Improved Differentially Private Euclidean Distance Approximation

Nina Mesing Stausholm
nimn@itu.dk
IT University of Copenhagen
BARC

Abstract

This work shows how to privately and more accurately estimate Euclidean distance between pairs of vectors. Input vectors x and y are mapped to differentially private sketches x' and y', from which one can estimate the distance between x and y. Our estimator relies on the Sparser Johnson-Lindenstrauss constructions by Kane & Nelson (Journal of the ACM 2014), which for any $0 < \alpha, \beta < 1/2$ have optimal output dimension $k = \Theta(\alpha^{-2} \log(1/\beta))$ and sparsity $s = O(\alpha^{-1} \log(1/\beta))$. We combine the constructions of Kane & Nelson with either the Laplace or the Gaussian mechanism from the differential privacy literature, depending on the privacy parameters ε and δ. We also suggest a differentially private version of Fast Johnson-Lindenstrauss Transform (FJLT) by Ailon & Chazelle (SIAM Journal of Computing 2009) which offers a tradeoff in speed for variance for certain parameters. We answer an open question by Kenthapadi et al. (Journal of Privacy and Confidentiality 2013) by analyzing the privacy and utility guarantees of an estimator for Euclidean distance, relying on Laplacian rather than Gaussian noise. We prove that the Laplace mechanism yields lower variance than the Gaussian mechanism whenever $\delta < \beta^{O(1/\alpha)}$. Thus, our work poses an improvement over the work of Kenthapadi et al. by giving a more efficient estimator with lower variance for sufficiently small δ. Our sketch also achieves *pure* differential privacy as a neat side-effect of the Laplace mechanism rather than the *approximate* differential privacy guarantee of the Gaussian mechanism, which may not be sufficiently strong for some settings.

Our main result is a special case of more general, technical results proving that one can generally construct unbiased estimators for Euclidean distance with a high level of utility even under the constraint of differential privacy. The bulk of our analysis is proving that the variance of the estimator does not suffer too much in the presence of differential privacy.

1 Introduction

The well-known Johnson-Lindenstrauss Lemma [26] is a fundamental tool in dimensionality reduction and has applications in a variety of fields. The lemma allows for significant speed-ups in applications such as nearest-neighbor search [2, 24], computational geometry [10], document comparison [43], data streams [23], clustering [6, 13], graph sparsification [41], low-rank approximation [12], numerical linear algebra [46, 11, 16] and many more. The Johnson-Lindenstrauss Lemma, or JL lemma for short, states that for any $0 < \alpha, \beta < 1/2$ and input dimension $d > 0$, there exists a random $k \times d$-projection matrix S such that S preserves Euclidean norm of any input vector $x \in \mathbb{R}^d$ up to a factor $(1 \pm \alpha)$ with probability at least $1 - \beta$. Classic examples of projections satisfying the lemma include the constructions from [24, 2, 1, 14, 28]. Jayram & Nelson [25], and later Kane et al. [27], proved the remarkable result that the optimal output dimension k is independent of the input dimension d. In particular, they showed that $k = \Theta(\alpha^{-2} \log(1/\beta))$ is optimal. In the case where n input vectors are known in advance, the *Johnson-Lindenstrauss Flattening Lemma* states that there exist projections preserving Euclidean distance for *all* pairs of these vectors within a factor $(1 \pm \alpha)$, simultaneously. We concern ourselves with a distributed setting, where data is held by several parties and may not be present at the same time. Hence, in our setting the length preserving projection S must be public, so any party holding input vector $x \in \mathbb{R}^d$ can compute and release Sx. For inputs $x, y \in \mathbb{R}^d$ held by different parties, one can estimate the Euclidean distance as $\|Sx - Sy\|_2^2 = \|S(x - y)\|_2^2$. By the Johnson-Lindenstrauss Lemma, this estimate is within a factor $(1 \pm \alpha)$ of $\|x - y\|_2^2$ with high probability. We henceforth use *transform* and *projection* interchangeably and refer to random projections satisfying the Johnson-Lindenstrauss Lemma as *Johnson-Lindenstrauss projections*, or simply *JL projections*. We will even

PODS '21, June 20–25, 2021,
Virtual Event, China

misuse this convention slightly, as we also use this name for projections that preserve Euclidean norm in *expectation*, as defined in Definition 4.

As the input x may contain sensitive information, the released projection of x must preserve *privacy* to prevent third parties from learning the input x. Privacy has often been obtained through simple anonymization by removing obvious identifiers, but several cases have shown that this approach is insufficient [3, 15, 36, 42]. Due to its stringent definition and provable guarantees, we concern ourselves with *differential privacy* [18], which is usually achieved by perturbing the result of a query, to obfuscate the true result slightly. Thus, we analyze the privacy and utility guarantees when adding noise to the projection Sx. That is, for a noise distribution $\mathcal{D}$ and noise vectors $\eta, \mu \in \mathcal{D}^k$, we analyze whether we can privately and accurately estimate $\|x-y\|_2^2$ from $Sx+\eta$ and $Sy+\mu$. The main questions of interest are: *How much noise do we need to add?* and *What utility guarantees can we achieve?* We define differential privacy and mention common choices for noise distribution $\mathcal{D}$ in Section 3.2.

1.1 Differentially Private Random Projections

This work improves on the work of Kenthapadi et al. [29], in which it was shown how to construct an (ε,δ)-differentially private version of a JL transform allowing for high accuracy estimators for squared Euclidean distance. The idea applied by Kenthapadi et al. is simple: Let P be the i.i.d. normally distributed JL transform where each entry is drawn from the standard Normal distribution. For input vector $x \in [0,1]^d$, add Gaussian noise to each entry of Px.

Kenthapadi et al. prove Theorems 1 and 2. We remark that these results extend naturally to $x, y \in \mathbb{R}^d$.

Theorem 1 ([29])**.** *Let P be a $k \times d$-projection matrix with i.i.d. entries from the standard Normal distribution*

and let $x, y \in [0,1]^d$ be input vectors. Let $\eta, \mu \sim \mathcal{N}(0,\sigma^2)^k$ be noise vectors. If $\sigma \geq 4/\varepsilon\sqrt{\log(1/\delta)}$, $\varepsilon < \ln(1/\delta)$ and $k > 2(\ln(d) + \ln(2/\delta))$, then $Px+\eta$ is (ε,δ)-differentially private.

Note 1. *Kenthapadi et al. show that for $k > 2\ln(d) + 2\ln(1/\delta')$, the ℓ_2-sensitivity of P is greater than 2 with probability at most δ'. We will assume that the ℓ_2-sensitivity of P is computed exactly in an initializing step, as discussed in Section 2.1.1, and hence avoid this assumption on k. From [25, 27] we know that for any $\alpha, \beta \in (0,1/2)$ $k = \Theta\left(\alpha^{-2}\log(1/\beta)\right)$ is optimal in the non-private case. We use this value of k and discuss the optimal k for the noisy construction in Section 6.2.1. We also remark that the σ in Theorem 1 can be exchanged with $\sigma \geq \Delta_2\varepsilon^{-1}\sqrt{2\log(1.25/\delta)}$ by a later result from [19] (See Lemma 2), where Δ_2 is the (exact) ℓ_2-sensitivity of P.*

Theorem 2 ([29])**.** *Let P be a $k \times d$-projection matrix with i.i.d. entries from the standard Normal distribution and $x, y \in [0,1]^d$ be input vectors. Let $\eta, \mu \sim \mathcal{N}(0,\sigma^2)^k$ be noise vectors, where σ is independent of the realization of P. Define*

$$\hat{E}_{iid} := \|(Px+\eta) - (Py+\mu)\|_2^2 - 2k\sigma^2.$$

Then

1. *$\hat{E}_{iid}$ is an unbiased estimator for $\|x-y\|_2^2$.*

2. $\operatorname{Var}\left[\hat{E}_{iid}\right] = \frac{2}{k}\|x-y\|_2^4 + 8\sigma^2\|x-y\|_2^2 + 8\sigma^4 k.$

Note 2. *Letting σ be independent of the realization of P might lead to complete loss of privacy if the ℓ_2-sensitivity of P is much higher than 1, as argued in Section 2.1.1. Hence, we let σ be a function of the exact ℓ_2-sensitivity of P, Δ_2, as discussed in Note 1.*

1.2 New Contributions

An immediate idea to achieve a speed-up is to apply the techniques of Kenthapadi et al. to a JL transform, which is faster than the i.i.d. normally distributed JL transform. We show such a result for a private Fast Johnson-Lindenstrauss Transform (FJLT) [2] in Section 5.2, but remark that the privacy issue of Kenthapadi et al. mentioned in Note 2 carries over, if we simply exchange the i.i.d. normally distributed JL transform for the FJLT. We discuss how to address this issue in Section 5 to obtain a differentially private version of FJLT, where σ does not depend on the ℓ_2-sensitivity of the transform (which could be very large). Kenthapadi et al. leave open the question of whether we can obtain better results with Laplacian noise. We answer this question by proving that we can indeed obtain an ε-differentially private estimator for squared Euclidean distances, which has better variance for certain parameters. Specifically, we show the following main theorem:

Theorem 3. *For any $0 < \alpha, \beta < 1/2$ and any integer $d > 0$ there exists a random $k \times d$-projection S for $k = \Theta\left(\alpha^{-2}\log(1/\beta)\right)$ with sparsity $s = O\left(\alpha^{-1}\log(1/\beta)\right)$ and a distribution $\mathcal{D}$ over $\mathbb{R}$ such that for any $x, y \in \mathbb{R}^d$ and $\eta, \mu \sim \mathcal{D}^k$ we define:*

$$\hat{E}_{SJLT} := \|(Sx+\eta) - (Sy+\mu)\|_2^2 - \frac{2ks}{\varepsilon^2}.$$

Then

1. *$\hat{E}_{SJLT}$ is an unbiased estimator for $\|x-y\|_2^2$.*
2. $\operatorname{Var}\left[\hat{E}_{SJLT}\right] \leq \frac{2}{k}\|x-y\|_2^4 + O\left(\frac{s}{\varepsilon^2}\|x-y\|_2^2 + \frac{s^2}{\varepsilon^4}k\right)$.
3. *The sketch $(S, Sx+\eta)$ is ε-differentially private.*
4. *For a data stream, we can update the sketch $(S, Sx+\eta)$ in time $O(s)$.*
5. *$Sx+\eta$ can be computed in time $O(s\|x\|_0 + k)$. Given $Sx+\eta$ and $Sy+\mu$, $\hat{E}_{SJLT}$ can be computed in time $O(k)$.*

The noise distribution $\mathcal{D}$ will depend on the sparsity of S but it is crucial that $\mathcal{D}$ is otherwise independent of S. We state our improvements over the work of Kenthapadi et al. [29]:

- Recall that the projection P of Kenthapadi et al. has constant ℓ_2-sensitivity with high probability. Under this assumption, we combine Theorems 1, 2 and Note 1 to see that

$$\operatorname{Var}[\hat{E}_{iid}] = \frac{2}{k}\|x-y\|_2^4 + O\left(\frac{\log(1/\delta)}{\varepsilon^2}\|x-y\|_2^2 + \frac{\log^2(1/\delta)}{\varepsilon^4}k\right),$$

  and so $\hat{E}_{SJLT}$ improves over $\hat{E}_{iid}$ in terms of variance whenever $\delta < e^{-s} = \beta^{O(1/\alpha)}$ (see Section 7). In the case where P has higher sensitivity, our results give an even better improvement.

- Kenthapadi et al. have an additional initialization cost of $O(dk)$ to compute the sensitivity of the projection matrix. We refer the reader to Section 2.1.1 for a detailed discussion.

- Our estimator $\hat{E}_{SJLT}$ is more efficient as the update time, i.e., time to compute $Sx+\eta$, is $O(s\|x\|_0 + k)$ rather than $O(k\|x\|_0 + k)$ for $s = o(k)$.

- Rather than *approximate* differential privacy, which may be insufficient for some applications, we achieve *pure* differential privacy.

Our improved efficiency in Theorem 3 relies on the sparsity of the Sparser JL transforms by Kane & Nelson [28], henceforth referred to as *the SJLT*. We remark that the results of Kenthapadi et al. extend naturally to these JL transforms, and thus they would obtain the same efficiency for $\delta > \beta^{O(1/\alpha)}$. We do, although, give the analysis proving that these transforms can indeed be used. Using a SJLT instead of the i.i.d. normally distributed transform, the work of Kenthapadi et al. would also avoid the initialization cost.

Related to our analysis for the SJLT, we remark that our main result is, in fact, a special case of an even more general result: we give a class of length preserving linear transformations that allow for efficient, private estimators for Euclidean distance with a high level of utility. The FJLT and SJLT are merely examples of such linear transformations. We define what is meant by *length preserving* in section 3.3 and prove our general, technical results in Section 4. In Section 5 we give two differentially private versions of FJLT and in Section 6, we prove Theorem 3 by applying the technical results to the SJLT with noise from the Laplace distribution. Finally, we compare the work of Kenthapadi et al. with our private FJLT and SJLT in Section 7.

