Private sampling: a noiseless approach for generating differentially private synthetic data[*]

March Boedihardjo[†], Thomas Strohmer[‡], and Roman Vershynin[§]

Abstract. In a world where artificial intelligence and data science become omnipresent, data sharing is increasingly locking horns with data-privacy concerns. Differential privacy has emerged as a rigorous framework for protecting individual privacy in a statistical database, while releasing useful statistical information about the database. The standard way to implement differential privacy is to inject a sufficient amount of noise into the data. However, in addition to other limitations of differential privacy, this process of adding noise will affect data accuracy and utility. Another approach to enable privacy in data sharing is based on the concept of synthetic data. The goal of synthetic data is to create an as-realistic-as-possible dataset, one that not only maintains the nuances of the original data, but does so without risk of exposing sensitive information. The combination of differential privacy with synthetic data has been suggested as a best-of-both-worlds solutions. In this work, we propose the first noisefree method to construct differentially private synthetic data; we do this through a mechanism called "private sampling". Using the Boolean cube as benchmark data model, we derive explicit bounds on accuracy and privacy of the constructed synthetic data. The key mathematical tools are hypercontractivity, duality, and empirical processes. A core ingredient of our private sampling mechanism is a rigorous "marginal correction" method, which has the remarkable property that importance reweighting can be utilized to exactly match the marginals of the sample to the marginals of the population.

1. Introduction.

In a world where artificial intelligence and data science are penetrating more and more aspects of our life, data sharing is increasingly locking horns with data-privacy concerns. This conflict is playing out around the globe, as private and public organizations are trying to find ways to share data without compromising sensitive personal information.

There exist various attempts to protect sensitive information in data. Historically the way to share private information without betraying privacy was through *anonymization* [46], i.e., by stripping away enough identifying information from a dataset, so that the so-modified data could be shared freely. Anonymization, however, proved to be a fragile means to protect data privacy. In actuality, identifying individuals using seemingly non-unique identifiers is far easier than proponents of data anonymization expected. For instance, Netflix and AOL customers were all accurately identified from purportedly anonymized data. De-identification requires precise definitions of "unique identifiers". Furthermore, de-identification suffers from an aging problem: it is already quite difficult enough to determine exactly what data identifies information that needs to be protected (say, the identity of individuals), but it is even more

[*]Submitted to the editors September 30, 2021

Funding: M.B. acknowledges support from NSF DMS-2140592. T.S. acknowledges support from NSF-DMS-1737943, NSF DMS-2027248, NSF CCF-1934568 and a CeDAR Seed grant. R.V. acknowledges support from NSF DMS-1954233, NSF DMS-2027299, U.S. Army 76649-CS, and NSF+Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning.

[†]Department of Mathematics, University of California Irvine, CA (marchb@uci.edu).

[‡]Center of Data Science and Artificial Intelligence Research University of California, Davis and Department of Mathematics, University of California Davis, CA (strohmer@math.ucdavis.edu).

[§]Department of Mathematics, University of California Irvine, CA (rvershyn@uci.edu).

difficult to accurately predict what potential auxiliary information could be available in the future. This leads to an arms race between de-identification and re-identification.

The well-documented failures of anonymization have prompted aggressive research on data sanitization, ranging from k-anonymity [39, 5] to today's highly acclaimed differential privacy [21]. The concept of k-anonymity was introduced to address the risk of re-identification of anonymized data through linkage to other datasets. The idea behind *k-anonymity* is to maintain privacy by guaranteeing that for every record in a database there are k of indistinguishable copies.

Differential privacy is a framework to quantify the extent to which individual privacy in a statistical database is preserved while releasing useful statistical information about the database [21]. Differential privacy is a popular and robust method that comes with a rigorous mathematical framework and provable guarantees. Differential privacy can protect aggregate information, but not sensitive information in general. Also, if enough identical queries are asked, the protection provided by differential privacy is diluted. Additionally, if the query being asked requires high specificity, then it is more difficult to uphold differential privacy. In any case, in all the aforementioned methods the basic tradeoff between utility and privacy represents a serious limitation.

Synthetic data provide a promising concept to solve this conundrum [7]. The goal of synthetic data is to create an as-realistic-as-possible dataset, one that not only maintains the nuances of the original data, but does so without risk of exposing sensitive information. Synthetic datasets are generated from existing datasets and maintain the statistical properties of the original dataset. Since (ideally) synthetic data contain no protected information, the datasets can be shared freely among investigators in academia or industry, without security and privacy concerns.

It has been frequently recommended that synthetic data may be combined with differential privacy to achieve a best-of-both-worlds scenario [23, 7, 27, 29, 10]. As observed in [7], "The most ideal data to use in any analysis will always be original data. But when that option is not available, synthetic data plus differential privacy offers a great compromise." Synthetic data are not only a succinct way of representing the answers to large numbers of queries, but they also permit one to carry out other data analysis tasks, such as visualization or regression.

