious noise oracle $\mathcal{O}_{\alpha,\sigma,f}$, a $(O(\eta\sigma), \epsilon)$-inexact-learner $\mathcal{A}_G$ running in time T_G, there exists an algorithm that takes* $\mathrm{poly}((1/\alpha)^{1/\eta^2}, (O(1)/\eta)^{1/\eta}, \log(T_G/\delta\eta\alpha))$ *samples, runs in time* $T_G \cdot \mathrm{poly}(d, (1/\alpha)^{1/\eta^2}, (O(1)/\eta)^{1/\eta}, \log(1/\eta\alpha\delta))$*, and with probability* $1 - \delta$ *returns a list $\mathcal{L}$ of size $\tilde{O}((1/\alpha)^{1/\eta^2})$ such that* $\min_{x \in \mathcal{L}} \|\nabla f(x)\| \leq O(\eta\sigma) + \epsilon$*. Additionally, the exponential dependence on $1/\eta$ in the list size is necessary.*

Proof. This follows by substituting the guarantees of Fact 2.1 for the algorithm $\mathcal{A}_{ME}$ in Theorem 1.4. ☐