2 Related Work

Differential privacy is usually achieved by adding random noise to the output of a query to obfuscate the exact result, before publishing the result. This idea is easily extended to vector outputs by simply adding noise to each entry of the output vector. This technique has been studied extensively in previous work, see for example [33, 38, 22, 31].

We consider a distributed setting, where party i adds noise $\eta_i \sim \mathcal{D}^k$ to the projection Sx_i of input vector x_i and releases the noisy projection $Sx_i + \eta_i$ for future distance estimation. All parties must use the same randomized matrix S and noise drawn from the same distribution $\mathcal{D}$. It is crucial that the projection matrix is public, and only the noise be kept secret.

2.1 Versions of Johnson-Lindenstrauss Transformations

We refer to the classical JL transform by Indyk & Motwani [24] as the *i.i.d. normally distributed JL transform*. As the name suggests, the random projection matrix consists of i.i.d. entries from the standard Normal distribution.

The *sparsity* of the random projection, i.e., the number of non-zero entries per column is an important tool in speeding up dimensionality reduction. Ailon & Chazelle [2] presented a JL transform with a sparser projection matrix with a mixture of normally distributed entries and 0s. This transform is commonly known as *The Fast Johnson-Lindenstrauss Transform* or in short, *FJLT*. We describe the transform in detail in Section 5.1.

The sparsity not only affects the sensitivity of the transformation (see Section 3.2.1 for the definition of sensitivity), but also the time required to compute the projection of an input vector x. For a random projection S with sparsity s, we can compute Sx in time $O(s\|x\|_0)$. Kane & Nelson [28] show that the JL transform of Dasgupta et al. [14] requires sparsity $s = \tilde{\Omega}(\alpha^{-1} \log^2(1/\beta))$, and Nelson & Nguyen showed that this sparsity is optimal up to a factor $O(\log(1/\alpha))$ [37]. Kane & Nelson [28] also give two sparser constructions with $s = \Theta(\alpha^{-1} \log(1/\beta))$ for embedding into $k = \Theta(\alpha^{-2} \log(1/\beta))$ dimensions. These transformations are commonly known as *The Sparser JL Transforms* and we will henceforth refer to them as *SJLT*. We describe SJLT in Section 6.1.

2.1.1 Differentially Private JL Construction

Kenthapadi et al. [29], which was also discussed in Section 1.1, give a private estimator for Euclidean distance relying on the i.i.d. normally distributed JL transform. A drawback of their construction is that the ℓ_2-sensitivity is only 1 in *expectation*, so the sensitivity *might* not be small. This is the case if the random projection has even a single very large entry. The authors suggest drawing noise calibrated to a low sensitivity projection matrix independently of the actual projection matrix P. However, with a small probability, P *does not* have low sensitivity, in which case the noise is not ensured to provide differential privacy. Kenthapadi et al. "hide" the probability of drawing a high-sensitivity projection under δ, but for a fixed P, either the noise provides privacy, or certain inputs would always be distinguishable, even in the presence of noise calibrated to low sensitivity. An alternative solution is to compute the sensitivity of the fixed P and calibrate the noise to the actual sensitivity. Hence, initialization requires time $O(dk)$. Kenthapadi et al. state without proof that their results extend to the JL transformations from [1, 14]. Xu et al. [47] extend the work of [29] with experimental comparisons with JTree [9], PrivBayes [48], PriView [39] and PrivateSVM [40].

2.2 Differentially Private Linear Transformations

Mir et al. (PODS11) [33] suggest a general framework for generating pan-private linear transformations by initializing with noise from the exponential mechanism. The work argues how to create a ε-pan private estimator for (squared) Euclidean distance with multiplicative error $(1+\gamma)$ and additive error $\text{poly}(\log d, \varepsilon^{-1}, \gamma^{-1}, \log(q^{-1})) + O(Z)$, with probability at least $1-q$, where Z is an upper bound on the entries of the input vector. The technique used by Mir et al. can be used for private dimensionality reduction, but is computationally inefficient as the sketch relies on the exponential mechanism for noise addition.

In an earlier (unpublished) version of the same work, [32], Mir et al. analyze the *cropped* second moment for a parameter τ, defined for input vector $x \in \mathbb{Z}^d$ as $\sum_{i=1}^{d} \min\{x_i^2, \tau\}$. In this work, Mir et al. show a 2ε-differentially private estimator with additive error $O_\varepsilon(\tau\sqrt{d})$ with high probability. Differential privacy is achieved by an application of Randomized Response [45]. As our error depends on $\|x - y\|_2$ and $\sqrt{k} < \sqrt{d}$, we see an improvement when x and y are sparse. The problems are not directly comparable as the cropped second moment of Mir et al. applies to integer inputs, whereas we consider inputs over the reals.

2.3 When Data is Known in Advance

If input data is known in advance, there are other techniques to achieve differential privacy. A central unit with access to all data can compute the *exact* distances (up to the error incurred by the JL embedding) and add noise specifically calibrated to this distance. This technique often incurs less noise, but is not applicable in our setting, as data is split among several parties and may not all be available at once.

Blocki et al. [5] show that, as long as the projection matrix is kept secret, the i.i.d. normally distributed JL transform allows for differentially private estimates of distances with the accuracy guarantees from the Johnson-Lindenstrauss Lemma. Upadhyay [44] proves that this technique does not generally work to preserve privacy for sparser JL projections. As we consider a distributed setting, keeping the projection matrix secret is unattainable. Bhaskar et al. [4] introduce *noiseless privacy* where the output is always exact, rather than a noisy approximation. The privacy guarantees are of a similar form as differential privacy but rely on assumptions about the distribution of the data and auxiliary information, whereas differential privacy aims for a higher level of generality.

2.3.1 Representing Noise from Continuous Noise Distributions

We will assume that noise is drawn from either the continuous Laplace or Gaussian distribution, which, however, may introduce practical issues. Mironov [34] described how privacy may be lost due to floating-point error when sampling noise from a continuous distribution. As an alternative to the continuous Laplace distribution, Mironov suggests the *Snapping mechanism*, which incurs an additional error of approximately Δ_1/ε compared to noise from $\text{Lap}(\Delta_1/\varepsilon)$, where Δ_1 is the ℓ_1-sensitivity of the query.

[20] improve over the Snapping Mechanism, by drawing noise from a discrete distribution, differing from the Laplace distribution by at most a factor $(1 + \frac{1+2/\varepsilon}{2^k})$ for a fixed integer k, which controls the accuracy of the discretization. It suffices to use $k \in [10, 45]_{\mathbb{Z}}$.

A discrete, "hole-free" alternative to the Gaussian distribution, requiring only expected constant time is suggested in [20]. The distribution builds on the Binomial distribution with parameters n and $p = 1/2$ and the work of [7] to give a distribution which for large n differs from the Gaussian distribution by at most $O(\log^{1.5}(n)/\sqrt{n})$.

In a very recent work, Canonne et al. [8] describe a discretization of the Gaussian distribution supported on $\mathbb{Z}$ whose variance is at most that of the corresponding continuous Gaussian distribution, and hence allows for identical or slightly better utility. Simultaneously, the discretization has sub-Gaussian tails compared to the corresponding continuous Gaussian distribution and essentially the same privacy guarantees. We refer to the discussion in [8] for further reading on discretizations of the Laplace and Gaussian distributions.

2.4 Lower bounds

McGregor et al. [30] show that any protocol for estimating Hamming distance (and so for inner product, which again leads to a protocol for estimating squared Euclidean distance) of two binary k-dimensional vectors in a differentially private manner incurs an additive error of $\tilde{\Omega}(\sqrt{k})$, which is contrasted by the observation that simple Randomized Response [45] allows for error $O(\sqrt{k})$. The error lower bound implies a $\tilde{\Omega}(k)$ lower bound for the variance of the noisy estimator. In contrast, our variance of the noise added (we may disregard the variance introduced by the JL projection, as this error occurs even in the non-private version) depends on $\|x - y\|_2^2 \leq d$ and k (for binary input vectors).

Independently from the work of McGregor et al., Mir et al. [33] show a similar lower bound of additive error $\Omega(\sqrt{k})$ for estimating inner product for binary vectors in a pan-private setting. The lower bound by McGregor et al. implies a lower bound for pan-private algorithms, which is weaker than the lower bound of Mir et al. in the case of single-pass algorithms and dynamic data. Hardt & Talwar [21] show that an ε-differentially private algorithm for the second frequency moment F_2 requires an additive error factor of $\Omega(1/\varepsilon)$, which is comparable to our result (up to polynomial and logarithmic factors).

3 Preliminaries

3.1 Notation

Let $x \in \mathbb{R}^d$ be an input vector. For a $k \times d$-matrix S, let Sx be the linear transformation of x under S. Let $\eta \sim \mathcal{D}^k$ for a noise distribution $\mathcal{D}$. We denote by $Sx + \eta$ the noisy counter-part to Sx and let $\eta_* \sim \mathcal{D}$ denote a random variable drawn according to $\mathcal{D}$. We use *transformation of x* and *projection of x* interchangeably as our main focus will be on random projections. Unless otherwise specified *projection* always refers to a *random projection.*

Denote by $\|x\|_p$ the ℓ_p-norm of x and let $1[p]$ denote the indicator variable for predicate p.

3.2 Differential Privacy

Intuitively, differential privacy guarantees that one cannot (confidently) distinguish between whether an output is the result generated from a specific input vector or from a *neighboring* vector:

Definition 1 (Neighboring inputs)**.** *Vectors $x, y \in \mathbb{R}^d$ are called* neighboring*, sometimes also* adjacent*, if*

$$\|x - y\|_1 \leq 1.$$

We remark that this definition is a generalization of the natural attribute-level privacy for binary input vectors, where privacy is preserved for a single bit-flip. For user-level privacy, we suppose that the contribution of a single user affects the ℓ_1-norm of the input vector by at most 1. This is the case when we consider example histograms. More general user-level privacy is out of scope of this work.

Definition 2 (Differential Privacy [18, 17])**.** *A randomized mechanism $\mathcal{M}$ preserves (ε, δ)-differential privacy, or* approximate *differential privacy, if for any neighboring input vectors x and y, and for all subsets $S \subset \mathrm{Range}(\mathcal{M})$, we have*

$$\Pr[\mathcal{M}(x) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(y) \in S] + \delta.$$

where the probability is over the random choices of $\mathcal{M}$. If $\delta = 0$ we say that $\mathcal{M}$ preserves ε-differential privacy or pure *differential privacy.*

A common interpretation of approximate differential privacy is that we get pure differential privacy except with probability δ [35].

3.2.1 Sensitivity

Dwork et al. showed that we can obtain differential privacy by adding noise calibrated to the *sensitivity* of a function [18]. We define the sensitivity of a linear transformation:

Definition 3 (ℓ_p-sensitivity of transformation [29])**.** *For $p \geq 1$, the ℓ_p-sensitivity of a linear transformation $S : \mathbb{R}^d \to \mathbb{R}^k$ is*

$$\Delta_p(S) = \max_{\substack{x,y \in \mathbb{R}^d \\ \|x-y\|_1 \leq 1}} \|Sx - Sy\|_p = \max_{1 \leq j \leq d} \left(\sum_{i=1}^{k} |S_{i,j}|^p \right)^{1/p} = \max_{1 \leq j \leq d} \|S_{\cdot,j}\|_p$$

where $S_{\cdot,j}$ is the j^{th} column of S.

Note 3. *The definition follows from the observation that any vector of ℓ_1-norm 1 (which is the case for neighboring vectors) can be represented as a convex linear combinations of basis vectors.*

3.2.2 Techniques In Differential Privacy

We present two fundamental techniques in differential privacy that we use extensively in our analysis.

Lemma 1 (Laplace Mechanism [18])**.** *For linear transformation $S \in \mathbb{R}^{k\times d}$ and input $x \in \mathbb{R}^d$, the Laplace Mechanism with parameter b outputs $Sx + \eta$ for $\eta \sim \mathrm{Lap}(0, b)^k$. Let Δ_1 be the ℓ_1-sensitivity of S. The Laplace Mechanism with parameter $\Delta_1\varepsilon^{-1}$ preserves ε-differential privacy.*

Lemma 2 (Gaussian Mechanism [17, 19])**.** *For linear transform $S \in \mathbb{R}^{k\times d}$ and input $x \in \mathbb{R}^d$, the Gaussian Mechanism with parameter σ outputs $Sx + \eta$ for $\eta \sim \mathcal{N}(0, \sigma^2)^k$. Let Δ_2 be the ℓ_2-sensitivity of S. The Gaussian Mechanism with parameter $\sigma \geq \Delta_2\varepsilon^{-1}\sqrt{2\log(1.25/\delta)}$ preserves (ε, δ)-differential privacy.*

3.3 Length Preserving Property

Our technical results in Section 4 rely on linear transforms with the *Length Preserving Property (LPP)*:

Definition 4 (Length Preserving Property (LPP))**.** *A random $k \times d$-projection S satisfies the* Length Preserving Property *if for any $x \in \mathbb{R}^d$ we have*

$$\mathrm{E}_S\left[\|Sx\|_2^2\right] = \|x\|_2^2.$$

Note that if S satisfies LPP, then S also preserves Euclidean distances and inner products, as $\langle x, y\rangle = \frac{\|x\|_2^2+\|y\|_2^2-\|x-y\|_2^2}{2}$.