On a high level, differential privacy is achieved via randomness. The standard way to introduce randomness in differential privacy is to add noise, either to the data queries, the data themselves, or in case of synthetic data during the data generation process. For a small sample of work see e.g. [21, 23, 24, 3, 29, 16]. Unfortunately, noise will negatively affect utility and can inject systematic errors—hence bias—into the data [37, 48, 22]. To illustrate these issues, assume the dataset under consideration consists of images, each depicting the face of a person. We can attempt to generate a differentially private synthetic dataset by adding a sufficient amount of noise to each image (e.g., by adding random noise [33] or by distorting or blurring the images [38, 45]), such that the persons in the images can no longer be identified. Ignoring for the moment the possibility of re-identifying a person by applying denoising or deblurring techniques to the distorted images, it is clear that utility of this dataset can decrease significantly during this process of adding noise, perhaps to the point that many of the nuances one might be interested in are no longer present.

To illuminate the effect of introducing systematic error when adding noise to ensure differential privacy, we just need to look at the issues reported with differentially private US Census 2020 demonstration data, which have resulted in diminished quality of statistics for small populations such as tribal nations [43, 37, 22].

These considerations raise a fundamental question:

Can we generate differentially private synthetic data without adding noise?

In this paper, we give a positive and constructive answer. Using the Boolean cube as our data model, we will develop a noiseless method to generate synthetic data, which approximately preserve low-dimensional marginals of the original dataset. Our method is based on a *private sampling* framework and comes with explicit bounds on privacy and accuracy. The key mathematical tools are hypercontractivity, duality, and empirical processes. A core ingredient of our private sampling framework is a rigorous "marginal correction" method, which has the remarkable property that importance reweighting can be utilized to *exactly* match the marginals of the sample to the marginals of the population.

There exist other methods to generate differentially private synthetic data without adding noise, such as those based on generative adversarial networks [30, 1, 12, 47, 17]. However, these methods are just empirical and do not come with any rigorous bounds regarding accuracy or privacy. Those deep learning based methods that do come with privacy guarantees—but still without any accuracy guarantees—require injecting noise into the synthetic data generation process [44, 26, 6].

2. Synthetic data and differential privacy.

Differential privacy has emerged as the de facto standard for guaranteeing privacy in data sharing. Recall the definition of differential privacy:

Definition 2.1 (Differential Privacy [21]). A randomized mechanism $\mathcal{M} : \mathcal{S}^N \to \mathcal{R}$ satisfies ε-differential privacy if for any two adjacent datasets $X_1, X_2 \in \mathcal{S}^N$ differing by one element, and any output subset $\mathcal{O} \in \mathcal{R}$ it holds that

$$\mathbb{P}[\mathcal{M}(X_1) \in \mathcal{O}] \le e^{\varepsilon} \cdot \mathbb{P}[\mathcal{M}(X_2) \in \mathcal{O}].$$

Numerous techniques have been proposed for generating privacy-preserving synthetic data (e.g. [2, 13, 1, 15, 32]), but without providing formal privacy guarantees. Almost all existing mechanisms to implement differential privacy inject some sort of noise into the data or the data queries, see e.g. the Laplacian mechanism [19]. This is also the case for differentially private synthetic data, see for instance [28, 4].

Obviously, we want our synthetic data to be similar to the original data. To that end we need some metrics to measure similarity. A common and natural choice is to try to (approximately) preserve low-dimensional marginals [4, 40]. A marginal of the data X is the fraction of the elements x_i with specified values of specified parameters. On the one hand, marginals are important in their own right as a tool of statistical analysis. On the other hand, if the synthetic data preserve e.g. two-dimensional marginals (i.e., covariance matrices) with sufficient accuracy, the synthetic dataset is expected to inherit other significant properties from the original dataset, such as similar behavior with respect to clustering, classification or regression[1].

[1]So far this expectation has only been verified empirically in various papers, while a rigorous mathematical

However, we are immediately met with a remarkable *no-go* theorem due to Ullman and Vadhan [41]. They proved the surprising result that (under standard cryptographic assumptions) there is no polynomial-time differentially private algorithm that takes a dataset $X \in (\{0,1\}^p)^n$ and outputs a synthetic dataset $Y \in (\{0,1\}^p)^k$ such that all two-dimensional marginals of Y are approximately equal to those of X.

There is an extensive literature on privately releasing answers to linear queries, but without producing synthetic data, see e.g. [4, 25, 24, 23, 40, 9, 34, 20] for a small sample. The paper [9] gives an ϵ-differentially private synthetic data algorithm whose accuracy scales logarithmically with the number of queries, but the complexity scales exponentially with p. In [4], Barak et al. derive a method for producing accurate and private synthetic Boolean data based on contingency table releases and linear programming; their method scales with 2^p, and thus is exponential in p. In [24, 23] the authors propose methods for producing private synthetic data with an error bound of about $\tilde{\mathcal{O}}(\sqrt{n}p^{1/4})$ per query. However, the associated algorithms have running time that is at least exponential in p. This computational inefficiency is not surprising in light of [41].

