# Tight Lower Bounds for $\ell_2$ Sampling

WILLIAM SWARTWORTH, Carnegie Mellon University, USA
DAVID P. WOODRUFF, Carnegie Mellon University, USA

We study linear sketches and data stream algorithms for the $\ell_p$-sampling problem. Here one chooses an $r \times n$ matrix $A$ and computes $A \cdot v$ for an underlying vector $v$. From $A \cdot v$, one should output a coordinate $i \in \{1, 2, \ldots, n\}$ with probability $(1 \pm \epsilon)\frac{\|v_i\|_p}{\|v\|_p^p}$. One is also allowed to output FAIL with a small constant probability. In the estimation version of the problem, one would also like to output a $(1 \pm \epsilon)$-approximation to $\frac{|v_i|^p}{\|v\|_p^p}$, which is the probability for which $i$ was sampled. For $0 < p < 2$, work of Jayaram and Woodruff (FOCS, 2018) showed an upper bound of $r = O(\log n)$ while for $p = 2$ the best known upper bound is $r = O(\log^2 n)$. They posed it as an open question to resolve this gap. We prove an $\Omega(\log^2 n)$ sketching dimension lower bound for $p = 2$, thereby separating the complexity for $0 < p < 2$ from that of $p = 2$, and showing that their upper bound is optimal. Further, we show an $\Omega(\epsilon^{-2} \log n)$ sketching dimension lower bound for the estimation problem, improving the previous $\Omega(\epsilon^{-2})$ bound and showing optimality of their algorithm. These results give sketching dimension lower bounds for real-valued inputs, but by applying a result of Gribelyuk, Lin, Woodruff, Yu, and Zhou (STOC, 2025), we obtain the same sketching dimension lower bound for poly$(n)$-valued integer sketches on poly$(n)$-bounded integer-valued streams, thereby showing the optimality of Jayaram and Woodruff's sketch for turnstile streams.

CCS Concepts: • **Theory of computation → Sketching and sampling**; **Lower bounds and information complexity**.

Additional Key Words and Phrases: l2 sampling, lower bounds, streaming algorithms

**ACM Reference Format:**
William Swartworth and David P. Woodruff. 2026. Tight Lower Bounds for $\ell_2$ Sampling. *Proc. ACM Manag. Data* 4, 2 (PODS), Article 119 (May 2026), 15 pages. https://doi.org/10.1145/3801915

## 1 Introduction

The streaming model of computation is a standard framework for analyzing massive datasets that are too large to be stored in available memory, such as database logs, financial market data, physical sensor readings, and social network activity. In this model, an algorithm typically makes a single pass over a sequence of updates to an underlying dataset, and must use a small amount of memory to solve a given task.

Linear sketches are one of the most common and powerful techniques for designing streaming algorithms, especially in the so-called turnstile streaming model which allows for positive and negative updates to the underlying dataset. One reason is that they are easy to maintain under such operations. Formally, for an underlying data vector $v \in \mathbb{R}^n$, a linear sketch maintains a much smaller vector $A \cdot v \in \mathbb{R}^m$ (where $m \ll n$) via a *sketching matrix* $A \in \mathbb{R}^{r \times n}$. Given an update $v \leftarrow v + w$ in the stream, for some vector $w$, one can update the sketch $A(v + w) \leftarrow Av + Aw$ without knowledge of $v$. It is known that under certain conditions, an optimal algorithm in the turnstile streaming

Authors' Contact Information: William Swartworth, wswartworth@gmail.com, Carnegie Mellon University, Pittsburgh, PA, USA; David P. Woodruff, dwoodruf@cs.cmu.edu, Carnegie Mellon University, Pittsburgh, PA, USA.


ACM 2836-6573/2026/5-ART119
https://doi.org/10.1145/3801915

model for any problem can be implemented as a linear sketch, making lower bounds for linear sketches broadly applicable to understanding the limits of streaming computation [1, 8, 9].

Within this context, one of the most widely studied problems is that of $\ell_p$ sampling. The objective of an $\ell_p$ sampler is to select an index $i$ with probability proportional to its contribution to the $p$-th power $\|v\|_p^p$ of the $p$-norm of $v$, that is, with probability approximately $\frac{|v_i|^p}{\|v\|_p^p}$. These were introduced in [10] and followed by a number of works which achieved improved bounds [2, 7]. These papers introduce a relative error $\nu$ and output an index with probability $(1 \pm \nu)\frac{|v_i|^p}{\|v\|_p^p}$ with space complexity depending polynomially on $\nu^{-1}$. This powerful primitive is a key subroutine in algorithms for numerous problems, including heavy hitter detection, norm estimation, and finding duplicates. A significant breakthrough in this area was the development of *perfect samplers* in [6] for the range $p \in (0, 2)$, which achieve this exact probability distribution, i.e., $\nu = 0$, up to overall additive $\frac{1}{\text{poly}(n)}$ variation distance, and using a provably optimal sketching dimension of $r = O(\log n)$. Note that with constant probability these samplers are allowed to output a special FAIL symbol These sketches imply a streaming algorithm for poly$(n)$-bounded integer vectors $v$ using $O(\log^2 n)$ bits of memory [6]. .

Despite this progress for $p \in (0, 2)$, the sketching complexity of $\ell_2$-sampling remained open. The best-known upper bound for perfect $\ell_2$-sampling has sketching dimension $r = O(\log^2 n)$, leaving a logarithmic gap compared to the $\Omega(\log n)$ sketching dimension lower bound that holds for all $p \geq 0$ (see, e.g., [7]; this is also implicit in the 1-sparse compressed sensing lower bounds of [3]). In terms of bit complexity for poly$(n)$-bounded turnstile streams this gap translates to an $O(\log^3 n)$ upper bound in [6] versus an $\Omega(\log^2 n)$ bit lower bound in [7]. This poses the central question, which is an open question in Section 8 of [6]:

*Is $\ell_2$ sampling an inherently harder problem than $\ell_p$ sampling for $p < 2$?*

We note that $\ell_2$-sampling is especially important, as it can for example be used to obtain perfect $\ell_p$-sampling for all $p > 2$ via a black box reduction which in turn has applications to $\ell_p$-moment estimation [13]. They also have applications in randomized numerical linear algebra to sampling by leverage scores in turnstile streams [4].

### 1.1 Our Results

We answer the above question in the affirmative, both in the real-valued sketching model and in the finite precision turnstile streaming model. Namely, we prove an $\Omega(\log^2 n)$ sketching dimension lower bound for $p = 2$, thereby separating the complexity for $0 < p < 2$ from that of $p = 2$ and showing that the upper bound of Jayaram and Woodruff [6] is optimal. Further, we show an $\Omega(\epsilon^{-2} \log n)$ sketching dimension lower bound for the frequency estimation problem, improving the previous $\Omega(\epsilon^{-2})$ bound and showing the optimality of the algorithm of [6]. While these results give sketching dimension lower bounds for real-valued inputs, by applying a recent result of Gribelyuk, Lin, Woodruff, Yu, and Zhou [5] we obtain the same sketching dimension lower bound for poly$(n)$-valued integer sketches on poly$(n)$-bounded integer-valued input streams, thereby showing the optimality of Jayaram and Woodruff's sketch for turnstile streams as well.

