Profile Reconstruction from Private Sketches

Hao WU [1] **Rasmus Pagh** [1]

Abstract

Given a multiset of n items from $\mathcal{D}$, the *profile reconstruction* problem is to estimate, for $t = 0, 1, \ldots, n$, the fraction $\vec{f}[t]$ of items in $\mathcal{D}$ that appear exactly t times. We consider differentially private profile estimation in a distributed, space-constrained setting where we wish to maintain an updatable, private sketch of the multiset that allows us to compute an approximation of $\vec{f} = (\vec{f}[0], \ldots, \vec{f}[n])$. Using a histogram privatized using discrete Laplace noise, we show how to "reverse" the noise, using an approach of Dwork et al. (ITCS '10). We show how to speed up their LP-based technique from polynomial time to $O(d + n \log n)$, where $d = |\mathcal{D}|$, and analyze the achievable error in the ℓ_1, ℓ_2 and ℓ_∞ norms. In all cases the dependency of the error on d is $O(1/\sqrt{d})$ — we give an information-theoretic lower bound showing that this dependence on d is asymptotically optimal among all private, updatable sketches for the profile reconstruction problem with a high-probability error guarantee.

1. Introduction

The *profile* is a fundamental statistic of a data set. Given a multiset of n items from a finite domain $\mathcal{D}$, the profile is a vector $\vec{f}$ whose $t^{(th)}$ entry equals the fraction $\vec{f}[t]$ of items in $\mathcal{D}$ that appear exactly t times. The concept has arisen in many contexts under various names, including *rarity*, *incidence*, *fingerprint*, *prevalence*, *collision statistics*, *anonymous histogram*, *unattributed histogram*, and *histogram of histograms*. Besides succinctly summarizing frequency information in a dataset, for example the degree distribution in a graph, a profile is sufficient to compute any *symmetric* function of the dataset histogram (Chen et al., 2024).

In settings where space, communication or privacy constraints prevents us from computing the exact profile it is of interest to *approximate* the profile $\vec{f}$ by a vector $\vec{r}$ such that $\vec{r} - \vec{f}$ has small ℓ_p norm for some parameter p.

Streaming algorithms. Datar & Muthukrishnan (2002) considered estimating the profile in a streaming setting, showing that a small sample of the domain $\mathcal{D}$ suffices to ensure small error (they focus on ℓ_1 error, but the technique applies to general ℓ_p norms). The sample can be described by a compact hash function, making it suitable for use in distributed settings. More recently, Chen, Indyk, and Woodruff (2024) improved the error bound, and showed matching space lower bounds for streaming algorithms. General estimators for symmetric functions based on samples were presented and analyzed by Acharya, Das, Orlitsky, and Suresh (2017).

Differentially private algorithms. Differential privacy (Dwork et al., 2006) is a leading approach to ensure that the output of an algorithm, from simple statistics to machine learning models, does not reveal too much about the inputs. It is one of the central techniques for enabling collection of statistics, training of machine learning models, and other computations on data that may hold sensitive information. We will focus on the privacy of individual elements in the multiset, that is, our differential privacy guarantee is defined in terms of pairs of neighboring datasets where the number of occurrences of a particular element differs by 1, and that are otherwise identical.

In the *central model* of differential privacy (where a trusted curator holds the input data), privately releasing a *profile* with small ℓ_2 error was considered by Hay, Li, Miklau, and Jensen (2009); Hay, Rastogi, Miklau, and Suciu (2010), the latter focusing on the application of graph degree distributions. Private synthetic graphs with specific degree distribution were also studied by Karwa and Slavkovic (2012).

Motivated by estimating repetitions in password datasets, Blocki, Datta, and Bonneau (2016) studied differentially private protocols for profile estimation, focusing on ℓ_2 error.

Suresh (2019) considered an alternative (incomparable) error metric, where the profile is represented in the "verbatim" form of a vector encoding all frequencies in sorted order. They presented an algorithm to release verbatim profiles with a small ℓ_1 error. Recently Manurangsi (2022) showed improved tight upper and lower bounds on the privacy/ℓ_1 error trade-off for verbatim profiles.

[1]Department of Computer Science, University of Copenhagen, Denmark. Correspondence to: Hao WU <hawu@di.ku.dk>, Rasmus Pagh <pagh@di.ku.dk>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

Differentially private distributed/streaming algorithms. Dwork, Naor, Pitassi, Rothblum, and Yekhanin (2010) studied profile estimation (referred to as *t-incidence*) in the pan-private streaming setting, where the internal state of the streaming algorithm must satisfy a differential privacy guarantee. Pan-private algorithms protect against *intrusions*, i.e., unintended release of the internal state of the algorithm. Unlike the other works on private profile estimation mentioned above, the neighboring relation considered by Dwork et al. (2010) is defined on the profile itself, i.e., datasets are neighboring if they differ *by any amount* on the count of one item (so the profiles differ in two entries). In a nutshell[1], their algorithm works by maintaining a *histogram* that stores the number of occurrences of each item perturbed by additive noise. Since the noisy histogram is a linear function of the histogram it is *updatable* when the count of an item changes. Dwork et al. (2010) finally show how to derive an estimate of the profile from the noisy histogram by solving a linear programming problem. The size of the sketch required for achieving ℓ_∞ error α on the profile estimate with high probability is stated as $\mathrm{poly}(\log n, 1/\varepsilon, 1/\alpha)$ with unspecified exponents. Unfortunately, the paper does not analyze for what values of α it is possible to solve the linear program (with high probability), meaning that the privacy-utility trade-off obtained remains unclear.

1.1. Our contributions

In this paper we study how to approximate the profile from a noisy histogram of the data. Specifically, we consider the setting where *discrete Laplace* noise has been added to the counts in the histogram. (This is different from the noise distribution used by Dwork et al. (2010), which is nearly uniform modulo a small prime p_i.) Computing the empirical profile directly from the noisy histogram can lead to large errors. For example, consider the situation where $\mathcal{D}$ has n items that each appear once in the dataset and let $\varepsilon > 0$ be a small constant. After adding noise to the histogram to make it ε-differentially private (and possibly rounding up negative values to 0), the fraction of items with a count of 1 will be $\approx \varepsilon$, far from the true fraction of 1.

Instead we use an approach similar to that of Dwork et al. (2010) to approximately invert the effect of the noise. Since we use a known noise distribution, the *expected* value of the profile computed from the noisy histogram is a linear function of the true profile. Due to concentration we can argue that the observed values in the profile based on the noisy histogram are in fact close to the expected ones, and thus it makes sense to find a histogram whose expected profile after noise would be close to the observed noisy profile. Unlike Dwork et al. (2010) who formulate this as a feasibility question of a linear program, we show that it can be reduced to solving a linear ℓ_p-regression problem with a circulant matrix, paving the way for a near-linear time algorithm. We also give a careful analysis of the error $\mathcal{E}rr_p$ in ℓ_p norm, for $p \in \{1, 2, \infty\}$, showing the following upper bound:

Theorem 1.1. *Let $\eta \in (0,1)$, $\varepsilon > 0$. Denote $B \doteq \frac{1}{\varepsilon} \ln\left(\max\left\{\frac{2d}{\eta(e^\varepsilon+1)}, \frac{8\,e^\varepsilon}{e^{2\varepsilon}-1}\right\}\right)$, and assume that $n \geq B$. There is an algorithm that, given a private version $\tilde{h}$ of a d-dimensional histogram $\vec{h}$ published by the ε-DP discrete Laplace mechanism and a parameter $p = 1, 2$ or ∞, returns an approximate $\vec{r}$ for the profile $\vec{f}$ of $\vec{h}$ in $O(d + n \log n)$ time. Moreover, with probability at least $1 - \eta$, the estimation error $\mathcal{E}rr_p \doteq ||\vec{r} - \vec{f}||_p$ satisfies*

$$\mathcal{E}rr_1 \in O\left(\left(\frac{e^\varepsilon+1}{e^\varepsilon-1}\right)^2 \left(\frac{\sum_{t=-B}^{n+B}\sqrt{\mathbb{E}\left[\tilde{f}[t]\right]}}{\sqrt{d}} + \sqrt{\frac{2\ln\frac{1}{\eta}}{d}}\right)\right) \quad (1)$$

$$\mathcal{E}rr_2 \in O\left(\left(\frac{e^\varepsilon+1}{e^\varepsilon-1}\right)^2 \sqrt{\frac{\ln\frac{1}{\eta}}{d}}\right) \quad (2)$$

$$\mathcal{E}rr_\infty \in \left(\left(\frac{e^\varepsilon+1}{e^\varepsilon-1}\right)^2 \left(\sqrt{\frac{2\,(e^\varepsilon-1)}{d\,(e^\varepsilon+1)}\ln\frac{n}{\eta}} + \frac{\ln\frac{n}{\eta}}{3\,d}\right)\right) \quad (3)$$

where $\tilde{f}[t] = \frac{1}{d}|\{\ell \in \mathcal{D} : \tilde{h}[\ell] = t\}|$, $t = -B, \ldots, n+B$.

Notice that the error bounds in all three norms are proportional to $O(1/\sqrt{d})$, where d is the domain size. As we will discuss below, this dependence is optimal under ε-differential privacy for updatable sketches. On the other hand, when $\varepsilon \in O(1)$, the $O(1/\varepsilon^{1.5})$ dependence for $\mathcal{E}rr_\infty$ on the privacy parameter does not match the $\Omega(1/\varepsilon)$ lower bound on the error of releasing a single counter (Hardt & Talwar, 2010). We note that the bound on $\mathcal{E}rr_1$ is a refinement of the bound implied by the bound on $\mathcal{E}rr_2$ and the fact that ℓ_1 distances in n dimensions are at most $\sqrt{n}$ times larger than ℓ_2 distances. Since errors are normalized by a $1/d$ factor one might hope to achieve errors that decrease more rapidly with d than those of Theorem 1.1. However, we show the following lower bound:

Theorem 1.2. *For every ε-DP and updatable algorithm $\mathcal{A} : [0\,..\,n]^d \to \mathcal{Y}$, and every profile estimation algorithm $\mathcal{R} : \mathcal{Y} \to [0,1]^{n+1}$, if the sensitivity $\Delta_{\mathcal{R}\circ\mathcal{A}} \in o\left(\frac{1}{\sqrt{d}} \cdot \frac{\varepsilon}{e^{c\cdot\varepsilon}}\right)$ (for some universal constant $c > 0$), then there exists an input $\vec{h} \in [0\,..\,n]^d$ such that*

$$\mathbb{E}\left[\left\|\vec{f} - \mathcal{R}\left(\mathcal{A}(\vec{h})\right)\right\|_\infty\right] \in \Omega\left(\frac{1}{\sqrt{d}\cdot e^{c\cdot\varepsilon}}\right). \quad (4)$$

This lower bound is not limited to algorithms reconstructing the profile from a perturbed histogram published by the

[1] Since there is no good bound on the difference between histograms with their notion of neighboring datasets, Dwork et al. (2010) actually work with histograms of counts modulo several small random primes p_i, and use these to estimate profiles on the counts modulo p_i, which are in turn used to estimate the original profile. In our context this complication goes away.

discrete Laplace mechanism. Instead, it applies to a broader class of algorithms that attempt to reconstruct the profile from outputs of *updatable* ε-differentially private algorithms. The formal definition of an "updatable" algorithm and the sensitivity $\Delta_{\mathcal{R}\circ\mathcal{A}}$ will be given in Section 5. Informally, the former denotes the property that when the input histogram $\vec{h}$ is updated, we can directly modify $\mathcal{A}(\vec{h})$ to reflect the change, and the latter represents the maximum distance between the profiles reconstructed from the outputs of $\mathcal{A}$ for neighboring histograms. The lower bound shows that an error of $\Omega(1/\sqrt{d})$ is unavoidable for such sketches.

1.2. Technical overview

As noted by Datar & Muthukrishnan (2002) and Dwork et al. (2010) it suffices to estimate the profile of a sample of elements from the domain, since the profile of the sample will closely resemble the true profile. Thus we will assume that this sampling step has already been done, and consider the full histogram of (sampled) items. The final error bounds add the sampling error and the error from privacy using the triangle inequality, but here we focus on the latter.

Given a noisy histogram $\tilde{h}$, we can compute an empirical profile $\tilde{f}$. As the noisy histogram is obtained by adding independent noise to each coordinate of the full histogram $\vec{h}$, the empirical profile can be viewed as a function of a set of independent random variables and should concentrate around its expectation $\mathbb{E}[\tilde{f}]$. Hence, we regard the empirical profile as a reliable representative of its expectation.

Although the expectation $\mathbb{E}[\tilde{f}]$ itself may not be a good approximation of the original profile $\vec{f}$ of $\vec{h}$, there is a subtle connection between them: $\mathbb{E}[\tilde{f}]$ can be computed via a linear transform of the original profile $\vec{f}$. Due the symmetry of discrete Laplace noise, we can write the linear transform as $\boldsymbol{A}\vec{f}$ for a circulant matrix $\boldsymbol{A}$ with non-zero eigenvalues.

Since the inversion for such linear transforms can be computed in $O(n \log n)$ time via Fast Fourier transform (FFT), it motivates us to compute $\boldsymbol{A}^{-1}\tilde{f}$ as an approximation of the true profile. However, such an approximation may not be a valid profile: it may contain coordinates outside $[0, 1]$, and its coordinates do not sum up to 1. We will design a systematic procedure for fixing these violations, which also relies on FFT and runs in $O(n \log n)$ time.

The approximation error comes from three sources: the deviation of $\tilde{f}$ from its expectation, the inverse linear transform, and the violation fixing procedure. Each part will be analyzed separately. The second part, in particular, might be of independent interest as it involves examining various matrix norms of $\boldsymbol{A}^{-1}$, which naturally arises from the discrete Laplace mechanism.

To establish a lower bound, we employ a reduction from the d-dimensional inner product estimation problem in the framework of two-party differential privacy where an error of $\Omega(\sqrt{d})$ is inevitable (McGregor et al., 2010; Haitner et al., 2022).

2. Problem Description

Let $\mathcal{D} \doteq \{1, \ldots, d\}$ be a set of d items. Given a multiset of n items from $\mathcal{D}$, for each item $\ell \in \mathcal{D}$, let $\vec{h}[\ell]$ be the number of times it appears in the multiset. The *histogram* is a vector $\vec{h} \doteq (\vec{h}[1], \ldots, \vec{h}[d]) \in [0\,..\,n]^d$. The *profile* of the histogram $\vec{h}$ is a frequency vector $\vec{f} \doteq (\vec{f}[0], \ldots, \vec{f}[n]) \in [0, 1]^{n+1}$, where $\vec{f}[t] = (1/d) \cdot |\{\ell \in \mathcal{D} : \vec{h}[\ell] = t\}|$ is the fraction of elements in $\mathcal{D}$ which appears t times in the multiset, for each $t \in [0\,..\,n]$.

Our goal is to design an efficient and accurate algorithm for estimating the profile $\vec{f}$ from a private histogram $\tilde{h}$ output by the *discrete Laplace mechanism* (or geometric mechanism) due to Ghosh, Roughgarden, and Sundararajan (2009).

Discrete Laplace Mechanism. Given an input histogram $\vec{h}$, the mechanism aims at outputting a perturbed version $\tilde{h}$ of $\vec{h}$ that is insensitive to the addition/removal of an arbitrary item in the multiset. Specifically, the mechanism creates $\tilde{h}$ by adding independent noises to each coordinate of $\vec{h}$, such that

$$\tilde{h}[\ell] = \vec{h}[\ell] + \mathbb{DL}\mathrm{ap}\left(e^{-\varepsilon}\right), \forall \ell \in \mathcal{D},$$

where $\mathbb{DL}\mathrm{ap}\left(e^{-\varepsilon}\right)$ denotes a random variable following discrete Laplace distribution: $\Pr\left[\mathbb{DL}\mathrm{ap}\left(e^{-\varepsilon}\right) = t\right] = \frac{1-e^{-\varepsilon}}{1+e^{-\varepsilon}} \cdot e^{-\varepsilon \cdot |t|}$, $\forall t \in \mathbb{Z}$. In practice, a clipping step can be applied after coordinate perturbation:

$$\tilde{h}[\ell] = \mathrm{clip}\Big[\vec{h}[\ell] + \mathbb{DL}\mathrm{ap}\left(e^{-\varepsilon}\right), 0, n\Big], \forall \ell \in \mathcal{D}.$$

where $\mathrm{clip}[\vec{h}[\ell] + \mathbb{DL}\mathrm{ap}\left(e^{-\varepsilon}\right), 0, n] \doteq \max\{0, \min\{\vec{h}[\ell] + \mathbb{DL}\mathrm{ap}\left(e^{-\varepsilon}\right), n\}\}$ ensures the returned value is in $[0\,..\,n]$.

Privacy Guarantee. The privacy guarantee provided by the mechanism is measured by the framework of *differential privacy*. Formally, we call two histograms $\vec{h}$ and $\vec{h}'$ *neighboring*, written as $\vec{h} \sim \vec{h}'$, if $\exists \ell \in \mathcal{D}$, s.t., $|\vec{h}[\ell] - \vec{h}'[\ell]| \leq 1$, and $\forall \ell' \in \mathcal{D} \setminus \{\ell\}$, it holds that $\vec{h}[\ell'] = \vec{h}'[\ell']$. The discrete Laplace mechanism ensures that the output histograms have similar distributions for neighboring inputs.

Definition 2.1 (ε-Differentially Privacy (Dwork & Roth, 2014))**.** Given $\varepsilon > 0$, a randomized algorithm $\mathcal{A} : \mathbb{N}^d \to \mathcal{Z}$ is called ε-differentially private (DP), if for every $\vec{h}, \vec{h}' \in \mathbb{N}^d$ such that $\vec{h} \sim \vec{h}'$, and all (measurable) $Z \subseteq \mathcal{Z}$,

$$\Pr\left[\mathcal{A}(\vec{h}) \in Z\right] \leq e^{\varepsilon} \cdot \Pr[\mathcal{A}(\vec{h}') \in Z]\,. \tag{5}$$

Ghosh et al. (2009) showed that the discrete Laplace mechanism is ε-DP (with or without the clipping step).

3. Preliminaries

Definition 3.1 (Circulant Matrix)**.** Let $p \in \mathbb{N}^+$ and $\vec{c} = \left[c_0, c_1, \cdots, c_{p-1}\right] \in \mathbb{C}^p$ be a row vector. Define the (right) circular shift operator $\sigma : \mathbb{C}^p \to \mathbb{C}^p$ by $\sigma(\vec{c}) = \left[c_{p-1}, c_0, \cdots, c_{p-2}\right]$. The circulant matrix $\mathbf{circ}(\vec{c})$ is a matrix whose rows $\vec{c}, \sigma(\vec{c}), \ldots, \sigma^{p-1}(\vec{c})$ are generated by iteratively applying the operator σ on $\vec{c}$.