Already a slightly relaxed formulation of the worst-case no-go result in [41] already leads to computationally feasible algorithms. For example, if we relax "all marginals" to "most marginals", it is shown in [10] that there exists a polynomial-time differentially private algorithm generating synthetic data $Y \in (\{0,1\}^p)^k$ such that the error between the marginals of Y and X is small. Remarkably, the result does not only hold for two-dimensional marginals, but for marginals of *all dimensions*. The downside is that the guaranteed accuracy is rather low (although it is essentially optimal for microaggregation-based methods). If we relax "worst data" to "typical data", generating accurate differentially private synthetic Boolean (or other domain constrained) data becomes tractable [29, 11].

The paper [40] proposes an algorithm with complexity $np^{\mathcal{O}(\sqrt{d})}$ that returns ϵ-differentially private d-dimensional marginals under the assumption $n \geq p^{\mathcal{O}(\sqrt{d})}$. However, that algorithm does not yield synthetic data, in contrast to the algorithm proposed in this paper.

Another line of important work deals with with privacy-preserving data analysis in a statistical framework [18, 14], but they are also not concerned with synthetic data.

Yet, in *all* of the aforementioned papers differential privacy is achieved by *adding noise* during the data generation process. In this paper we propose an alternative, *noise-free*, mechanism called *private sampling*.

3. Main result.

We model the true data $X = (x_1, \ldots, x_n)$ as a sequence of n points from the Boolean cube $\{0,1\}^p$, which is a standard benchmark data model [4, 41, 23, 36, 29, 8]. For example, X might represent the health records of n patients, where each health record consists of p parameters. These parameters are $0/1$ numbers that represent the answers to the standard health history questionnaire, such as "does the patient smoke?", "does the patient have diabetes?". We can also represent categorical data (gender, occupation, etc.) or numerical data (by splitting them into intervals) on the Boolean cube via binary or one-hot encoding.

We would like to manufacture a synthetic dataset $Y = (y_1, \ldots, y_k)$, another sequence of

verification is an important open problem.

k elements of the cube. Our two desiderata are *privacy* and *accuracy*. Specifically, we would like the synthetic data to be differentially private, and all low-dimensional marginals of Y to exactly or approximately match those of X.

In our derivations it is more convenient to work on the Boolean cube $\{-1,1\}^p$ instead of $\{0,1\}^p$. Note that it is straightforward to translate our results from one cube to the other. We recall that on the Boolean cube, a marginal of a function $f : \{-1,1\}^p \to \mathbb{R}$ is defined as a sum of values of f on the points of the cube that have specified values of specified parameters. For example, a two-dimensional marginal of f is $\sum_{x\in\{-1,1\}^p} f(x)\mathbf{1}_{\{x(1)=x(2)=1\}}(x)$. If f is a density, a marginal can be interpreted as the probability that a random point Z drawn from the cube according to f has specified values of specified parameters; in the example below it is $\mathbb{P}\left\{Z(1) = Z(2) = 1\right\}$. Marginals of the data $X = (x_1, \ldots, x_n)$ can be interpreted as marginals of the uniform density $f_n = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{x_i}$ on X. An example of a two-dimensional marginal is the fraction of elements x_i whose first and second parameters equal 1, i.e. $\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{x_i(1)=x_i(2)=1\}}$. This could represent for example the number of patients who smoke and have diabetes.

Here we explore a new *noiseless* approach: take a new sample $S = (s_1, \ldots, s_m)$ uniformly from the cube, reweight S to make the marginals match those of the true data X, and resample from the weighted sample S.

But is this even possible? Let us assume the dataset $X = (x_1, \ldots, x_n)$ is drawn from the cube independently and according to some unknown density. Draw a new sample $S = (s_1, \ldots, s_m)$ according to some known density, for example uniformly from the cube[2]. Can we reweight S so that the reweighted sample has approximately the same marginals as X? Note that there are precisely $\binom{p}{\le d}$ marginals of degree at most d, where $\binom{p}{\le d} := \binom{p}{0} + \binom{p}{1} + \cdots + \binom{p}{d}$. Surprisingly, we can even match all marginals *exactly*.

Let us state it this result informally; a rigorous, non-asymptotic and more general statement is given in Theorem 8.1.

Theorem 3.1 (Matching marginals). *Consider two regularly varying densities*[3] *on the cube* $\{0,1\}^p$, *and draw two independent samples X and S from the cube according to these two distributions. If* $\min(|X|, |S|) \gg e^{2d}\binom{p}{\le d}$, *then with probability* $1 - o(1)$ *there exists a density on S that has exactly the same marginals up to dimension d as the uniform distribution on X.*

Remark 3.2. *To match all $\binom{p}{\le d}$ marginals of dimension at most d, it makes sense to have at least as many data points. This explains the requirement on n in the theorem heuristically (but not rigorously). The prefactor e^{2d} is negligible compared to $\binom{p}{\le d}$ if $d \ll p$.*

The path towards proving (a rigorous version of) Theorem 3.1 leads through Lemma 3.3, which introduces the concept of *private sampling*. The main technical challenges—which occupy most of this paper—is then to show that the assumptions of Lemma 3.3 can be satisfied under competitive conditions on the sample complexity, cardinality of X and S, and the number of queries, while still maintaining high accuracy.