Furthermore, we extend our analysis to the problem of providing a $(1 \pm \epsilon)$ relative error estimate of the value $v_i$ for a sampled index $i$. For $p < 2$, this task is known to require $O(\epsilon^{-p})$ additional sketching dimension or $O(\epsilon^{-p} \log n)$ additional bits in the turnstile streaming model. For $p = 2$, prior algorithms required an additional $O(\epsilon^{-2} \log n)$ sketching dimension, or $O(\epsilon^{-2} \log^2 n)$ bits in the turnstile streaming model, which is yet another instance of a mysterious $\log n$ factor gap from $0 < p < 2$. We close this gap as well, showing yet another phase transition at $p = 2$ by proving a matching $\Omega(\epsilon^{-2} \log n \log \frac{1}{\delta})$ sketching dimension lower bound for this estimation task when using an $\ell_2$ sampler and a matching $\Omega(\epsilon^{-2}\log^2 n \log\frac{1}{\delta})$ bit complexity lower bound in turnstile streams, thereby fully characterizing the source of the additional complexity. Combining our two main results, we show:

**Theorem 1.1.** *An $(\epsilon,\delta)$-$\ell_2$-sampler with frequency estimation requires a sketching dimension of at least*

$$\Omega(\log^2 n + \frac{1}{\epsilon^2}\log n).$$

*Moreover, storing such a sketch for vectors with poly$(n)$ sized entries requires*

$$\Omega(\log^3 n + \frac{1}{\epsilon^2}\log^2 n)$$

*bits of space.*

## 1.2 Preliminaries

*$\ell_2$ samplers.* We use the following definition for an $\ell_2$ sampling sketch. For us, we only consider sketches with constant failure probability. Additionally, while $\ell_2$ samplers exist, our lower bounds hold against samplers that perform $\ell_2$ sampling up to constants. So we use this less restrictive definition here.

**Definition 1.1.** *An $\ell_2$ sampling sketch is an algorithm that observes a linear sketch $Sv$ of $v \in \mathbb{R}^n$ with $S : \mathbb{R}^n \to \mathbb{R}^m$, and chooses to either output* FAIL, *with probability at most $\frac{1}{2}$, or output a coordinate $\ell$. Conditioned on not outputting* FAIL, *the outputted coordinate must satisfy*

$$\Pr(\ell = i) = \frac{(1 \pm \frac{1}{2})v_i^2}{\|v\|^2}.$$

Existing $\ell_2$ sampling sketches are also able to output accurate frequency estimates for the sampled coordinate. We will provide a lower bound against such sketches as well.

**Definition 1.2.** *An $(\epsilon,\delta) - \ell_2$ sampling sketch with frequency estimation is an algorithm that observes a linear sketch $Sv$ of $v \in \mathbb{R}^n$ with $S : \mathbb{R}^n \to \mathbb{R}^m$, and chooses to either output* FAIL, *with probability at most $\frac{1}{2}$, or output a coordinate $\ell$. Conditioned on not outputting* FAIL, *the outputted coordinate must satisfy*

$$\Pr(\ell = i) = \frac{(1 \pm \frac{1}{2})v_i^2}{\|v\|^2}.$$

*Additionally, conditioned on not failing, the algorithm must out an estimate $\hat{v}_i$ satisfying*

$$\Pr(|\hat{v}_i - v_i| > \epsilon v_i) \leq \delta.$$

*Leverage scores.* We will use leverage scores several times throughout our arguments. The $i$th leverage score of an orthogonal projection $\Pi$ is by definition $\tau_i^2 := \|\Pi e_i\|^2$.The following is the well-known fact that the sum of leverage scores of a matrix equals its rank.

**Proposition 1.1.** *Let $\Pi \in \mathbb{R}^{n\times n}$ be an orthogonal projection of rank $m$. Then $\sum_{i=1}^n \|\Pi e_k\|^2 = m$.*

Proof. We have

$$\sum_{i=1}^n \|\Pi e_k\|^2 = \mathrm{Tr}(\Pi^T\Pi) = \mathrm{Tr}(\Pi^2) = \mathrm{Tr}(\Pi) = m.$$

□

*Heavy Hitters.* We will use the language of heavy-hitters throughout. Given a frequency vector $v$, we say that coordinate $i$ is $(\epsilon,\ell_p)$-heavy if $\|v_i\|^2 \geq \epsilon\|v\|_p^p$. Unless otherwise stated, $\epsilon$-heavy will always mean $(\epsilon,\ell_2)$-heavy.

### 1.3 Technical Overview

*$\ell_2$-sampling.* We show that an $\ell_2$ sampler requires a sketching dimension of at least $\Omega(\log^2 n)$. To do this, we first construct a game that can be solved by an $\ell_2$ sampler, and then show that any sketch which solves the game requires $\Omega(\log^2 n)$ sketching dimension.

To construct the game, first observe that an $\ell_2$ sampling sketch can be applied to find an $\Omega(1)$-heavy-hitter with constant probability. If $v_i^2 \geq \frac{1}{2}\|v\|_2^2$, then by definition, the sampler chooses coordinate $i$ with constant probability. Unfortunately this isn't enough to give an $\Omega(\log^2 n)$ lower bound, since CountSketch already finds heavy-hitters using $O(\log n)$ sketching dimension.

An $\ell_2$ sampler is more powerful though, and can find smaller "spikes", albeit with lower probability. To make this precise we take the input of our game to be $v = \mathcal{N}(0, I_n) + 2^{r/2}e_k$ where $k$ is uniform over coordinates and $r$ is uniform over $[\log n]$. At scale $r$, an $\ell_2$ sampler recovers coordinate $k$ with probability roughly $2^r/n$. To codify this in the game, we design the game to give a payout of $n/2^r$ when the chosen scale is $r$ and the sketch correctly identifies coordinate $k$. An $\ell_2$ sampler achieves expected payout of $\Omega(1)$ conditionally for each $r$, and therefore an $\Omega(1)$ expected payout overall. Intuitively, achieving a constant expected score for each $r$ should require a sketching dimension of $\Omega(\log n)$, and there are $\log n$ scales $r$, so this suggests a sketching lower bound of $\Omega(\log^2 n)$.

The situation at each scale is reminiscent of 1-sparse recovery which is known to have $\Omega(\log n)$ sketching lower bound [11]. One way to prove this is via the Shannon-Hartley theorem, and indeed this would generalize to show that obtaining constant expected payout at any fixed scale $r$ requires $\Omega(\log n)$ rows. Unfortunately, for us this does not give enough information about what the sketch must look like at each scale. For example, if one could just use the same sketch for each $r$, then the sketching dimension would not need to increase with the number of levels.

We show that the situation is different for different values of $r$. For recovering spike of size $\sqrt{n}$ planted within a standard Gaussian vector, a dense Gaussian sketch would suffice to attain constant probability of success. However for smaller spikes, our argument will show that it cannot. We roughly show that finding a spike of size $2^{r/2}$ with $n/2^r$ probability requires scaling up $n/2^r$ coordinates by a factor of $\sqrt{n}/2^{r/2}$, so that if scaled, the large coordinate becomes nearly $\Omega(1)$ heavy. In other words we show that the rescaling that all known $\ell_2$ sampling sketches perform is necessary for recovering smaller spikes.

The discussion above is for a fixed scale $r$. For each scale $r$, there are $n/2^{r/2}$ coordinates rescaled to $O(\sqrt{n})$. Since we do not know $r$ ahead of time, all of these rescalings must be performed simultaneously, and so we end up with 1 coordinate rescaled by $\sqrt{n}$, 2 coordinates rescaled by $\sqrt{n/2}$, 4 coordinates rescaled by $\sqrt{n/4}$, and so on. The net effect is that the rescaling scales the squared norm of the input vector by a factor of $\log n$. Thus at scale $r$, when rescaling produces a heavy-hitter, it is not $O(1)$ heavy, but rather $O(1/\log n)$ heavy. This is the source of our additional $\log n$ factor.