4 Technical Results

We now show our general, technical lemmas which will be useful for proving Theorem 3. Let S be a random $k \times d$-matrix with LPP as defined in Definition 4 and let $x, y \in \mathbb{R}^d$. Let $\mathcal{D}$ be a zero-mean distribution and $\eta, \mu \sim \mathcal{D}^k$ noise vectors. Let $\eta_* \sim \mathcal{D}$. We define

$$\hat{E}_{gen} := \|(Sx + \eta) - (Sy + \mu)\|_2^2 - 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^2].$$

Our technical results are as follows:

Lemma 3. *We have*

1. *$\hat{E}_{gen}$ is an unbiased estimator for $\|x - y\|_2^2$.*
2. *The variance of $\hat{E}_{gen}$ is*

$$\mathrm{Var}\left[\hat{E}_{gen}\right] = \mathrm{Var}\left[\|Sx - Sy\|_2^2\right] + 8\,\mathrm{E}_{\mathcal{D}}[\eta_*^2]\|x - y\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^2]^2$$

Proof. See Appendix A. □

Hence, the variance of $\hat{E}_{gen}$ is close to the variance of the non-private estimator, but has an additional noise term depending on the output dimension k and the Euclidean distance of the input vectors. The following result describes the privacy guarantees of $\hat{E}_{gen}$:

Lemma 4. *Let Δ_1 and Δ_2 be the ℓ_1- and ℓ_2-sensitivities of S, respectively. Let $\delta > 0$ be given and define*

$$m := \min\left\{\Delta_1, \Delta_2\sqrt{\ln(1/\delta)}\right\}.$$

There is a distribution $\mathcal{D}$ such that

1. *The sketch $(S, Sx + \eta)$ is differentially private.*
2. $\mathrm{Var}\left[\hat{E}_{gen}\right] = \mathrm{Var}\left[\|Sx - Sy\|_2^2\right] + O\left(\frac{m^2}{\varepsilon^2}\|x - y\|_2^2 + \frac{m^4}{\varepsilon^4}k\right)$.

3. Given Sx and Sy, the estimate $\hat{E}_{gen}$ can be computed in time $O(k)$.

Proof. We show that it suffices to let $\mathcal{D}$ be either the Normal or Laplace distribution for well-chosen parameters. We start with the following useful note:

Note 4. *Let $n!!$ be the product of the numbers $1, ..., n$ that have the same parity as n. For $L \sim \mathrm{Lap}(b)$ and $G \sim \mathcal{N}(0, \sigma^2)$, we have*

$$\forall n \in \mathbb{N}: \ \mathrm{E}_{\mathcal{D}}[L^n] = \frac{n!}{(b^{-1})^n}$$

$$\textit{for even } n\textit{:} \ \mathrm{E}_{\mathcal{D}}[G^n] = (n-1)!!\sigma^n.$$

By Lemma 2, the noisy projection $Sx + \eta$ is (ε, δ)-differentially private for $\mathcal{D} = \mathcal{N}(0, \sigma^2)$ with $\sigma \geq \frac{\Delta_2}{\varepsilon}\sqrt{2\ln(1.25/\delta)}$. By the post-processing property of differential privacy, $\hat{E}_{gen}$ is also (ε, δ)-differentially private. From Lemma 3 and Note 4

$$\mathrm{Var}\left[\hat{E}_{gen}\right] = \mathrm{Var}\left[\|Sx - Sy\|_2^2\right] + O\left(\frac{\Delta_2^2 \ln\left(\frac{1}{\delta}\right)}{\varepsilon^2}\|x - y\|_2^2 + \frac{\Delta_2^4 \ln^2\left(\frac{1}{\delta}\right)}{\varepsilon^4}k\right). \tag{1}$$

Similarly, by Lemma 1 $\hat{E}_{gen}$ is ε-differentially private for $\mathcal{D} = \mathrm{Lap}(\Delta_1/\varepsilon)$, and from Lemma 3 and Note 4 we get

$$\mathrm{Var}\left[\hat{E}_{gen}\right] = \mathrm{Var}\left[\|Sx - Sy\|_2^2\right] + O\left(\frac{\Delta_1^2}{\varepsilon^2}\|x - y\|_2^2 + \frac{\Delta_1^4}{\varepsilon^4}k\right). \tag{2}$$

Finally, we can draw noise from the Laplace or the Normal distribution in constant time. □

Note 5. *As seen in the proof of Lemma 4, letting $\mathcal{D} = Lap(\Delta_1/\varepsilon)$ gives $m = \Delta_1$ and letting $\mathcal{D} = \mathcal{N}(0, \sigma^2)$ for $\sigma \geq \Delta_2\varepsilon^{-1}\sqrt{2\ln(1.25/\delta)}$ gives $m = \Delta_2\sqrt{\ln(1/\delta)}$. We wish to choose the $\mathcal{D}$ which minimizes $\mathrm{Var}[\hat{E}_{gen}]$. Ignoring constants, (2) is upper bounded by (1) when*

$$\Delta_1 < \Delta_2\sqrt{\ln(1/\delta)} \quad \Leftrightarrow \quad \delta < e^{-\Delta_1^2/\Delta_2^2}. \tag{3}$$

Hence, when (3) is satisfied, we use $\mathcal{D} = \mathrm{Lap}(\Delta_1/\varepsilon)$ and otherwise let $\mathcal{D} = \mathcal{N}(0, \sigma^2)$ with $\sigma \geq \Delta_2\varepsilon^{-1}\sqrt{2\log(1.25/\delta)}$.

5 Private Fast Johnson-Lindenstrauss Transform

We now discuss a private version of the Fast Johnson-Lindenstrauss transform (FJLT) by Ailon & Chazelle [2]. We first remind the reader of the non-private transform in Section 5.1 and then give two private versions in Section 5.2.

5.1 Description of (non-private) Fast Johnson-Lindenstrauss Transform (FJLT)

We are concerned only with the transform preserving ℓ_2-distances, but refer the reader to [2] for the transform preserving ℓ_1-distances as well as the analysis for the transforms.

FJLT is a random distribution of linear mappings $\Phi : \mathbb{R}^d \to \mathbb{R}^k$ with $k = O(\log(1/\beta)/\alpha^2)$, such that for $\alpha, \beta \in (0, 1/2)$, with probability at least $1 - \beta$

$$(1 - \alpha)k\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \alpha)k\|x\|_2^2.$$

For given values of d, α, β, we describe how to obtain the random mapping Φ as the product of three real valued matrices, P, H and D:

- D is a random $d \times d$-diagonal matrix with D_{jj} drawn independently from $\{-1, +1\}$ with probability 1/2.

- H is a $d \times d$-normalized Hadamard matrix such that for $f, j \in [d]$

$$H_{fj} = \frac{1}{\sqrt{d}}(-1)^{\langle f-1, j-1 \rangle}$$

where $\langle f, j \rangle$ is the dot-product between vectors expressing f and j in binary representation.

- P is a random $k \times d$-matrix whose entries are independently either normally distributed or 0. Specifically, for

$$q = \min\left\{\Theta\left(\frac{\log^2(1/\beta)}{d}\right), 1\right\}$$

we let P_{if} be drawn (independently) from $\mathcal{N}(0, q^{-1})$ with probability q and $P_{if} = 0$ with probability $1-q$ for $i \in [k]$ and $f \in [d]$.

The transform Φ is defined as

$$\Phi := PHD.$$

To formalize, we get the following lemma:

Lemma 5 (Lemma 2.1 from [2]). *Let $\alpha, \beta \in (0, 1/2)$ and let Φ be a random $k \times d$-projection matrix as described above. Let $x \in \mathbb{R}^d$. With probability at least $1-\beta$, the following two events occur:*

- $(1-\alpha)k\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\alpha)k\|x\|_2^2$.
- *The mapping $\Phi : \mathbb{R}^d \to \mathbb{R}^k$ requires time*

$$O\left(d \log d + dq\frac{\log(1/\beta)}{\alpha^2}\right)$$

for $q = \min\left\{\Theta\left(\frac{\log^2(1/\beta)}{d}\right), 1\right\}$.

Proof. See [2]. □

We will henceforth concern ourselves with the *normalized* FJLT, $1/\sqrt{k} \cdot \Phi$, such that

$$(1-\alpha)\|x\|_2^2 \leq 1/k\|\Phi x\|_2^2 \leq (1+\alpha)\|x\|_2^2.$$

Lemma 6. *The normalized FJLT satisfies LPP (see Definition 4).*

Proof. See Appendix B.2. □

Lemma 7. *Let $x, y \in \mathbb{R}^d$ and let Φ be the FJLT as described above. Then*

$$\mathrm{Var}[1/k\|\Phi x - \Phi y\|_2^2] \leq \frac{3}{k}\|x-y\|_2^4.$$

Proof. The proof follows directly from Lemma 11 in Appendix B.3. □

5.2 Private FJLT

In this section, we argue how to construct a differentially private version of FJLT by adding Gaussian noise to the input.

If we simply exchange the i.i.d. normally distributed JL transform for FJLT in the work of Kenthapadi et al. [29], we get the following result. Note that the ℓ_2-sensitivity of the (normalized) projection is concentrated around 1, which justifies the choice of Gaussian noise.

Corollary 1. *Let Φ be a random $k \times d$-FJLT and let $x, y \in \mathbb{R}^d$ be input vectors. Let Δ_2 be the ℓ_2-sensitivity of Φ and let $\eta, \mu \sim \mathcal{N}(0, \sigma^2)^k$ for $\sigma \geq \Delta_2\varepsilon^{-1}\sqrt{2\log(1.25/\delta)}$ be noise vectors. Define*

$$\hat{E}_{FJLT_o} := \frac{1}{k}\|(\Phi x + \eta) - (\Phi y + \mu)\|_2^2 - 2k\sigma^2$$

- *$\hat{E}_{FJTL_o}$ is an unbiased estimator for $\|x-y\|_2^2$.*
- *The estimator has variance*

$$\mathrm{Var}\left[\hat{E}_{FJLT_o}\right] \le \frac{3}{k}\|x-y\|_2^4 + O\left(k\sigma^4 + \sigma^2\|x-y\|_2^2\right).$$

- *The sketch $(\Phi, \Phi x + \eta)$ is (ε, δ)-differentially private.*
- *The sketch $\Phi x + \eta$ can be computed in time*

$$O\left(\max\left\{d\log d, \frac{dq\log(1/\beta)}{\alpha^2}\right\}\right)$$

  for $q = \min\{\Theta(\log^2(1/\beta)/d), 1\}$.

Proof. That the estimator is unbiased and the variance follow from Lemmas 3, 6 and 7. Privacy follows from Lemmas 2 and 4. Running time follows from Lemmas 5 and 4. □

Note 6. *Although the ℓ_2-sensitivity of the normalized FJLT is concentrated around 1, the sensitivity of* Φ could *(with a small probability) be very large, so the sketch $\Phi x + \eta$ suffers from the same initialization cost as the work of Kenthapadi et al. (see Section 2.1.1).*

We now introduce a private version of FJLT, where we perturb the *input*. This version avoids the issue described in Note 6, but will inevitably introduce error depending on the input size.

Lemma 8. *Let Φ be a random $k \times d$-FJLT and let $x, y \in \mathbb{R}^d$ be input vectors. Let $\eta, \mu \sim \mathcal{N}(0, \sigma^2)^d$ for $\sigma \ge \varepsilon^{-1}\sqrt{2\log(1.25/\delta)}$ be noise vectors. Define*

$$\hat{E}_{FJLT_i} := \frac{1}{k}\|\Phi(x+\eta) - \Phi(y+\mu)\|_2^2 - 2d\sigma^2$$

- *$\hat{E}_{FJTL_i}$ is an unbiased estimator for $\|x-y\|_2^2$.*
- *The estimator has variance*

$$\mathrm{Var}\left[\hat{E}_{FJLT_i}\right] \le \frac{3}{k}\|x-y\|_2^4 + O\left(\frac{d^2\sigma^4}{k} + d\sigma^2\|x-y\|_2^2\right).$$

- *The sketch $(\Phi, \Phi(x+\eta))$ is (ε, δ)-differentially private.*
- *The sketch $\Phi(x+\eta)$ can be computed in time*

$$O\left(\max\left\{d\log d, \frac{dq\log(1/\beta)}{\alpha^2}\right\}\right)$$

  for $q = \min\{\Theta(\log^2(1/\beta)/d), 1\}$.