[2]Since the cardinality of S will be chosen to be smaller than that of the dataset X, we call S also the reduced space.

[3]A density f is *regularly varying* if $\sup f(x)/f(y) = O(1)$ where the supremum is over all points x and y in the cube. Our results are more general; as we will see shortly, the regularity assumption can be relaxed.

As a "non-example" for Theorem 3.1, consider a probability measure supported on the set of patients whose first parameter equals 0, and a different probability measure supported on the set of patients whose first parameter equals 1. Then even a one-dimensional marginal – the distribution of the first parameter – will be different for X and Y, no matter how Y is reweighted. This example shows that some form of regularity assumption will be required in the theorem.

The density h^* on S that is guaranteed by Theorem 3.1 can be *computed efficiently*. Indeed, this task can be set up as a linear program with $|S|$ variables (the values of the density on S), $\binom{p}{\le d}$ linear equations (to match the marginals to those of X), and $|S|$ linear inequalities (to ensure the density is nonnegative on S).

Once this density h^* is computed, we can generate synthetic data $Y = (y_1, \dots, y_k)$ by drawing independent points from S according to the density h^*.

3.1. Private sampling. Is such synthetic data Y private? Here is a general tool that basically says: yes, Y is private as long as the density h^* has bounded sensitivity.

Lemma 3.3 (Private sampling). *Let Ω be a finite set. Let f be a mapping that takes a dataset X as input and returns a probability mass function $f(X)$ on Ω. Suppose $\varepsilon > 0$ and $k \in \mathbb{N}$ are chosen so that*

$$\left\| f(X_1)/f(X_2) \right\|_\infty \le \exp(\varepsilon/k)$$

for all datasets X_1 and X_2 that differ on a single element. Then the algorithm that takes X as input and returns a sample of k points drawn from Ω independently and according to the distribution $f(X)$ is ε-differentially private.

Proof. The probability that a given k-tuple of points $\omega_1, \dots, \omega_k \in \Omega$ is drawn when sampled from distribution $f(X_1)$ equals $\prod_{i=1}^k f(X_1)(\omega_i)$. Similarly, the probability that this same tuple is drawn when sampled from distribution $f(X_2)$ equals $\prod_{i=1}^k f(X_2)(\omega_i)$. If the databases X_1 and X_2 differ on a single element, the assumption implies that the ratio of these probabilities is bounded by $\prod_{i=1}^k \exp(\varepsilon/k) = \exp(\varepsilon)$. This means that the sampling mechanism is ε-differentially private. ■

3.2. Difficulties and their resolution. Unfortunately, the density h^* guaranteed by Theorem 3.1 is too sensitive. Indeed, the sensitivity bound in Lemma 3.3 needs to be proved for *arbitrary* input data, while Theorem 3.1 only works with high probability. For some input data X, a suitable density exists, and for another input data Z, no suitable density exists. Moving from X toward Z by changing one data point at a time, we can find a pair of datasets X_1 and X_2 that differ in a single data point so that the algorithm succeeds to find a density for X_1 and fails for X_2. This means that the algorithm is non-private.

The other issue is that there can be (and usually are) many suitable densities h^*. Which one to chose? How to devise a selection rule that upholds privacy?

In other words, we need to work around the possible non-existence and non-uniqueness of the solution. We resolve both issues here. To ensure existence, we employ *shrinking*: we move the solution space (the set of all functions on S, possibly negative-valued, that have the same marginals as X) toward the uniform density on S until the resulting set contains a nonnegative function (thus a density). For the selection rule, we choose the closest solution to the uniform density on S in the L^2 metric.

Furthermore, while S is chosen randomly, we do need S to be *well-conditioned* in a sense that will be discussed in detail in Section 9. At this point suffice it to say that (i) the well-conditionedness of S can be expressed in terms of a bound on the smallest singular value $\sigma_{\min}(M)$ of the $m \times \binom{p}{\leq d}$ matrix M with entries $w(s)$, where $s \in S$ and w is a Walsh function[4] of degree at most d; (ii) the well-conditionedness of M can be easily achieved and easily verified.

This leads us to the algorithm outlined in the next subsection.

3.3. Algorithm. We provide a high-level description of our proposed method in Algorithm 3.1. See Remark 12.1 regarding the computational complexity of this algorithm.

Algorithm 3.1 Private sampling synthetic data algorithm

Input: a sequence X of n points in $\{-1,1\}^p$ (true data); m: cardinality of S; d: the degree of the marginals to be matched; parameters δ, Δ with $\Delta > \delta > 0$.

1. Draw m points from $\{-1,1\}^p$ independently and uniformly, and call this set S (reduced space).
2. Form the $m \times \binom{p}{\leq d}$ matrix M with entries $w(s)$, where $s \in S$ and w is a Walsh function of degree at most d. If the smallest singular value of M is bounded below by $\sqrt{m}/2e^d$, call S well conditioned and proceed. Otherwise return "Failure" and stop.
3. Consider the affine space H consisting of all densities on S that have exactly the same marginals up to dimension d as the true data X.
4. If necessary, shrink H toward the uniform density on S just so the resulting affine space $\tilde{H}$ contains a density that is lower bounded by $2\delta/m$ and upper bounded by $(\Delta-\delta)/m$.
5. Among all densities in $\tilde{H}$ that are lower bounded by δ/m and upper bounded by Δ/m, pick one closest to the uniform density in the L^2 norm.