Fact 3.2 (Eigenvectors and Eigenvalues (Gray, 2005)). Let $w_t \doteq e^{-\frac{2\pi t \cdot i}{p}}$, for each $t \in [0 \mathinner{.\,.} p-1]$. Then the eigenvalues $\varphi_0, \ldots, \varphi_{p-1}$ and the corresponding eigenvectors $\vec{v}_0, \ldots, \vec{v}_{p-1}$ of the matrix $\mathbf{circ}(\vec{c})$ are given by

$$\varphi_t \doteq c_0 + c_1(w_t)^1 + \cdots + c_{p-1}(w_t)^{p-1}, \tag{6}$$

$$\vec{v}_t \doteq 1/\sqrt{p} \cdot \left[1, (w_t)^1, \cdots, (w_t)^{p-1}\right]^\top. \tag{7}$$

Fact 3.3 (Fast Fourier Transform (FFT) (Cormen et al., 2009)). Let $\boldsymbol{V} \doteq [\vec{v}_0 \cdots \vec{v}_{p-1}] \in \mathbb{C}^{p \times p}$ be the matrix composed of column vectors $\vec{v}_t$ (defined in Equation (7)) for $t = 0, 1, \ldots, p-1$. Then for each $\vec{x} \in \mathbb{C}^p$, $\boldsymbol{V}\vec{x}, \boldsymbol{V}^{-1}\vec{x}, \vec{x}^\top \boldsymbol{V}$ and $\vec{x}^\top \boldsymbol{V}^{-1}$ can be computed in $O(p \log p)$ time.

4. Near-linear Time Algorithm

In this section, we introduce an algorithm that achieves the properties stated in Theorem 1.1. Our presentation is structured as follows. In Section 4.1, we demonstrate that we can transform the histograms published by the discrete Laplace mechanisms with the clipping step into ones without clipping steps, ensuring they follow the same distribution. This allows us to address perturbed histograms with and without clipping in a unified approach. Subsequently, in Section 4.2, we formulate the task of searching for a good estimate of the profile as an optimization problem. Section 4.3 presents a near-linear algorithm for systematically finding an approximate solution to the optimization problem. Sections 4.4 and 4.5 analyze the time complexity and the approximation error of the proposed algorithms, respectively. Finally, Section 4.6 provides a more detailed comparison of our algorithm with the previous approach.

4.1. Unfolding Clipped Histograms

Assume that $\tilde{h}$ has been clipped, so that each entry has value between 0 and n. Based on the clipped histogram, we can construct a new histogram that has the same distribution as the an unclipped one. Consider an entry $\tilde{h}[\ell] = \text{clip}\left[\vec{h}[\ell] + \mathbb{DL}\text{ap}\left(e^{-\varepsilon}\right), 0, n\right]$ for some $\ell \in \mathcal{D}$, and the recovery procedure:

$$\tilde{h}'[\ell] \doteq \begin{cases} \tilde{h}[\ell], & \text{if } 0 < \tilde{h}[\ell] < n, \\ \tilde{h}[\ell] - \mathbb{DE}\text{xp}\left(e^{-\varepsilon}\right), & \text{if } \tilde{h}[\ell] = 0, \\ \tilde{h}[\ell] + \mathbb{DE}\text{xp}\left(e^{-\varepsilon}\right), & \text{if } n = \tilde{h}[\ell], \end{cases} \tag{8}$$

where $\mathbb{DE}\text{xp}\left(e^{-\varepsilon}\right)$ denotes a random variable following the geometric distribution: $\Pr\left[\mathbb{DE}\text{xp}\left(e^{-\varepsilon}\right) = t\right] = (1 - e^{-\varepsilon}) \cdot e^{-\varepsilon \cdot t}$, $\forall t \in \mathbb{N}$.

Lemma 4.1. *$\tilde{h}'[\ell]$ has the same distribution as $\vec{h}[\ell] + Z$, where Z is an independent copy of $\mathbb{DL}\text{ap}\left(e^{-\varepsilon}\right)$.*

Appendix B contains the detailed proof of the theorem. Here, we present a high-level overview of the proof's intuition: when restricting the two-directional discrete Laplace random variable $\mathbb{DL}\text{ap}\left(e^{-\varepsilon}\right)$ to a single direction, it reduces to a random variable that follows the geometric distribution. It is a well-known fact that for a geometric random variable, truncating it at a certain value and adding an independent copy of the random variable at the truncated value preserves its geometric distribution. This property is commonly referred to as the "memorylessness" property of the geometric distribution.

4.2. Searching for Profiles

A natural attempt would be estimating the profile $\vec{f}$ with $\tilde{f}[0 \mathinner{.\,.} n] \doteq (\tilde{f}[0], \ldots, \tilde{f}[n])$, where $\tilde{f}[t] = (1/d) \cdot |\{\ell \in \mathcal{D} : \tilde{h}[\ell] = t\}|, \forall t \in \mathbb{Z}$. There are several issues with such an estimate. First, $\tilde{f}[0 \mathinner{.\,.} n]$ is not an empirical profile, as its entries do not sum up to 1. An even more significant concern is that such an estimate can incur an error of $\Omega(1)$.

Example. Consider the example where $\vec{h}[\ell] = 1, \forall \ell \in \mathcal{D}$. Hence, $\vec{f}[1] = 1$ and $\vec{f}[t] = 0$ for all $t \neq 1$. But, it can be shown that $\mathbb{E}[\tilde{f}[1]] = \frac{1-e^{-\varepsilon}}{1+e^{-\varepsilon}}$ and $\mathbb{V}ar[\tilde{f}[1]] \leq \frac{1}{d} \cdot \frac{1-e^{-\varepsilon}}{1+e^{-\varepsilon}}$. By Chebyshev's inequality, when $\varepsilon \in \Theta(1)$, with constant probability, $\|\tilde{f}[0 \mathinner{.\,.} n] - \vec{f}\|_p \geq |\tilde{f}[1] - \vec{f}[1]| \in \Omega(1)$, for $p = 1, 2, \infty$.

To deepen our understanding of the relationship between $\tilde{f}$ and $\vec{f}$, let us consider a fixed $t \in [0 \mathinner{.\,.} n]$. There are $\vec{f}[t] \cdot d$ items in $\mathcal{D}$ such that $\vec{h}[\ell] = t$. For each of these items, the random variable $\tilde{h}[\ell] = \vec{h}[\ell] + \mathbb{DL}\text{ap}\left(e^{-\varepsilon}\right)$ is independent and identically distributed. Our *analysis* considers an algorithm, outlined in Algorithm 1, for generating $\tilde{f}$ directly from $\vec{f}$. We emphasize that our final algorithm is the combination of Algorithms 2 and 3, and does not make use of Algorithm 1.

Algorithm 1 Private Profile Generator $\mathcal{A}(\vec{r})$

Input: $\vec{r}$
Initialize an all zero profile $\vec{v}$
for $t \in [0 \mathinner{.\,.} n]$ **do**
  for $j \leftarrow 1$ to $\vec{r}[t] \cdot d$ **do**
    $I \leftarrow t + \mathbb{DL}\text{ap}\left(e^{-\varepsilon}\right)$
    $\vec{v}[I] \leftarrow \vec{v}[I] + 1/d$
return $\vec{v}$ ▷ the perturbed profile

To search for a good approximation $\vec{r}$ of $\vec{f}$, our strategy shifts towards searching for a $\vec{r}$ such that $\mathcal{A}(\vec{r})$ (the vector returned by Algorithm 1 given input $\vec{r}$) is close to $\tilde{f}$. The search is framed as an optimization problem. Our goal is

then to design efficient algorithm for solving the optimization problem, and to prove that the proximity between $\mathcal{A}(\vec{r})$ and $\tilde{f}$ indeed implies the proximity between $\vec{r}$ and $\vec{f}$.

Optimization Problem. Let $B \in \mathbb{R}^+$ be the parameter (as defined in Theorem 1.1), and $\mathcal{E}_B$ be the event that the all noises introduced by Algorithm 1 are absolutely bounded by B, and $\text{LB} \doteq -B$, $\text{UB} \doteq n + B$, $\tilde{f} \doteq (\tilde{f}[\text{LB}], \ldots, \tilde{f}[\text{UB}])$. Given $p = 1, 2$ of ∞, the optimization problem we rely on is given by

$$\begin{aligned} \min_{\vec{r} \in \mathbb{R}^{n+1}} \quad & \left\| \mathbb{E}_{\mathcal{A}} \left[\mathcal{A}(\vec{r}) \mid \mathcal{E}_B \right] - \tilde{f} \right\|_p \\ \text{s.t.} \quad & \sum_{t=0}^{n} \vec{r}[t] = 1, \\ & 0 \le \vec{r}[t] \le 1, \quad \forall t \in [0 \mathinner{..} n]. \end{aligned} \tag{P$_1$}$$

To simplify the notation, when it is clear from the context, we write $\mathbb{E}_{\mathcal{A}} \left[\mathcal{A}(\vec{r}) \mid \mathcal{E}_B \right]$ as $\mathbb{E} \left[\mathcal{A}(\vec{r}) \mid \mathcal{E}_B \right]$. In what follows, we provide motivation for this formulation.

Conditioning on $\mathcal{E}_B$. Without the event $\mathcal{E}_B$, $\mathcal{A}(\vec{r})$ can have unbounded support, as the noises introduced by Algorithm 1 are unbounded. Conditioned on event $\mathcal{E}_B$, only $\mathcal{A}(\vec{r})[\text{LB}], \ldots, \mathcal{A}(\vec{r})[\text{UB}]$ process nonzero values. Therefore, $\mathbb{E} \left[\mathcal{A}(\vec{r}) \mid \mathcal{E}_B \right]$ can be viewed as a vector indexed from LB to UB. Further, by proper choice of B, the event $\overline{\mathcal{E}_B}$ only occurs with sufficiently small probability. Therefore, the event $\mathcal{E}_B$ encompasses the majority of cases.

Taking the Expectation. Without the expectation, the error $\|\mathcal{A}(\vec{r}) - \tilde{f}\|_p$ is itself a random variable (even conditioned on $\mathcal{E}_B$). The objective in (P$_1$) is a relaxation of the expected error, as it provides a lower bound for the latter. To see this, note that $\|\cdot\|_p$ is convex for $p \ge 1$, and according to Jensen's inequality, we have

$$\left\| \mathbb{E}_{\mathcal{A}} \left[\mathcal{A}(\vec{r}) \mid \mathcal{E}_B \right] - \tilde{f} \right\|_p \le \mathbb{E}_{\mathcal{A}} \left[\left\| \mathcal{A}(\vec{r}) - \tilde{f} \right\|_p \mid \mathcal{E}_B \right].$$

Matrix Form. We now reformulate the above in matrix form. The primary advantage of this formulation lies in its potential for expressing the objective in special matrix form, facilitating the development of efficient optimization algorithms. Let $\vec{c} \in \mathbb{R}^{n+2B+1}$ be a vector defined as follows:

- The first B entries are set to $1, e^{-\varepsilon}, \ldots, e^{-\varepsilon \cdot B}$.
- The last B entries are set to $e^{-\varepsilon \cdot B}, e^{-\varepsilon \cdot (B-1)}, \ldots, e^{-\varepsilon}$.
- All other entries are set to 0.

Define $P_{norm} \doteq \frac{1 + e^{-\varepsilon} - 2\left(e^{-\varepsilon}\right)^{B+1}}{1 - e^{-\varepsilon}}$, and

$$\boldsymbol{A} \doteq \frac{1}{P_{norm}} \textbf{circ}(\vec{c}). \tag{9}$$

To maintain consistency with $\tilde{f}$, we assume that both rows and columns of $\boldsymbol{A}$ are indexed by $[\text{LB} \mathinner{..} \text{UB}]$. The following lemma states that, we can express the objective in (P$_1$) in matrix form.

Lemma 4.2. *Let $\mathcal{A}(\cdot)$ be the Algorithm 1, $\boldsymbol{A}$ be matrix defined as Equation (9), and $\vec{r} = (\vec{r}[0], \ldots, \vec{r}[n]) \in \mathbb{R}_{\ge 0}^{n+1}$ be a vector such that $\sum_{t \in [0 \mathinner{..} n]} \vec{r}[t] = 1$. Assume that we expand the vector $\vec{r}$ to one in $\mathbb{R}^{n+2B+1}$ by padding $\vec{r}[-B] = \cdots = \vec{r}[-1] = 0$ to the front and $\vec{r}[n+1] = \cdots = \vec{r}[n+B] = 0$ to the back of $\vec{r}$. The following claim holds: $\mathbb{E} \left[\mathcal{A}(\vec{r}) \mid \mathcal{E}_B \right] = \boldsymbol{A} \times \vec{r}$.*

Based on the above lemma, the optimization problem (P$_1$) is equivalent to the following one:

$$\begin{aligned} \min_{\vec{r} \in \mathbb{R}^{n+2B+1}} \quad & \left\| \boldsymbol{A} \times \vec{r} - \tilde{f} \right\|_p \\ \text{s.t.} \quad & \sum_{t=0}^{n} \vec{r}[t] = 1, \\ & 0 \le \vec{r}[t] \le 1, \forall t \in [0 \mathinner{..} n], \\ & \vec{r}[t] = 0, \forall t \in [\text{LB} \mathinner{..} \text{UB}] \setminus [0 \mathinner{..} n]. \end{aligned} \tag{P$_2$}$$

An important advantage lies in the potential efficiency gained from the circulant matrix $\boldsymbol{A}$, as detailed in Section 4.4, enabling $\boldsymbol{A}^{-1}\vec{x}$ to be computed in $O(n \log n)$ time for each vector $\vec{x}$. This motivates the choice of computing $\boldsymbol{A}^{-1}\tilde{f}$ as the approximate profile $\vec{r}$, resulting in zero error in the objective function:

$$\left\| \boldsymbol{A} \times \boldsymbol{A}^{-1} \times \tilde{f} - \tilde{f} \right\|_p = 0.$$

However, it's crucial to note that such a solution may potentially violate other constraints outlined in (P$_2$). This observation motivates the development of an algorithm for systematically solving (P$_2$), which will be discussed in the next section.

4.3. Optimization Algorithms

In this section, we present effective algorithms to address the optimization problem defined by (P$_2$). Our approach consists of two steps: initially, we concentrate on a relaxed version of (P$_2$) and introduce an algorithm that optimally solves it. Subsequently, building upon the solution obtained in the first step, we introduce a rounding procedure designed to yield an approximate solution for (P$_2$), satisfying all the specified constraints. The proofs of all lemmas in this section are included in the *Appendix D*.

Relaxation. There are two categories of constraints on $\vec{r}$ in optimization problem (P$_2$): its coordinates should sum up to 1; and each of its coordinates should be within a given range. Optimizing (P$_2$) simultaneously with two constraints is non-trivial, so we initially focus on the first constraint, leading to optimization problem (P$_3$):

$$\min \quad \left\| \boldsymbol{A} \times \vec{r} - \tilde{f} \right\|_p \quad \text{s.t.} \quad \textstyle\sum_{t=0}^{n} \vec{r}[t] = 1. \tag{P$_3$}$$

The optimization problem (P$_3$) has a simple geometric interpretation. Since $\boldsymbol{A}$ is invertible (as will be shown in

Algorithm 2 Fast Inversion $\mathcal{A}_{fst\text{-}inv}$

Input: $\tilde{h}$ **and** $p \in \{1, 2, \infty\}$
Compute $\tilde{f}$ based on $\tilde{h}$
$\vec{u} \leftarrow \boldsymbol{A}^{-1} \times \tilde{f}$
Let $\vec{\mathbb{1}}_{0:n}$ be a vector, s.t., $\vec{\mathbb{1}}_{0:n}[t] = \begin{cases} 1, & \forall t \in [0\,..\,n] \\ 0, & \text{otherwise} \end{cases}$
Let $\vec{a} \leftarrow \arg\max_{\|\vec{a}'\|_p = 1} \langle \vec{\mathbb{1}}_{0:n} \boldsymbol{A}^{-1} \times \vec{a}' \rangle$
$\vec{r} \leftarrow \vec{u} - \frac{\langle \vec{\mathbb{1}}_{0:n}, \vec{u} \rangle - 1}{\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a} \rangle} \cdot \boldsymbol{A}^{-1} \times \vec{a}$
return $\vec{r}$ ▷ the estimated profile

Lemma 4.6), the function $\|\cdot\|_{\boldsymbol{A},p} : \mathbb{R}^{[0\,..\,n]} \to \mathbb{R}$, defined by

$$\|\vec{v}\|_{\boldsymbol{A},p} \doteq \|\boldsymbol{A} \times \vec{v}\|_p, \quad \forall \vec{v} \in \mathbb{R}^{[0\,..\,n]}.$$

constitutes a norm. The objective of (P_3) is equivalent to

$$\|\boldsymbol{A} \times (\vec{r} - \boldsymbol{A}^{-1} \times \tilde{f})\|_p$$

Hence, (P_3) seeks to find the point nearest to $\boldsymbol{A}^{-1} \times \tilde{f}$ (measured by the distance induced by the norm $\|\cdot\|_{\boldsymbol{A},p}$) within the hyperplane specified by $\sum_{t=0}^{n} \vec{r}[t] = 1$.

This inspires Algorithm 2 for solving (P_3). First, it computes a vector $\vec{u} = \boldsymbol{A}^{-1} \times \tilde{f}$. Let $\vec{\mathbb{1}}_{0:n}$ be a binary vector such that $\vec{\mathbb{1}}_{0:n}[t] = 1$ only when $t \in [0\,..\,n]$. If $\langle \vec{\mathbb{1}}_{0:n}, \vec{u} \rangle \neq 1$, the algorithm rectifies this by adjusting $\vec{u}$ along the direction $\boldsymbol{A}^{-1} \times \vec{a}$ (by an appropriate amount), where $\vec{a}$ is the unit vector (in terms of the norm $\|\cdot\|_p$) which maximizes $\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a} \rangle$. In essence, adjusting $\vec{u}$ along $\boldsymbol{A}^{-1} \times \vec{a}$ is the most cost-effective way to correct the violation $\langle \vec{\mathbb{1}}_{0:n}, \vec{u} \rangle \neq 1$. Formally, Lemma 4.3 holds.

Lemma 4.3. *Algorithm 2 returns an optimal solution for the optimization problem* (P_3).