To describe our argument a bit more precisely, note that the sketching matrix can without loss of generality be taken an orthogonal projection $\Pi : \mathbb{R}^n \to \mathbb{R}^n$ where $\Pi$ has rank $m$, corresponding to the sketching dimension. By Yao's minimax principle [14], for fixed $m$, the expected payout of the game above is maximized for a deterministic choice of $\Pi$. In this picture, the coordinate-wise rescalings correspond to the *leverage scores* $\tau_i^2 := \|\Pi e_i\|^2$ of $\Pi$ [12]. Let $x_k^{(r)} = \Pi(2^{r/2}e_k)$ be the sketches of the possible spikes. Our key technical observation is as follows:

Lemma 1. *Let $v_1, \ldots, v_m$ be vectors in $\mathbb{R}^d$, with $\|v_i\| \leq \frac{1}{10}\sqrt{\ln m}$ for all $i$. Let $I$ be drawn uniformly from $[m]$, and let $X$ be distributed as $\mathcal{N}(v_I, I_d)$. The probability of recovering $I$ from an observation of $X$ is at most $1/poly(m)$.*

This implies that the spikes that project into the ball $B$ of radius $c\sqrt{\log n}$ cannot be differentiated from one another under the background $\mathcal{N}(0, I_m)$ noise. At each scale $r$, we must therefore have $n(2^r/n) = 2^r$ vectors that project outside of $B$. The squared norm of each projection is $\|x_k^{(r)}\|^2 = \tau_k^2 2^r$, so there must be $2^r$ leverage scores that are least $\log n/2^r$. This holds for all $r$, and we show that this implies that $m$, which is the sum of the leverage scores, is at least $\log^2 n$.

We point out here that our lemma has a short heuristic derivation. Suppose that we would like to be able to recover $I$ under the constraint that the $v_i$'s all lie in a ball of radius $\frac{1}{10}\log n$. The best-case scenario should plausibly be when the $v_i$'s are maximally separated, i.e., when they are at the vertices of a regular simplex in $m-1$ dimensions. Up to a constant-factor rescaling, this simplex is isometric to the set of $m$ coordinate vectors in $\mathbb{R}^m$, rescaled by $\frac{1}{10}\sqrt{\log n}$. But it is easy to understand this corresponding mixture of Gaussians. The MAP recovery algorithm simply chooses the coordinate vector corresponding to the largest coordinate $X$'. But the largest coordinate of $\frac{1}{10}\sqrt{\log n}e_k + \mathcal{N}(0, I_n)$ is typically not $e_k$, since a $\mathcal{N}(0, 1)$ random variable is at least $\frac{1}{10}\sqrt{\log n}$ with at least $n^{-0.5}$ probability, say. Unfortunately, this argument is difficult to make rigorous, so we supply a different argument in the section below.

*$\ell_2$-sampling with frequency estimation.* We also consider lower bounds for samplers that output a multiplicative approximation $(1 \pm \epsilon)$ with failure probability at most $\delta$ to the coordinate it samples. The idea is similar to frequency estimation, although the details are a bit different. Here we consider a distinguishing game where the player observes the sketch, is told $r$ and $k$ afterwards, and must decide whether the spike had size $2^{r/2}$ or $(1+\epsilon)2^{r/2}$. The game pays out 1 for a correct, and $-\frac{1}{\delta}$ to penalize incorrect answers. We observe that an $\ell_2$ sampler can achieve $\Omega(1)$ expected payout on this game. On the other hand, we show that achieving $\Omega(1)$ expected payout requires a sketching dimension of $\Omega(\frac{1}{\epsilon^2}\log n \log \frac{1}{\delta})$. The idea is that projection of $v_k e_k$ is distributed as either $\mathcal{N}(\|\Pi e_k\| v_k, 1)$ or $\mathcal{N}((1+\epsilon)\|\Pi e_k\| v_k, 1)$. We bound the number of samples needed to confidently distinguish between these Gaussians, and consequently bound the expected score of the game at each scale. Then we sum across coordinates $k$ and scales $r$, and carefully bound the resulting sum. This turns out to give the correct lower bound of $\Omega(\frac{1}{\epsilon^2}\log n \log \frac{1}{\delta})$ for for achieving an $\Omega(1)$ expected score in the above game.

*Bit lower bounds.* Finally we show how to apply the recent lifting framework of [5] to obtain bit lower bounds for sketches. The main idea from [5] that we apply is that under certain technical conditions on the sketching matrix (we can assume without loss generality), the sketch of a discrete Gaussian is close in total variation distance to that of a continuous Gaussian, rounded to the nearest point in the lattice induced by $A$. Since our games our unaffected by this small amount of running, this implies that the payout on discrete Gaussian inputs cannot be higher than for continuous Gaussian inputs. Thus we obtain a lower bound for discrete distributions as well, which generally require $\Omega(\log n)$ bits per entry on input vectors with polynomailly sized inputs.

## 2 Sampling Lower Bound

### 2.1 Main Probabilistic Lemma

The main idea is to bound the probability of recovering the sampled index from a gaussian mixture, when all means lie in a ball of radius $O(\sqrt{\log n})$. We start with two propositions that we use below.

**Proposition 2.1.** *Let $X \sim \mathcal{N}(0, I_d)$ be a random vector in $\mathbb{R}^d$, with probability density function (PDF) $f(x)$. Let $f_1(x)$ be the PDF of a $\mathcal{N}(v, I_d)$ distribution, where $v \in \mathbb{R}^d$ is a fixed non-zero vector. Then for* *any $r > 0$, the probability $\Pr(f_1(X)/f(X) \le r)$, where $X \sim \mathcal{N}(0, I_d)$, is given by*

$$\Pr\left(\frac{f_1(X)}{f(X)} \le r\right) = \Phi\left(\frac{\ln r + \frac{1}{2}\|v\|^2}{\|v\|}\right),$$

*where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution $\mathcal{N}(0,1)$ and $\|v\|$ denotes the Euclidean norm of $v$.*

Proof. The PDF of $X \sim \mathcal{N}(0, I_d)$ is

$$f(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}x^T x\right).$$

The PDF of a $\mathcal{N}(v, I_d)$ distribution is

$$f_1(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(x-v)^T(x-v)\right).$$

We are interested in the probability $\Pr(f_1(X)/f(X) \le r)$, where $X \sim \mathcal{N}(0, I_d)$. The ratio of the PDFs is

$$\begin{aligned}\frac{f_1(X)}{f(X)} &= \frac{\frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(X-v)^T(X-v)\right)}{\frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}X^T X\right)} \\ &= \exp\left(-\frac{1}{2}(X-v)^T(X-v) + \frac{1}{2}X^T X\right) \\ &= \exp(X^T v - \frac{1}{2}\|v\|^2).\end{aligned}$$

The probability we want to calculate is

$$\Pr\left(\exp\left(X^T v - \frac{1}{2}\|v\|^2\right) \le r\right) = \Pr(X^T v \le \ln r + \frac{1}{2}\|v\|^2).$$