Output: a sequence Y of k points from S according to this density.

The well-conditionedness of S in Algorithm 3.1 defined via the condition $\sigma_{\min}(M) > \sqrt{m}/2e^d$ essentially says that the subsampled Walsh basis is almost orthogonal. The scaling $\sqrt{m}$ is natural: the entries of M all have absolute value 1, hence the columns of M have Euclidean norm $\sqrt{m}$. If we had $\sigma_{\min}(M) = \sqrt{m}$, this would imply that the columns of M (the subsampled Walsh functions) are mutually orthogonal. We require a relaxed (by a factor $2e^d$) version of this orthogonality.

What if S fails the desired condition? We can simply resample S until it is well conditioned. But this is only a useful strategy if the chances of success are sufficiently high. Under some mild conditions (see Section 9) success happens with probability $> 1/2$, hence the expected number or trials until success is ≤ 2. This way Algorithm 3.1 succeeds deterministically, but its running time becomes random (albeit with the rather modest expected overhead time ≤ 2).

Definition 3.4. *We say that the synthetic dataset Y is δ-*accurate *if each of its marginals up to degree (or dimension) d is within δ from the corresponding marginal of the true dataset X.*

[4]See Section 4 for basic definitions related to Fourier analysis of the Boolean cube.

The following theorem guarantees the accuracy and privacy of the algorithm. We state it informally here, and more accurately in Theorems 12.3 and 12.5.

Theorem 3.5 (Privacy and accuracy). *Let the size of the reduced space S satisfy $m \asymp e^{2d}\binom{p}{\le d}$.*

(a) Algorithm 3.1 succeeds (i.e. does not return "Failure") with high probability.

(b) If the size of the synthetic data satisfies $k \ll \sqrt{n}/m$, then Algorithm 3.1 is $o(1)$-differentially private.

(c) Suppose $n \gg e^{2d}\binom{p}{\le d}$, $k \gg \log\binom{p}{\le d}$, and the true data points X are sampled independently from some density that is upper bounded by $\Delta/2^p$. Then, with high probability, the synthetic data generated via Algorithm 3.1 is $o(1)$-accurate up to dimension d.

For a more formal presentation of Algorithm 3.1, see Algorithm 12.1 below. A formal version of part (a) of Theorem 3.5 is shown in Proposition 9.3; part (b) is shown in Theorem 12.3 and Remark 12.4; part (c) is shown in Theorem 12.5. The mathematical techniques to prove these results revolve around Fourier analysis of Boolean functions and empirical processes, see Sections 4–7.

In case the true data X is sampled form a regular density, the algorithm will not apply any shrinkage, since in this case Theorem 3.1 guarantees the existence of a solution. (We make this rigorous in Remark 12.6.) In this case, the private synthetic data Y will be sampled in an *unbiased way* from the density h^* that has *exactly* the same marginals as the true data X.

3.4. Further remarks.

There is a one-sample version of Theorem 3.1. Let us state it here informally; a more accurate statement is given in Theorem 8.2.

Theorem 3.6 (Marginal correction). *Consider a regularly varying density f on the cube $\{0,1\}^p$ and draw an independent sample S from the cube according to this distribution. If $|S| \gg e^{2d}\binom{p}{\le d}$, then with probability $1-o(1)$ there exists a density h on S that has exactly the same marginals as f up to dimension d. Moreover, h is within a $1+o(1)$ factor of the uniform density on S.*

The law of large numbers tells us that the sample S must have *approximately* the same marginals as the density f from which S was drawn. Theorem 3.6 tells us that we can make the marginals *exactly* the same by a slight reweighting of S, i.e. by weights that are all $1+o(1)$.

4. Fourier analysis.

The proof of Theorem 3.1 is based on hypercontractivity, duality, and empirical processes.

Let us start by recalling the basic Fourier analysis on the Boolean cube $\{-1,1\}^p$ [35].

The *Walsh functions* $w_J : \{-1,1\}^p \to \{-1,1\}$ are indexed by subsets $J \subset [p]$ and are defined as

$$w_J(x) = \prod_{j \in J} x(j), \tag{4.1}$$

with the convention $w_\emptyset = 1$.