Proof. For proofs that the estimator is unbiased and for the variance, see Appendix C.1. We remark that the factor d on the last term in the variance is a by-product of applying Φ to the noise. Privacy follows directly from the Gaussian mechanism (see Lemma 2), as the ℓ_2-sensitivity is at most 1 (clearly, as we perturb the input vectors). As noise can be added in time $O(d)$, the time required to compute the sketch follows from Lemma 5. □

Note 7. *By spherical symmetry of the Normal distribution, $\Phi(x+\eta)$ and $\Phi x + P\eta$, where P is defined in Section 5.1, are identically distributed. Hence, one could add the same amount of noise after the Hadamard transform to get a differentially private sketch, that is, compute $P(HDx+\eta)$. Thus, for a given projection P, suppose column j is all zeros, then we can immediately set $\eta_j = 0$. This way, we may save a bit of randomness.*

6 Private Sparser Johnson-Lindenstrauss Transform

In this section, we turn to the question of perturbation using Laplacian noise rather than Gaussian noise. We present and analyze a private sketch based on the SJLT, and conclude Theorem 3 in Section 6.2.3. The main observation about this sketch is that we perturb the *output* vectors rather than the input vectors while avoiding the initialization cost that was inherent to the work of Kenthapadi et al. as well as Corollary 1. We compare the work of Kenthapadi et al., our private FJLT from Lemma 8 and our private SJLT from Theorem 3 in Section 7.

Theorem 3 is proven by combining the technical Lemmas 3 and 4 with the SJLT. Due to their sparsity, these transforms are more efficient than the suggestions from [29]. We remark that this is just one example of linear transformations that our results can be applied to. It should also be noted that the results of Kenthapadi et al. are directly transferable to the SJLT, although the results were only proven for the i.i.d. normally distributed JL transform, whereas we give the analysis here.

6.1 Description of (non-private) Sparser Johnson-Lindenstrauss Transforms (SJLT)

We first describe the SJLT from [28]. We focus on the c)-construction and remark that similar arguments applies for the b)-construction. Let $k = \Theta(\alpha^{-2}\log n)$ and let $x \in \mathbb{R}^d$ be an input vector. Let $h_1, ..., h_s : [d] \to [k/s]$ and $\varphi_1, ..., \varphi_s : [d] \to \{-1, +1\}$ be independent, random hash functions from $O(\log(1/\beta))$-wise independent families. Define $\xi_{ri}(j) = 1[h_r(j) = i]$. Then $\mathrm{E}[\xi_{ri}(j)^2] = \mathrm{E}[\xi_{ri}(j)] = \frac{s}{k}$. The projection matrix S is defined by

$$S_{(i,r),j} = \frac{1}{\sqrt{s}}\varphi_r(j)\xi_{ri}(j)$$

for $i = 1, ..., k/s$ and $r = 1, ..., s$. Hence, entry $i' = i \cdot r \in [k]$ in the resulting embedding Sx can be described as

$$(Sx)_{i'} = (Sx)_{(i,r)} = \frac{1}{\sqrt{s}}\sum_{j=1}^{d}\varphi_r(j)\xi_{ri}(j)x_j.$$

We can think of Sx as a vector consisting of s blocks, each of length k/s. The ith block describes the projection of x under h_i and φ_i.

Lemma 9. *The SJLT as described above satisfy LPP from Definition 4.*

Proof. The proof is a simple calculation and can be found in Appendix D.1. □

Lemma 10. *Let $x, y \in \mathbb{R}^d$ and let S be the SJLT as described above. Then*

$$\mathrm{Var}\left[\|Sx - Sy\|_2^2\right] \leq \frac{2}{k}\|x - y\|_2^4.$$

Proof. The proof can be found in Appendix D.2. □

6.2 Private SJLT

We now turn to proving our main theorem, Theorem 3. Combining Lemmas 3, 9 and 10, we obtain the following corollary.

Corollary 2. *Let S be a random $k \times d$-SJLT and let $x, y \in \mathbb{R}^d$ be input vectors. Let $\eta, \mu \sim \mathcal{D}^k$ be noise vectors where each entry is drawn from a zero-mean distribution $\mathcal{D}$. Then*

$$\hat{E}_{SJLT_{\mathcal{D}}} := \|(Sx + \eta) - (Sy + \mu)\|_2^2 - 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^2]$$

is an unbiased estimator for $\|x - y\|_2^2$ with variance

$$\mathrm{Var}\left[\hat{E}_{SJLT_{\mathcal{D}}}\right] \leq \frac{2}{k}\|x - y\|_2^4 + 8\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|x - y\|_2^2 + 2k\left(\mathrm{E}_{\mathcal{D}}[\eta_*^4] + \mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2\right). \quad (4)$$

We have yet to choose $\mathcal{D}$ to ensure differential privacy of this estimator as well as argue about the efficiency. We first discuss the value of k.

6.2.1 Optimal Projection Dimension

For the non-private SJLT, the optimal projection dimension is $k = \Theta\left(\alpha^{-2}\log(1/\beta)\right)$. One may ask what k is optimal in the private case. The analysis and our optimal k are very similar to the findings in [29]: we see that the variance in (4) is minimized for $k = \Theta\left(\frac{\|x-y\|_2^2}{\sqrt{\mathrm{E}[\eta_*^4]+\mathrm{E}[\eta_*^2]^2}}\right)$. By the same argument as in [29], generally, no fixed value of k will be optimal for the entire input domain, although there might be exceptions, when certain properties of the data are known. As in the work of Kenthapadi et al., if we have input domain X, then we may let $\nu = \max_{x\in X}\{\|x\|_2^2\}$ to obtain $k = \Theta(\nu\varepsilon^2/\Delta_1^2)$ for $\mathcal{D} = \mathrm{Lap}(\Delta_1/\varepsilon)$. Note that k might not be optimal for all input vectors. We assume that ν is unknown and may be very large – in particular, we consider vectors over the reals – and thus proceed with $k = \Theta(\alpha^{-2}\log(1/\beta))$.

6.2.2 Efficiency

Let S be a SJLT with sparsity $s = O\left(\alpha^{-1}\log(1/\beta)\right)$ and let input $x \in \mathbb{R}^d$ be given. The embedding Sx can be computed in time $O(s\|x\|_0)$. Assuming that we sample from $\mathcal{N}(0,\sigma^2)$ and $\mathrm{Lap}(b)$ in constant time, random noise vector $\eta \sim \mathcal{D}$ for $\mathcal{D} = \mathrm{Lap}(\Delta_1/\varepsilon)$ or $\mathcal{D} = \mathcal{N}(0,\sigma^2)$ for $\sigma \geq \Delta_2\varepsilon^{-1}\sqrt{2\log(1.25/\delta)}$ can be added in time $O(k)$ to give $Sx+\eta$. From [20] we know that we can at least sample from discretizations in expected constant time, so this assumption seems reasonable. For given $Sx+\eta$ and $Sy+\mu$, the estimator $\hat{E}_{SJLT}$ can be computed in time $O(k)$.

6.2.3 Summing Up

The SJLT as described in Section 6.1, where $k = \Theta(\alpha^{-2}\log(1/\beta))$ and $s = O(\alpha^{-1}\log(1/\beta))$, has ℓ_1-sensitivity $\Delta_1 = \sqrt{s}$ and ℓ_2-sensitivity $\Delta_2 = 1$. Hence, consider Corollary 2 with $\mathcal{D} = \mathrm{Lap}(\sqrt{s}/\varepsilon)$. Lemma 1 ensures that $\hat{E}_{SJLT}$ is ε-differentially private. Combining with Section 6.2.2 finishes the proof of Theorem 3. If instead we let $\mathcal{D} = \mathcal{N}(0,\sigma^2)$ for $\sigma \geq \varepsilon^{-1}\sqrt{2\ln(1.25/\delta)}$ in Corollary 2, $\hat{E}_{SJLT}$ is (ε,δ)-differentially private and achieves the same variance as the work of Kenthapadi et al., while we gain a speed-up as well as avoid the initialization cost. Finally, we remark that by Note 5, we minimize the variance of $\hat{E}_{SJLT}$ by letting $\mathcal{D} = \mathrm{Lap}(\sqrt{s}/\varepsilon)$ whenever $\delta < e^{-s}$.

7 Comparison

We now compare Lemma 8 and Theorem 3 with the work of Kenthapadi et al.

We first compare the running times to see for what parameters the private FJLT is faster than the private SJLT and then compare the variances for the two methods to get the speed-variance trade-off. Finally, we compare to the results of Kenthapadi et al.

Recall that our private FJLT can be computed in time

$$O\left(\max\left\{d\log d, \frac{\log^3(1/\beta)}{\alpha^2}\right\}\right),$$

and the private SJLT can be computed in time bounded by $O(sd)$ (for dense vectors) where $s = O\left(\log(1/\beta)\alpha^{-1}\right)$. Observing that

$$O(sd) > O(d\log d) \quad \Leftrightarrow \quad d < e^{O(s)} = \frac{1}{\beta^{O(1/\alpha)}}$$

and

$$O(sd) > O\left(\frac{\log^3(1/\beta)}{\alpha^2}\right) \quad \Leftrightarrow \quad d > O\left(\frac{\log^2(1/\beta)}{\alpha}\right),$$

we conclude that our private FJLT is indeed faster than the private SJLT whenever

$$O\left(\frac{\log^2(1/\beta)}{\alpha}\right) < d < \frac{1}{\beta^{O(1/\alpha)}}. \tag{5}$$

We now turn to comparing the variances of the private versions of FJLT and SJLT: Recall from Lemma 8 that the private FJLT has variance

$$\mathrm{Var}\left[\hat{E}_{FJLT_i}\right] \leq \frac{3}{k}\|x-y\|_2^4 + O\left(d\sigma^2\|x-y\|_2^2 + \frac{d^2\sigma^4}{k}\right).$$

while, as seen in Theorem 3, the private SJLT has variance

$$\mathrm{Var}\left[\hat{E}_{SJLT}\right] \leq \frac{2}{k}\|x-y\|_2^4 + O\left(\frac{s}{\varepsilon^2}\|x-y\|_2^2 + \frac{s^2}{\varepsilon^4}k\right).$$

For the sake of simplicity, we will disregard the variance incurred by the transforms and limit ourselves to considering the terms incurred by the noise addition. The private SJLT (in particular) achieves a better variance than the private FJLT whenever

$$\begin{aligned} &O\left(\frac{d^2\sigma^4}{k}\right) = O\left(\frac{d^2\log^2(1/\delta)}{\varepsilon^4 k}\right) > O\left(\frac{s^2 k}{\varepsilon^4}\right) \qquad \text{and} \\ &O\left(d\sigma^2\|x-y\|_2^2\right) = O\left(\frac{d\log(1/\delta)}{\varepsilon^2}\|x-y\|_2^2\right) > O\left(\frac{s}{\varepsilon^2}\|x-y\|_2^2\right). \end{aligned}$$

Treating each of the inequalities separately, we analyze for what values of δ this is the case:

$$O\left(\frac{d^2\log^2(1/\delta)}{\varepsilon^4 k}\right) > O\left(\frac{s^2 k}{\varepsilon^4}\right) \quad \Leftrightarrow \quad \log(1/\delta) > O\left(\frac{sk}{d}\right) \quad \Leftrightarrow \quad \frac{1}{e^{O(sk/d)}} > \delta$$

and

$$O\left(\frac{d\log(1/\delta)}{\varepsilon^2}\|x-y\|_2^2\right) > O\left(\frac{s}{\varepsilon^2}\|x-y\|_2^2\right) \quad \Leftrightarrow \quad \log(1/\delta) > O\left(\frac{s}{d}\right) \quad \Leftrightarrow \quad \frac{1}{e^{O(s/d)}} > \delta.$$

Hence, in particular, the private SJLT has smaller variance than the private FJLT whenever

$$\delta < \min\left\{1/e^{O(s/d)}, 1/e^{O(sk/d)}\right\} = 1/e^{O(sk/d)} = 1/e^{O\left(\frac{\log^2(1/\beta)}{\alpha^3 d}\right)} = \beta^{O\left(\frac{\log(1/\beta)}{\alpha^3 d}\right)}.$$

The variance of the estimator from Theorem 2 by Kenthapadi et al. was

$$\mathrm{Var}[\hat{E}_{iid}] = \frac{2}{k}\|x-y\|_2^4 + O\left(\sigma^2\|x-y\|_2^2 + \sigma^4 k\right).$$

An argument similar to the one above proves that the variance of our private SJLT improves over the variance of Kenthapadi et al. when $\delta < e^{-s} = \beta^{O(1/\alpha)}$. Clearly, Kenthapadi et al. always achieves better variance than our private FJLT, due to the dependence on d which was inherent from perturbing the input rather than the output, and we may assume $k < d$.

Hence, we see a trade-off in running time versus variance, for certain values of input dimension d.

To sum up the above discussion, suppose that $\delta < \beta^{O(1/\alpha)}$. Then the private SJLT obtains the best variance out of all the methods. If d satisfies (5), then the private FJLT achieves the best running time, and otherwise, the private SJLT improves over the private FJLT in terms of both variance and running time.

Acknowledgements

This work was supported by Investigator Grant 16582, Basic Algorithms Research Copenhagen (BARC), from the VILLUM Foundation. I would like to thank my advisor Rasmus Pagh for the support, great discussions and pointing out several interesting limitations of previous work, leading to this work. I would also like to thank the reviewers for their excellent comments helping to improve this paper.

References

[1] Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. Syst. Sci.*, 66(4):671–687, 2003.

[2] Nir Ailon and Bernard Chazelle. The fast johnson–lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009.