Note that $Y := X^T v$ is distributed as $\mathcal{N}(0, \|v\|^2)$. We can standardize $Y$ by defining $W = Y/\|v\|$. Then $W \sim \mathcal{N}(0,1)$. The probability becomes

$$\Pr\left(W\|v\| \le \ln r + \frac{1}{2}\|v\|^2\right) = \Pr\left(W \le \frac{\ln r + \frac{1}{2}\|v\|^2}{\|v\|}\right),$$

from which the claim follows by the definition of $\Phi$. □

**Proposition 2.2.** *Let $\mathcal{D}_1, \ldots, \mathcal{D}_m$ be probability distributions on $\mathbb{R}^d$ with PDFs $f_1, \ldots, f_m$. Let $I$ be uniform on $[n]$ and let $X$ be drawn from $\mathcal{D}_I$. Suppose that for all $i$*

$$\Pr_{Z \sim \mathcal{D}_i}\left(\frac{f_i(Z)}{f_1(Z) + \ldots + f_m(Z)} \ge p\right) \le \delta.$$

*Let $\hat{I} : \mathbb{R}^d \to [m]$ be fixed. Then*

$$\Pr(\hat{I}(X) = I) \le \delta + p.$$

Proof. The posterior probability of $I = k$ given $X = x$ is $q_k(x) = \Pr(I = k | X = x)$. Using Bayes' theorem and $\Pr(I = k) = 1/m$:

$$q_k(x) = \frac{\Pr(X = x | I = k)\Pr(I = k)}{\Pr(X = x)} = \frac{f_k(x)(1/m)}{\sum_{j=1}^m f_j(x)(1/m)} = \frac{f_k(x)}{\sum_{j=1}^m f_j(x)}.$$

Let $S(x) = \sum_{j=1}^m f_j(x)$. So $q_k(x) = f_k(x)/S(x)$.

Let $A_k = \{x \mid \hat{I}(x) = k\}$ be the set of $x$ values for which we estimate $I$ to be $k$. Let $B_k = \{x \mid q_k(x) > p\}$. The probability of correctly recovering $I$ is:

$$\Pr(\hat{I}(X) = I) = \frac{1}{m}\sum_{k=1}^{m}\int_{A_k} f_k(x)dx = \frac{1}{m}\sum_{k=1}^{m}\left(\int_{A_k\cap B_k} f_k(x)dx + \int_{A_k\cap B_k^c} f_k(x)dx\right)$$

To bound the first sum,

$$\frac{1}{m}\sum_{k=1}^{m}\int_{A_k\cap B_k} f_k(x)dx \le \frac{1}{m}\sum_{k=1}^{m}\int_{B_k} f_k(x)dx \le \delta,$$

by the stated condition.

For the second sum, note that $f_k(x) \le pS(x)$ for all $x \in B_k^c$. Therefore

$$\begin{aligned}
\frac{1}{m}\sum_{k=1}^{m}\int_{A_k\cap B_k^c} f_k(x)dx &\le \frac{1}{m}\sum_{k=1}^{m}\int_{A_k\cap B_k^c} pS(x)dx \\
&= p\sum_{k=1}^{m}\int_{A_k\cap B_k^c} \frac{S(x)}{m}dx \\
&= p\sum_{k=1}^{m}\int_{A_k\cap B_k^c} f_X(x)dx \\
&= p\int_{\cup_k(A_k\cap B_k^c)} f_X(x)dx,
\end{aligned}$$

which is at most $p$ as desired.

□

Lemma 2. *Let $v_1, \ldots, v_m$ be vectors in $\mathbb{R}^d$, with $\|v_i\| \le \frac{1}{10}\sqrt{\ln m}$ for all i. Let I be drawn uniformly from $[m]$, and let X be distributed as $\mathcal{N}(v_I, I_d)$. The probability of recovering I from an observation of X is at most $2m^{-0.5}$.*

Proof. Fix $i$ and suppose that $x$ is distributed according to $\mathcal{N}(v_i, I_d)$. Let $f_i$ be corresponding PDF. By Proposition 2.1, we have

$$\begin{aligned}
\Pr\left(\frac{f_j(x)}{f_i(x)} \le m^{-0.5}\right) &= \Phi\left(\frac{\ln(m^{-0.5}) + \frac{1}{2}\|v_i - v_j\|^2}{\|v_i - v_j\|}\right) \\
&= \Phi\left(\frac{-0.5\ln(m) + \frac{1}{2}\|v_i - v_j\|^2}{\|v_i - v_j\|}\right) \\
&\le \Phi\left(\frac{-0.5\ln(m) + \frac{1}{50}\ln m}{\frac{1}{5}\sqrt{\ln m}}\right) \\
&= \Phi(-2.4\sqrt{\ln m}) \\
&\le \exp((-2.4\sqrt{\ln m})^2/2) \\
&= m^{-2.88}.
\end{aligned}$$

Here we bounded $\|v_i - v_j\|$ by $\|v_i\| + \|v_j\| \le \frac{1}{5}\sqrt{\ln m}$, and used that the expression in the second line is monotonic in $\|v_i - v_j\|$.

It follows from this that

$$\Pr\left(\frac{f_i(x)}{f_1(x)+\ldots+f_m(x)} \geq m^{-0.5}\right) = \Pr\left(\frac{f_1(x)+\ldots+f_m(x)}{f_i(x)} \leq m^{0.5}\right)$$
$$\leq \Pr\left(\min_j \frac{f_j(x)}{f_i(x)} \leq m^{-0.5}\right)$$
$$\leq m^{-1.88},$$

where we applied a union bound in the last step. Then by Proposition 2.2, the probability of recovering $I$ is at most $m^{-0.5} + m^{-1.88} \leq 2m^{-0.5}$.

□

## 2.2 Lower Bound for $\ell_2$ Samplers

The problem we consider is identifying the location of a planted spike in a $\mathcal{N}(0, I_d)$ random vector. The size of the spike is chosen to be a random power of $\sqrt{2}$ between 1 and $\sqrt{n}$. When the spike is of size $R$, we would like to formulate the problem so that the index of the spike must be recovered with probability $R^2/n$. However we would like to use Yao's principle to reduce to deterministic sketches. A probability guarantee depending on the choice of scale does not directly allow for applying Yao's principle. So instead, we introduce a game with a payoff scheme, where the payoff is inversely proportional to the success probabilities that we would like to enforce at each scale.

**Problem 2.1.** *Let $n$ be a power of 2. Let $r$ be drawn uniformly from $[\log n]$, and let $k$ be drawn uniformly from $[n]$. We define the distribution $\mathcal{D}$ as the distribution of $N(0, I_n) + 2^{r/2}e_k$. Given a linear sketch $\Pi v$ of a sample $v \sim \mathcal{D}$, the goal is to output the coordinate $k$. Correctly outputting the coordinate $k$ earns a score of $\max(n/2^r, n^{1/4})$.*

First we observe that an $\ell_2$ sampling sketch gets a good score on this game.

**Proposition 2.3.** *An $\ell_2$-sampling sketch that samples coordinate $i$ with probability $(1 \pm \frac{1}{2})\frac{f_i^2}{\|f\|_2^2}$ achieves a score of $\Theta(1)$ in Problem 2.1.*

Proof. Suppose that the spike has size $2^{r/2}$. Note that the norm of a $\mathcal{N}(0, I_n)$ sample is $O(\sqrt{n})$ with good probability, so the sampler chooses coordinate $k$ with probability $\Omega(2^r/n)$ by the $\ell_2$ sampling guarantee. It follows that the expected payoff is $\Omega(1)$ conditioned on $n/2^r > n^{1/4}$, which occurs with constant probability. So the overall expected payoff is $\Omega(1)$. □

**Theorem 2.1.** *A randomized sketch for Problem 2.1 that achieves a score of $\Theta(1)$ must have sketching dimension at least $\Omega(\log^2 n)$.*

Proof. By Yao's principle, it suffices to show the lower bound for a deterministic sketch and recovery strategy. Without loss of generality, suppose that the sketch is an orthogonal projection $\Pi$ of rank $m$.