The canonical inner product on the space of real-valued functions on $\{-1,1\}^p$ is defined as

$$\langle f, g \rangle_{L^2} = \frac{1}{2^p} \sum_{x \in \{-1,1\}^p} f(x)\, g(x).$$

This inner product defines the space $L^2 = L^2(\{-1,1\}^p)$. More generally, for $1 \le q < \infty$, the $L^q = L^q(\{-1,1\}^p)$ is the space of real-valued functions on the cube with the norm

$$\|f\|_{L^q} = \Big(\frac{1}{2^p} \sum_{x \in \{-1,1\}^p} \big|f(x)\big|^q \Big)^{1/q}.$$

Walsh functions form an orthonormal basis of L^2, so any function $f : \{-1,1\}^p \to \mathbb{R}$ admits a Fourier expansion

$$f = \sum_{J \subset [p]} \hat{f}_J w_J, \quad \text{where } \hat{f}_J = \langle f, w_J \rangle \text{ are Fourier coefficients.}$$

Thus, any function f on the cube can be orthogonally decomposed into low and high frequencies:

$$f = f^{\le d} + f^{>d},$$

where

$$f^{\le d} = \sum_{J \subset [p], |J| \le d} \langle f, w_J \rangle w_J \quad \text{and} \quad f^{>d} = \sum_{J \subset [p], |J| > d} \langle f, w_J \rangle w_J.$$

Clearly, the function $f^{\le d}$ is determined by the Fourier coefficients of f up to dimension d, and vice versa.

We say that a function f on the cube has *degree at most* d if $f = f^{\le d}$. Such functions form the "low-frequency" space

$$W^{\le d} = \Big\{ f : \ f = f^{\le d} \Big\} = \operatorname{span}\{ w_J : |J| \le d \},$$

and it has dimension $\binom{p}{\le d}$. The orthogonal complement to this subspace in L^2 is the "high-frequency" subspace

$$W^{>d} = \Big\{ f : \ f = f^{>d} \Big\} = \operatorname{span}\{ w_J : |J| > d \}.$$

The following result is well known, see [35, Theorem 9.22]:

Theorem 4.1 (Hypercontractivity). *For any $d \le p$ and any function $f : \{-1,1\}^p \to \mathbb{R}$ of degree at most d, we have*

$$\|f\|_{L^2} \le e^d \|f\|_{L^1}.$$

4.1. Connection to marginals. The low-degree Fourier coefficients of $f : \{-1,1\}^p \to \mathbb{R}$ determine the low-dimensional marginals of f. More precisely, $f^{\le d}$ determines the values of all marginals of f up to dimension (or degree) d.

To see this, consider the example of the two-dimensional marginal in which the first parameter is set to 1 and the second is set fo -1. The value of such marginal of f is $\sum_{x \in \{-1,1\}^p} f(x) \mathbf{1}_{\{x(1)=1,\, x(2)=-1\}}$. Now,

$$\mathbf{1}_{\{x(1)=1,\, x(2)=-1\}}(x) = \mathbf{1}_{\{x(1)=1\}}(x) \mathbf{1}_{\{x(2)=-1\}} = \Big(\frac{1 + x(1)}{2} \Big) \Big(\frac{1 - x(2)}{2} \Big),$$

so expanding the right hand side and using the definition of Walsh functions, we see that

$$\mathbf{1}_{\{x(1)=1,\,x(2)=-1\}} = \frac{1}{4}\left(w_\emptyset + w_{\{1\}} - w_{\{2\}} - w_{\{1,2\}}\right).$$

Thus, the marginal can be written as

$$\sum_{x\in\{-1,1\}^p} f(x)\mathbf{1}_{\{x(1)=1,\,x(2)=-1\}} = \frac{1}{4}\left(\hat{f}_\emptyset + \hat{f}_{\{1\}} - \hat{f}_{\{2\}} - \hat{f}_{\{1,2\}}\right),$$

and so it depends only on the Fourier coefficients on f up to degree 2, or equivalently only on $f^{\le 2}$.

5. Empirical processes.

Let μ be a probability measure on $\{-1,1\}^p$, and let

$$\mu_m = \frac{1}{m}\sum_{i=1}^m \delta_{\theta_i}$$

be the corresponding (random) *empirical measure*, i.e., the uniform probability measure on the sample $\{\theta_1,\dots,\theta_m\}$ of points drawn from the cube independently according to the distribution μ. These two measures define the population and empirical L^q norms of functions on the cube:

$$\|F\|^q_{L^q(\mu)} := \mathbb{E}\big|F(\theta_1)\big|^q; \qquad \|F\|^q_{L^q(\mu_m)} := \frac{1}{m}\sum_{i=1}^m \big|F(\theta_i)\big|^q. \tag{5.1}$$

We clearly have $\mathbb{E}\|F\|_{L^1(\mu_m)} = \|F\|_{L^1(\mu)}$. The following result provides a uniform deviation inequality.

Proposition 5.1 (Deviation of the empirical L^1 norm). *Let μ be a probability measure on $\{-1,1\}^p$ and μ_m be the empirical counterpart. Then*

$$\mathbb{E}\sup_{F\in W^{\le d},\,\|F\|_{L^2}=1}\Big|\|F\|_{L^1(\mu_m)} - \|F\|_{L^1(\mu)}\Big| \le 2\sqrt{\frac{1}{m}\binom{p}{\le d}}.$$

The L^2 norm on the left side is with respect to the uniform probability measure on the cube.