[3] Michael Barbaro and Tom Zeller. A face is exposed for aol searcher no. 4417749. *New York Times*, 01 2006.

[4] Raghav Bhaskar, Abhishek Bhowmick, Vipul Goyal, Srivatsan Laxman, and Abhradeep Thakurta. Noiseless database privacy. In *17th International Conference on the Theory and Application of Cryptology and Information Security, ASIACRYPT*, volume 7073 of *Lecture Notes in Computer Science*, pages 215–232, 2011.

[5] Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. The Johnson-Lindenstrauss transform itself preserves differential privacy. In *53rd Symposium on Foundations of Computer Science, FOCS*, pages 410–419, 2012.

[6] Christos Boutsidis, Anastasios Zouzias, Michael W. Mahoney, and Petros Drineas. Randomized dimensionality reduction for k-means clustering. *IEEE Trans. Inf. Theory*, 61(2):1045–1062, 2015.

[7] Karl Bringmann, Fabian Kuhn, Konstantinos Panagiotou, Ueli Peter, and Henning Thomas. Internal DLA: efficient simulation of a physical growth model. In *Automata, Languages, and Programming, ICALP*, volume 8572 of *Lecture Notes in Computer Science*, pages 247–258, 2014.

[8] Clément L. Canonne, Gautam Kamath, and Thomas Steinke. The discrete gaussian for differential privacy. *CoRR*, abs/2004.00010, 2020.

[9] Rui Chen, Qian Xiao, Yu Zhang, and Jianliang Xu. Differentially private high-dimensional data publication via sampling-based inference. In *21st International Conference on Knowledge Discovery and Data Mining*, pages 129–138, 2015.

[10] Kenneth L. Clarkson. Tighter bounds for random projections of manifolds. In *24th Symposium on Computational Geometry*, pages 39–48, 2008.

[11] Kenneth L. Clarkson and David P. Woodruff. Numerical linear algebra in the streaming model. In *41st Symposium on Theory of Computing, STOC*, pages 205–214, 2009.

[12] Kenneth L. Clarkson and David P. Woodruff. Low rank approximation and regression in input sparsity time. In *Symposium on Theory of Computing Conference, STOC*, pages 81–90, 2013.

[13] Michael B. Cohen, Sam Elder, Cameron Musco, Christopher Musco, and Madalina Persu. Dimensionality reduction for k-means clustering and low rank approximation. In *47th Symposium on Theory of Computing, STOC*, pages 163–172, 2015.

[14] Anirban Dasgupta, Ravi Kumar, and Tamás Sarlós. A sparse Johnson-Lindenstrauss transform. In *42nd Symposium on Theory of Computing, STOC*, pages 341–350, 2010.

[15] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3:1376, 2013.

[16] Petros Drineas, Michael W. Mahoney, S. Muthukrishnan, and Tamás Sarlós. Faster least squares approximation. *Numerische Mathematik*, 117(2):219–249, 2011.

[17] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, volume 4004 of *Lecture Notes in Computer Science*, pages 486–503, 2006.

[18] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In *3rd Theory of Cryptography Conference, TCC*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284, 2006.

[19] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

[20] Differential Privacy Team Google. Secure noise generation. Technical report, Google, 2020.

[21] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In Leonard J. Schulman, editor, *42nd Symposium on Theory of Computing, STOC*, pages 705–714, 2010.

[22] Justin Hsu, Sanjeev Khanna, and Aaron Roth. Distributed private heavy hitters. In *Automata, Languages, and Programming, ICALP*, volume 7391 of *Lecture Notes in Computer Science*, pages 461–472, 2012.

[23] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006.

[24] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *30th Annual Symposium on the Theory of Computing, STOC*, pages 604–613, 1998.

[25] T. S. Jayram and David P. Woodruff. Optimal bounds for johnson-lindenstrauss transforms and streaming problems with sub-constant error. In *22nd Symposium on Discrete Algorithms, SODA*, pages 1–10, 2011.

[26] William B Johnson and Joram Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.

[27] Daniel M. Kane, Raghu Meka, and Jelani Nelson. Almost optimal explicit johnson-lindenstrauss families. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 14th International Workshop, APPROX and 15th International Workshop, RANDOM*, volume 6845 of *Lecture Notes in Computer Science*, pages 628–639, 2011.

[28] Daniel M. Kane and Jelani Nelson. Sparser Johnson-Lindenstrauss transforms. *J. ACM*, 61(1):4:1–4:23, 2014.

[29] Krishnaram Kenthapadi, Aleksandra Korolova, Ilya Mironov, and Nina Mishra. Privacy via the Johnson-Lindenstrauss transform. *J. Priv. Confidentiality*, 5(1), 2013.

[30] Andrew McGregor, Ilya Mironov, Toniann Pitassi, Omer Reingold, Kunal Talwar, and Salil Vadhan. The limits of two-party differential privacy. In *51st Annual Symposium on Foundations of Computer Science*, pages 81–90, 2010.

[31] Frank McSherry and Ilya Mironov. Differentially private recommender systems: Building privacy into the netflix prize contenders. In *15th International Conference on Knowledge Discovery and Data Mining*, pages 627–636, 2009.

[32] Darakhshan J. Mir, S. Muthukrishnan, Aleksandar Nikolov, and Rebecca N. Wright. Pan-private algorithms: When memory does not help. *CoRR*, abs/1009.1544, 2010.

[33] Darakhshan J. Mir, S. Muthukrishnan, Aleksandar Nikolov, and Rebecca N. Wright. Pan-private algorithms via statistics on sketches. In *30th Symposium on Principles of Database Systems, PODS*, pages 37–48, 2011.

[34] Ilya Mironov. On significance of the least significant bits for differential privacy. In *Conference on Computer and Communications Security, CCS*, pages 650–661, 2012.

[35] Ilya Mironov. Rényi differential privacy. In *30th Computer Security Foundations Symposium, CSF*, pages 263–275. IEEE Computer Society, 2017.

[36] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *Symposium on Security and Privacy*, pages 111–125, 2008.

[37] Jelani Nelson and Huy L. Nguyen. OSNAP: faster numerical linear algebra algorithms via sparser subspace embeddings. In *54th Symposium on Foundations of Computer Science, FOCS*, pages 117–126. IEEE Computer Society, 2013.

[38] Rasmus Pagh and Nina Mesing Stausholm. Efficient differentially private f_0 linear sketching. *CoRR*, abs/2001.11932, 2020.

[39] Wahbeh H. Qardaji, Weining Yang, and Ninghui Li. Priview: practical differentially private release of marginal contingency tables. In *International Conference on Management of Data, SIGMOD*, pages 1435–1446, 2014.

[40] Benjamin I. P. Rubinstein, Peter L. Bartlett, Ling Huang, and Nina Taft. Learning in a large function space: Privacy-preserving mechanisms for SVM learning. *J. Priv. Confidentiality*, 4(1), 2012.

[41] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM J. Comput.*, 40(6):1913–1926, 2011.

[42] Latanya Sweeney. Only you, your doctor, and many others may know. *Technology Science*, 2015092903(9):29, 2015.

[43] Pang-Ning Tan, Michael S. Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Addison-Wesley, 2005.

[44] Jalaj Upadhyay. Randomness efficient fast-Johnson-Lindenstrauss transform with applications in differential privacy and compressed sensing. *arXiv preprint arXiv:1410.2470*, 2014.

[45] Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

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

[47] Chugui Xu, Ju Ren, Yaoxue Zhang, Zhan Qin, and Kui Ren. Dppro: Differentially private high-dimensional data release via random projection. *IEEE Transactions on Information Forensics and Security*, 12(12):3081–3093, 2017.

[48] Jun Zhang, Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava, and Xiaokui Xiao. Privbayes: private data release via bayesian networks. In *International Conference on Management of Data, SIGMOD*, pages 1423–1434, 2014.

A Omitted Proofs for Technical Lemmas

Lemma 3. *We have*

1. $\hat{E}_{gen}$ *is an unbiased estimator for* $\|x-y\|_2^2$.

2. *The variance of* $\hat{E}_{gen}$ *is*

$$\mathrm{Var}\left[\hat{E}_{gen}\right] = \mathrm{Var}\left[\|Sx-Sy\|_2^2\right] + 8\,\mathrm{E}_{\mathcal{D}}[\eta_*^2]\|x-y\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^2]^2$$

Proof. We start by showing 1). For simpler notation, we define $z := x - y$. By independence and since $\mathrm{E}_{\mathcal{D}}[\eta_i] = 0$ for all i,

$$\begin{aligned}
\mathrm{E}_{S,\mathcal{D}}\left[\left\|(Sx+\eta)-(Sy+\mu)\right\|_2^2\right] &= \mathrm{E}_{S,\mathcal{D}}\left[\sum_{i=1}^{k}\left((Sx+\eta)_i-(Sy+\mu)_i\right)^2\right] \\
&= \mathrm{E}_{S,\mathcal{D}}\left[\sum_{i=1}^{k}\left((Sz)_i^2+(\eta_i-\mu_i)^2+2(\eta_i-\mu_i)(Sz)_i\right)\right] \\
&= \mathrm{E}_S\left[\sum_{i=1}^{k}(Sz)_i^2\right] + 2\sum_{i=1}^{k}\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right] = \|z\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]
\end{aligned}$$

where we in the last step used that S has the LPP. So clearly, re-inserting $z = x - y$

$$\mathrm{E}_{S,\mathcal{D}}\left[\left\|(Sx+\eta)-(Sy+\mu)\right\|_2^2 - 2k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\right] = \|x-y\|_2^2.$$

We turn to proving 2):

$$\mathrm{Var}\left[\hat{E}_{gen}\right] = \mathrm{E}_{S,\mathcal{D}}\left[\hat{E}_{gen}^2\right] - \mathrm{E}_{S,\mathcal{D}}\left[\hat{E}_{gen}\right]^2 = \mathrm{E}_{S,\mathcal{D}}\left[\hat{E}_{gen}^2\right] - \|x-y\|_2^4, \tag{6}$$

so we analyze the first term:

$$\begin{aligned}
\mathrm{E}_{S,\mathcal{D}}\left[\hat{E}_{gen}^2\right] &= \mathrm{E}_{S,\mathcal{D}}\left[\left(\left\|(Sx+\eta)-(Sy+\mu)\right\|_2^2 - 2k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\right)^2\right] \\
&= \mathrm{E}_{S,\mathcal{D}}\left[\left\|(Sx+\eta)-(Sy+\mu)\right\|_2^4\right] + 4k^2\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2 \\
&\quad - 4k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\mathrm{E}_{S,\mathcal{D}}\left[\left\|(Sx+\eta)-(Sy+\mu)\right\|_2^2\right]
\end{aligned} \tag{7}$$

The last term in (7) equals

$$4k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\left(\|x-y\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\right) = 4k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|x-y\|_2^2 + 8k^2\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2 \tag{8}$$

The first term in (7) equals

$$\begin{aligned}
\mathrm{E}_{S,\mathcal{D}}\left[\left\|(Sx+\eta)-(Sy+\mu)\right\|_2^4\right] &= \mathrm{E}_S\left[\|S(x-y)\|_2^4\right] + 4(k+2)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|x-y\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] \\
&\quad + 2k(1+2k)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2
\end{aligned}$$

The proof of the claim straightforward but tedious and thus left out here. It is proven formally in Appendix A.0.1.