Let $v_k^{(r)} = 2^{r/2}\Pi e_k$, be the projection of the $k$-th spike at scale $r$. Note that we observe $\mathcal{N}(0, I) + v$ where $v$ is uniform from the set $\{v_k^{(r)}\}$.

Let $P_{kr}$ be the probability of winning conditioned on fixed values for $k, r$. Let $Q$ be the random variable representing the payout of our strategy. Let $B = \{v_k^{(r)} : \|v_k^{(r)}\| \leq \frac{1}{10}\sqrt{\log n}\}$. For fixed $r$,

$$\sum_k \|v_k^{(r)}\|^2 = \sum_k 2^r \|\Pi e_k\|^2 = 2^r m.$$

So at most $\frac{100 \cdot 2^r m}{\log n}$ of these lie outside the ball of radius $\frac{1}{10}\sqrt{\log r}$. This implies that $|B^c| \leq \frac{200mn}{\log n}$. So either $m \geq \Omega(\log^2 n)$, in which case we are done, or $|B| \geq n$.

By Lemma 2 the probability of correctly recovering $k$ and $r$ from the sketch conditioned on $v_k^{(r)} \in B$ is at most $2n^{-0.5}$ for any strategy. To win however, we only need to correctly identify $k$. There are at most $\log n$ values of $r$ though, so

$$\Pr(\text{win}|v_k^{(r)} \in B) \leq 2n^{-0.5} \log n$$

since a strategy that recovers $k$ can recover $r$ with $\frac{1}{\log n}$ probability, simply by uniform guessing.

Since the maximum payout is $n^{1/4}$, we have

$$\frac{1}{|B|} \sum_{v_k^{(r)} \in B} \frac{n}{2^r} P_{kr} = \mathbb{E}(Q|v_k^{(r)} \in B) \leq n^{1/4}(2n^{-0.5} \log n) \leq 2n^{-0.25} \log n.$$

Then we have

$$\begin{aligned}
\mathbb{E}(Q) &= \frac{4}{n \log n} \sum_{k,r} \left(\frac{n}{2^r}\right) P_{kr} \\
&= \frac{4}{n \log n} \sum_{v_k^{(r)} \in B} \left(\frac{n}{2^r}\right) P_{kr} + \frac{4}{n \log n} \sum_{v_k^{(r)} \notin B} \left(\frac{n}{2^r}\right) P_{kr} \\
&\leq \frac{4}{n \log n} |B| (2n^{-0.25} \log n) + \frac{4}{n \log n} \sum_{v_k^{(r)} \notin B} \left(\frac{n}{2^r}\right) P_{kr} \\
&\leq 2n^{-0.25} \log n + \frac{4}{n \log n} \sum_{v_k^{(r)} \notin B} \left(\frac{n}{2^r}\right) P_{kr}.
\end{aligned}$$

We are left with bounding the remaining sum. To this end we write

$$\begin{aligned}
\frac{4}{n \log n} \sum_{v_k^{(r)} \notin B} \left(\frac{n}{2^r}\right) P_{kr} &= \frac{4}{n \log n} \sum_{k,r} \mathbb{1}\left[\tau_k^2 2^r \geq \frac{1}{100} \log n\right] \left(\frac{n}{2^r}\right) P_{kr} \\
&\leq \frac{4}{\log n} \sum_{k,r} \mathbb{1}\left[\tau_k^2 2^r \geq \frac{1}{100} \log n\right] \left(\frac{1}{2^r}\right) \\
&= \frac{4}{\log n} \sum_{r} \frac{1}{2^r} \sum_{k} \mathbb{1}\left[\tau_k^2 2^r \geq \frac{1}{100} \log n\right] \\
&= \frac{4}{\log n} \sum_{r} \frac{1}{2^r} \#\{k : \tau_k^2 \geq \frac{1}{100} \frac{\log n}{2^r}\} \\
&= \frac{4}{\log n} \sum_{r} \frac{L_r}{2^r},
\end{aligned}$$

where we set $L_r = \#\{k : \tau_k^2 \geq \frac{1}{100} \frac{\log n}{2^r}\}$. Note that

$$L_{r+1} - L_r = \#\{k : \tau_k^2 \in \left[\frac{1}{100} \frac{\log n}{2^{r+1}}, \frac{1}{100} \frac{\log n}{2^r}\right)\},$$

so

$$m = \sum_{i=1}^{n} \tau_i^2 \geq \sum_{r=-\infty}^{\infty} (L_{r+1} - L_r) \frac{1}{100} \frac{\log n}{2^{r+1}}$$
$$= \frac{\log n}{100} \left( \sum_{r=-\infty}^{\infty} \frac{L_{r+1}}{2^{r+1}} - \frac{1}{2} \sum_{r=-\infty}^{\infty} \frac{L_r}{2^r} \right)$$
$$= \frac{\log n}{200} \sum_{r=-\infty}^{\infty} \frac{L_r}{2^r}.$$

It follows that $\sum_{r=-\infty}^{\infty} \frac{L_r}{2^r} \leq \frac{200m}{\log n}$. Plugging in to our calculations above gives

$$\mathbb{E}(Q) \leq 2n^{-0.25} \log n + \frac{4}{\log n} \frac{200m}{\log n} = 2n^{-0.25} \log n + \frac{800m}{\log^2 n},$$

and the result follows. □

## 3 Frequency Estimation

We consider the following problem.

**Problem 3.1.** *Let $n$ be a power of 2. Let the distribution $\mathcal{D}$ be defined as $\mathcal{N}(0, I_n) + (1 + z\epsilon)2^{r/2}e_k$ where $k$ is uniform on $[n]$, $r$ is uniform on $[\log n]$, and $z$ is uniform on $\{0, 1\}$.*

*Alice chooses an orthogonal projection $\Pi$, and observes $\Pi v$ where $v \sim \mathcal{D}$. Alice is then told $r$ and $k$, and must choose to either pass or output a guess $\hat{z}$ of $z$. If Alice passes, she receives a score of 0. If $\hat{z} = z$ she receives a score of $n/2^r$, and if $\hat{z} \neq z$ she receives a score of $-\alpha n/2^r$.*

Next we show that an $\ell_2$ sampler with frequency estimation can get a good expected score on this game.

Lemma 3. *Let $\Pi$ is an $\ell_2$-sampling sketch that passes with probability at most $1/2$ and conditioned on not passing, samples item $i$ with probability $(1 \pm \frac{1}{2}) \frac{f_i}{\|f\|_2^2}$, and conditioned on sampling item $i$ outputs a frequency estimate $\hat{f}_i$ satisfying $\hat{f}_i = (1 \pm \epsilon) f_i$ with probability at least $1 - \delta$.*

*Given such a sampler, Alice can achieve an expected score of $\Omega(1 - (\alpha + 1)\delta)$ on Problem 3.1.*

Proof. Alice simply uses the sketch to perform $\ell_2$ sampling on the frequency vector. If the sampler fails, then she passes. Otherwise she checks if her sample matches $k$, if it does she uses the sampler to compute a $(1 \pm \epsilon)$ multiplicative estimate to $v_k$.