Proof. Any function $F \in W^{\le d}$ is a linear combination of low-degree Walsh functions,

$$F = \sum_{|J|\le d} a_J w_J.$$

Without loss of generality (by rescaling) we can assume that

$$\|F\|^2_{L^2} = \sum_{|J|\le d} a_J^2 = 1. \tag{5.2}$$

By definition of the $L^1(\mu)$ norm in (5.1), we have

$$\|F\|_{L^1(\mu)} = \mathbb{E}\left|\sum_{|J|\le d} a_J w_J(\theta_1)\right| = \mathbb{E}\big|\langle w(\theta_1), a\rangle\big|\,,$$

where, for every θ in the cube, $w(\theta) := \big(w_J(\theta)\big)_{|J|\le d}$ is a vector in $\mathbb{R}^{\binom{p}{\le d}}$, and similarly $a = (a_J)_{|J|\le d}$ denotes the coefficient vector in $\mathbb{R}^{\binom{p}{\le d}}$. By (5.2), a is a unit vector, i.e. $a \in S^{\binom{p}{\le d}-1}$. In a similar way, the definition of the empirical L^1 norm in (5.1) yields

$$\|F\|_{L^1(\mu_m)} = \frac{1}{m}\sum_{i=1}^m \left|\sum_{|J|\le d} a_J w_J(\theta_i)\right| = \frac{1}{m}\sum_{i=1}^m \big|\langle w(\theta_i), a\rangle\big|\,.$$

Then

$$\begin{aligned} E &:= \mathbb{E} \sup_{F\in W^{\le d},\,\|F\|_{L^2}=1} \Big|\|F\|_{L^1(\mu_m)} - \|F\|_{L^1(\mu)}\Big| \\ &= \mathbb{E} \sup_{a\in S^{\binom{p}{\le d}-1}} \left|\frac{1}{m}\sum_{i=1}^m \big|\langle w(\theta_i), a\rangle\big| - \mathbb{E}\big|\langle w(\theta_1), a\rangle\big|\right|. \end{aligned}$$

Applying a symmetrization inequality for empirical processes (see e.g. [42, Exercise 8.3.24]), we get

$$E \le 2\,\mathbb{E} \sup_{a\in S^{\binom{p}{\le d}-1}} \left|\frac{1}{m}\sum_{i=1}^m \varepsilon_i \big|\langle w(\theta_i), a\rangle\big|\right|,$$

where $(\varepsilon_i)_{i=1}^m$ denote i.i.d. Rademacher random variables, which are independent of the sample points $(\theta_i)_{i=1}^m$.

The exterior absolute value can be removed using the symmetry of the Rademacher random variables, and the interior absolute values can be removed using Talagrand's contraction principle, see [42, Exercise 6.7.7], thus continuing our bound as

$$\begin{aligned} E &\le 2\,\mathbb{E} \sup_{a\in S^{\binom{p}{\le d}-1}} \frac{1}{m}\sum_{i=1}^m \varepsilon_i \langle w(\theta_i), a\rangle \\ &= 2\,\mathbb{E}\left\|\frac{1}{m}\sum_{i=1}^m \varepsilon_i w(\theta_i)\right\|_2 \le \frac{2}{m}\left(\mathbb{E}\left\|\sum_{i=1}^m \varepsilon_i w(\theta_i)\right\|_2^2\right)^{1/2} = \frac{2}{m}\left(\sum_{i=1}^m \mathbb{E}\big\|w(\theta_i)\big\|_2^2\right)^{1/2} \end{aligned}$$

where the last step follows by conditioning on (θ_i). Since all $\binom{p}{\le d}$ coordinates of all vectors $w(\theta_i)$ equal ± 1, we have $\big\|w(\theta_i)\big\|_2^2 = \binom{p}{\le d}$ deterministically. Substituting this bound, we complete the proof. ∎

6. Enforcing a uniform bound and sparsity. We will now prove that for any function F on the Boolean cube, there is another function that simultaneously satisfies the three desiderata: (a) it has the same marginals (or Fourier coefficients) as F up to dimension d; (b) it is very sparse – in fact, it is supported on a random set of a given cardinality; and (c) it is uniformly bounded. The following result guarantees the existence of such function $F - w$.

Theorem 6.1. *Let μ be a probability measure on the cube $\{-1,1\}^p$ whose density is bounded below by $\alpha/2^p$, and let μ_m be the empirical counterpart. If $m \geq 16(\alpha\gamma)^{-2}e^{2d}\binom{p}{\leq d}$, then the following holds with probability at least $1-\gamma$. For any function $F : \{-1,1\}^p \to \mathbb{R}$, we have*

$$\inf\left\{\|F-w\|_\infty : \; w \in W^{>d}, \; F - w \subset S_{\mu_m}\right\} \leq \frac{2e^d 2^p}{\alpha m}\left\|F^{\leq d}\right\|_{L^2}$$

where S_{μ_m} denotes the set of the functions supported on $\mathrm{supp}(\mu_m)$.