Rounding. Given a solution $\vec{r}$ returned by Algorithm 2, Algorithm 3 post-processes it to ensure satisfaction of all constraints in (P_2).

Algorithm 3 unfolds in three phases: Phase 1 (Line 1), wherein the algorithm sets $\vec{r}[t]$ to 0 for each $t \in [\mathrm{LB}\,..\,\mathrm{UB}] \setminus [0\,..\,n]$; Phase 2 (Line 2 to Line 7), where the algorithm clips $\vec{r}[t]$ to the range of $[0, 1]$; and Phase 3 (Line 8 to Line 12), during which the algorithm adjusts $\vec{r}[t]$ to ensure their sum equals 1. It is essential to observe that Phase 3 will not generate any coordinate outside the range $[0, 1]$. Moreover, in Phase 3, it is adequate to solely consider decreasing the coordinate values of $\vec{r}$, which will be formally explained in the proof of Lemma 4.4.

Lemma 4.4. *The vector returned by Algorithm 3 satisfies all constraints in the optimization problem* (P_2).

Algorithm 3 Rounding $\mathcal{A}_{rnd}$

Input: $\vec{r}$
1: $\vec{r}[t] \leftarrow 0, \forall t \in [\mathrm{LB}\,..\,\mathrm{UB}] \setminus [0\,..\,n]$
2: $T_{>1} \leftarrow \{t \in [0\,..\,n] : \vec{r}[t] > 1\}$
3: $T_{<0} \leftarrow \{t \in [0\,..\,n] : \vec{r}[t] < 0\}$
4: $s_{>1} \leftarrow \sum_{t \in T_{>1}} (1 - \vec{r}[t])$
5: $s_{<0} \leftarrow \sum_{t \in T_{<0}} (0 - \vec{r}[t])$
6: $\vec{r}[t] \leftarrow 1, \forall t \in T_{>1}$
7: $\vec{r}[t] \leftarrow 0, \forall t \in T_{<0}$
8: $s \leftarrow s_{>1} + s_{<0}$. ▷ we must have $s \geq 0$
9: **if** $s > 0$ **then**
10: Let $\tau \geq 0$ be the solution for $\sum_{t=0}^{n} \min\{\tau, \vec{r}[t]\} = s$
11: $\vec{r}[t] \leftarrow \vec{r}[t] - \min\{\tau, \vec{r}[t]\}, \forall t \in [0\,..\,n]$
12: **end if**
13: **return** $\vec{r}$ ▷ the rounded profile

4.4. Time Analysis

In this section, we study the running time of both Algorithm 2 (Theorem 4.5) and Algorithm 3 (Theorem 4.7).

Theorem 4.5. *Algorithm 2 runs in $O(d + n \log n)$ time.*

The proof of Theorem 4.5 relies heavily on the properties of $\boldsymbol{A}$ as a circulant matrix.

Lemma 4.6. *The eigenvalues $\varphi_t, t = -B, \ldots, n + B$ of $\boldsymbol{A}$ can be computed in $O(n)$ time. Further, $\boldsymbol{A}$ is invertible and*

$$\boldsymbol{A}^{-1} = \boldsymbol{V} \times \mathbf{diag}\left(\varphi_{-B}^{-1}, \varphi_{-B+1}^{-1}, \ldots, \varphi_{n+B}^{-1}\right) \times \boldsymbol{V}^{-1},$$

where $\mathbf{diag}\left(\varphi_{-B}^{-1}, \varphi_{-B+1}^{-1}, \ldots, \varphi_{n+B}^{-1}\right)$ *is a diagonal matrix, and $\boldsymbol{V} \doteq [\vec{v}_{-B} \cdots \vec{v}_{n+B}]$ is the matrix comprising column vectors $\vec{v}_t = \frac{1}{\sqrt{n+2B+1}} \cdot \left[1, (w_t)^1, \cdots, (w_t)^{n+2B}\right]^\top$, $w_t \doteq e^{-\frac{2\pi \cdot (t+B) \cdot i}{n+2B+1}}$, for $t = -B, \ldots, n + B$.*

The proof of Lemma 4.6 is included in the Appendix E. Now we are ready to prove Theorem 4.5.

Proof of Theorem 4.5. First, computing the empirical profile $\tilde{f}$ from $\tilde{h}$ takes $O(d)$ time. Next, by fast Fourier transform (Fact 3.3), multiplying a vector with both $\boldsymbol{V}$ and $\boldsymbol{V}^{-1}$ can be performed in $O((n + 2B + 1)\log(n + 2B + 1)) = O(n \log n)$ time. Based on the decomposition of $\boldsymbol{A}^{-1}$ (Lemma 4.6), we see that multiplying $\boldsymbol{A}^{-1}$ with a vector takes $O(n \log n)$ time. Therefore, it suffices to show that $\vec{a}$ can also be found in $O(n \log n)$ time.

For consistency, assume that the coordinates of $\vec{a}$ are indexed from $-B$ to $n + B$ and write $\vec{a} = [a_{-B}, \ldots, a_{n+B}]$. Then $\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a} \rangle = \sum_{t=-B}^{n+B} \langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1}[\cdot, t] \rangle \cdot a_t$, where $\boldsymbol{A}^{-1}[\cdot, t]$ is the $t^{(th)}$ column of $\boldsymbol{A}^{-1}$. For convenience, let c_t be a short hand for $\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1}[\cdot, t] \rangle$. Observe that $\vec{c} \doteq [c_{-B}, \ldots, c_{n+B}] = \left(\vec{\mathbb{1}}_{0:n}\right)^\top \times \boldsymbol{A}^{-1}$ can be computed in $O(n \log n)$ time. Once the c_t are known, finding $\vec{a}$

is equivalent to solving the following optimization problem: $\max_{\|\vec{a}\|_p=1} \langle \vec{c}, \vec{a} \rangle$.

It is straightforward to see that: i) $\vec{a} = \text{sign}(c_t) \cdot \vec{e}_t$ where $\vec{e}_t$ is a standard basis vector and $t = \arg\max_t |c_t|$ when $p = 1$; ii) $\vec{a} = \frac{\vec{c}}{\|\vec{c}\|_2}$ when $p = 2$; iii) $\vec{a} = [\text{sign}(c_{-B}), \ldots, \text{sign}(c_{n+B})]^\top$ when $p = \infty$. □

Theorem 4.7. *Algorithm 3 has an implementation which runs in $O(n \log n)$ time.*

The detailed proof of the theorem is included in the Appendix E. At a high level, we can initially sort the $\vec{r}[t]$ in non-decreasing order in $O(n \log n)$ time. Subsequently, we identify the smallest integer t^* for which $\vec{r}[t^*] \geq \tau$. It can be demonstrated that $\tau = \frac{s - \sum_{t=0}^{t^*-1} \vec{r}[t]}{n+1-t^*}$.

4.5. Error Analysis

In this section, we analyze the estimation error when we combined algorithms 2 and 3, and show that this results in the estimation error stated in Theorem 1.1.

The approximation errors arise from two sources: the difference between the optimal solution of (P_3) and the actual profile; and the error introduced during the rounding procedure. We will demonstrate that the errors resulting from the former are of the same order of magnitudes as those indicated in Theorem 1.1, while the latter introduces only additional constants to the error of the former. Formally, the following two lemmas hold.

Lemma 4.8 (Error from Optimization (P_3))**.** *Given the profile $\tilde{f}$ of a private version $\tilde{h}$ of a histogram $\vec{h}$ published by the ε-DP discrete Laplace mechanism and a parameter $p = 1, 2$ or ∞, with probability at least $1 - \eta$, Algorithm 2 returns a vector $\vec{r}$ with the estimation error $\mathcal{E}rr_p \doteq \|\vec{r} - \vec{f}\|_p$ satisfying Inequalities (1,2,3).*

Lemma 4.9 (Error from Rounding)**.** *Given a vector $\vec{r}$ returned by Algorithm 2, let $\mathcal{E}rr_p$ denote the error before running Algorithm 3, and $\mathcal{E}rr'_p$ denote the error afterwards. Then $\mathcal{E}rr'_p \leq \mathcal{E}rr_p$ for $p = 1, 2$, and $\mathcal{E}rr'_\infty \leq 2\,\mathcal{E}rr_\infty$.*

Combing Lemma 4.8 and 4.9 immediately proves the error bounds in Theorem 1.1. In the rest of the section, we present the proofs for Lemma 4.8 and Lemma 4.9 separately.

Error from Optimization (P_3)

Proof of Lemma 4.8. We need to prove that, the closeness between $\boldsymbol{A} \times \vec{r}$ and $\tilde{f}$ implies the closeness between $\vec{r}$ and $\vec{f}$. To bound $\|\vec{r} - \vec{f}\|_p$, we observe that

$$\begin{aligned}\left\|\vec{r} - \vec{f}\right\|_p &= \left\|\boldsymbol{A}^{-1} \times \boldsymbol{A} \times \left(\vec{r} - \vec{f}\right)\right\|_p \leq \|\boldsymbol{A}^{-1}\|_p \cdot \left\|\boldsymbol{A} \times \left(\vec{r} - \vec{f}\right)\right\|_p \\ &\leq \|\boldsymbol{A}^{-1}\|_p \cdot \left(\left\|\boldsymbol{A} \times \vec{r} - \tilde{f}\right\|_p + \left\|\tilde{f} - \boldsymbol{A} \times \vec{f}\right\|_p\right) \\ &\leq \|\boldsymbol{A}^{-1}\|_p \cdot \left(2 \cdot \left\|\tilde{f} - \boldsymbol{A} \times \vec{f}\right\|_p\right),\end{aligned}$$

where the first inequality comes from the definition of matrix norm $\|\boldsymbol{A}^{-1}\|_p \doteq \max_{\vec{v}:\|\vec{v}\|_p=1} \|\boldsymbol{A}^{-1} \times \vec{v}\|_p$; the last one from that $\vec{r}$ is the solution for optimization problem (P_3). It remains to bound $\|\boldsymbol{A}^{-1}\|_p$ and $\|\tilde{f} - \boldsymbol{A} \times \vec{f}\|_p$ separately.

Lemma 4.10. *Let $\varphi_t, t = -B, \ldots, n + B$ be the eigenvalues of $\boldsymbol{A}$. Then*

$$\begin{aligned}\|\boldsymbol{A}^{-1}\|_1 = \|\boldsymbol{A}^{-1}\|_\infty &\leq \frac{2 + e^{-\varepsilon} + e^{\varepsilon}}{e^{\varepsilon} - e^{-\varepsilon} - 4 \cdot e^{-\varepsilon \cdot B}} \cdot P_{norm}, \\ \|\boldsymbol{A}^{-1}\|_2 \leq \max_t \frac{1}{|\varphi_t|} &\leq \frac{P_{norm} \cdot (1 + e^{-\varepsilon})}{\left(1 - e^{-\varepsilon} - 2 \cdot e^{-\varepsilon \cdot (B+1)}\right)}.\end{aligned}$$

Lemma 4.11. *With probability at least $1 - \eta$, it holds that*

$$\begin{aligned}\left\|\tilde{f} - \boldsymbol{A} \times \vec{f}\right\|_1 &\leq \frac{1}{\sqrt{d}} \cdot \sum_{t=-B}^{n+B} \sqrt{\mathbb{E}\left[\tilde{f}[t]\right]} + \sqrt{\frac{2 \cdot \ln \frac{1}{\eta}}{d}}. \\ \left\|\tilde{f} - \boldsymbol{A} \times \vec{f}\right\|_2 &\leq \sqrt{\frac{1}{d}} + \sqrt{\frac{\ln \frac{1}{\eta}}{d}}. \\ \left\|\tilde{f} - \boldsymbol{A} \times \vec{f}\right\|_\infty &\leq \frac{1}{\sqrt{d}} \cdot \sqrt{2 \cdot \frac{1}{P_{norm}} \cdot \ln \frac{n}{\eta}} + \frac{1}{3d} \cdot \ln \frac{n}{\eta}.\end{aligned}$$

The proof of Lemma 4.10 and Lemma 4.11 are included in the Appendix F. Finally, combining these lemmas, the assumption of B in Theorem 1.1, and that $P_{norm} = \frac{1 + e^{-\varepsilon} - 2 \cdot (e^{-\varepsilon})^{B+1}}{1 - e^{-\varepsilon}}$ finishes the proof. □

Error from Rounding. The full proof for Lemma 4.9 is provided in the Appendix F. Here is a brief outline of the proof for the case of $p = 1$ to offer some insight. The proof for the cases of $p = 2, \infty$ are similar but much more technical. The algorithm is divided into three phases:

- Phrase 1 (Line 1): In this phase, the algorithm rounds $\vec{r}[t]$ to 0 for each $t \in [-B \,..\, n + B] \setminus [0 \,..\, n]$.
- Phase 2 (Line 2 to Line 7): In this phase, the algorithm clips the $\vec{r}[t]$ to the range of $[0, 1]$.
- Phase 3 (Line 8 to Line 12): In this phase, the algorithm adjusts the $\vec{r}[t]$ so that they sum up to 1.

It is evident that the quantity $\left|\vec{r}[t] - \vec{f}[t]\right|$ only decreases in Phases 1 and 2, as $\vec{f}[t] \in [0, 1]$. Further it can be shown that in Phase 2, $\|\vec{r} - \vec{f}\|_1$ decreases by an amount of $-s_{>1} + s_{<0}$, and in Phase 3, $\|\vec{r} - \vec{f}\|_1$ can increase by at most $s_{>1} + s_{<0}$. Hence overall, $\|\vec{r} - \vec{f}\|_1$ decreases.

4.6. Comparison with Previous Approach

Dwork, Naor, Pitassi, Rothblum, and Yekhanin (2010) also studied the profile estimation (referred to as the *t-incidence*) problem, with a key difference from our work: their approach aimed for pan-privacy, which resulted in significantly more noise added to the histogram $\vec{h} \in [0..n]^d$. To simplify the comparison, we assume a sampling step involved in their method has already been executed, and focus on the

full histogram of sampled items. Let $\tilde{h}$ denote the noisy histogram. For each $\ell \in [d]$, their requirement was that

$$\big(\max_{t\in[0..n]} \Pr[\tilde{h}[\ell]=t]\big)/\big(\min_{t\in[0..n]} \Pr[\tilde{h}[\ell]=t]\big) \le e^{\varepsilon}.$$

Specifically, they used noise scaling with n: $\tilde{h}[\ell] \leftarrow \vec{h}[\ell]+Z \pmod{n+1}$, where Z is a noise variable such that $\Pr[Z = t] \propto e^{-\varepsilon\cdot t/(n+1)}$.

In contrast, our paper considers adding symmetric discrete Laplace noise $\mathbb{DL}ap(e^{-\varepsilon})$ to each entry, which is fundamental and widely used for publishing privatized histograms. Our technique is applicable even in the pan-private setting, with appropriately scaled noise.

Another minor difference lies in the optimization problem: their linear program (LP) aims to achieve solutions with ℓ_∞ error, whereas our algorithm supports $\ell_1/\ell_2/\ell_\infty$ errors. Moreover, they did not provide an almost linear-time algorithm for solving their LP. However, the almost linear-time algorithm developed in our paper can solve their LP as well, as it also involves a circulant matrix.

5. Lower Bound

In this section, we discuss the lower bound in Theorem 1.2. We will begin by providing the formal definition of *updatable* histogram publishing algorithms and the *sensitivity* of the profile reconstruction algorithms when composed with the private histogram publishing algorithm. To establish the lower bound, we will then present a reduction from the inner product estimation problem in the two-party setting.

Our lower bound applies to an algorithm $\mathcal{A}$ that is updatable, meaning that there exists an update algorithm $\mathcal{U}$, s.t., given the output $\mathcal{A}(\vec{h})$ and some other histogram $\vec{h}'$, $\mathcal{U}$ can adjust the output directly (by computing $\mathcal{U}(\mathcal{A}(\vec{h}), \tilde{h})$), as if the input to $\mathcal{A}$ is $\vec{h}+\vec{h}'$. Formally,

Definition 5.1 (Updatable Algorithm)**.** A (randomized) algorithm $\mathcal{A} : [0\,..\,n]^d \to \mathcal{Y}$ is called updatable, if there exists another (randomized) algorithm $\mathcal{U} : \mathcal{Y} \times [0\,..\,n]^d \to \mathcal{Y}$, the *update algorithm*, such that for every $\vec{h}, \vec{h}' \in [0\,..\,n]^d$, $\mathcal{U}(\mathcal{A}(\vec{h}), \vec{h}')$ has the same distribution as $\mathcal{A}(\vec{h}+\vec{h}')$.

Example 5.2. If $\mathcal{A}$ is the discrete Laplace mechanism, then it is updatable. Let $\mathcal{U}$ be the vector addition operation: $\mathcal{U}(\mathcal{A}(\vec{h}), \vec{h}') = \vec{h} + \mathbb{DL}\mathrm{ap}\left(e^{-\varepsilon}\right)^d + \vec{h}' = \mathcal{A}(\vec{h}+\vec{h}')$, where $\mathbb{DL}\mathrm{ap}\left(e^{-\varepsilon}\right)^d \in \mathbb{Z}^d$ is a vector composed of independent discrete Laplace random variables.

When $\mathcal{A}$ represents a randomized algorithm, it can be conceptualized as a function, denoted as F_A, acting on both the input of $\mathcal{A}$ and the randomness inherent in its execution (commonly referred to as "coin tosses" in the literature). This randomness is formally represented by the random variable C_A. In what follows, denote $\mathcal{R} : \mathcal{Y} \to [0,1]^{n+1}$ as an algorithm trying to reconstruct the profile of $\vec{h}$ when given $\mathcal{A}(\vec{h})$.

Definition 5.3. The sensitivity of $\mathcal{R}\circ\mathcal{A}$ is defined as the maximum deviation of $\mathcal{R}\circ F_A$, over all possible realization c_A of C_A and all possible pair of neighboring inputs of $\mathcal{A}$:

$$\Delta_{\mathcal{R}\circ\mathcal{A}} \doteq \max_{c_A}\max_{\vec{h}\sim\vec{h}'} \left\|\mathcal{R}\Big(F_A\Big(\vec{h}, c_A\Big)\Big) - \mathcal{R}\Big(F_A\Big(\vec{h}', c_A\Big)\Big)\right\|_\infty,$$

where the second maximum is over all $\vec{h}, \vec{h}' \in [0\,..\,n]^d$ such that $\|\vec{h}-\vec{h}'\|_1 \le 1$

In Appendix G, we show that when $\mathcal{A}$ is the discrete Laplace mechanism, $\mathcal{R}$ is the profile reconstruction algorithm stated in Theorem 1.1, and $\frac{1}{\sqrt{d}} \in o\big(\big(\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}\big)^2 \cdot \frac{\varepsilon}{e^{c\cdot\varepsilon}}\big)$, $\Delta_{\mathcal{R}\circ\mathcal{A}}$ satisfies $\Delta_{\mathcal{R}\circ\mathcal{A}} \in o\big(\frac{1}{\sqrt{d}} \cdot \frac{\varepsilon}{e^{c\cdot\varepsilon}}\big)$. Hence, the lower bound in Theorem 1.2 applies to our algorithms.