Inserting (8) and Claim A into (7), we get

$$\begin{aligned}
\mathrm{E}_{S,\mathcal{D}}\left[\hat{E}_{gen}^2\right] &= \mathrm{E}_{S,\mathcal{D}}\left[\left(\left\|(Sx+\eta)-(Sy+\mu)\right\|_2^2 - 2k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\right)^2\right] \\
&= \mathrm{E}_S\left[\|S(x-y)\|_2^4\right] + 4(k+2)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|x-y\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 2k(1+2k)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2 \\
&\quad + 4k^2\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2 - 4k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|x-y\|_2^2 - 8k^2\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2 \\
&= \mathrm{E}_S\left[\|S(x-y)\|_2^4\right] + 8\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|x-y\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 2k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2.
\end{aligned}$$

Inserting this expression into (6) proves that the variance is

$$\begin{aligned}
\mathrm{Var}\left[\hat{E}_{gen}\right] &= \mathrm{E}_S\left[\|S(x-y)\|_2^4\right] + 8\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|x-y\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 2k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2 - \|x-y\|_2^4 \\
&= \mathrm{Var}\left[\|S(x-y)\|_2^2\right] + 8\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|x-y\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 2k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2,
\end{aligned}$$

again using that S satisfies LPP. □

A.0.1 Proof of Claim A

We repeat the claim for convenience: The first term in (7) equals

$$\mathrm{E}_{S,\mathcal{D}}\left[\left\|(Sx+\eta)-(Sy+\mu)\right\|_2^4\right] = \mathrm{E}_S\left[\|S(x-y)\|_2^4\right] + 4(k+2)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|x-y\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 2k(1+2k)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2$$

Proof. For a simpler notation, we define $z := x - y$. By simply unfolding the expression, we see that

$$\begin{aligned}
&= \sum_{i,\ell=1}^{k} \mathrm{E}_S\left[(Sz)_i^2(Sz)_\ell^2\right] + \sum_{i,\ell=1}^{k}\left(\mathrm{E}_{\mathcal{D}}\left[\eta_i^2\right]+\mathrm{E}_{\mathcal{D}}\left[\mu_i^2\right]\right)\mathrm{E}_S\left[(Sz)_\ell^2\right] \\
&+ 0 + \sum_{i,\ell=1}^{k}\mathrm{E}_S\left[(Sz)_i^2\right]\left(\mathrm{E}_{\mathcal{D}}\left[\eta_\ell^2\right]+\mathrm{E}_{\mathcal{D}}\left[\mu_\ell^2\right]\right) + \sum_{i=1}^{k}\mathrm{E}_{\mathcal{D}}\left[(\eta_i-\mu_i)^4\right] + \sum_{i\neq\ell}\mathrm{E}_{\mathcal{D}}\left[(\eta_i-\mu_i)^2\right]\mathrm{E}_{\mathcal{D}}\left[(\eta_\ell-\mu_\ell)^2\right] \\
&+ 2\sum_{i=1}^{k}\mathrm{E}_S\left[(Sz)_i\right]\mathrm{E}_{\mathcal{D}}\left[(\eta_i-\mu_i)^3\right] + 0 + 2\sum_{\ell=1}^{k}\mathrm{E}_{\mathcal{D}}\left[(\eta_\ell-\mu_\ell)^3\right]\mathrm{E}_S\left[(Sz)_\ell\right] + 4\sum_{i=1}^{k}\mathrm{E}_{\mathcal{D}}\left[(\eta_i-\mu_i)^2\right]\mathrm{E}_S\left[(Sz)_i^2\right] \\
&+ 4\sum_{i\neq\ell}\underbrace{\mathrm{E}_{\mathcal{D}}\left[(\eta_i-\mu_i)(\eta_\ell-\mu_\ell)\right]}_{=0}\mathrm{E}_S\left[(Sz)_i(Sz)_\ell\right]
\end{aligned}$$

where we used that $\mathrm{E}[\eta_i] = \mathrm{E}[\mu_i] = 0$ for all $i = 1, ..., k$ and that the noise is drawn independently of S.

Recalling that $\mathrm{E}_{\mathcal{D}}[\eta_i^2] = \mathrm{E}_{\mathcal{D}}[\mu_i^2] = \mathrm{E}_{\mathcal{D}}[\eta_*^2]$ for all i, we obtain

$$\begin{aligned}
&= \sum_{i,\ell=1}^{k}\mathrm{E}_S\left[(Sz)_i^2(Sz)_\ell^2\right] + 4k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|z\|_2^2 + k\left(2\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 6\,\mathrm{E}_{\mathcal{D}}[\eta_*^2]^2\right) + \sum_{i\neq\ell}4\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2 \\
&+ 2\sum_{i=1}^{k}\mathrm{E}_S\left[(Sz)_i\right]\left(\mathrm{E}_{\mathcal{D}}\left[\eta_*^3\right]-\mathrm{E}_{\mathcal{D}}\left[\eta_*^3\right]\right) + 2\sum_{\ell=1}^{k}\mathrm{E}_S\left[(Sz)_\ell\right]\left(\mathrm{E}_{\mathcal{D}}\left[\eta_*^3\right]-\mathrm{E}_{\mathcal{D}}\left[\eta_*^3\right]\right) \\
&+ 4\sum_{i=1}^{k}2\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\mathrm{E}_S\left[(Sz)_i^2\right]
\end{aligned}$$

which simplifies to

$$\begin{aligned}
&= \mathrm{E}_S\left[\|Sz\|_2^4\right] + 4k\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|z\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 6k\,\mathrm{E}_{\mathcal{D}}[\eta_*^2]^2 + 4(k^2-k)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2 + 8\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|z\|_2^2 \\
&= \mathrm{E}_S\left[\|Sz\|_2^4\right] + 4(k+2)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|z\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 2k(1+2k)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2
\end{aligned}$$

Re-inserting $z = x - y$, we conclude that

$$\mathrm{E}_{S,\mathcal{D}}\left[\left\|(Sx+\eta)-(Sy+\mu)\right\|_2^4\right] = \mathrm{E}_S\left[\|S(x-y)\|_2^4\right] + 4(k+2)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]\|x-y\|_2^2 + 2k\,\mathrm{E}_{\mathcal{D}}[\eta_*^4] + 2k(1+2k)\,\mathrm{E}_{\mathcal{D}}\left[\eta_*^2\right]^2.$$

□

B Omitted Proofs for FJLT

B.1 Primitives

This section will give some primitives that will be useful in the next section. Non-trivial arguments can be found in Section B.1.1. We let $\Phi = PHD$ be the FJLT transform as described in Section 5.1 and $x, y \in \mathbb{R}^d$ be any real vectors. Let $X \sim \mathcal{N}(0, q^{-1})$. Then for any $i, n \in [k]$ and $j, \ell \in [d]$

$$\mathrm{E}_P[P_{ij}] = 0, \qquad \mathrm{E}_P\left[P_{ij}^2\right] = q \cdot \mathrm{E}_X\left[X^2\right] = 1$$

$$\mathrm{E}_P\left[P_{ij}^4\right] = q \cdot \mathrm{E}_X\left[X^4\right] = q \cdot 3q^{-2} = \frac{3}{q}.$$

$$\mathrm{E}_D[D_{jj}] = 0, \qquad \mathrm{E}_D[D_{jj}^2] = D_{jj}^2 = 1.$$

$$\mathrm{E}[\Phi_{ij}] = \sum_{f=1}^{d} \mathrm{E}_P\left[P_{if}\right] H_{fj}\, \mathrm{E}_D\left[D_{jj}\right] = 0$$

$$\mathrm{E}_\Phi[\Phi_{ij}\Phi_{n\ell}] = \begin{cases} \mathrm{E}_\Phi[\Phi_{ij}]\,\mathrm{E}_\Phi[\Phi_{n\ell}] = 0, & i \neq n,\ j \neq \ell \\ \mathrm{E}_P\left[(PH)_{ij}\right]\mathrm{E}_P\left[(PH)_{nj}\right] = 0, & i \neq n,\ j = \ell \\ \mathrm{E}_\Phi[\Phi_{ij}]\,\mathrm{E}_\Phi[\Phi_{i\ell}] = 0, & i = n,\ j \neq \ell \\ \mathrm{E}_\Phi[\Phi_{ij}^2] = 1, & i = n,\ j = \ell \end{cases}$$

$$\mathrm{E}_\Phi[\Phi_{ij}^2\Phi_{n\ell}^2] = \begin{cases} 1, & i \neq n \\ \frac{3}{qd} + 1 - \frac{3}{d}, & i = n,\ j \neq \ell \\ \frac{3}{qd} + 3 - \frac{3}{d}, & i = n,\ j = \ell \end{cases} \tag{9}$$

$$\mathrm{E}_\Phi[\Phi_{ij}\Phi_{i\ell}\Phi_{nv}\Phi_{nw}] = \begin{cases} \mathrm{E}_\Phi[\Phi_{ij}^2\Phi_{nv}^2], & j = \ell,\ v = w \\ \mathrm{E}_\Phi[\Phi_{ij}^2\Phi_{i\ell}^2], & i = n,\ (j = v,\ \ell = w) \vee (j = w,\ \ell = v) \\ 0, & \text{otherwise} \end{cases}$$

$$\mathrm{E}_\Phi[(\Phi x)_i(\Phi y)_n] = \sum_{j,\ell=1}^{d} x_j y_\ell\, \mathrm{E}_\Phi[\Phi_{ij}\Phi_{n\ell}] = \begin{cases} 0, & i \neq n \\ \sum_{j=1}^{d} x_j y_j, & i = n \end{cases}$$

$$\mathrm{E}_\Phi[(\Phi x)_i^2(\Phi y)_n^2] = \|x\|_2^2\|y\|_2^2, \qquad i \neq n \tag{10}$$

$$\mathrm{E}_\Phi[(\Phi x)_i^2(\Phi y)_i^2] = \frac{3}{d}\left(\frac{d}{3} + \left(\frac{1}{q} - 1\right)\right)\left(\|x\|_2^2\|y\|_2^2 + 2\langle x, y\rangle^2\right) - \frac{6}{d}\left(\frac{1}{q} - 1\right)\sum_{j=1}^{d} x_j^2 y_j^2 \tag{11}$$

B.1.1 Arguments for Primitives

Argument for (9)

We use that

$$\mathrm{E}_\Phi[\Phi_{ij}^2\Phi_{n\ell}^2] = \sum_{f,g,h,s=1}^{d} \mathrm{E}_P[P_{if}P_{ig}P_{nh}P_{ns}]H_{fj}H_{gj}H_{h\ell}H_{s\ell}\,\mathrm{E}_D[D_{jj}^2D_{\ell\ell}^2]$$

and so for $i \neq n$

$$\sum_{f,h=1}^{d} \mathrm{E}_P\left[P_{if}^2\right]\mathrm{E}_P\left[P_{nh}^2\right] H_{fj}^2 H_{h\ell}^2 = 1$$

and for $i = n$

$$\begin{aligned} \sum_{f,g,h,s=1}^{d} \mathrm{E}_P[P_{if}P_{ig}P_{nh}P_{ns}]H_{fj}H_{gj}H_{h\ell}H_{s\ell} &= \sum_{f=1}^{d} \mathrm{E}_P[P_{if}^4]H_{fj}^2H_{f\ell}^2 + \sum_{f\neq h=1}^{d} \mathrm{E}_P[P_{if}^2]\,\mathrm{E}_P[P_{ih}^2]H_{fj}^2H_{h\ell}^2 \\ &\quad + 2\sum_{f\neq\ell=1}^{d} \mathrm{E}_P[P_{if}^2]\,\mathrm{E}_P[P_{ig}^2]H_{fj}H_{gj}H_{f\ell}H_{g\ell} \\ &= \mathrm{E}_P[P_{if}^4]/d + (d^2 - d)/d^2 + 2(\langle H_j, H_\ell\rangle^2 - 1/d) \\ &= \mathrm{E}_P[P_{if}^4]/d + 1 - 3/d + 2\langle H_j, H_\ell\rangle^2 \\ &= \begin{cases} \mathrm{E}_P[P_{if}^4]/d + 1 - 3/d, & j \neq \ell \\ \mathrm{E}_P[P_{if}^4]/d + 3 - 3/d, & j = \ell \end{cases} \end{aligned}$$

because $\langle H_j, H_\ell \rangle = \begin{cases} 0, & j \neq \ell \\ 1, & j = \ell. \end{cases}$

Argument for (10) and (11)

We used that

$$\mathrm{E}_\Phi[(\Phi x)_i^2 (\Phi y)_n^2] = \sum_{j,v=1}^{d} x_j^2 y_v^2 \, \mathrm{E}_\Phi[\Phi_{ij}^2 \Phi_{nv}^2] + 2 \sum_{j \neq \ell = 1}^{d} x_j^2 y_\ell^2 \, \mathrm{E}_\Phi[\Phi_{ij}\Phi_{i\ell}\Phi_{nj}\Phi_{n\ell}]$$

where we for $i \neq n$ get $\|x\|_2^2 \|y\|_2^2$ and for $i = n$ get

$$\begin{aligned}
&\sum_{j=1}^{d} x_j^2 y_j^2 \, \mathrm{E}_\Phi[\Phi_{ij}^4] + \sum_{j \neq v=1}^{d} x_j^2 y_v^2 \, \mathrm{E}_\Phi[\Phi_{ij}^2 \Phi_{iv}^2] + 2 \sum_{j \neq \ell = 1}^{d} x_j x_\ell y_j y_\ell \, \mathrm{E}_\Phi[\Phi_{ij}^2 \Phi_{i\ell}^2] \\
&= \sum_{j=1}^{d} x_j^2 y_j^2 \left(\frac{3}{qd} + 3 - \frac{3}{d} \right) + \sum_{j \neq v=1}^{d} x_j^2 y_v^2 \left(\frac{3}{qd} + 1 - \frac{3}{d} \right) + 2 \sum_{j \neq \ell = 1}^{d} x_j x_\ell y_j y_\ell \left(\frac{3}{qd} + 1 - \frac{3}{d} \right) \\
&= \sum_{j=1}^{d} x_j^2 y_j^2 \left(\frac{3}{qd} + 3 - \frac{3}{d} \right) + \left(\|x\|_2^2 \|y\|_2^2 - \sum_{j=1}^{d} x_j^2 y_j^2 \right) \left(\frac{3}{qd} + 1 - \frac{3}{d} \right) + 2 \left(\langle x, y \rangle^2 - \sum_{j=1}^{d} x_j^2 y_j^2 \right) \left(\frac{3}{qd} + 1 - \frac{3}{d} \right) \\
&= \frac{3}{d} \left(\frac{d}{3} + \left(\frac{1}{q} - 1 \right) \right) \left(\|x\|_2^2 \|y\|_2^2 + 2\langle x, y \rangle^2 \right) - \frac{6}{d} \left(\frac{1}{q} - 1 \right) \sum_{j=1}^{d} x_j^2 y_j^2
\end{aligned}$$

B.2 Proof of FJLT Satisfying LPP

Lemma 6. *The normalized FJLT satisfies LPP (see Definition 4).*

Proof. Applying the primitives from Appendix B.1, we get

$$\mathrm{E}_\Phi \left[\frac{1}{k} \|\Phi x\|_2^2 \right] = \frac{1}{k} \mathrm{E}_\Phi \left[\sum_{i=1}^{k} (\Phi x)_i^2 \right] = \frac{1}{k} \sum_{i=1}^{k} \sum_{j,\ell=1}^{d} \mathrm{E}_\Phi \left[\Phi_{ij} \Phi_{i\ell} \right] x_j x_\ell = \frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{d} \mathrm{E}_\Phi \left[\Phi_{ij}^2 \right] x_j^2 = \|x\|_2^2.$$

□

B.3 Variance under FJLT

For convenience, we prove the following result, as it will be useful in this form for several other proofs. Note that Lemma 7 follows directly from Lemma 11.