The probability that she samples coordinate $k$ is $\Theta(2^r/n)$. Given this, the probability that she correctly guesses $z$ is at least $1 - \delta$ by correctness of the sampler. Thus up to constants, her expected score is

$$\frac{2^r}{n} [(1 - \delta) \frac{n}{2^r} + \delta(-\frac{\alpha n}{2^r})],$$

as claimed. □

We will use the following simple bound below.

**Proposition 3.1.** *We have the bound $\exp(-1/x) \leq \min(1, x^2)$ for $x \geq 0$.*

Proof. It is clear that $\exp(-1/x) \leq 1$. For the other bound it suffices to show that $f(x) := x^2 \exp(1/x) \geq 1$. Note that $f'(x) = (2x - 1) \exp(1/x)$, and so $f$ is minimized when $x = 1/2$ where it achieves a minimum of $(e/2)^2 > 1$. □

The following propositions bounds the probability that we can confidently identify the part of a Gaussian mixture that a sample arises from.

**Proposition 3.2.** *Let $\mathcal{D}_0$ and $\mathcal{D}_1$ be $\mathcal{N}(0,1)$ and $\mathcal{N}(-T,1)$ respectively where $T \geq 0$. Suppose that $k$ is uniform on $\{0,1\}$ and that $X$ is drawn from the mixture distribution $\frac{1}{2}\mathcal{D}_0 + \frac{1}{2}\mathcal{D}_1$. Let $\mathcal{E}$ be the event that $\{\Pr(k=1|X=t) \geq \beta$ or $\Pr(k=0|X=t) \geq \beta\}$ or in other words that the MAP estimate for $k$ has $\beta$ confidence. Then*

$$\Pr(\mathcal{E}) \leq \mathbb{1}_{T \geq \sqrt{C_\beta}} + \min(1, 128T^4/C_\beta^4),$$

*where $C_\beta = \log \frac{\beta}{1-\beta}$.*

Proof. The posterior probability of $k=0$ conditioned on observing $X=t$ is

$$\Pr(k=0|X=t) = \frac{e^{-t^2/2}}{e^{-t^2/2} + e^{-(t+T)^2/2}}.$$

This quantity is at least $\beta$ precisely when

$$e^{-Tt-T^2/2} \leq \frac{1}{\beta} - 1,$$

or equivalently when

$$t \geq -\frac{T}{2} + \frac{1}{T}\log(\frac{\beta}{1-\beta}) := -\frac{T}{2} + \frac{C_\beta}{T}.$$

To bound this, consider two cases. If $C_\beta/T \geq 2(T/2)$ (or equivalently when $T \leq \sqrt{C_\beta}$),

$$\Pr(X \geq -\frac{T}{2} + \frac{C_\beta}{T}) \leq \Pr(X \geq \frac{C_\beta}{2T}) \leq \exp(-C_\beta^2/(8T^2)),$$

by the Chernoff tail bound.

In the second case where $T \geq \sqrt{C_\beta}$, we simply bound $\Pr(X \geq -\frac{T}{2} + \frac{C_\beta}{T}) \leq 1$. Combining these bounds and considering both $k=0$ and $k=1$ gives

$$\Pr(\mathcal{E}) \leq \mathbb{1}_{T \geq \sqrt{C_\beta}} + 2\exp(-C_\beta^2/(8T^2)),$$

and combining with Proposition 3.1 yields the claim. □

Lemma 4. *Let $\tau_i^2 = \|\Pi e_i\|_2^2$, and set $f_\beta(t) = \mathbb{1}_{T \geq \sqrt{C_\beta}} + \min(1, 128T^4/C_\beta^4)$, where $C_\beta$ is as defined above. Then if Alice uses the sketch $\Pi$, her expected payout is at most*

$$\frac{1}{n\log_2 n}\sum_{r=1}^{\log_2 n}\sum_{i=1}^{n}\frac{n}{2^r} f_{\frac{\alpha}{1+\alpha}}(\epsilon 2^{r/2}\|\Pi e_i\|).$$

Proof. Note that Alice observes $\Pi(1+z\epsilon)2^{r/2}e_k) + g$ where $g$ is the projection of a $\mathcal{N}(0, I_n)$ random variable onto $\mathrm{Im}(\Pi) = S$. By choosing an orthonormal basis of $S$ containing $\Pi(1+z\epsilon)2^{r/2}e_k)$, this is equivalent to Alice distinguishing between $\mathcal{N}(0,1)$ and $\mathcal{N}(\epsilon 2^{r/2}\|\Pi e_k\|, 1)$.

If Alice can only estimate $z$ with confidence less than $\frac{\alpha}{1+\alpha}$, then she should pass on her turn, otherwise her expected payout is negative. However, conditioned on $r$ and $k$ the probability that Alice observes a sample that causes her to be $\frac{\alpha}{1+\alpha}$ confident is at most

$$p_{r,k} := f_{\frac{\alpha}{1+\alpha}}(\epsilon 2^{r/2}\|\Pi e_k\|)$$

by Proposition 3.2. Then by the payout rule of the game, Alice's expect payout is at most

$$\frac{1}{n\log_2 n}\sum_{r=1}^{\log_2 n}\sum_{i=1}^{n} p_{r,i}\frac{n}{2^r} = \frac{1}{n\log_2 n}\sum_{r=1}^{\log_2 n}\sum_{i=1}^{n}\frac{n}{2^r} f_{\frac{\alpha}{1+\alpha}}(\epsilon 2^{r/2}\|\Pi e_i\|).$$

□

The final step is to bound the sum above.

Lemma 5. *Using our previous notation, we have the bound*

$$\frac{1}{n\log_2 n}\sum_{r=1}^{\log_2 n}\sum_{i=1}^{n}\frac{n}{2^r}f_{\frac{\alpha}{1+\alpha}}(\epsilon 2^{r/2}\|\Pi e_i\|) \leq \frac{8}{\log_2 n}\sum_{i=1}^{n}\frac{\sqrt{128}\epsilon^2\tau_i^2}{C_{\alpha/(1+\alpha)}^2}+\frac{2}{\log n}\sum_{i=1}^{n}\frac{\epsilon^2}{C_{\alpha/(1+\alpha)}}\tau_i^2.$$

Proof. We consider the two terms of $f$ separately.

*First sum.* We first bound

$$\frac{1}{n\log_2 n}\sum_{r=1}^{\log_2 n}\sum_{i=1}^{n}\frac{n}{2^r}\min\left(1,\frac{128\epsilon^4 2^{2r}\tau_i^4}{C_{\alpha/(1+\alpha)}^4}\right).$$

To cut down on notation, set $\sigma_i^4=\frac{128\epsilon^4\tau_i^4}{C_{\alpha/(1+\alpha)}^4}$, so that we wish to bound

$$\frac{1}{\log_2 n}\sum_{r=1}^{\log_2 n}\frac{1}{2^r}\sum_{i=1}^{n}\min(1,2^{2r}\sigma_i^4).$$

We divide into dyadic sets. Let $B_k=\{i:\sigma_i^2\in(\frac{1}{2^{k+1}},\frac{1}{2^k}]\}$. Then

$$\sum_{i=1}^{n}\min(1,2^{2r}\sigma_i^4)\leq\sum_{k=0}^{r-1}|B_k|+\sum_{k=r}^{\infty}|B_k|(\frac{1}{2^k})^2 2^{2r}=\sum_{k=0}^{r-1}|B_k|+\sum_{k=r}^{\infty}\frac{1}{4^{k-r}}|B_k|.$$