Two Party Differential Privacy. We will prove the lower bound by establishing a connection between the profile estimation problem and the inner product estimation problem in the two party differential privacy setting (McGregor et al., 2010).

Two players, Alice and Bob, equipped with unrestricted computational capabilities, collaborate to compute the inner product $\langle\vec{x},\vec{y}\rangle$, where $\vec{x} \in \{-1,1\}^d$ is the binary string held by Alice, and $\vec{y} \in \{-1,1\}^d$ is one held by Bob. The collaboration follows a pre-agreed protocol, denoted as Π. In this protocol, they alternate turns to communicate with each other. During each player's turn, the communicated information is determined by their own vector, the cumulative communication up to that point, and possibly their randomness. The sequences of messages, which collectively form the protocol's transcript, are denoted as Π. To protect the data of both parties, it is required that each party's view of Π, as a function of the other party's data, should be differentially private.

Formally, the view of Alice, denoted as $\Pi_{\vec{x}} : \{-1,1\}^d \to \mathcal{Z}$ (where $\mathcal{Z}$ is the set of all possible sequences of messages), is regarded as a (randomized) algorithm with input string $\vec{y}$ when $\vec{x}$ is fixed. The view $\Pi_{\vec{x}}$ is ε-DP, if for each measurable $Z \subseteq \mathcal{Z}$ and every pair of neighboring vector $\vec{y}$ and $\vec{y}'$ s.t. $\|\vec{y}-\vec{y}'\| = 1$, it holds that

$$\Pr\left[\Pi_{\vec{x}}(\vec{y}) \in Z\right] \le e^{\varepsilon}\cdot \Pr\left[\Pi_{\vec{x}}(\vec{y}') \in Z\right].$$

We can define the view of Bob $\Pi_{\vec{y}} : \{-1,1\}^d \to \mathcal{Z}$ and the ε-DP property of $\Pi_{\vec{y}}$ in a similar manner. The protocol Π is called ε-DP, if $\Pi_{\vec{x}}$ is ε-DP for all possible $\vec{x}$, and $\Pi_{\vec{y}}$ is ε-DP for all possible $\vec{y}$. McGregor et al. (2010) established a lower bound on inner product estimation, which was further refined by Haitner et al. (2022).

Algorithm 4 Protocol P

Alice: **send** $m_A \leftarrow \mathcal{A}(\vec{x} + \vec{\mathbb{1}})$ to Bob.

Bob: ∘ on receiving m_A, compute $\vec{r} \leftarrow \mathcal{R}(\mathcal{U}(m_A, \vec{y} + \vec{\mathbb{1}}))$
∘ **publish** $d(\vec{r}[4] + \vec{r}[0] - \vec{r}[2] + 3\Delta_{\mathcal{R}\circ\mathcal{A}} \cdot \mathbb{L}\mathrm{ap}\,(1/\varepsilon))$

Fact 5.4 (The Two Party Lower Bound (Haitner et al., 2022)). An ε-DP two-party protocol that, for some $\ell \geq \log d$, estimate the inner product over $\{\text{-}1, 1\}^d \times \{\text{-}1, 1\}^d$ up to an additive error ℓ with probability $c \cdot e^{c\cdot\epsilon} \cdot \ell/\sqrt{d}$ (for some universal constant $c > 0$), can be used to construct a key-agreement protocol.

Since an information theoretically secure key agreement does not exist (Haitner et al., 2022), Fact 5.4 implies that such protocols do not exist either.

Sketch Proof of Theorem 1.2. Based on Fact 5.4, we present a sketch proof for the theorem. The complete proof is included in the Appendix G.

We will prove the theorem by contraction. Assume there exists an ε-DP and updatable algorithm $\mathcal{A} : [0\,..\,n]^d \to \mathcal{Y}$, profile estimation algorithm $\mathcal{R} : \mathcal{Y} \to [0, 1]^{n+1}$, such that, for every input $\vec{h} \in [0\,..\,n]^d$,

$$\mathbb{E}[\|\vec{f} - \mathcal{R}(\mathcal{A}(\vec{h}))\|_\infty] \in o\big(1/(\sqrt{d} \cdot e^{c\cdot\varepsilon})\big).$$

We will construct an ε-DP protocol P such that for every $\vec{x}$ and $\vec{y}$, Bob's output (denoted by m_B) satisfies

$$\Pr\left[|m_B - \langle\vec{x}, \vec{y}\rangle| \geq \mathcal{E}rr\right] \in o(1), \tag{10}$$

where $\mathcal{E}rr = \sqrt{d}/(2 \cdot c \cdot e^{c\cdot\varepsilon})$, contradicting Fact 5.4.

The protocol is outlined by Algorithm 4, where $\mathbb{L}\mathrm{ap}\,(1/\varepsilon)$ denotes a random variable following the Laplace distribution with probability density $p(z) = \frac{\varepsilon}{2} \cdot \exp\left(-\varepsilon \cdot |z|\right), \forall z \in \mathbb{R}$.

Since $\vec{x}, \vec{y} \in \{\text{-}1, 1\}^d$ and $\vec{x} + \vec{\mathbb{1}}, \vec{y} + \vec{\mathbb{1}} \in \{0, 2\}^d$, it can be verified that

$$\begin{aligned}\langle\vec{x}, \vec{y}\rangle &= |\{\ell \in [d] : \vec{x}[\ell] = \vec{y}[\ell]\}| - |\{\ell \in [d] : \vec{x}[\ell] \neq \vec{y}[\ell]\}| \\ &= \left|\left\{\ell \in [d] : (\vec{x} + \vec{\mathbb{1}})[\ell] = (\vec{y} + \vec{\mathbb{1}})[\ell]\right\}\right| \\ &\quad - \left|\left\{\ell \in [d] : (\vec{x} + \vec{\mathbb{1}})[\ell] \neq (\vec{y} + \vec{\mathbb{1}})[\ell]\right\}\right| \\ &= d \cdot (\vec{f}[4] + \vec{f}[0] - \vec{f}[2]).\end{aligned}$$

Via the accuracy assumption of $\mathcal{R}$ and the sensitivity assumption of $\Delta_{\mathcal{R}\circ\mathcal{A}}$ (in Theorem 1.2), it holds that $m_B \doteq d \cdot (\vec{r}[4] + \vec{r}[0] - \vec{r}[2] + 3 \cdot \Delta_{\mathcal{R}\circ\mathcal{A}} \cdot \mathbb{L}\mathrm{ap}\,(1/\varepsilon))$ is an estimate of $d \cdot (\vec{f}[4] + \vec{f}[0] - \vec{f}[2])$ with expected error $o(\sqrt{d}/e^{c\cdot\varepsilon})$:

$$\mathbb{E}\left[\left|m_B - d \cdot (\vec{f}[4] + \vec{f}[0] - \vec{f}[2])\right|\right] \in o(\sqrt{d}/e^{c\cdot\varepsilon}).$$

Combing $d \cdot (\vec{f}[4] + \vec{f}[0] - \vec{f}[2]) = \langle\vec{x}, \vec{y}\rangle$ with Markov inequality proves Equation (10).

Finally, we verify that P is ε-DP. Clearly $\mathcal{A}(\vec{x} + \vec{\mathbb{1}})$ is ε-DP. The output of Bob, $d \cdot (\vec{r}[4] + \vec{r}[0] - \vec{r}[2] + 3 \cdot \Delta_{\mathcal{R}\circ\mathcal{A}} \cdot \mathbb{L}\mathrm{ap}\,(1/\varepsilon))$ is also ε-DP. This is because, by the definition of $\Delta_{\mathcal{R}\circ\mathcal{A}}$, it serves as an upper bound on the sensitivity of $\vec{r}[4], \vec{r}[0]$ and $\vec{r}[2]$. □

We provide some further comments on the lower bound. First, it is also possible to prove a lower bound of

$$\mathbb{E}\left[\left\|\vec{f} - \mathcal{R}(\mathcal{A}(\vec{h}))\right\|_\infty\right] \in o\Big(\tfrac{1}{\sqrt{d}\cdot\log d} \cdot \tfrac{1}{e^\varepsilon}\Big), \tag{11}$$

for Theorem 1.2, via a similar reduction to the two-party inner product lower bound accuracy lower bound by McGregor et al. (2010). In particular, they showed that with probability $1 - \delta$, inner product estimation in a two-party setting incurs an additive error of $\Omega\Big(\frac{\sqrt{d}}{\log d} \cdot \frac{\delta}{e^\varepsilon}\Big)$. Compared to Equation (4), Equation (11) avoids a constant c blow up in the exponent, at the cost of an addition factor of $\log d$ in the denominator.

Second, Theorem 1.2 assumes $\Delta_{\mathcal{R}\circ\mathcal{A}} \in o\big(\frac{1}{\sqrt{d}} \cdot \frac{\varepsilon}{e^{c\cdot\varepsilon}}\big)$. We considered removing this assumption because it appears that if $\Delta_{\mathcal{R}\circ\mathcal{A}} \in \Omega\big(\frac{1}{\sqrt{d}} \cdot \frac{\varepsilon}{e^{c\cdot\varepsilon}}\big)$, it would imply that $\mathbb{E}[\|\vec{f} - \mathcal{R}(\mathcal{A}(\vec{h}))\|_\infty] \in \Omega\big(\frac{1}{\sqrt{d}\cdot e^{c\cdot\varepsilon}}\big)$ for some $\vec{h}$. However, this is not the case. Since $\mathcal{A}$ is a randomized algorithm, $\Delta_{\mathcal{R}\circ\mathcal{A}}$ is defined over all randomness c_A of $\mathcal{A}$ as per Definition 5.3. It is possible that $\Delta_{\mathcal{R}\circ\mathcal{A}} \in \Omega\big(\frac{1}{\sqrt{d}} \cdot \frac{\varepsilon}{e^{c\cdot\varepsilon}}\big)$ for the c_A's with very small probabilities. Therefore, we cannot derive a lower bound for $\mathbb{E}[\|\vec{f} - \mathcal{R}(\mathcal{A}(\vec{h}))\|_\infty]$.

Finally, our current lower bound applies to an algorithm $\mathcal{A}$ that protects the number of occurrences of a single element by 1. It extends naturally to an algorithm $\mathcal{A}$ that protects the number of occurrences of a single element by an arbitrary amount, as long as it satisfies the other assumptions in Theorem 1.2.

Impact Statement

The goal of this work is to advance privacy-preserving data analysis and machine learning in distributed and large-scale settings. Generally speaking this work adds to the toolbox of ways in which we can incorporate privacy into system designs. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

Acknowledgements

We thank the anonymous reviewers for their feedback which helped improve the paper. The authors are supported by Providentia, a Data Science Distinguished Investigator grant from Novo Nordisk Fonden and affiliated with Basic Algorithms Research Copenhagen (BARC), supported by the VILLUM Foundation grant 16582.

References

Acharya, J., Das, H., Orlitsky, A., and Suresh, A. T. A unified maximum likelihood approach for estimating symmetric properties of discrete distributions. In *International Conference on Machine Learning (ICML)*, volume 70 of *Proceedings of Machine Learning Research*, pp. 11–21. PMLR, 2017.

Audibert, J., Munos, R., and Szepesvári, C. Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. *Theor. Comput. Sci.*, 410(19):1876–1902, 2009.

Blocki, J., Datta, A., and Bonneau, J. Differentially private password frequency lists. In *Network and Distributed System Security Symposium (NDSS)*. The Internet Society, 2016.

Chen, J. Y., Indyk, P., and Woodruff, D. P. Space-Optimal Profile Estimation in Data Streams with Applications to Symmetric Functions. In Guruswami, V. (ed.), *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, volume 287 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 32:1–32:22, 2024.

Chung, F. and Lu, L. Concentration inequalities and martingale inequalities: a survey. *Internet Math.*, 3(1):79–127, 2006.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. *Introduction to Algorithms, 3rd Edition*. MIT Press, 2009. ISBN 978-0-262-03384-8.

Datar, M. and Muthukrishnan, S. Estimating rarity and similarity over data stream windows. In *Proceedings of European Symposium on Algorithms (ESA)*, pp. 323–335. Springer, 2002.

Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. D. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, Proceedings*, volume 3876 of *Lecture Notes in Computer Science*, pp. 265–284. Springer, 2006.

Dwork, C., Naor, M., Pitassi, T., Rothblum, G. N., and Yekhanin, S. Pan-private streaming algorithms. In *Proceedings of Innovations in Computer Science (ICS)*, pp. 66–80, 2010. Now known as ITCS.

Ghosh, A., Roughgarden, T., and Sundararajan, M. Universally utility-maximizing privacy mechanisms. In Mitzenmacher, M. (ed.), *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pp. 351–360. ACM, 2009.

Gray, R. M. Toeplitz and circulant matrices: A review. *Found. Trends Commun. Inf. Theory*, 2(3), 2005.

Haitner, I., Mazor, N., Silbak, J., and Tsfadia, E. On the complexity of two-party differential privacy. In Leonardi, S. and Gupta, A. (eds.), *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pp. 1392–1405. ACM, 2022.

Hardt, M. and Talwar, K. On the geometry of differential privacy. In Schulman, L. J. (ed.), *Proceedings of Symposium on Theory of Computing (STOC)*, pp. 705–714. ACM, 2010.

Hay, M., Li, C., Miklau, G., and Jensen, D. D. Accurate estimation of the degree distribution of private networks. In *International Conference on Data Mining (ICDM)*, pp. 169–178. IEEE Computer Society, 2009.

Hay, M., Rastogi, V., Miklau, G., and Suciu, D. Boosting the accuracy of differentially private histograms through consistency. *Proc. VLDB Endow.*, 3(1):1021–1032, 2010.

Karwa, V. and Slavkovic, A. B. Differentially private graphical degree sequences and synthetic graphs. In *Privacy in Statistical Databases (PSD)*, volume 7556 of *Lecture Notes in Computer Science*, pp. 273–285. Springer, 2012.

Manurangsi, P. Tight bounds for differentially private anonymized histograms. In *5th Symposium on Simplicity in Algorithms*, pp. 203–213. SIAM, 2022.

McGregor, A., Mironov, I., Pitassi, T., Reingold, O., Talwar, K., and Vadhan, S. P. The limits of two-party differential privacy. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, USA*, pp. 81–90. IEEE Computer Society, 2010.

Mitzenmacher, M. and Upfal, E. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.

Suresh, A. T. Differentially private anonymized histograms. In *Advances in Neural Information Processing Systems 32*, pp. 7969–7979, 2019.

Tolstikhin, I. O. Concentration inequalities for samples without replacement. *Theory of Probability & Its Applications*, 61(3):462–481, 2017.

A. Probability Inequalities

Definition A.1 (Lipschitz Condition)**.** Consider an arbitrary set $\mathcal{Z}$. A function $\Delta : \mathcal{Z}^d \to \mathbb{R}$ satisfies the Lipschitz condition with bound $c \in \mathbb{R}$ if, for every $i \in [d]$ and for every sequence of values $x_1, \ldots, x_d \in \mathcal{Z}$ and $y_i \in \mathbb{R}$,

$$|\Delta(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d) - \Delta(x_1, x_2, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_d)| \leq c.$$

Fact A.2 (McDiarmid's Inequality (Mitzenmacher & Upfal, 2017; Tolstikhin, 2017)). Let $\Delta : \mathcal{Z}^d \to \mathbb{R}$ be a function that satisfies the Lipschitz condition with bound $c \in \mathbb{R}$. Let $X_1, \ldots, X_d$ be independent random variables from the set $\mathcal{Z}$ Then for all $\lambda \geq 0$,

$$\Pr\left[\Delta(X_1, \ldots, X_d) - \mathbb{E}\left[\Delta(X_1, \ldots, X_d)\right] \geq \lambda\right] \leq \exp\left(-\frac{2\lambda^2}{dc^2}\right).$$

Fact A.3 (Bernstein's Inequality (Chung & Lu, 2006; Audibert et al., 2009)). Let $X_1, \ldots, X_d$ be independent real-valued random variables such that $|X_i| \leq c$ with probability one. Let $S_d = \sum_{i \in [d]} X_i$ and $\mathbb{V}\mathrm{ar}\left[S_d\right] = \sum_{i \in [d]} \mathbb{V}\mathrm{ar}\left[X_i\right]$. Then for all $\eta \in (0, 1)$,

$$|S_d - \mathbb{E}\left[S_d\right]| \leq \sqrt{2\,\mathbb{V}\mathrm{ar}\left[S_d\right] \ln \frac{2}{\eta}} + \frac{2c \ln \frac{2}{\eta}}{3},$$

with probability at least $1 - \eta$.