Lemma 11. *Let $k \times d$-matrix $\Phi = PHD$, where P_{ij} is $\mathcal{N}(0, q^{-1})$ with probability q and and 0 otherwise. For input vector $\eta \sim \mathcal{D}^d$ for a real-valued distribution $\mathcal{D}$:*

$$\mathrm{Var}[\|\Phi \eta\|_2^2] \leq \frac{3}{k} \mathrm{E}_\eta \left[\|\eta\|_2^2 \right].$$

For $x \in \mathbb{R}^d$, we get

$$\mathrm{Var}[\|\Phi x\|_2^2] \leq \frac{3}{k} \|x\|_2^2.$$

Proof.

$$\begin{aligned}
\operatorname{Var}[\|\Phi\eta\|_2^2] &= \mathrm{E}_{\Phi,\eta}[\|\Phi\eta\|_2^4] - \mathrm{E}_{\Phi,\eta}[\|\Phi\eta\|_2^2]^2 = \sum_{i,n=1}^{k} \mathrm{E}_{\Phi,\eta}\left[(\Phi\eta)_i^2(\Phi\eta)_n^2\right] - k^2\,\mathrm{E}_\eta\left[\|\eta\|_2^4\right] \\
&= \sum_{i=1}^{k} \mathrm{E}_{\Phi,\eta}\left[(\Phi\eta)_i^4\right] + \sum_{i\neq n=1}^{k} \mathrm{E}_{\Phi}\left[(\Phi\eta)_i^2(\Phi\eta)_n^2\right] - k^2\,\mathrm{E}_\eta\left[\|\eta\|_2^4\right] \\
&= \frac{9k}{d}\left(\frac{d}{3} + \left(\frac{1}{q} - 1\right)\right)\mathrm{E}_\eta\left[\|\eta\|_2^4\right] - \frac{6k}{d}\left(\frac{1}{q} - 1\right)\mathrm{E}_\eta\left[\|\eta\|_4^4\right] - k^2\,\mathrm{E}_\eta\left[\|\eta\|_2^4\right] \\
&= 3k\left(\frac{2}{3} + \frac{3}{d}\left(\frac{1}{q} - 1\right)\right)\mathrm{E}_\eta\left[\|\eta\|_2^4\right] - \frac{6k}{d}\left(\frac{1}{q} - 1\right)\mathrm{E}_\eta\left[\|\eta\|_4^4\right].
\end{aligned}$$

which again implies

$$\begin{aligned}
\operatorname{Var}\left[\frac{1}{k}\|\Phi\eta\|_2^2\right] &= \frac{3}{k}\left(\frac{2}{3} + \frac{3}{d}\left(\frac{1}{q} - 1\right)\right)\mathrm{E}_\eta\left[\|\eta\|_2^4\right] - \frac{6}{dk}\left(\frac{1}{q} - 1\right)\mathrm{E}_\eta\left[\|\eta\|_4^4\right] \\
&\leq \frac{3\,\mathrm{E}_\eta\left[\|\eta\|_2^4\right]}{k}\left(\frac{2}{3} + \frac{3}{d}\left(\frac{1}{q} - 1\right)\right) \leq \frac{3}{k}\,\mathrm{E}_\eta\left[\|\eta\|_2^4\right]
\end{aligned}$$

when $q \geq \frac{1}{d/9+1}$. □

C Omitted Proofs for Private FJLT

C.1 Estimator and Variance for Private FJLT

Lemma 12. *We have*

1. *$\hat{E}_{FJLT_i}$ is an unbiased estimator for $\|x-y\|_2^2$.*
2. *$\operatorname{Var}[\hat{E}_{FJLT_i}] \leq \frac{3}{k}\|x-y\|_2^4 + O\left(\frac{d^2\sigma^4}{k} + d\sigma^2\|x-y\|_2^2\right)$.*

Proof. We repeatedly apply the primitives of Section B.1 and Section B.3.

We start by proving 1). Observe that

$$\begin{aligned}
\mathrm{E}\left[\|\Phi(x+\eta) - \Phi(y+\mu)\|_2^2\right] &= \mathrm{E}\left[\|\Phi(x-y) + \Phi(\eta-\mu)\|_2^2\right] \\
&= \mathrm{E}\left[\|\Phi(x-y)\|_2^2\right] + \mathrm{E}\left[\|\Phi(\eta-\mu)\|_2^2\right] \\
&= k\|x-y\|_2^2 + k\,\mathrm{E}_{\eta,\mu}\left[\|\eta-\mu\|_2^2\right]
\end{aligned}$$

Since $\eta, \mu \sim \mathcal{N}(0,\sigma^2)^d$, we have $\eta - \mu \sim \mathcal{N}(0, 2\sigma^2)^d$ and so

$$\mathrm{E}_\eta\left[\|\eta-\mu\|_2^2\right] = \sum_{j=1}^{d} \mathrm{E}_{\eta,\mu}[(\eta_j - \mu_j)^2] = 2d\sigma^2.$$

We conclude that

$$\hat{E}_{FJLT_i} = 1/k\|\Phi(x+\eta) - \Phi(y+\mu)\|_2^2 - 2d\sigma^2$$

is an unbiased estimator for $\|x-y\|_2^2$.

We turn to proving 2). Note that

$$\operatorname{Var}\left[1/k\|\Phi(x-y) + \Phi(\eta-\mu)\|_2^2 - 2d\sigma^2\right] = \frac{1}{k^2}\operatorname{Var}\left[\|\Phi(x-y) + \Phi(\eta-\mu)\|_2^2\right],$$

so it suffices to consider the RHS. For readability, we will do the analysis for x and η, and eventually substitute x for $x-y$ and η for $\eta-\mu$, recalling that if $\eta \sim \mathcal{N}(0,\sigma^2)$, then $\eta - \mu \sim \mathcal{N}(0, 2\sigma^2)$.

For any $x, \eta \in \mathbb{R}^d$

$$\begin{aligned}
\mathrm{E}_{\Phi,\eta}\left[\|\Phi x + \Phi\eta\|_2^2\right]^2 &= \left(\mathrm{E}_\Phi\left[\|\Phi x\|_2^2\right] + \mathrm{E}_\Phi\left[\|\Phi\eta\|_2^2\right]\right)^2 \\
&= \mathrm{E}_\Phi\left[\|\Phi x\|_2^2\right]^2 + \mathrm{E}_\Phi\left[\|\Phi\eta\|_2^2\right]^2 + 2\,\mathrm{E}_\Phi\left[\|\Phi x\|_2^2\right]\mathrm{E}_\Phi\left[\|\Phi\eta\|_2^2\right]
\end{aligned}$$

By the triangle inequality, we see that

$$\begin{aligned}
\mathrm{E}_{\Phi,\eta}\left[\|\Phi x + \Phi\eta\|_2^4\right] &= \mathrm{E}_{\Phi,\eta}\left[\left(\|\Phi x + \Phi\eta\|_2^2\right)^2\right] \\
&\leq \mathrm{E}_{\Phi,\eta}\left[\left(\|\Phi x\|_2^2 + \|\Phi\eta\|_2^2 + 2\|\Phi x\|_2\|\Phi\eta\|_2\right)^2\right] \\
&= \mathrm{E}_\Phi\left[\|\Phi x\|_2^4\right] + \mathrm{E}_{\Phi,\eta}\left[\|\Phi\eta\|_2^4\right] + 6\,\mathrm{E}_{\Phi,\eta}\left[\|\Phi x\|_2^2\|\Phi\eta\|_2^2\right]
\end{aligned}$$

Where the last equality follows from the zero-meaned η leading to a several terms cancelling out.

Hence, the variance is bounded by

$$\begin{aligned}
\mathrm{Var}\left[\|\Phi x + \Phi\eta\|_2^2\right] \leq\ & \mathrm{E}_\Phi\left[\|\Phi x\|_2^4\right] + \mathrm{E}_{\Phi,\eta}\left[\|\Phi\eta\|_2^4\right] + 6\,\mathrm{E}_{\Phi,\eta}\left[\|\Phi x\|_2^2\|\Phi\eta\|_2^2\right] - \mathrm{E}_\Phi\left[\|\Phi x\|_2^2\right]^2 - \mathrm{E}_{\Phi,\eta}\left[\|\Phi\eta\|_2^2\right]^2 \\
& - 2\,\mathrm{E}_\Phi\left[\|\Phi x\|_2^2\right]\mathrm{E}_{\Phi,\eta}\left[\|\Phi\eta\|_2^2\right]
\end{aligned}$$

which again implies

$$\begin{aligned}
\mathrm{Var}\left[1/k\|\Phi x + \Phi\eta\|_2^2\right] \leq\ & \mathrm{Var}_\Phi\left[1/k\|\Phi x\|_2^2\right] + \mathrm{Var}_{\Phi,\eta}\left[1/k\|\Phi\eta\|_2^2\right] + \frac{6}{k^2}\,\mathrm{E}_{\Phi,\eta}\left[\|\Phi x\|_2^2\|\Phi\eta\|_2^2\right] \\
& - \frac{2}{k^2}\,\mathrm{E}_\Phi\left[\|\Phi x\|_2^2\right]\mathrm{E}_{\Phi,\eta}\left[\|\Phi\eta\|_2^2\right]
\end{aligned} \tag{12}$$

For the last term we have

$$\mathrm{E}_\Phi\left[\|\Phi x\|_2^2\right]\mathrm{E}_{\Phi,\eta}\left[\|\Phi\eta\|_2^2\right] = 2k^2\|x\|_2^2\,\mathrm{E}_\eta[\|\eta\|_2^2] = 2k^2 d\sigma^2\|x\|_2^2$$

and for the second to last term we get:

$$\begin{aligned}
\mathrm{E}_{\Phi,\eta}\left[\|\Phi x\|_2^2\|\Phi\eta\|_2^2\right] &= \sum_{i,n=1}^{k}\mathrm{E}_{\Phi,\eta}[(\Phi x)_i^2(\Phi\eta)_n^2] = \sum_{i=1}^{k}\mathrm{E}_{\Phi,\eta}[(\Phi x)_i^2(\Phi\eta)_i^2] + \sum_{i\neq n=1}^{k}\mathrm{E}_{\Phi,\eta}[(\Phi x)_i^2(\Phi\eta)_n^2] \\
&= \frac{3k}{d}\left(\frac{d}{3} - \left(1 - \frac{1}{q}\right)\right)\left(\|x\|_2^2\,\mathrm{E}\left[\|\eta\|_2^2\right] + 2\,\mathrm{E}\left[\langle x, \eta\rangle^2\right]\right) + \frac{6k}{d}\left(1 - \frac{1}{q}\right)\sum_{j=1}^{d} x_j^2\,\mathrm{E}\left[\eta_j^2\right] \\
&\quad + (k^2 - k)\|x\|_2^2\,\mathrm{E}_\eta[\|\eta\|_2^2] \\
&= \frac{3k}{d}\left(\frac{d}{3} - \left(1 - \frac{1}{q}\right)\right)\left(\|x\|_2^2\,\mathrm{E}\left[\|\eta\|_2^2\right] + 2\|x\|_2^2\,\mathrm{E}_\eta[\eta_*^2]\right) + \frac{6k}{d}\left(1 - \frac{1}{q}\right)\|x\|_2^2\,\mathrm{E}\left[\eta_*^2\right] \\
&\quad + (k^2 - k)\|x\|_2^2\,\mathrm{E}_\eta[\|\eta\|_2^2] \\
&= k\left(k - \frac{3}{d}\left(1 - \frac{1}{q}\right)\right)\|x\|_2^2 d\sigma^2 + 2k\|x\|_2^2\sigma^2 \\
&= k\|x\|_2^2\sigma^2\left(kd + 2 - 3\left(1 - \frac{1}{q}\right)\right)
\end{aligned}$$