Then we calculate

$$\begin{aligned}
\sum_{r=1}^{\log_2 n}\frac{1}{2^r}\sum_{i=1}^{n}\min(1,2^{2r}\sigma_i^4) &\leq \sum_{r=1}^{\log n}\frac{1}{2^r}\left(\sum_{k=0}^{r-1}|B_k|+\sum_{k=r}^{\infty}\frac{1}{4^{k-r}}|B_k|\right)\\
&=\sum_{r=1}^{\log n}\frac{1}{2^r}\sum_{k=0}^{\infty}\mathbb{1}_{k\leq r-1}|B_k|+\sum_{r=1}^{\log n}\frac{1}{2^r}\sum_{k=0}^{\infty}\mathbb{1}_{k\geq r}\frac{1}{4^{k-r}}|B_k|\\
&=\sum_{k=0}^{\infty}|B_k|\left(\sum_{r=1}^{\log n}\frac{1}{2^r}\mathbb{1}_{k\leq r-1}+\sum_{r=1}^{\log n}\frac{1}{2^r}\mathbb{1}_{k\geq r}\frac{1}{4^{k-r}}\right)\\
&=\sum_{k=0}^{\infty}|B_k|\left(\sum_{r=1}^{k}\frac{1}{2^{2k-r}}+\sum_{r=k+1}^{\log n}\frac{1}{2^r}\right)\\
&\leq\sum_{k=0}^{\infty}|B_k|\left(\frac{1}{2^{k-1}}+\frac{1}{2^k}\right)\\
&\leq 8\sum_{k=0}^{\infty}\frac{1}{2^{k+1}}|B_k|\\
&\leq 8\sum_{i=1}^{n}\sigma_i^2.
\end{aligned}$$

This implies that

$$\frac{1}{n\log_2 n}\sum_{r=1}^{\log_2 n}\sum_{i=1}^{n}\frac{n}{2^r}\min\left(1,\frac{128\epsilon^4 2^{2r}\tau_i^4}{C_{\alpha/(1+\alpha)}^4}\right)\leq\frac{8}{\log_2 n}\sum_{i=1}^{n}\sigma_i^2=\frac{8}{\log_2 n}\sum_{i=1}^{n}\frac{\sqrt{128}\epsilon^2\tau_i^2}{C_{\alpha/(1+\alpha)}^2}.$$

*Second sum.* Set $\beta = \frac{\alpha}{1+\alpha}$. We would like to bound

$$\frac{1}{n\log_2 n}\sum_{r=1}^{\log_2 n}\sum_{i=1}^{n}\frac{n}{2^r}\mathbb{1}_{\epsilon^2 2^r\|\Pi e_i\|^2\geq C_\beta} = \frac{1}{\log_2 n}\sum_{r=1}^{\log_2 n}\frac{1}{2^r}\#\{i : \frac{\epsilon^2}{C_\beta}\|\Pi e_i\|^2 \geq \frac{1}{2^r}\}.$$

To bound this sum, let $D_k = \{i : \frac{\epsilon^2}{C_\beta}\|\Pi e_i\|^2 \in (\frac{1}{2^{k+1}}, \frac{1}{2^k}]\}$. Then the sum above is at most

$$\begin{aligned}\frac{1}{\log_2 n}\sum_{r=0}^{\log_2 n}\frac{1}{2^r}(|D_0| + \ldots + |D_r|) &\leq \frac{1}{\log_2 n}(2|D_0| + |D_1| + \frac{1}{2}|D_2| + \ldots)\\ &\leq \frac{2}{\log n}\sum_{i=1}^{n}\frac{\epsilon^2}{C_\beta}\|\Pi e_i\|^2\end{aligned}$$

□

Putting the pieces together gives our main result of this section.

**Theorem 3.1.** *An algorithm that performs $\ell_2$ sampling with frequency estimation requires $\Omega(\frac{1}{\epsilon^2}\log n \log\frac{1}{\delta})$ space. Here $\delta$ is the probability that the frequency estimate is not a $(1\pm\epsilon)$ multiplicative approximation.*

Proof. Assume $\delta$ is bounded by fixed constant, say $1/4$ and set $\alpha = \frac{1}{2\delta} - 1$ in Problem 3.1 so that by Lemma 3 allows Alice to achieve an expected score of $\Omega(1)$.

On the other hand, Lemma 4 and Lemma 5 give a bound on the expected payout of

$$\frac{8}{\log_2 n}\sum_{i=1}^{n}\frac{\sqrt{128}\epsilon^2\tau_i^2}{C_{\alpha/(1+\alpha)}^2} + \frac{2}{\log_2 n}\sum_{i=1}^{n}\frac{\epsilon^2}{C_{\alpha/(1+\alpha)}}\tau_i^2.$$

Let the sketching dimension be $m$. Combining with 1.1, and noting that $C_{\alpha/(1+\alpha)} = \log\alpha = \Omega(\log\frac{1}{\delta})$, this bound is

$$O\left(\frac{\epsilon^2}{\log n\log^2\frac{1}{\delta}}m + \frac{\epsilon^2}{\log n\log\frac{1}{\delta}}m\right) = O\left(\frac{\epsilon^2}{\log n\log\frac{1}{\delta}}m\right).$$

So $m$ must be at least $\Omega(\frac{1}{\epsilon^2}\log n\log\frac{1}{\delta})$ as claimed. □

## 4 Bit Lower Bounds

First recall the definition of the discrete Gaussian.

**Definition 4.1.** *For $\mu \in \mathbb{R}^n$ and PSD $\Sigma \in \mathbb{R}^{n\times n}$, the discrete Gaussian $\mathcal{D}(\mu,\Sigma)$ on $\mathbb{Z}^n$ is the distribution on $\mathbb{Z}^n$ such that there is a normalizing constant $\lambda$ with $\Pr_{x\sim\mathcal{D}(\mu,\Sigma)}(x = y) = \lambda f_{\mathcal{N}(\mu,\Sigma)}(y)$, where $f_{\mathcal{N}(\mu,\Sigma)}$ is the density function for a (continuous) normal distribution. In other words the discrete Gaussian is a distribution on $\mathbb{Z}^n$ with density proportional to that of the corresponding continuous Gaussian over its support.*

We recall the following result from [5]. As the next lemma will remove the first condition, we defer to [5] for a definition of $\lambda^\perp(A)$.

**Theorem 4.1.** *(Lemma 5.3 of [5]) Let $C > 0$ be a fixed constant, and let $S \in \mathbb{R}^{n\times n}$ be a full rank matrix. Let $A \in \mathbb{R}^{r\times n}$ be a matrix with rational entries and $r \leq n$ such that:*

(1) $\lambda^\perp_{max}(A) \leq \frac{\sigma_n(S)}{10C\log n}$
(2) $n^{6C} \geq \sigma_1(S) \geq \sigma_n(S) \geq n^{5C+3}$

*Let $\mathcal{D}_1$ be the distributions of $Ax$ where $x \sim \mathcal{D}(0, S^TS)$ and let $\mathcal{D}_2$ be the distribution of $y$ where $y$ is obtained by rounding $Ax$ for a sample $x \sim \mathcal{N}(0, S^TS)$ to the lattice point of its unit cell in $A\mathbb{Z}^n$. Then $\mathcal{D}_1$ and $\mathcal{D}_2$ differ by at most $\frac{1}{n^C}$ in total variation distance.*

The first condition in Theorem 4.1 turns out not to be important, as we can assume it without loss of generality by increasing the size of the sketch by at most a constant factor.