B. Proofs for Section 4.1

Proof of Lemma 4.1. Our goal is to prove for all $t \in \mathbb{Z}$,

$$\Pr\left[\tilde{h}'[\ell] = t\right] = \Pr\left[\vec{h}[\ell] + Z = t\right]. \tag{12}$$

The equation holds trivially for each integer $t \in (0, n)$. By symmetry, we prove that the equation holds for each integer $t \geq n$. In this case,

$$\begin{aligned}
\Pr\left[\tilde{h}'[\ell] = t\right] &= \Pr\left[\mathbb{DE}\mathrm{xp}\left(e^{-\varepsilon}\right) = t - n\right] \cdot \Pr\left[\mathbb{DL}\mathrm{ap}\left(e^{-\varepsilon}\right) \geq n\right] \\
&= \left(\left(1 - e^{-\varepsilon}\right) \cdot e^{-\varepsilon \cdot (t-n)}\right) \cdot \left(\sum_{j \geq n - \vec{h}[\ell]} \frac{1 - e^{-\varepsilon}}{1 + e^{-\varepsilon}} \cdot e^{-\varepsilon \cdot j}\right) \\
&= \left(1 - e^{-\varepsilon}\right) \cdot e^{-\varepsilon \cdot (t-n)} \cdot \frac{1 - e^{-\varepsilon}}{1 + e^{-\varepsilon}} \cdot \frac{e^{-\varepsilon \cdot \left(n - \vec{h}[\ell]\right)}}{1 - e^{-\varepsilon}} \\
&= \frac{1 - e^{-\varepsilon}}{1 + e^{-\varepsilon}} \cdot e^{-\varepsilon \cdot \left(t - \vec{h}[\ell]\right)} \\
&= \Pr\left[\vec{h}[\ell] + Z = t\right].
\end{aligned}$$

□

C. Proofs for Section 4.2

Proof of Lemma 4.2. Let $\mathbb{TDL}\mathrm{ap}\left(e^{-\varepsilon}, B\right)$ denotes a random variable following truncated Laplace distribution:

$$\Pr\left[\mathbb{TDL}\mathrm{ap}\left(e^{-\varepsilon}, B\right) = t\right] = \frac{1}{P_{norm}} \cdot e^{-\varepsilon \cdot |t|}, \forall t \in [-B \,..\, B],$$

where

$$P_{norm} = 1 + 2 \cdot \sum_{j=1}^{B} e^{-\varepsilon \cdot j} = 1 + 2 \cdot \frac{e^{-\varepsilon} - \left(e^{-\varepsilon}\right)^{B+1}}{1 - e^{-\varepsilon}} = \frac{1 + e^{-\varepsilon} - 2 \cdot \left(e^{-\varepsilon}\right)^{B+1}}{1 - e^{-\varepsilon}}. \tag{13}$$

Conditioned on $\mathcal{E}_B$, all random noises introduced by Algorithm 1 follows truncated Laplace distribution. Therefore,

$$\mathcal{A}(\vec{r})[t] = \frac{1}{d} \cdot \sum_{k=0}^{t+B} \sum_{j=1}^{\vec{r}[k]\cdot d} \mathbb{1}_{[k+\mathbb{TDL}\mathrm{ap}(e^{-\varepsilon},B)=t]}, \quad \forall t \in [-B\,..\,-1].$$

$$\mathcal{A}(\vec{r})[t] = \frac{1}{d} \cdot \sum_{k=t-B}^{t+B} \sum_{j=1}^{\vec{r}[k]\cdot d} \mathbb{1}_{[k+\mathbb{TDL}\mathrm{ap}(e^{-\varepsilon},B)=t]}, \quad \forall t \in [0\,..\,n].$$

$$\mathcal{A}(\vec{r})[t] = \frac{1}{d} \cdot \sum_{k=t-B}^{n} \sum_{j=1}^{\vec{r}[k]\cdot d} \mathbb{1}_{[k+\mathbb{TDL}\mathrm{ap}(e^{-\varepsilon},B)=t]}, \quad \forall t \in [n+1\,..\,n+B],$$

where $\mathbb{1}_{[\cdot]}$ is the indicator for some event. Since by assumption, we have $\vec{r}[t] = 0$ for all $t \in [-B\,..\,n+B] \setminus [0\,..\,n]$, the above equations can be written in a unified manner:

$$\mathcal{A}(\vec{r})[t] = \frac{1}{d} \cdot \sum_{k=t-B}^{t+B} \sum_{j=1}^{\vec{r}[k]\cdot d} \mathbb{1}_{[k+\mathbb{TDL}\mathrm{ap}(e^{-\varepsilon},B)\equiv t]}, \tag{14}$$

where for the ease of notations, we omit the modular operations on the indexes and the events of the indicators. In particular, we assume $\vec{r}[k] \doteq \vec{r}[\mathsf{mi}(k)]$, and $k + \mathbb{TDL}\mathrm{ap}\,(e^{-\varepsilon}, B) \equiv t$ represents the event that $\mathsf{mi}(k + \mathbb{TDL}\mathrm{ap}\,(e^{-\varepsilon}, B)) = t$, where

$$\mathsf{mi}(x) \doteq \begin{cases} x, & \text{if } x \in [-B\,..\,n+B], \\ ((x+B) \mod (n+2B+1)) - B & \text{otherwise .} \end{cases}$$

Combining the definition of $\boldsymbol{A}$, we can verify that

$$\begin{aligned} \mathbb{E}\left[\mathcal{A}(\vec{r})[t] \mid \mathcal{E}_B\right] &= \frac{1}{d} \cdot \frac{1}{P_{norm}} \cdot \sum_{k=-B}^{B} e^{-\varepsilon\cdot|k|} \cdot (\vec{r}[t+k] \cdot d) \\ &= \sum_{k=-B}^{B} \frac{1}{P_{norm}} \cdot e^{-\varepsilon\cdot|k|} \cdot \vec{r}[t+k] \\ &= \boldsymbol{A}[t,\cdot] \times \vec{r}. \end{aligned}$$

□

D. Proofs for Section 4.3

Proof of Lemma 4.3. The proof consists of two parts. First, we show that the returned solution $\vec{r}$ satisfies the constraint $\left\langle \vec{\mathbb{1}}_{0:n}, \vec{r} \right\rangle = 1$:

$$\begin{aligned} \left\langle \vec{\mathbb{1}}_{0:n}, \vec{r} \right\rangle &= \left\langle \vec{\mathbb{1}}_{0:n}, \vec{u} - \frac{\left\langle \vec{\mathbb{1}}_{0:n}, \vec{u} \right\rangle - 1}{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a} \right\rangle} \cdot \boldsymbol{A}^{-1} \times \vec{a}' \right\rangle \\ &= \left\langle \vec{\mathbb{1}}_{0:n}, \vec{u} \right\rangle - \frac{\left\langle \vec{\mathbb{1}}_{0:n}, \vec{u} \right\rangle - 1}{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a} \right\rangle} \cdot \left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a} \right\rangle \\ &= 1. \end{aligned}$$

Next, we need to prove that $\vec{r}$ is an optimal solution. Consider an arbitrary vector $\vec{v}$, s.t., $\left\langle \vec{\mathbb{1}}_{0:n}, \vec{v} \right\rangle = 1$. We want to prove that $\left\| \boldsymbol{A} \times \vec{v} - \tilde{f} \right\|_p \geq \left\| \boldsymbol{A} \times \vec{r} - \tilde{f} \right\|_p$. The claim is trivial if $\left\| \boldsymbol{A} \times \vec{r} - \tilde{f} \right\|_p = 0$. Otherwise, based on Algorithm 2, observe

that

$$\boldsymbol{A} \times \vec{r} - \tilde{f} = \boldsymbol{A} \times (\vec{r} - \vec{u}) = \boldsymbol{A} \times \left(\frac{\left\langle \vec{\mathbb{1}}_{0:n}, \vec{u} \right\rangle - 1}{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a} \right\rangle} \cdot \boldsymbol{A}^{-1} \times \vec{a}. \right) = \frac{\left\langle \vec{\mathbb{1}}_{0:n}, \vec{u} \right\rangle - 1}{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a} \right\rangle} \cdot \vec{a}.$$

Since $\|\vec{a}\|_p = 1$, it holds that $\frac{\boldsymbol{A} \times (\vec{r} - \vec{u})}{\|\boldsymbol{A} \times (\vec{r} - \vec{u})\|_p} = \vec{a}$. Therefore,

$$\begin{aligned}
1 - \left\langle \vec{\mathbb{1}}_{0:n}, \vec{u} \right\rangle &= \left\langle \vec{\mathbb{1}}_{0:n}, \vec{v} - \vec{u} \right\rangle \\
&= \|\boldsymbol{A} \times (\vec{v} - \vec{u})\|_p \cdot \left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \frac{\boldsymbol{A} \times (\vec{v} - \vec{u})}{\|\boldsymbol{A} \times (\vec{v} - \vec{u})\|_p} \right\rangle, \\
1 - \left\langle \vec{\mathbb{1}}_{0:n}, \vec{u} \right\rangle &= \left\langle \vec{\mathbb{1}}_{0:n}, \vec{r} - \vec{u} \right\rangle \\
&= \|\boldsymbol{A} \times (\vec{r} - \vec{u})\|_p \cdot \left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \frac{\boldsymbol{A} \times (\vec{r} - \vec{u})}{\|\boldsymbol{A} \times (\vec{r} - \vec{u})\|_p} \right\rangle \\
&= \|\boldsymbol{A} \times (\vec{r} - \vec{u})\|_p \cdot \left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a} \right\rangle.
\end{aligned}$$

By the definition of $\vec{a}$, we have $\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a} \right\rangle \geq \left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \frac{\boldsymbol{A} \times (\vec{v} - \vec{u})}{\|\boldsymbol{A} \times (\vec{v} - \vec{u})\|_p} \right\rangle$, which proves that

$$\begin{aligned}
\|\boldsymbol{A} \times (\vec{r} - \vec{u})\|_p &= \left\| \boldsymbol{A} \times \vec{r} - \tilde{f} \right\|_p \\
&\leq \left\| \boldsymbol{A} \times \vec{v} - \tilde{f} \right\|_p = \|\boldsymbol{A} \times (\vec{v} - \vec{u})\|_p.
\end{aligned}$$

□

Proof of Lemma 4.4. When $\vec{r}$ is inputted to Algorithm 3, it satisfies $\sum_{t \in [0\,..\,n]} \vec{r}[t] = 1$. After line 1, it satisfies $\vec{r}[t] = 0$, for all $t \in [-B\,..\,n+B] \setminus [0\,..\,n]$. After line 7, it satisfies $\vec{r}[t] \in [0, 1]$, for all $t \in [0\,..\,n]$. However, it may violate is that $\sum_{t \in [0\,..\,n]} \vec{r}[t]$ should sum up to 1.

Since the cumulative change to the sum $\sum_{t \in [0\,..\,n]} \vec{r}[t]$ is $s_{>1} + s_{<0}$, after line 7, we have $\sum_{t \in [0\,..\,n]} \vec{r}[t] - 1 = s_{>1} + s_{<0}$. We claim that $s_{>1} + s_{<0} \geq 0$.

Since $s_{<0} \geq 0$, the claim holds trivially if $s_{>1} = 0$. On the other hand, if $s_{>1} < 0$, then there exists some $t \in [0\,..\,n]$, such that $\vec{r}[t] = 1$ after Line 6. However, after Line 6 and 7, we must have that $\vec{r}[t] \in [0, 1], \forall t \in [0\,..\,n]$. Therefore, $\sum_{t \in [0\,..\,n]} \vec{r}[t] \geq 1$, and $s_{>1} + s_{<0} = \sum_{t \in [0\,..\,n]} \vec{r}[t] - 1 \geq 0$.

In case that $s_{>1} + s_{<0} > 0$, the algorithm needs to decreases the $\vec{r}[t]$ to fix the constraint that $\sum_{t \in [0\,..\,n]} \vec{r}[t]$ should sum up to 1. Define the function $g(\tau) \doteq \sum_{t \in [0\,..\,n]} \min\{\tau, \vec{r}[t]\}$. The algorithm searches for the solution τ for which $g(\tau) = s_{>1} + s_{<0}$. By intermediate value theorem, the solution always exists, as

- g is a continuous function with respect to τ;
- $0 = g(0) < s_{>1} + s_{<0}$;
- $1 + s_{>1} + s_{<0} = \sum_{t \in [0\,..\,n]} \vec{r}[t] = g(1) > s_{>1} + s_{<0}$.

Finally the algorithm decreases each $\vec{r}[t]$ by $\min\{\tau, \vec{r}[t]\}$ to fix the constraint so that they sum up to 1. □

E. Proofs for Section 4.4

Proof of Lemma 4.6. The proof relies on Fact 3.2. It's worth emphasizing that $\boldsymbol{A}$ is a matrix with row and column indices starting from $-B$ (up to $n + B$), in contrast to the matrix described in Fact 3.2, where indices commence from 0. With appropriate shifting, we observe that the eigenvalues of $\boldsymbol{A}$ are given by

$$\varphi_t \doteq \left\langle \boldsymbol{A}[-B,:], \sqrt{n+2B+1}\cdot \vec{v}_t \right\rangle = \frac{1}{P_{norm}} \cdot \left(\sum_{j=-B}^{0} e^{-\varepsilon\cdot|j+B|}\cdot w_t^{j+B} + \sum_{j=1}^{n} 0\cdot w_t^{j+B} + \sum_{j=n+1}^{n+B} e^{-\varepsilon\cdot|j-(n+B+1)|}\cdot w_t^{j+B} \right)$$
$$= \frac{w_t^B}{P_{norm}} \cdot \left(1 + \sum_{j=1}^{B} e^{-\varepsilon\cdot j}\cdot w_t^{j} + \sum_{j=1}^{B} e^{-\varepsilon\cdot j}\cdot w_t^{-j}\cdot w_t^{n+2B+1} \right)$$
$$= \frac{w_t^B}{P_{norm}} \cdot \left(1 + \frac{(e^{-\varepsilon}\cdot w_t)^1 - (e^{-\varepsilon}\cdot w_t)^{B+1}}{1-e^{-\varepsilon}\cdot w_t} + \frac{(e^{-\varepsilon}\cdot w_t^{-1})^1 - (e^{-\varepsilon}\cdot w_t^{-1})^{B+1}}{1-e^{-\varepsilon}\cdot w_t^{-1}} \right).$$

Observe that

$$1 + \frac{(e^{-\varepsilon}\cdot w_t)^1 - (e^{-\varepsilon}\cdot w_t)^{(B+1)}}{1-e^{-\varepsilon}\cdot w_t} + \frac{(e^{-\varepsilon}\cdot w_t^{-1})^1 - (e^{-\varepsilon}\cdot w_t^{-1})^{(B+1)}}{1-e^{-\varepsilon}\cdot w_t^{-1}}$$
$$= 1 + \frac{e^{-\varepsilon}\cdot w_t - e^{-\varepsilon\cdot(B+1)}\cdot w_t^B\cdot w_t^1}{1-e^{-\varepsilon}\cdot w_t} + \frac{e^{-\varepsilon}\cdot w_t^{-1} - e^{-\varepsilon\cdot(B+1)}\cdot w_t^{-B}\cdot w_t^{-1}}{1-e^{-\varepsilon}\cdot w_t^{-1}}$$
$$= 1 + \frac{e^{-\varepsilon}\cdot w_t - e^{-2\varepsilon} - e^{-\varepsilon\cdot(B+1)}\cdot w_t^B\cdot\left(w_t^1 - e^{-\varepsilon}\right) + e^{-\varepsilon}\cdot w_t^{-1} - e^{-2\varepsilon} - e^{-\varepsilon\cdot(B+1)}\cdot w_t^{-B}\cdot\left(w_t^{-1} - e^{-\varepsilon}\right)}{1-e^{-\varepsilon}\cdot w_t - e^{-\varepsilon}\cdot w_t^{-1} + e^{-2\varepsilon}}$$
$$= \frac{1 - e^{-2\varepsilon} - e^{-\varepsilon\cdot(B+1)}\cdot\left(w_t^{B+1} + w_t^{-(B+1)} - e^{-\varepsilon}\cdot w_t^B - e^{-\varepsilon}\cdot w_t^{-B}\right)}{1-e^{-\varepsilon}\cdot w_t - e^{-\varepsilon}\cdot w_t^{-1} + e^{-2\varepsilon}}.$$

It is easy to see that each φ_t can be computed in $O(1)$ time, therefore all the eigenvalues can be computed in $O(n)$ time. This proves the first part of the lemma. It follows that

$$\boldsymbol{A}\times\boldsymbol{V} = \boldsymbol{V}\,\mathbf{diag}\left(\varphi_{-B},\varphi_{-B+1},\ldots,\varphi_{n+B}\right),$$
$$\boldsymbol{A} = \boldsymbol{V}\times\mathbf{diag}\left(\varphi_{-B},\varphi_{-B+1},\ldots,\varphi_{n+B}\right)\times(\boldsymbol{V})^{-1}.$$

Since $\varphi_t \neq 0$ for all $t\in[-B\,..\,n+B]$, the inverse of $\mathbf{diag}\left(\varphi_{-B},\varphi_{-B+1},\ldots,\varphi_{n+B}\right)$ is given by

$$\mathbf{diag}\left(\varphi_{-B},\varphi_{-B+1},\ldots,\varphi_{n+B}\right)^{-1} = \mathbf{diag}\left(\varphi_{-B}^{-1},\varphi_{-B+1}^{-1},\ldots,\varphi_{n+B}^{-1}\right).$$

Therefore, it can be verified that

$$\boldsymbol{A}^{-1} = \boldsymbol{V}\times\mathbf{diag}\left(\varphi_{-B}^{-1},\varphi_{-B+1}^{-1},\ldots,\varphi_{n+B}^{-1}\right)\times\boldsymbol{V}^{-1}.$$

□

Proof of Theorem 4.7. All steps in Algorithm 3 clearly run in linear time, with the exception of the search for the τ such that $\sum_{t\in[0\,..\,n]}\min\{\tau,\vec{r}[t]\} = s$.

We show that it can be found in $O(n\log n)$ time. First, sort the $\vec{r}[t]$ in non-decreasing order in $O(n\log n)$ time. Now, assume that $\vec{r}[0]\leq\cdots\leq\vec{r}[n]$. After this step, our goal is to find the smallest $t^*\in[0\,..\,n]$ such that $\vec{r}[t^*]\geq\tau$. Since

$$\sum_{t=0}^{n}\min\{\tau,\vec{r}[t]\} = \sum_{t=0}^{t^*-1}\vec{r}[t] + \sum_{t=t^*}^{n}\tau$$
$$= \sum_{t=0}^{t^*-1}\vec{r}[t] + (n+1-t^*)\tau = s,$$

t^* is the smallest number in $[0\,..\,n]$ satisfying $\sum_{t=0}^{t^*-1}\vec{r}[t] + (n+1-t^*)\vec{r}[t^*]\geq s$ and therefore can be identified in $O(n)$ time. Once t^* is known, τ is determined by $\tau = \left(s - \sum_{t=0}^{t^*-1}\vec{r}[t]\right)/(n+1-t^*)$.

□

F. Proofs for Section 4.5

Proof of Lemma 4.9. Since $\vec{f}$ is a vector in $\mathbb{R}^{n+1}$, we expand it into one in $\mathbb{R}^{n+2B+1}$ by padding $\vec{f}[-B] = \vec{f}[-B+1] = \cdots = \vec{f}[-1] = 0$ to the front and $\vec{f}[n+1] = \vec{f}[n+2] = \cdots = \vec{f}[n+B] = 0$ to the back of the vector. After this, $\vec{f}$ has the same dimension as $\vec{r}$, therefore $\left\|\vec{r} - \vec{f}\right\|_p$ is well-defined.