Inserting into (12) and applying Section B.3, we see that

$$\begin{aligned}
\mathrm{Var}[1/k\|\Phi x + \Phi\eta\|_2^2] &\leq \frac{3}{k}\|x\|_2^4 + O\left(\frac{d^2\sigma^4}{k}\right) + \frac{6}{k}\|x\|_2^2\sigma^2\left(kd + 2 - 3\left(1 - \frac{1}{q}\right)\right) - 4d\sigma^2\|x\|_2^2 \\
&= \frac{3}{k}\|x\|_2^4 + O\left(\frac{d^2\sigma^4}{k}\right) + \frac{2}{k}\|x\|_2^2\sigma^2\left(kd + 6 + 9\left(\frac{1}{q} - 1\right)\right)
\end{aligned}$$

Substituting x for $x - y$ and η for $\eta - \mu$ proves that

$$\mathrm{Var}[1/k\|\Phi(x + \eta) - \Phi(y + \mu)\|_2^2 - 2d\sigma^2] \leq \frac{3}{k}\|x - y\|_2^4 + O\left(\frac{d^2\sigma^4}{k} + d\sigma^2\|x\|_2^2 + \frac{\sigma^2}{qk}\|x\|_2^2\right)$$

Recalling that $q = \min\left\{\Theta\left(\frac{\log k}{d}\right), 1\right\}$ we get

$$\frac{3}{k}\|x-y\|_2^4 + O\left(\frac{d^2\sigma^4}{k} + d\sigma^2\|x\|_2^2\right)$$

concluding the proof. □

D Omitted Proofs for SJLT

D.1 Proof of SJLT Satisfying LPP

Lemma 9. *The SJLT as described above satisfy LPP from Definition 4.*

Proof. We show the result here for the c)-construction. A similar proof shows the result for the b)-construction.

$$\begin{aligned}
\mathrm{E}_S\left[\|Sx\|_2^2\right] &= \mathrm{E}_S\left[\sum_{i=1}^{k/s}\sum_{r=1}^{s}(Sx)_{(i,r)}^2\right] = \frac{1}{s}\,\mathrm{E}_{h,\varphi}\left[\sum_{i=1}^{k/s}\sum_{r=1}^{s}\left(\sum_{j=1}^{d}\varphi_r(j)\xi_{ri}(j)x_j\right)^2\right] \\
&= \frac{1}{s}\,\mathrm{E}_S\left[\sum_{i=1}^{k/s}\sum_{r=1}^{s}\sum_{j,\ell=1}^{d}\varphi_r(j)\varphi_r(\ell)\xi_{ri}(j)\xi_{ri}(\ell)x_jx_\ell\right] \\
&= \frac{1}{s}\sum_{j=1}^{d}x_j^2\sum_{i=1}^{k/s}\sum_{r=1}^{s}\mathrm{E}_h\left[\xi_{ri}(j)\right] = \|x\|_2^2
\end{aligned}$$

because $\varphi_r(j)$ and $\varphi_r(\ell)$ are independent for $j \neq \ell$ and $\mathrm{E}_\varphi[\varphi_r(j)] = 0$. □

D.2 Proof of Variance of (non-private) SJLT

The following lemma will be useful throughout this appendix. The proof is immediate from the definition of ξ.

Lemma 13.

$$\begin{aligned}
\mathrm{E}_\xi[\xi_{ri}(j)\xi_{tn}(\ell)] &= \begin{cases}
\mathrm{E}_\xi[\xi_{ri}(j)]\,\mathrm{E}_\xi[\xi_{tn}(\ell)], & j \neq \ell \\
\mathrm{E}_\xi[\xi_{ri}(j)^2], & r = t,\ i = n,\ j = \ell \\
\mathrm{E}_\xi[\xi_{ri}(j)]\,\mathrm{E}_\xi[\xi_{ti}(j)], & r \neq t,\ i = n,\ j = \ell \\
0, & r = t,\ i \neq n,\ j = \ell \\
\mathrm{E}_\xi[\xi_{ri}(j)]\,\mathrm{E}_\xi[\xi_{tn}(j)], & r \neq t,\ i \neq n,\ j = \ell
\end{cases} \\
&= \begin{cases}
s^2/k^2, & j \neq \ell \\
s/k, & r = t,\ i = n,\ j = \ell \\
s^2/k^2, & r \neq t,\ i = n,\ j = \ell \\
0, & r = t,\ i \neq n,\ j = \ell \\
s^2/k^2, & r \neq t,\ i \neq n,\ j = \ell
\end{cases}
\end{aligned}$$

where we recalled that

$$h_r(j) = i \wedge h_r(j) = n, \qquad \Leftrightarrow \qquad i = n.$$

We now prove Lemma 10. We state it here for convenience.

Lemma 10. *Let $x, y \in \mathbb{R}^d$ and let S be the SJLT as described above. Then*

$$\mathrm{Var}\left[\|Sx - Sy\|_2^2\right] \leq \frac{2}{k}\|x-y\|_2^4.$$

Proof. Throughout the proof, we will apply Lemma 13 without further comment. By linearity of S and since S satisfies LPP, it is sufficient to show that for $x \in \mathbb{R}^d$

$$\operatorname{Var}\left[\|Sx\|_2^2\right] = \mathrm{E}_S\left[\|Sx\|_2^4\right] - \left(\mathrm{E}_S\left[\|Sx\|_2^2\right]\right)^2 = \mathrm{E}_S\left[\|Sx\|_2^4\right] - \|x\|_2^4 \leq \frac{2}{k}\|x\|_2^4.$$

We will consider the first term:

$$\begin{aligned}\mathrm{E}_S\left[\|Sx\|_2^4\right] &= \mathrm{E}_S\left[\left(\sum_{i=1}^{k/s}\sum_{r=1}^{s}(Sx)_{(i,r)}^2\right)^2\right] = \mathrm{E}_S\left[\left(\sum_{i=1}^{k/s}\sum_{r=1}^{s}\left(\sum_{j=1}^{d}\frac{1}{\sqrt{s}}x_j\varphi_r(j)\xi_{ri}(j)\right)^2\right)^2\right] \\ &= \frac{1}{s^2}\mathrm{E}_S\left[\left(\sum_{i=1}^{k/s}\sum_{r=1}^{s}\sum_{j,\ell=1}^{d}x_jx_\ell\varphi_r(j)\varphi_r(\ell)\xi_{ri}(j)\xi_{ri}(\ell)\right)^2\right] \qquad (13)\end{aligned}$$

Letting

$$a = \sum_{i=1}^{k/s}\sum_{r=1}^{s}\sum_{j=1}^{d}x_j^2\xi_{ri}(j)$$

and

$$b = \sum_{i=1}^{k/s}\sum_{r=1}^{s}\sum_{j\neq\ell}x_jx_\ell\varphi_r(j)\varphi_r(\ell)\xi_{ri}(j)\xi_{ri}(\ell).$$

we can express (13) as

$$\frac{1}{s^2}\mathrm{E}_S\left[(a+b)^2\right] = \frac{1}{s^2}\mathrm{E}_S\left[a^2+b^2+2ab\right] \qquad (14)$$

The proofs of the following claims are straightforward but tedious and thus we leave them out here. They can be found in Appendix D.2.1.

$$\mathrm{E}_S\left[a^2\right] = s^2\|x\|_2^4.$$

$$\mathrm{E}_S\left[b^2\right] = \frac{2s^2}{k}\left(\|x\|_2^4 - \|x\|_4^4\right).$$

$$2\,\mathrm{E}_S\left[ab\right] = 0$$

Inserting Claims D.2-D.2 into (14), we conclude that

$$\mathrm{E}_S\left[\|Sx\|_2^4\right] = \|x\|_2^4 + \frac{2}{k}\left(\|x\|_2^4 - \|x\|_4^4\right)$$

finally proving that

$$\operatorname{Var}\left[\|Sx\|_2^2\right] = \frac{2}{k}\left(\|x\|_2^4 - \|x\|_4^4\right) \leq \frac{2}{k}\|x\|_2^4.$$

□

D.2.1 Proof of Claims

Throughout this section, we apply Lemma 13 repeatedly without further comment.

$$\mathrm{E}_S\left[a^2\right] = s^2\|x\|_2^4.$$

Proof.

$$\begin{aligned}
\mathrm{E}_h\left[\left(\sum_{i=1}^{k/s}\sum_{r=1}^{s}\sum_{j=1}^{d} x_j^2\xi_{ri}(j)\right)^2\right] &= \sum_{i,n=1}^{k/s}\sum_{r,t=1}^{s}\sum_{j,\ell=1}^{d} x_j^2x_\ell^2\,\mathrm{E}_h\left[\xi_{ri}(j)\xi_{tn}(\ell)\right] \\
&= \sum_{i,n=1}^{k/s}\sum_{r,t=1}^{s}\sum_{j=1}^{d} x_j^4\,\mathrm{E}_h\left[\xi_{ri}(j)\xi_{tn}(j)\right] + \sum_{i,n=1}^{k/s}\sum_{r,t=1}^{s}\sum_{j\neq\ell=1}^{d} x_j^2x_\ell^2\,\mathrm{E}_h\left[\xi_{ri}(j)\xi_{tn}(\ell)\right] \\
&= \sum_{i=1}^{k/s}\sum_{r=1}^{s}\sum_{j=1}^{d} x_j^4\frac{s}{k} + \sum_{i=1}^{k/s}\sum_{r\neq t=1}^{s}\sum_{j=1}^{d} x_j^4\frac{s^2}{k^2} + \sum_{i\neq n=1}^{k/s}\sum_{r=1}^{s}\sum_{j=1}^{d} x_j^4\cdot 0 \\
&\quad + \sum_{i\neq n=1}^{k/s}\sum_{r\neq t=1}^{s}\sum_{j=1}^{d} x_j^4\frac{s^2}{k^2} + \sum_{i,n=1}^{k/s}\sum_{r,t=1}^{s}\sum_{j\neq\ell=1}^{d} x_j^2x_\ell^2\frac{s^2}{k^2} \\
&= s\|x\|_4^4 + \frac{s(s^2-s)}{k}\|x\|_4^4 + \left(1-\frac{s}{k}\right)(s^2-s)\|x\|_4^4 + s^2\left(\|x\|_2^4 - \|x\|_4^4\right) \\
&= s^2\|x\|_2^4
\end{aligned}$$

□

$$\mathrm{E}_S\left[b^2\right] = \frac{2s^2}{k}\left(\|x\|_2^4 - \|x\|_4^4\right).$$

Proof.

$$\begin{aligned}
\mathrm{E}_{h,\varphi}\left[\left(\sum_{i=1}^{k/s}\sum_{r=1}^{s}\sum_{j\neq\ell} x_jx_\ell\varphi_r(j)\varphi_r(\ell)\xi_{ri}(j)\xi_{ri}(\ell)\right)^2\right] &= \sum_{i,n=1}^{k/s}\sum_{r,t=1}^{s}\sum_{\substack{j\neq\ell\\ v\neq w}} x_jx_\ell x_vx_w\,\mathrm{E}_\varphi\left[\varphi_r(j)\varphi_r(\ell)\varphi_t(v)\varphi_t(w)\right] \\
&\qquad\cdot\mathrm{E}_h\left[\xi_{ri}(j)\xi_{ri}(\ell)\xi_{tn}(v)\xi_{tn}(w)\right]
\end{aligned}$$

We remark that, as $j\neq\ell$, we have

$$\mathrm{E}_\varphi\left[\varphi_r(j)\varphi_r(\ell)\varphi_t(v)\varphi_t(w)\right] = \begin{cases} 1, & r=t\wedge((j=v\wedge\ell=w)\vee(j=w\wedge\ell=v)) \\ 0, & \text{otherwise.}\end{cases}$$

This leaves us with

$$\begin{aligned}
2\sum_{i,n=1}^{k/s}\sum_{r=1}^{s}\sum_{j\neq\ell} x_j^2x_\ell^2\,\mathrm{E}_h\left[\xi_{ri}(j)\xi_{ri}(\ell)\xi_{rn}(j)\xi_{rn}(\ell)\right] &= 2\sum_{i=1}^{k/s}\sum_{r=1}^{s}\sum_{j\neq\ell} x_j^2x_\ell^2\,\mathrm{E}_h\left[\xi_{ri}(j)^2\right]\mathrm{E}_h\left[\xi_{ri}(\ell)^2\right] \\
&= 2\sum_{i=1}^{k/s}\sum_{r=1}^{s}\sum_{j\neq\ell} x_j^2x_\ell^2\frac{s^2}{k^2} = \frac{2s^2}{k}\left(\|x\|_2^4 - \|x\|_4^4\right)
\end{aligned}$$

where we used that $\mathrm{E}_h[\xi_{ri}(j)\xi_{rn}(j)] = 0$ if $i\neq n$. □

$$2\,\mathrm{E}_S\left[ab\right] = 0$$

Proof. Observe that

$$\mathrm{E}_S[ab] = \sum_{i,v=1}^{k/s}\sum_{r,t=1}^{s}\sum_{\substack{j=1\\ v\neq w}}^{d} x_j^2x_vx_w\,\mathrm{E}_\varphi\left[\varphi_t(v)\varphi_t(w)\right]\mathrm{E}_h\left[\xi_{ri}(j)\xi_{tn}(v)\xi_{tn}(w)\right] = 0.$$

because the signs are independent. □