Lemma 6. *(Lemma 3.1 of [5]) Let $A \in \mathbb{Z}^{r\times n}$ for $r \le 0.25n$ be a full rank integer matrix with entries bounded by $M$ for some $M \ge n$. There exists a pre-processing that adds rows to $A$ to form a matrix $A' \in \mathbb{Z}^{m\times n}$ for $m \le 4r$ such that $\lambda_{\max}(\mathcal{L}^{\perp}(A') \le \sqrt{nM^2}$.*

The next lemma shows that we may take $A$ to have integer entries bounded by $n$ in magnitude. This is convenient because it means we do not have to worry about the size of $A$'s entries being chosen adaptively to the length of the stream.

Lemma 7. *A sketching matrix $A \in \mathbb{R}^{m\times n}$ with integer entries bounded in magnitude by $n^{\ell}$ may be simulated by a sketching matrix $A' \in \mathbb{R}^{m\ell\times n}$ with integer entries bounded in magnitude by $n$. Moreover, storing $A'v$ is no more expensive that storing $Av'$ as integer vectors, when $v$ when the entries of $v$ are constrained to be at most $n^c$ for a constant $c$.*

Proof. To construct $A'$ simply write each element of $A$ as its $n$-ary expansion. For each place value $k$, construct a new matrix $A^{(k)}$ where $A^{(k)}_{ij}$ is the $k$th $n$-ary digit of $A_{ij}$. Then let $A'$ be the vertical concatenation of the matrices $A^{(k)}$.

Note that the entries of $Av$ can be as large as $n^{c+\ell}$, so storing $Av$ requires $\Omega(m(1+\ell)\log n)$ space. On the other hand, the entries of $A'v$ are bounded by $n^{c+1}$ so $A'v$ may be stored with $O(ml\log n)$ space. □

**Theorem 4.2.** *A linear sketch for $\ell_2$ sampling requires at least $\Omega(\log^3 n)$ space to store when $v$ has poly$(n)$ sized entries. Similarly, a linear sketch for $(\epsilon,\delta)$-$\ell_2$-sampling with frequency estimation requires at least $\Omega(\frac{1}{\epsilon^2}\log^2 n\log\frac{1}{\delta})$ bits.*

Proof. We give the proof for $\ell_2$ sampling sketches, as the proof for frequency estimation is nearly identical, except that we use Problem 3.1 instead. From the lemma above, we may assume that the entries of $A$ are integers bounded by $n$ in magnitude. Consider the distribution $\mathcal{D}_0$ defined to be $\mathcal{D}(0, n^{2c}I_n) + n^c 2^{r/2}e_k$ for a constant $c$, where $r, k$ are distributed as in Problem 2.1, and where the payout scheme is the same as in Problem 2.1. Note that this is just a rescaled and discretized version of Problem 2.1. Also let $\mathcal{D}_1$ be the continuous version of this distribution: $\mathcal{N}(0, n^{2c}I_n) + n^c\lfloor 2^{r/2}\rfloor e_k$ where $r$ and $k$ are as above. Also let $\mathcal{D}_{i,r,k}$ denote the distribution $\mathcal{D}_i$ conditioned on the values of $r$ and $k$.

Let $g$ be the (deterministic) recovery function such that $g(Av)$ is the output of the $\ell_2$ sampler on the input vector $v$. By Lemma 6, we may assume without loss of generality that condition 1 of Theorem 4.1 holds, and we may ensure that condition 2 holds by choosing the constant $c$ appropriately.

Now, let $A\mathcal{D}_i$ be the distribution gotten by applying $A$ to a sample from $\mathcal{D}_i$, and let round$(\cdot)$ be the function that rounds to the nearest lattice point of the lattice defined by $A$.

By Theorem 4.1, we have that $\mathrm{TVD}(\mathrm{round}(A\mathcal{D}_{1,r,k}), A\mathcal{D}_{0,r,k}) \le n^{-C}$, and therefore

$$\mathrm{TVD}(g(\mathrm{round}(A\mathcal{D}_{1,r,k})), g(A\mathcal{D}_{0,r,k})) \le n^{-C},$$

for all $r, k$. Since the payout is bounded by $n$, this means that for any $r, k$ the expected payout differs by at most $1/\mathrm{poly}(n)$ between the two distributions. So if there is an algorithm to obtain expected payout $\Omega(1)$ on the second distribution, then it must be possible to obtain $\Omega(1)$ expected payout on the first distribution. So by our lower bound Theorem 2.1, $A$ must have $\Omega(\log^2 n)$ rows, and so storing the sketch $Av$ requires $\Omega(\log^3 n)$ bits. □

## Acknowledgments

Both authors were supported in part by Office of Naval Research award number N000142112647 and a Simons Investigator Award.

## References

[1] Yuqing Ai, Wei Hu, Yi Li, and David P. Woodruff. [n. d.]. New Characterizations in Turnstile Streams with Applications. In *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan (LIPIcs, Vol. 50)*, Ran Raz (Ed.). 20:1–20:22.

[2] Alexandr Andoni, Robert Krauthgamer, and Krzysztof Onak. 2011. Streaming algorithms via precision sampling. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*. IEEE, 363–372.

[3] Khanh Do Ba, Piotr Indyk, Eric Price, and David P. Woodruff. 2010. Lower Bounds for Sparse Recovery. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, Moses Charikar (Ed.). SIAM, 1190–1197.

[4] Petros Drineas, Malik Magdon-Ismail, Michael W. Mahoney, and David P. Woodruff. 2011. Fast approximation of matrix coherence and statistical leverage. *CoRR* abs/1109.3843 (2011).

[5] Elena Gribelyuk, Honghao Lin, David P Woodruff, Huacheng Yu, and Samson Zhou. 2025. Lifting linear sketches: Optimal bounds and adversarial robustness. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*. 395–406.

[6] Rajesh Jayaram and David Woodruff. 2021. Perfect l_p sampling in a data stream. *SIAM J. Comput.* 50, 2 (2021), 382–439.

[7] Hossein Jowhari, Mert Sağlam, and Gábor Tardos. 2011. Tight bounds for lp samplers, finding duplicates in streams, and related problems. In *Proceedings of the thirtieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. 49–58.

[8] John Kallaugher and Eric Price. 2020. Separations and equivalences between turnstile streaming and linear sketching. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy (Eds.). ACM, 1223–1236.

[9] Yi Li, Huy L Nguyen, and David P Woodruff. 2014. Turnstile streaming algorithms might as well be linear sketches. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. 174–183.

[10] Morteza Monemizadeh and David P Woodruff. 2010. 1-pass relative-error lp-sampling with applications. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, 1143–1160.

[11] Eric Price and David P Woodruff. 2012. Applications of the shannon-hartley theorem to data streams and sparse recovery. In *2012 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2446–2450.

[12] David P Woodruff et al. 2014. Sketching as a tool for numerical linear algebra. *Foundations and Trends® in Theoretical Computer Science* 10, 1–2 (2014), 1–157.

[13] David P. Woodruff, Shenghao Xie, and Samson Zhou. 2025. Perfect Sampling in Turnstile Streams Beyond Small Moments. arXiv:2504.07237 [cs.DS] https://arxiv.org/abs/2504.07237

[14] Andrew Chi-Chin Yao. 1977. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*. IEEE Computer Society, 222–227.

Received December 2025; accepted February 2026