We study how the error $\left\|\vec{r} - \vec{f}\right\|_p$ changes in Algorithm 3, for $p = 1, 2, \infty$. To begin with, we divide Algorithm 3 into 3 phrases:

- Phrase 1 (Line 1): in which the algorithm rounds $\vec{r}[t]$ to 0 for each $t \in [-B \mathinner{..} n+B] \setminus [0 \mathinner{..} n]$.
- Phase 2 (Line 2 to Line 7), in which the algorithm clips the $\vec{r}[t]$ to the range of $[0, 1]$.
- Phase 3 (Line 8 to Line 12), in which the algorithm adjusts the $\vec{r}[t]$ so that they sum up to 1.

The error only decreases in Phase 2 and 3:

- Since $\vec{f}[t] = 0$ for each $t \in [-B \mathinner{..} n+B] \setminus [0 \mathinner{..} n]$, rounding $\vec{r}[t]$ to 0 for each $t \in [-B \mathinner{..} n+B] \setminus [0 \mathinner{..} n]$ only decrease $\left\|\vec{r} - \vec{f}\right\|_p$.
- Since $\vec{f}[t] \in [0, 1]$, for each $t \in [0 \mathinner{..} n]$, clipping the $\vec{r}[t], t \in [0 \mathinner{..} n]$ to the range $[0, 1]$ also decreases $\left\|\vec{r} - \vec{f}\right\|_p$.

It remains to study the change of $\left\|\vec{r} - \vec{f}\right\|_p$ in Phase 3. We discuss this for $p = 1, 2, \infty$ separately. To distinguish the value of $\vec{r}$ at different phases of the algorithm, we denote it as $\vec{r}_0$ before Phase 1, $\vec{r}_1$ right after Phase 1, $\vec{r}_2$ right after Phase 2, and $\vec{r}_3$ right after Phase 3.

$\left\|\vec{r}_3 - \vec{f}\right\|_1 \leq \left\|\vec{r}_1 - \vec{f}\right\|_1$: observe that in Phase 2 $\left\|\vec{r} - \vec{f}\right\|_1$ decreases by an amount of $-s_{>1} + s_{<0}$:

$$\begin{aligned}
\sum_{t \in [0 \mathinner{..} n]} \left|\vec{r}_1[t] - \vec{f}[t]\right| &= \sum_{t \in T_{>1}} \left|\vec{r}_1[t] - \vec{f}[t]\right| + \sum_{t \in T_{<0}} \left|\vec{r}_1[1] - \vec{f}[t]\right| + \sum_{t \in [0 \mathinner{..} n] \setminus (T_{>1} \cup T_{<0})} \left|\vec{r}_2[t] - \vec{f}[t]\right| \\
&= \sum_{t \in T_{>1}} \left(\left|1 - \vec{f}[t]\right| + \vec{r}_1[t] - 1\right) + \sum_{t \in T_{<0}} \left(\left|0 - \vec{f}[t]\right| + 0 - \vec{r}_1[t]\right) + \sum_{t \in [0 \mathinner{..} n] \setminus (T_{>1} \cup T_{<0})} \left|\vec{r}_2[t] - \vec{f}[t]\right| \\
&= \sum_{t \in [0 \mathinner{..} n]} \left|\vec{r}_2[t] - \vec{f}[t]\right| - s_{>1} + s_{<0}.
\end{aligned}$$

In Phase 3, $\left\|\vec{r} - \vec{f}\right\|_1$ can increase by at most $s_{>1} + s_{<0}$. To see that, observe that $\sum_{t \in [0 \mathinner{..} n]} \vec{r}_2[t] - 1 = \sum_{t \in [0 \mathinner{..} n]} \vec{r}_2[t] - \sum_{t \in [0 \mathinner{..} n]} \vec{r}_1[t] = s_{>1} + s_{<0} = s$. After the adjustment, we have $\vec{r}_3[t] \leq \vec{r}_2[t]$ for each $t \in [0 \mathinner{..} n]$, and $\sum_{t \in [0 \mathinner{..} n]} \vec{r}_3[t] = 1$. Therefore,

$$\begin{aligned}
\sum_{t \in [0 \mathinner{..} n]} \left|\vec{r}_3[t] - \vec{f}[t]\right| - \sum_{t \in [0 \mathinner{..} n]} \left|\vec{r}_2[t] - \vec{f}[t]\right| &\leq \sum_{t \in [0 \mathinner{..} n]} \left|\vec{r}_2[t] - \vec{r}_3[t]\right| \\
&= \sum_{t \in [0 \mathinner{..} n]} \left(\vec{r}_2[t] - \vec{r}_3[t]\right) = \sum_{t \in [0 \mathinner{..} n]} \vec{r}_2[t] - 1 = s_{>1} + s_{<0}.
\end{aligned}$$

It follows that

$$\sum_{t \in [0 \mathinner{..} n]} \left|\vec{r}_3[t] - \vec{f}[t]\right| - \sum_{t \in [0 \mathinner{..} n]} \left|\vec{r}_1[t] - \vec{f}[t]\right| \leq s_{>1} + s_{<0} - (-s_{>1} + s_{<0}) = 2 \cdot s_{>1} \leq 0.$$

$\left\|\vec{r}_3 - \vec{f}\right\|_\infty \leq 2 \cdot \left\|\vec{r}_2 - \vec{f}\right\|_\infty$: Let $b_t \doteq \left|\vec{r}_2[t] - \vec{f}[t]\right|$ for each $t \in [0\,..\,n]$, and $b_{\max} \doteq \max_{t\in[0\,..\,n]} b_t$. It is easy to see that

$$\sum_{t\in[0\,..\,n]} (\vec{r}_2[t] - \min\{b_{\max}, \vec{r}_2[t]\}) \leq \sum_{t\in[0\,..\,n]} \vec{f}[t] = 1.$$

On the other hand, we have

$$\sum_{t\in[0\,..\,n]} (\vec{r}_2[t] - \min\{\tau, \vec{r}_2[t]\}) = 1.$$

It follows that $\sum_{t\in[0\,..\,n]} \min\{\tau, \vec{r}_2[t]\} \leq \sum_{t\in[0\,..\,n]} \min\{b_{\max}, \vec{r}_2[t]\}$, and hence $\tau \leq b_{\max}$.

$$\begin{aligned}\left\|\vec{r}_3 - \vec{f}\right\|_\infty &= \max_{t\in[0\,..\,n]} \left|\vec{r}_2[t] - \min\{\tau, \vec{r}_2[t]\} - \vec{f}[t]\right| \\ &\leq \max_{t\in[0\,..\,n]} \left(\left|\vec{r}_2[t] - \vec{f}[t]\right| + \min\{\tau, \vec{r}_2[t]\}\right) \\ &\leq 2 \cdot \left\|\vec{r}_2 - \vec{f}\right\|_\infty .\end{aligned}$$

$\left\|\vec{r}_3 - \vec{f}\right\|_2 \leq \left\|\vec{r}_2 - \vec{f}\right\|_2$: without lose of generality, assume that $\vec{r}_2[0] \leq \vec{r}_2[1] \leq \ldots \leq \vec{r}_2[n]$. The adjustment step (line 11) can be replaced by Algorithm 5.

Algorithm 5 Iterated Adjustment

```
t ← 0
while t ≤ n and s > 0 do
    τ' ← min(r⃗[t], s/(n−t+1))
    r⃗[j] ← r⃗[j] − τ', ∀j ∈ [t..n]
    s ← s − τ' · (n − t + 1)
    t ← t + 1
end while
```

We claim that, $\left\|\vec{r} - \vec{f}\right\|_2$ does not increase at each iteration. Fix some iteration t. Note that $\vec{r}[j]$ does not change for $j \in [0\,..\,t-1]$, whereas $\vec{r}[j]$ decreases by τ' for $j \in [t\,..\,n]$. Therefore, the total change of $\left\|\vec{r} - \vec{f}\right\|_2^2$ is given by

$$\begin{aligned}&\sum_{j\in[t\,..\,n]} \left|\vec{r}[t] - \tau' - \vec{f}[t]\right|^2 - \sum_{j\in[t\,..\,n]} \left|\vec{r}[t] - \vec{f}[t]\right|^2 \\ &= \sum_{j\in[t\,..\,n]} -\tau' \cdot \left(2 \cdot \vec{r}[t] - 2 \cdot \vec{f}[t] - \tau'\right) \\ &= -\tau' \cdot \left(2 \cdot \sum_{j\in[t\,..\,n]} \vec{r}[t] - 2 \cdot \sum_{j\in[t\,..\,n]} \vec{f}[t] - \sum_{j\in[t\,..\,n]} \tau'\right) \\ &= -\tau' \cdot \left(2 \cdot \left(\sum_{j\in[t\,..\,n]} \vec{r}[t] - 1\right) - (n - t + 1) \cdot \tau'\right) \\ &\leq 0,\end{aligned}$$

where the last inequality follows since $0 \leq \tau' \leq \vec{r}[j]$ for each $j \in [t\,..\,n]$. □

We present the proof for Lemma 4.11 first, then the proof for Lemma 4.10.

Proof of Lemma 4.11. Since $\vec{f}$ is a vector in $\mathbb{R}^{n+1}$, we expand it into one in $\mathbb{R}^{n+2B+1}$ by padding $\vec{f}[-B] = \vec{f}[-B+1] = \cdots = \vec{f}[-1] = 0$ to the front and $\vec{f}[n+1] = \vec{f}[n+2] = \cdots = \vec{f}[n+B] = 0$ to the back of the vector, so that $\boldsymbol{A} \times \vec{f}$ is well-defined.

Recall that $\tilde{h}[\ell] = \vec{h}[\ell] + \mathbb{DLap}\left(e^{-\varepsilon}\right), \forall \ell \in \mathcal{D}$. Since $B \geq \frac{1}{\varepsilon} \cdot \ln \frac{2 \cdot d}{\eta \cdot (e^{\varepsilon}+1)}$. Then

$$\Pr\left[\left|\mathbb{DLap}\left(e^{-\varepsilon}\right)\right| > B\right] = 2 \cdot \sum_{t \geq B} \frac{1-e^{-\varepsilon}}{1+e^{-\varepsilon}} \cdot e^{-\varepsilon \cdot |t|} = 2 \cdot \frac{1-e^{-\varepsilon}}{1+e^{-\varepsilon}} \cdot \frac{e^{-\varepsilon \cdot B}}{1-e^{-\varepsilon}} = \frac{2 \cdot e^{-\varepsilon \cdot B}}{1+e^{-\varepsilon}} \leq \frac{\eta}{d}.$$

By union bound,

$$\Pr\left[\max_{\ell \in \mathcal{D}} \left|\tilde{h}[\ell] - \vec{h}[\ell]\right| > B\right] \leq \eta.$$

Assuming that $\max_{\ell \in \mathcal{D}} \left|\tilde{h}[\ell] - \vec{h}[\ell]\right| \leq B$, $\tilde{f} = \mathcal{A}(\vec{f})$ can be equivalent viewed as the outcome of running Algorithm 1 with input $\vec{f}$, (conditioned on the event $\mathcal{E}_B$ that all noises added by the algorithm are absolutely bounded by B). It follows by Lemma 4.2 that

$$\mathbb{E}\left[\tilde{f}\right] = \mathbb{E}\left[\mathcal{A}\left(\vec{f}\right) \mid \mathcal{E}_B\right] = \boldsymbol{A} \times \vec{f}.$$

Bounding $\left\|\tilde{f} - \boldsymbol{A} \times \vec{f}\right\|_1$. Under these assumptions,

$$\begin{aligned}
\mathbb{E}\left[\left\|\tilde{f} - \boldsymbol{A} \times \vec{f}\right\|_1\right] &= \mathbb{E}\left[\left\|\tilde{f} - \mathbb{E}\left[\tilde{f}\right]\right\|_1\right] = \mathbb{E}\left[\sum_{t=-B}^{n+B} \left|\tilde{f}[t] - \mathbb{E}\left[\tilde{f}[t]\right]\right|\right] \\
&= \sum_{t=-B}^{n+B} \mathbb{E}\left[\left|\tilde{f}[t] - \mathbb{E}\left[\tilde{f}[t]\right]\right|\right] = \sum_{t=-B}^{n+B} \mathbb{E}\left[\sqrt{\left|\tilde{f}[t] - \mathbb{E}\left[\tilde{f}[t]\right]\right|^2}\right] \\
&\leq \sum_{t=-B}^{n+B} \sqrt{\mathbb{E}\left[\left|\tilde{f}[t] - \mathbb{E}\left[\tilde{f}[t]\right]\right|^2\right]}.
\end{aligned}$$

Consider a $t \in [-B \,..\, n+B]$. Conditioned on the event that $\max_{\ell \in \mathcal{D}} \left|\tilde{h}[\ell] - \vec{h}[\ell]\right| \leq B$, the noises added to the $\tilde{h}[\ell]$ follows independent truncated Laplace distribution. For each $\ell \in \mathcal{D}$, define the indicator $\mathbb{1}_{\left[\tilde{h}[\ell]=t\right]}$ for the event that $\tilde{h}[\ell] = t$. We have

$$\Pr\left[\mathbb{1}_{\left[\tilde{h}[\ell]=t\right]} = 1\right] = \frac{e^{-\varepsilon \cdot \left|t - \vec{h}[\ell]\right|}}{P_{norm}},$$

where $P_{norm} = \frac{1+e^{-\varepsilon}-2\cdot(e^{-\varepsilon})^{B+1}}{1-e^{-\varepsilon}}$. Since $\tilde{f}[t] = \frac{1}{d} \cdot \sum_{\ell \in \mathcal{D}} \mathbb{1}_{\left[\tilde{h}[\ell]=t\right]}$, we have

$$\begin{aligned}
\mathbb{E}\left[\left|\tilde{f}[t] - \mathbb{E}\left[\tilde{f}[t]\right]\right|^2\right] &= \mathbb{V}\mathrm{ar}\left[\frac{1}{d} \cdot \sum_{\ell \in \mathcal{D}} \mathbb{1}_{\left[\tilde{h}[\ell]=t\right]}\right] \\
&= \frac{1}{d^2} \cdot \sum_{\ell \in \mathcal{D}} \mathbb{V}\mathrm{ar}\left[\mathbb{1}_{\left[\tilde{h}[\ell]=t\right]}\right] \\
&= \frac{1}{d^2} \cdot \sum_{\ell \in \mathcal{D}} \frac{e^{-\varepsilon \cdot \left|t - \vec{h}[\ell]\right|}}{P_{norm}} \cdot \left(1 - \frac{e^{-\varepsilon \cdot \left|t - \vec{h}[\ell]\right|}}{P_{norm}}\right) \\
&\leq \frac{1}{d} \cdot \mathbb{E}\left[\tilde{f}[t]\right].
\end{aligned}$$

Note that $1-\frac{e^{-\varepsilon\cdot|t-\vec{h}[\ell]|}}{P_{norm}} \geq 1-\frac{1}{P_{norm}} = 1-\frac{1-e^{-\varepsilon}}{1+e^{-\varepsilon}-2\cdot(e^{-\varepsilon})^{B+1}} = \frac{2e^{-\varepsilon}-2\cdot(e^{-\varepsilon})^{B+1}}{1+e^{-\varepsilon}-2\cdot(e^{-\varepsilon})^{B+1}}$. So the inequality is tight up to a factor of $\frac{2e^{-\varepsilon}-2\cdot(e^{-\varepsilon})^{B+1}}{1+e^{-\varepsilon}-2\cdot(e^{-\varepsilon})^{B+1}}$. It concludes that

$$\mathbb{E}\left[\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_1\right] \leq \sum_{t=-B}^{n+B}\sqrt{\mathbb{E}\left[\left|\tilde{f}[t]-\mathbb{E}\left[\tilde{f}[t]\right]\right|^2\right]} = \sum_{t=-B}^{n+B}\sqrt{\frac{1}{d}\cdot\mathbb{E}\left[\tilde{f}[t]\right]}.$$

Further, $\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_1$ is a function of the random variables $\tilde{h}[\ell]'s$:

$$\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_1 = \sum_{t\in[-B\,..\,n+B]}\left|\frac{1}{d}\cdot\sum_{\ell\in\mathcal{D}}\mathbb{1}_{[\tilde{h}[\ell]=t]}-\mathbb{E}\left[\tilde{f}[t]\right]\right|.$$

Changing the value of one $\tilde{h}[\ell]$ affects at most two terms on the right hand side of the equation, each by $1/d$. By Definition A.1, $\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_1$ satisfies the Lipschitz condition with bound $2/d$. Therefore, via Fact A.2,

$$\Pr\left[\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_1-\mathbb{E}\left[\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_1\right]\geq\sqrt{\frac{2\ln\frac{1}{\eta}}{d}}\right]\leq\exp\left(-\frac{2}{d\cdot(2/d)^2}\cdot\frac{2\ln\frac{1}{\eta}}{d}\right)=\eta.$$

Bounding $\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_2$. Similarly,

$$\begin{aligned}\mathbb{E}\left[\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_2\right] &= \mathbb{E}\left[\left\|\tilde{f}-\mathbb{E}\left[\tilde{f}\right]\right\|_2\right] = \mathbb{E}\left[\sqrt{\sum_{t=-B}^{n+B}\left|\tilde{f}[t]-\mathbb{E}\left[\tilde{f}[t]\right]\right|^2}\right]\\ &\leq\sqrt{\sum_{t=-B}^{n+B}\mathbb{E}\left[\left|\tilde{f}[t]-\mathbb{E}\left[\tilde{f}[t]\right]\right|^2\right]}\leq\sqrt{\sum_{t=-B}^{n+B}\frac{1}{d}\cdot\mathbb{E}\left[\tilde{f}[t]\right]}=\sqrt{\frac{1}{d}}.\end{aligned}$$

Further, $\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_2$ is a function of the random variables $\tilde{h}[\ell]$:

$$\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_2 = \sqrt{\sum_{t\in[-B\,..\,n+B]}\left|\frac{1}{d}\cdot\sum_{\ell\in\mathcal{D}}\mathbb{1}_{[\tilde{h}[\ell]=t]}-\mathbb{E}\left[\tilde{f}[t]\right]\right|^2},$$

We can view $\tilde{f} = \left((1/d)\cdot\sum_{\ell\in\mathcal{D}}\mathbb{1}_{[\tilde{h}[\ell]=-B]},\ldots,(1/d)\cdot\sum_{\ell\in\mathcal{D}}\mathbb{1}_{[\tilde{h}[\ell]=n+B]}\right)$. Assume that we change the value of one $\tilde{h}[\ell]$, to obtain a vector $\tilde{f}'$, which differs from $\tilde{f}$ by two coordinates, each by $1/d$. Therefore, $\left\|\tilde{f}'-\tilde{f}\right\|_2=\sqrt{2}/d$, and

$$\left|\left\|\tilde{f}'-\boldsymbol{A}\times\vec{f}\right\|_2-\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_2\right|\leq\left\|\tilde{f}'-\tilde{f}\right\|_2\leq\sqrt{2}/d.$$

By Definition A.1, $\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_2$ satisfies the Lipschitz condition with bound $\sqrt{2}/d$. Therefore, via Fact A.2,

$$\Pr\left[\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_2-\mathbb{E}\left[\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_2\right]\geq\sqrt{\frac{\ln\frac{1}{\eta}}{d}}\right]\leq\exp\left(-\frac{2}{d\cdot\left(\sqrt{2}/d\right)^2}\cdot\frac{\ln\frac{1}{\eta}}{d}\right)=\eta.$$

Bounding $\left\|\tilde{f}-\boldsymbol{A}\times\vec{f}\right\|_\infty$. To prove the last part, by Bernstein's inequality (Fact A.3), with probability at least $1-1/(n\eta)$,

we have

$$\begin{aligned}\left|\tilde{f}[t]-\mathbb{E}\left[\tilde{f}[t]\right]\right| &\le \sqrt{2\cdot\mathbb{V}\mathrm{ar}\left[\tilde{f}[t]\right]\cdot\ln\frac{n}{\eta}}+\frac{2}{3}\cdot\ln\frac{n}{\eta}\\ &\le \sqrt{\frac{2}{d}\cdot\left(\sum_{\ell\in\mathcal{D}}\frac{e^{-\varepsilon\cdot\left|t-\vec{h}[\ell]\right|}}{P_{norm}}\right)\cdot\ln\frac{n}{\eta}}+\frac{2}{3d}\cdot\ln\frac{n}{\eta}\\ &= \frac{1}{\sqrt{d}}\cdot\sqrt{2\cdot\mathbb{E}\left[\tilde{f}[t]\right]\cdot\ln\frac{n}{\eta}}+\frac{2}{3d}\cdot\ln\frac{n}{\eta}.\end{aligned}$$

It is not easy to have a non-trivial bound for $\mathbb{E}\left[\tilde{f}[t]\right]$, which depends on the histogram $\vec{h}$. On the other hand, a trivial bound can be given by

$$\mathbb{E}\left[\tilde{f}[t]\right]\le\frac{1}{d\cdot P_{norm}}\cdot\sum_{\ell\in\mathcal{D}}e^{-\varepsilon\cdot 0}=\frac{1}{P_{norm}}.$$

□

Proof of Lemma 4.10. We first bound $\left\|\boldsymbol{A}^{-1}\right\|_2$, since it is the easier part of the proof. According to Lemma 4.6,

$$\left\|\boldsymbol{A}^{-1}\right\|_2=\left\|\boldsymbol{V}\times\mathbf{diag}\left(\varphi_{-B}^{-1},\varphi_{-B+1}^{-1},\ldots,\varphi_{n+B}^{-1}\right)\times\boldsymbol{V}^{-1}\right\|_2 \tag{15}$$

$$\le\left\|\boldsymbol{V}\right\|_2\cdot\left\|\mathbf{diag}\left(\varphi_{-B}^{-1},\varphi_{-B+1}^{-1},\ldots,\varphi_{n+B}^{-1}\right)\right\|_2\cdot\left\|\boldsymbol{V}^{-1}\right\|_2. \tag{16}$$

Since $\boldsymbol{V}$ is the discrete Fourier transform matrix with orthonormal column and row vectors, it holds that $\left\|\boldsymbol{V}\right\|_2=\left\|\boldsymbol{V}^{-1}\right\|_2=1$. Moreover, $\left\|\mathbf{diag}\left(\varphi_{-B}^{-1},\varphi_{-B+1}^{-1},\ldots,\varphi_{n+B}^{-1}\right)\right\|_2=\max_{t\in[-B\,..\,n+B]}\frac{1}{|\varphi_t|}$. We proceed to show that to bounds $\max_{t\in[-B\,..\,n+B]}\frac{1}{|\varphi_t|}=1/\min_{t\in[-B\,..\,n+B]}|\varphi_t|$. For each $t\in[-B\,..\,n+B]$,

$$|\varphi_t|=\left|\frac{w_t^B}{P_{norm}}\cdot\frac{1-e^{-2\varepsilon}-e^{-\varepsilon\cdot(B+1)}\cdot\left(w_t^{B+1}+w_t^{-(B+1)}-e^{-\varepsilon}\cdot w_t^B-e^{-\varepsilon}\cdot w_t^{-B}\right)}{1-e^{-\varepsilon}\cdot w_t-e^{-\varepsilon}\cdot w_t^{-1}+e^{-2\varepsilon}}\right|$$

We have

$$\left|w_t^B\right|=1.$$

$$\left|1-e^{-\varepsilon}\cdot w_t-e^{-\varepsilon}\cdot w_t^{-1}+e^{-2\varepsilon}\right|\le\left|1+2\cdot e^{-\varepsilon}+e^{-2\varepsilon}\right|=\left(1+e^{-\varepsilon}\right)^2.$$

and

$$\begin{aligned}&\left|1-e^{-2\varepsilon}-e^{-\varepsilon\cdot(B+1)}\cdot\left(w_t^{B+1}+w_t^{-(B+1)}-e^{-\varepsilon}\cdot w_t^B-e^{-\varepsilon}\cdot w_t^{-B}\right)\right|\\ &\ge\left|1-e^{-2\varepsilon}\right|+e^{-\varepsilon\cdot(B+1)}\cdot\left|w_t^{B+1}+w_t^{-(B+1)}-e^{-\varepsilon}\cdot w_t^B-e^{-\varepsilon}\cdot w_t^{-B}\right|\\ &\ge 1-e^{-2\varepsilon}-e^{-\varepsilon\cdot(B+1)}\cdot 2\cdot\left(1+e^{-\varepsilon}\right)\\ &=\left(1-e^{-\varepsilon}-2\cdot e^{-\varepsilon\cdot(B+1)}\right)\cdot\left(1+e^{-\varepsilon}\right).\end{aligned}$$

Therefore

$$|\varphi_t|\ge\frac{1}{P_{norm}}\cdot\frac{\left(1-e^{-\varepsilon}-2\cdot e^{-\varepsilon\cdot(B+1)}\right)}{\left(1+e^{-\varepsilon}\right)}.$$

To bound $\left\|\boldsymbol{A}^{-1}\right\|_1$ and $\left\|\boldsymbol{A}^{-1}\right\|_\infty$, we will first show that there two values are equal; then we bound just $\left\|\boldsymbol{A}^{-1}\right\|_\infty$.

$\left\|\boldsymbol{A}^{-1}\right\|_1 = \left\|\boldsymbol{A}^{-1}\right\|_\infty$. We will show that, $\boldsymbol{A}^{-1}$ is also a circulant matrix. Since each column of a circulant matrix has the same ℓ_1-norm as each row of the matrix, and since

$$\left\|\boldsymbol{A}^{-1}\right\|_1 = \max_{t\in[-B\,..\,n+B]} \left\|\boldsymbol{A}^{-1}[\cdot,t]\right\|_1,$$
$$\left\|\boldsymbol{A}^{-1}\right\|_\infty = \max_{t\in[-B\,..\,n+B]} \left\|\boldsymbol{A}^{-1}[t,\cdot]\right\|_1,$$

it follows that $\left\|\boldsymbol{A}^{-1}\right\|_1 = \left\|\boldsymbol{A}^{-1}\right\|_\infty$. It remains to show that $\boldsymbol{A}^{-1}$ is also a circulant matrix. According to Lemma 4.6, $\boldsymbol{V}$ is the discrete Fourier transform matrix whose rows and columns are indexed from $-B$ to $n+B$. Therefore $\forall k,\ell \in [-B\,..\,n+B]$, it holds that

$$\begin{aligned}
\boldsymbol{V}[k,\ell] &= \frac{1}{\sqrt{n+2B+1}} \cdot \varphi_\ell^{k+B} \\
&= \frac{1}{\sqrt{n+2B+1}} \cdot e^{-\frac{2\pi(k+B)\cdot(\ell+B)\cdot i}{n+2B+1}}, \\
\boldsymbol{V}^{-1}[k,\ell] &= \frac{1}{\sqrt{n+2B+1}} \cdot \varphi_k^{-(\ell+B)} \\
&= \frac{1}{\sqrt{n+2B+1}} \cdot e^{\frac{2\pi(k+B)\cdot(\ell+B)\cdot i}{n+2B+1}}.
\end{aligned}$$

Since $\boldsymbol{A}^{-1} = \boldsymbol{V} \times \mathbf{diag}\left(\varphi_{-B}^{-1}, \varphi_{-B+1}^{-1}, \ldots, \varphi_{n+B}^{-1}\right) \times \boldsymbol{V}^{-1}$, we can verify that

$$\begin{aligned}
\boldsymbol{A}^{-1}[k,\ell] &= \boldsymbol{V}[k,\cdot] \times \mathbf{diag}\left(\varphi_{-B}^{-1}, \varphi_{-B+1}^{-1}, \ldots, \varphi_{n+B}^{-1}\right) \times \boldsymbol{V}^{-1}[\cdot,\ell] \\
&= \frac{1}{n+2B+1} \cdot \left\langle \left(e^{-\frac{2\pi(k+B)\cdot 0\cdot i}{n+2B+1}}\varphi_{-B}^{-1}, \ldots, e^{-\frac{2\pi(k+B)\cdot(n+2B)\cdot i}{n+2B+1}}\varphi_{n+B}^{-1}\right), \left(e^{\frac{2\pi 0\cdot(\ell+B)\cdot i}{n+2B+1}}, \ldots, e^{\frac{2\pi(n+2B)\cdot(\ell+B)\cdot i}{n+2B+1}}\right)\right\rangle \\
&= \frac{1}{n+2B+1} \cdot \sum_{t\in[-B\,..\,n+B]} e^{-\frac{2\pi(k-\ell)\cdot(t+B)\cdot i}{n+2B+1}} \varphi_t^{-1}.
\end{aligned}$$

Now it is easy to see that $\boldsymbol{A}^{-1}$ is a circulant matrix, since for each $-B \le k < n+B$, we have

$$\begin{aligned}
\boldsymbol{A}^{-1}[k+1,\ell+1] &= \boldsymbol{A}^{-1}[k,\ell], \quad \forall \ell \in [-B\,..\,n+B-1], \\
\boldsymbol{A}^{-1}[k+1,-B] &= \boldsymbol{A}^{-1}[k,n+B].
\end{aligned}$$

Bounding $\left\|\boldsymbol{A}^{-1}\right\|_\infty$. We will show that for each $\vec{v} \in \mathbb{R}^{n+2B+1}$, it holds that

$$\left\|\boldsymbol{A}^{-1}\vec{v}\right\|_1 \le \frac{2+e^{-\varepsilon}+e^{\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}-4\cdot e^{-\varepsilon\cdot B}} \cdot P_{norm} \cdot \|\vec{v}\|_1.$$

Since $\boldsymbol{A}$ is invertible, let $\vec{x} \doteq \boldsymbol{A}^{-1}\vec{v}$. It reduces to prove that

$$\|\vec{x}\|_1 \le \frac{2+e^{-\varepsilon}+e^{\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}-4\cdot e^{-\varepsilon\cdot B}} \cdot P_{norm} \cdot \|\boldsymbol{A}\vec{x}\|_1.$$

To finish the proof, we need the following lemma.

Lemma F.1. *Let $\vec{x} \in \mathbb{R}^{n+2B+1}$ and assume that for each $t \in [-B\,..\,n+B]$, there exists $b_t \in \mathbb{R}_{\ge 0}$, s.t. $|\boldsymbol{A}[t,\cdot] \times \vec{x}| \le b_t$. The for each $t \in [-B\,..\,n+B]$, we have*

$$|\vec{x}[t]| \le \frac{e^{-\varepsilon}+e^{\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}} \cdot \left(P_{norm} \cdot \left(b_t + \frac{b_{t-1}+b_{t+1}}{e^{-\varepsilon}+e^{\varepsilon}}\right) + \frac{e^{-\varepsilon\cdot B}}{e^{-\varepsilon}+e^{\varepsilon}} \cdot \left(|\vec{x}[t-B-1]| + |\vec{x}[t+B+1]| + e^{-\varepsilon}\cdot|\vec{x}[t-B]| + e^{-\varepsilon}\cdot|\vec{x}[t+B]|\right)\right),$$

where for simplicity we omit the modular operations on the subscripts and indexes. In particular, we define $b_{-B-1} \doteq b_{n+B}$, $b_{n+B+1} = b_{-B}$, and for each $t \notin [-B\,..\,n+B]$, we let $\vec{x}[t] \doteq \vec{x}[((t+B) \mod (n+2B+1)) + (-B)]$.

Based on Lemma F.1, we have

$$\begin{aligned}
\sum_{t\in[-B\,..\,n+B]} |\vec{x}[t]| &\le \sum_{t\in[-B\,..\,n+B]} \frac{e^{-\varepsilon}+e^{\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}} \cdot \left(\begin{array}{c} P_{norm}\cdot\left(b_t + \frac{b_{t-1}+b_{t+1}}{e^{-\varepsilon}+e^{\varepsilon}}\right) + \\ \frac{e^{-\varepsilon\cdot B}}{e^{-\varepsilon}+e^{\varepsilon}} \cdot (|\vec{x}[t-B-1]| + |\vec{x}[t+B+1]| + e^{-\varepsilon}\cdot|\vec{x}[t-B]| + e^{-\varepsilon}\cdot|\vec{x}[t+B]|) \end{array} \right) \\
&= \sum_{t\in[-B\,..\,n+B]} \frac{e^{-\varepsilon}+e^{\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}} \cdot P_{norm}\cdot b_t \cdot \left(1+\frac{2}{e^{-\varepsilon}+e^{\varepsilon}}\right) + 4\cdot \sum_{t\in[-B\,..\,n+B]} \frac{e^{-\varepsilon}+e^{\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}} \cdot \frac{e^{-\varepsilon\cdot B}}{e^{-\varepsilon}+e^{\varepsilon}} \cdot |\vec{x}[t]| \\
&= \sum_{t\in[-B\,..\,n+B]} \frac{2+e^{-\varepsilon}+e^{\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}} \cdot P_{norm}\cdot b_t + \sum_{t\in[-B\,..\,n+B]} \frac{4\cdot e^{-\varepsilon\cdot B}}{e^{\varepsilon}-e^{-\varepsilon}} \cdot |\vec{x}[t]|\,.
\end{aligned}$$

This implies that

$$\begin{aligned}
\frac{e^{\varepsilon}-e^{-\varepsilon}-4\cdot e^{-\varepsilon\cdot B}}{e^{\varepsilon}-e^{-\varepsilon}} \cdot \sum_{t\in[-B\,..\,n+B]} |\vec{x}[t]| &\le \sum_{t\in[-B\,..\,n+B]} \frac{2+e^{-\varepsilon}+e^{\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}} \cdot P_{norm}\cdot b_t, \\
\iff \sum_{t\in[-B\,..\,n+B]} |\vec{x}[t]| &\le \frac{2+e^{-\varepsilon}+e^{\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}-4\cdot e^{-\varepsilon\cdot B}} \cdot P_{norm} \cdot \sum_{t\in[-B\,..\,n+B]} b_t,
\end{aligned}$$

□

Proof of Lemma F.1. Fix a $t \in [-B\,..\,n+B]$. By assumption, we have

$$\begin{aligned}
P_{norm}\cdot b_t &\ge P_{norm}\cdot |\boldsymbol{A}[t,\cdot]\times\vec{x}| \\
&= \left| \sum_{j\in[B]} e^{-\varepsilon\cdot j}\cdot\vec{x}[t-j] + e^{-\varepsilon\cdot 0}\cdot\vec{x}[t] + \sum_{j\in[B]} e^{-\varepsilon\cdot j}\cdot\vec{x}[t+j] \right|.
\end{aligned}$$

On the other hand, since $\boldsymbol{A}$ is a circulant matrix,

$$\begin{aligned}
P_{norm}\cdot b_{t-1} &\ge P_{norm}\cdot |\boldsymbol{A}[t-1,\cdot]\times\vec{x}| \\
&= \left| \sum_{j\in[B+1]} e^{-\varepsilon\cdot(j-1)}\cdot\vec{x}[t-j] + e^{-\varepsilon}\cdot\vec{x}[t] + \sum_{j\in[B-1]} e^{-\varepsilon\cdot(j+1)}\cdot\vec{x}[t+j] \right|, \\
P_{norm}\cdot b_{t+1} &\ge P_{norm}\cdot |\boldsymbol{A}[t+1,\cdot]\times\vec{x}| \\
&= \left| \sum_{j\in[B-1]} e^{-\varepsilon\cdot(j+1)}\cdot\vec{x}[t-j] + e^{-\varepsilon}\cdot\vec{x}[t] + \sum_{j\in[B+1]} e^{-\varepsilon\cdot(j-1)}\cdot\vec{x}[t+j] \right|.
\end{aligned}$$

Therefore

$$\begin{aligned}
P_{norm}\cdot\left(b_t + \frac{b_{t-1}+b_{t+1}}{e^{-\varepsilon}+e^{\varepsilon}}\right) &\ge P_{norm}\cdot\left|\boldsymbol{A}[t,\cdot]\times\vec{x} - \frac{1}{e^{-\varepsilon}+e^{\varepsilon}}\cdot(\boldsymbol{A}[t-1,\cdot]\times\vec{x} + \boldsymbol{A}[t+1,\cdot]\times\vec{x})\right| \\
&= \left| \begin{array}{c} -\frac{e^{-\varepsilon\cdot B}}{e^{-\varepsilon}+e^{\varepsilon}}\cdot\vec{x}[t-B-1] + \left(1-\frac{e^{\varepsilon}}{e^{-\varepsilon}+e^{\varepsilon}}\right)\cdot e^{-\varepsilon\cdot B}\cdot\vec{x}[t-B] + \left(1-\frac{2e^{-\varepsilon}}{e^{-\varepsilon}+e^{\varepsilon}}\right)\cdot\vec{x}[t] \\ +\left(1-\frac{e^{\varepsilon}}{e^{-\varepsilon}+e^{\varepsilon}}\right)\cdot e^{-\varepsilon\cdot B}\cdot\vec{x}[t+B] - \frac{e^{-\varepsilon\cdot B}}{e^{-\varepsilon}+e^{\varepsilon}}\cdot\vec{x}[t+B+1] \end{array} \right| \\
&\ge \frac{e^{\varepsilon}-e^{-\varepsilon}}{e^{-\varepsilon}+e^{\varepsilon}}\cdot|\vec{x}[t]| - \frac{e^{-\varepsilon\cdot B}}{e^{-\varepsilon}+e^{\varepsilon}}\cdot\big(|\vec{x}[t-B-1]| + |\vec{x}[t+B+1]| + e^{-\varepsilon}\cdot|\vec{x}[t-B]| + e^{-\varepsilon}\cdot|\vec{x}[t+B]|\big).
\end{aligned}$$

It concludes that

$$|\vec{x}[t]| \le \frac{e^{-\varepsilon}+e^{\varepsilon}}{e^{\varepsilon}-e^{-\varepsilon}}\cdot\left(P_{norm}\cdot\left(b_t + \frac{b_{t-1}+b_{t+1}}{e^{-\varepsilon}+e^{\varepsilon}}\right) + \frac{e^{-\varepsilon\cdot B}}{e^{-\varepsilon}+e^{\varepsilon}}\cdot\big(|\vec{x}[t-B-1]| + |\vec{x}[t+B+1]| + e^{-\varepsilon}\cdot|\vec{x}[t-B]| + e^{-\varepsilon}\cdot|\vec{x}[t+B]|\big)\right).$$

□

G. Proof for Section 5

We show that, the lower bound in Theorem 1.2 applies to our algorithms, if $\frac{1}{d}\left(\frac{e^{\varepsilon}+1}{e^{\varepsilon}-1}\right)^2 \in o\left(\frac{1}{\sqrt{d}} \cdot \frac{\varepsilon}{c \cdot e^{c \cdot \varepsilon}}\right)$, i.e., $\frac{1}{\sqrt{d}} \in o\left(\left(\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}\right)^2 \cdot \frac{\varepsilon}{e^{c \cdot \varepsilon}}\right)$. We need the following lemma.

Lemma G.1. *Let $\mathcal{A}$ be the discrete Laplace mechanism, $C_A = (\mathbb{DL}\mathrm{ap}\,(e^{-\varepsilon}))^d$, $F_{\mathcal{A}}(\vec{h}, C_A) = \vec{h} + C_A$ and $\mathcal{R}$ be the profile reconstruction algorithm stated in Theorem 1.1. Then $\Delta_{\mathcal{R}\circ\mathcal{A}} \in O\left(\frac{1}{d} \cdot \left\|\boldsymbol{A}^{-1}\right\|_{\infty}\right)$.*

Combined with Lemma 4.10 on the matrix norm, and the assumption of B in Theorem 1.1, we have $1/d\,\|\boldsymbol{A}^{-1}\|_{\infty} \in O\left(\frac{1}{d}\left(\frac{e^{\varepsilon}+1}{e^{\varepsilon}-1}\right)^2\right)$, which proves our claim.

Proof of Lemma G.1. Let $\vec{h}, \vec{h}' \in [0\,..\,n]^d$ be two neighboring histograms that differ in just one coordinate by at most 1, i.e., $\left\|\vec{h} - \vec{h}'\right\|_1 \leq 1$. For every possible realization c_A of C_A, the histograms $\tilde{h} = F_A\left(\vec{h}, c_A\right), \tilde{h}' = F_A\left(\vec{h}', c_A\right)$ published by discrete Laplace mechanism also differ in just one coordinate by at most 1, i.e.,

$$\left\|F_A\left(\vec{h}, c_A\right) - F_A\left(\vec{h}', c_A\right)\right\|_1 = \left\|\vec{h} - \vec{h}'\right\|_1 \leq 1.$$

Let $\tilde{f}$ and $\tilde{f}'$ be the profiles that corresponds to $\tilde{h}$ and $\tilde{h}'$, respectively. Then $\tilde{f}$ and $\tilde{f}'$ differ in at most two coordinates, each by at most $1/d$. Applying $\mathcal{A}_{\textit{fst-inv}}$ (Algorithm 2) to $\tilde{f}$ and $\tilde{f}'$ the amplify their distance by:

$$\begin{aligned}
\left\|\mathcal{A}_{\textit{fst-inv}}(\tilde{f}) - \mathcal{A}_{\textit{fst-inv}}(\tilde{f}')\right\|_{\infty} &= \left\|\boldsymbol{A}^{-1} \times \left(\tilde{f} - \tilde{f}'\right) - \frac{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \left(\tilde{f} - \tilde{f}'\right)\right\rangle}{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a}\right\rangle} \cdot \boldsymbol{A}^{-1} \times \vec{a}\right\|_{\infty} \\
&\leq \left\|\boldsymbol{A}^{-1}\right\|_{\infty} \cdot \left(\left\|\tilde{f} - \tilde{f}'\right\|_{\infty} + \left|\frac{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \left(\tilde{f} - \tilde{f}'\right)\right\rangle}{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a}\right\rangle}\right| \cdot \|\vec{a}\|_{\infty}\right) \\
&\leq \left\|\boldsymbol{A}^{-1}\right\|_{\infty} \cdot \left(\left\|\tilde{f} - \tilde{f}'\right\|_{\infty} + \left|\frac{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \left(\tilde{f} - \tilde{f}'\right)\right\rangle}{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a}\right\rangle}\right|\right),
\end{aligned}$$

where the last inequality follows since $\|\vec{a}\|_{\infty} \leq \|\vec{a}\|_p = 1$ for $p = 1, 2, \infty$. We claim that

$$\left|\frac{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \left(\tilde{f} - \tilde{f}'\right)\right\rangle}{\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a}\right\rangle}\right| \leq \frac{2}{d}.$$

Since $\tilde{f} - \tilde{f}'$ is a vector of at most two non-zero coordinates, each bounded by $1/d$,

$$\begin{aligned}
\left|\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \left(\tilde{f} - \tilde{f}'\right)\right\rangle\right| &= \frac{1}{d} \cdot \max_{t,t'} \left|\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1}[\cdot, t]\right\rangle + \left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1}[\cdot, t']\right\rangle\right| \\
&\leq \frac{2}{d} \cdot \max_{t} \left|\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1}[\cdot, t]\right\rangle\right|.
\end{aligned}$$

On the other hand, since it holds for the standard basis vectors that $\|\vec{e}_t\|_p = 1$,

$$\left|\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{a}\right\rangle\right| \geq \max_{t} \left|\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{e}_t\right\rangle\right| = \max_{t} \left|\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1}[\cdot, t]\right\rangle\right|.$$

It remains to verify that $\max_t \left|\left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1}[\cdot, t]\right\rangle\right| \neq 0$. This is true, as

$$\left\langle \vec{\mathbb{1}}_{0:n}, \sum_t \boldsymbol{A}^{-1}[\cdot, t]\right\rangle = \left\langle \vec{\mathbb{1}}_{0:n}, \boldsymbol{A}^{-1} \times \vec{\mathbb{1}}\right\rangle = \left\langle \vec{\mathbb{1}}_{0:n}, \vec{\mathbb{1}}\right\rangle = n + 1,$$

where the equation $\boldsymbol{A}^{-1} \times \vec{\mathbb{1}} = \vec{\mathbb{1}}$ holds since $\boldsymbol{A} \times \vec{\mathbb{1}} = \vec{\mathbb{1}}$.

Combined, we have

$$\left\| \mathcal{A}_{\textit{fst-inv}}(\tilde{f}) - \mathcal{A}_{\textit{fst-inv}}(\tilde{f}') \right\|_\infty \in O\left(\frac{1}{d} \cdot \left\| \boldsymbol{A}^{-1} \right\|_\infty \right).$$

To ease the notation, from now on, denote $\vec{r} = \mathcal{A}_{\textit{fst-inv}}(\tilde{f})$ and $\vec{r}' \doteq \mathcal{A}_{\textit{fst-inv}}(\tilde{f}')$. We will finally show that, Applying $\mathcal{A}_{\textit{rnd}}$ (Algorithm 3) to $\vec{r}$ and $\vec{r}'$ the amplify their ℓ_∞ distance only by some constant:

$$\left\| \mathcal{A}_{\textit{rnd}}(\vec{r}) - \mathcal{A}_{\textit{rnd}}(\vec{r}') \right\|_\infty \in O(\left\| \vec{r} - \vec{r}' \right\|_\infty).$$

We investigate the steps of Algorithm 3. First, rounding the coordinates of $\vec{r}$ and $\vec{r}'$ to the interval $[0, 1]$ only decreases their ℓ_∞ distance.

Next, the algorithm searches for some thresholds τ and τ', such that

$$\sum_{t \in [0\,..\,n]} (\vec{r}[t] - \min\{\tau, \vec{r}[t]\}) = 1,$$
$$\sum_{t \in [0\,..\,n]} (\vec{r}'[t] - \min\{\tau', \vec{r}'[t]\}) = 1.$$

We claim that $\tau' \leq \tau + \left\| \vec{r} - \vec{r}' \right\|_\infty$, as

$$\sum_{t \in [0\,..\,n]} (\vec{r}'[t] - \min\{\tau + \left\| \vec{r} - \vec{r}' \right\|_\infty, \vec{r}'[t]\}) \leq \sum_{t \in [0\,..\,n]} (\vec{r}[t] - \min\{\tau, \vec{r}[t]\}) = 1.$$

Via symmetry, we can prove that $\tau \leq \tau' + \left\| \vec{r} - \vec{r}' \right\|_\infty$. As Algorithm 3 eventually update each $\vec{r}[t]$ and $\vec{r}'[t]$ by

$$\vec{r}[t] \leftarrow \vec{r}[t] - \min\{\tau, \vec{r}[t]\}, \quad \vec{r}'[t] \leftarrow \vec{r}'[t] - \min\{\tau', \vec{r}'[t]\},$$

after the update, it still holds that $|\vec{r}[t] - \vec{r}'[t]| \in O(\left\| \vec{r} - \vec{r}' \right\|_\infty)$, which finishes the proof. □

Proof of Theorem 1.2. We will prove the theorem by contraction. Assume there exists ε-DP and updatable algorithm $\mathcal{A} : [0\,..\,n]^d \to \mathcal{Y}$, profile estimation algorithm $\mathcal{R} : \mathcal{Y} \to [0, 1]^{n+1}$, such that, for every input $\vec{h} \in [0\,..\,n]^d$, s.t.

$$\mathbb{E}[\|\vec{f} - \mathcal{R}(\mathcal{A}(\vec{h}))\|_\infty] \in o\big(1/(\sqrt{d} \cdot e^{c \cdot \varepsilon})\big).$$

We will construct an ε-DP protocol P such that such that for every $\vec{x}$ and $\vec{y}$, Bob's output (denoted by m_B) satisfies

$$\Pr\left[|m_B - \langle \vec{x}, \vec{y} \rangle| \geq \mathcal{E}rr\right] \in o(1), \tag{17}$$

where $\mathcal{E}rr = \sqrt{d}/(2 \cdot c \cdot e^{c \cdot \varepsilon})$, contradicting Fact 5.4.

The protocol is given by Algorithm 4. Alice initializes the computation by sending $\mathcal{A}(\vec{x} + \vec{\mathbb{1}})$ to Bob. On receiving $\mathcal{A}(\vec{x} + \vec{\mathbb{1}})$, Bob updates it with their own data $\vec{y} + \vec{\mathbb{1}}$ by running algorithm $\mathcal{U}$, and the result is then passed to the profile estimation algorithm $\mathcal{R}$. Observe that by the property of updatable algorithm, $\mathcal{U}(\mathcal{A}(\vec{x} + \vec{\mathbb{1}}), \vec{y} + \vec{\mathbb{1}})) = \mathcal{A}(\vec{x} + \vec{\mathbb{1}} + \vec{y} + \vec{\mathbb{1}})$. Hence, $\mathcal{R}\Big(\mathcal{U}(\mathcal{A}(\vec{x} + \vec{\mathbb{1}}), \vec{y} + \vec{\mathbb{1}}))\Big)$ is essentially an estimate of the profile $\vec{f}$ of $\vec{h} \doteq \vec{x} + \vec{\mathbb{1}} + \vec{y} + \vec{\mathbb{1}}$. Finally, Bob publishes $d \cdot (\vec{r}[4] + \vec{r}[0] - \vec{r}[2] + 3 \cdot \Delta_{\mathcal{R} \circ \mathcal{A}} \cdot \mathbb{L}\mathrm{ap}\,(1/\varepsilon))$. where $\mathbb{L}\mathrm{ap}\,(1/\varepsilon) \in \mathbb{R}$ is a random variable following the Laplace distribution whose probability density function is given by $p(z) = \frac{\varepsilon}{2} \cdot \exp\,(-\varepsilon \cdot |z|)$, for all $z \in \mathbb{R}$.

It remains to analyze the privacy and utility guarantee of P respectively.

Privacy Guarantee. The view of Bob is $\mathcal{A}(\vec{x} + \vec{\mathbb{1}})$, which is ε-DP by the assumptions of $\mathcal{A}$. The view of Alice is $d \cdot (\vec{r}[4] + \vec{r}[0] - \vec{r}[2] + 3 \cdot \Delta_{\mathcal{R} \circ \mathcal{A}} \cdot \mathbb{L}\mathrm{ap}\,(1/\varepsilon))$. To prove this is ε-DP, it suffices to show that $\vec{r}[4] + \vec{r}[0] - \vec{r}[2]$ has sensitivity $\Delta_{\mathcal{R} \circ \mathcal{A}}$.

Assume the randomness C_A of $\mathcal{A}$ has realization c_A. The message $\mathcal{A}(\vec{x} + \vec{\mathbb{1}})$ received by Bob therefore can be written as $F_A(\vec{x} + \vec{\mathbb{1}}, c_A)$. After performing the update $\mathcal{U}(\mathcal{A}(\vec{x} + \vec{\mathbb{1}}), \vec{y} + \vec{\mathbb{1}})) = \mathcal{A}(\vec{x} + \vec{\mathbb{1}} + \vec{y} + \vec{\mathbb{1}})$, the

message becomes $F_A(\vec{x} + \vec{\mathbb{1}} + \vec{y} + \vec{\mathbb{1}}, c_A)$. Suppose that $\vec{y}$ is replaced by a neighbor binary vector $\vec{y}\,'$ such that $\|\vec{y} - \vec{y}\,'\|_1 \leq 1$. Then $\left\|\vec{x} + \vec{\mathbb{1}} + \vec{y} + \vec{\mathbb{1}} - (\vec{x} + \vec{\mathbb{1}} + \vec{y} + \vec{\mathbb{1}}\,')\right\|_1 \leq 1$, and via the definition of $\Delta_{\mathcal{R}\circ\mathcal{A}}$, we have $\left\|\mathcal{R}\Big(F_A(\vec{x} + \vec{\mathbb{1}} + \vec{y} + \vec{\mathbb{1}}, c_A)\Big) - \mathcal{R}\Big(F_A(\vec{x} + \vec{\mathbb{1}} + \vec{y}' + \vec{\mathbb{1}}, c_A)\Big)\right\|_\infty \leq \Delta_{\mathcal{R}\circ\mathcal{A}}$.

Accuracy Guarantee. Since $\vec{x}$ and $\vec{y}$ are binary vectors, $\langle \vec{x}, \vec{y} \rangle = |\{\ell \in [d] : (\vec{x} + \vec{y})[\ell] = 2\}| = d \cdot \vec{f}[2]$. Therefore

$$\left| d \cdot (\vec{r}[4] + \vec{r}[0] - \vec{r}[2] + 3 \cdot \Delta_{\mathcal{R}\circ\mathcal{A}} \cdot \mathbb{L}\mathrm{ap}\,(1/\varepsilon)) - d \cdot \left(\vec{f}[2] + \vec{f}[0] - \vec{f}[2]\right) \right| \tag{18}$$

$$\leq d \cdot \left(\left|\vec{r}[4] - \vec{f}[4]\right| + \left|\vec{r}[2] - \vec{f}[2]\right| + \left|\vec{r}[0] - \vec{f}[0]\right| + 3 \cdot \Delta_{\mathcal{R}\circ\mathcal{A}} \cdot |\mathbb{L}\mathrm{ap}\,(1/\varepsilon)| \right). \tag{19}$$

Since $\mathbb{E}\left[|\mathbb{L}\mathrm{ap}\,(1/\varepsilon)|\right] \in O(1/\varepsilon)$, and by assumption, $\mathbb{E}\left[\left\|\vec{r} - \vec{f}\right\|_\infty\right] \in o\left(\frac{1}{\sqrt{d}} \cdot \frac{1}{e^{c\cdot\varepsilon}}\right)$, and $\Delta_{\mathcal{R}\circ\mathcal{A}} \in o\left(\frac{1}{\sqrt{d}} \cdot \frac{\varepsilon}{e^{c\cdot\varepsilon}}\right)$, it concludes that

$$\mathbb{E}\left[\left| d \cdot (\vec{r}[4] + \vec{r}[0] - \vec{r}[2] + 3 \cdot \Delta_{\mathcal{R}\circ\mathcal{A}} \cdot \mathbb{L}\mathrm{ap}\,(1/\varepsilon)) - d \cdot \left(\vec{f}[2] + \vec{f}[0] - \vec{f}[2]\right) \right|\right] \in o\left(\frac{\sqrt{d}}{e^{c\cdot\varepsilon}}\right).$$

Combing $d \cdot (\vec{f}[4] + \vec{f}[0] - \vec{f}[2]) = \langle \vec{x}, \vec{y} \rangle$ with Markov inequality, we see

$$\Pr\left[|d \cdot (\vec{r}[4] + \vec{r}[0] - \vec{r}[2] + 3 \cdot \Delta_{\mathcal{R}\circ\mathcal{A}} \cdot \mathbb{L}\mathrm{ap}\,(1/\varepsilon)) - \langle \vec{x}, \vec{y} \rangle| \geq \frac{2 \cdot c_1 \cdot \sqrt{d}}{e^{c\cdot\varepsilon}} \right] \tag{20}$$

$$\leq \mathbb{E}\left[|d \cdot (\vec{r}[4] + \vec{r}[0] - \vec{r}[2] + 3 \cdot \Delta_{\mathcal{R}\circ\mathcal{A}} \cdot \mathbb{L}\mathrm{ap}\,(1/\varepsilon)) - \langle \vec{x}, \vec{y} \rangle|\right] \Bigg/ \left(\frac{\sqrt{d}}{2 \cdot c \cdot e^{c\cdot\varepsilon}}\right) \tag{21}$$

$$\in o\left(\frac{\sqrt{d}}{e^{c\cdot\varepsilon}} \Bigg/ \left(\frac{\sqrt{d}}{2 \cdot c \cdot e^{c\cdot\varepsilon}}\right)\right) \tag{22}$$

$$= o(1). \tag{23}$$

□