Throughout the proof, let us denote

$$S := \mathrm{supp}(\mu_m).$$

The L^1 norm of any function $F : \{-1,1\}^p \to \mathbb{R}$ naturally decomposes as

$$\|F\|_{L^1} = \|F\mathbf{1}_S\|_{L^1} + \|F\mathbf{1}_{S^c}\|_{L^1},$$

where $\mathbf{1}_S$ denotes the indicator function of S. Given $\delta > 0$, consider the weighted space L^1_δ where the norm is defined by

$$\|F\|_{L^1_\delta} := \|F\mathbf{1}_S\|_{L^1} + \delta\|F\mathbf{1}_{S^c}\|_{L^1}.$$

Lemma 6.2. *Consider the subspace $(W^{\leq d}, \|\,\|_{L^1_\delta})$ of L^1_δ. With probability at least $1-\gamma$, for every $\delta > 0$ we have*

$$\left\|\mathrm{Id} : \; (W^{\leq d}, \|\,\|_{L^1_\delta}) \to L^2\right\| \leq \frac{2e^d 2^p}{\alpha m}.$$

Proof. Proposition 5.1 combined with Markov's inequality and rescaling implies that, with probability $1-\gamma$, the following holds for all $F \in W^{\leq d}$:

$$\left|\|F\|_{L^1(\mu)} - \|F\|_{L^1(\mu_m)}\right| \leq \frac{2}{\gamma}\sqrt{\frac{1}{m}\binom{p}{\leq d}}\;\|F\|_{L^2} \leq \frac{\alpha}{2e^d}\|F\|_{L^2},$$

where in the last step we used the assumption on m.

Applying hypercontractivity (Theorem 4.1), the regularity assumption of μ, and the bound above, we obtain

$$\frac{1}{e^d}\|F\|_{L^2} \leq \|F\|_{L^1} \leq \frac{1}{\alpha}\|F\|_{L^1(\mu)} \leq \frac{1}{\alpha}\|F\|_{L^1(\mu_m)} + \frac{1}{2e^d}\|F\|_{L^2}.$$

Rearranging the terms, we obtain

$$\frac{1}{2e^d}\|F\|_{L^2} \leq \frac{1}{\alpha}\|F\|_{L^1(\mu_m)} = \frac{2^p}{\alpha m}\|F\mathbf{1}_S\|_{L^1} \leq \frac{2^p}{\alpha m}\|F\|_{L^1_\delta}$$

where in the middle step we used the definitions of S and of the norms in $L^1(\mu)$ and $L^1(\mu_m)$. Multiplying both sides by $2e^d$ completes the proof. ∎

Proof of Theorem 6.1. Let us dualize Lemma 6.2 with respect to the inner product on L^2. The identity operator is self-adjoint, and the adjoint operator has the same norm. So, with probability at least $1-\gamma$, for every $\delta > 0$ we have

$$\left\| \mathrm{Id} : \left(L^2\right)^* \to \left(W^{\le d}, \|\,\|_{L^1_\delta}\right)^* \right\| \le \frac{2e^d 2^p}{\alpha m} =: B.$$

The Hilbert space L^2 is self-dual. The dual to the weighted space L^1_δ is the weighted space $L^\infty_{1/\delta}$ defined as

$$\|F\|_{L^\infty_{1/\delta}} := \|F\mathbf{1}_S\|_{L^\infty} \vee \frac{1}{\delta}\|F\mathbf{1}_{S^c}\|_{L^\infty}. \tag{6.1}$$

The dual of a subspace is a quotient space of the dual:

$$\left(W^{\le d}, \|\,\|_{L^1_\delta}\right)^* = \left(L^1_\delta\right)^* / (W^{\le d})^\perp = L^\infty_\delta / W^{>d}.$$

Putting these considerations together, we get

$$\left\| \mathrm{Id} : L^2 \to L^\infty_\delta / W^{>d} \right\| \le B.$$

By definition of the quotient norm, this bound means that for every function $F : \{-1,1\}^p \to \mathbb{R}$ there exists $w \in W^{>d}$ such that

$$\|F - w\|_{L^\infty_\delta} \le B\|F\|_{L^2}.$$

By definition (6.1) of the weighted norm, this means that

$$\left\|(F-w)\mathbf{1}_S\right\|_\infty \le B\|F\|_{L^2} \quad \text{and} \quad \left\|(F-w)\mathbf{1}_{S^c}\right\|_\infty \le \delta B\|F\|_{L^2}. \tag{6.2}$$

Since the second bound holds for arbitrary $\delta > 0$, it follows that $\left\|(F-w)\mathbf{1}_{S^c}\right\|_\infty = 0$, i.e.

$$\mathrm{supp}(F-w) \subset S$$

as claimed in the theorem. Together with the first bound in (6.2), this proves that

$$\|F - w\|_\infty \le B\|F\|_{L^2}.$$

Thus, we showed every function $F : \{-1,1\}^p \to \mathbb{R}$ satisfies

$$\inf\left\{\|F-w\|_\infty : \ w \in W^{>d},\, F - w \subset S_{\mu_m}\right\} \le B\|F\|_{L^2}$$

Finally, note that the term $\|F\|_{L^2}$ on the right hand side can automatically be improved to $\left\|F^{\le d}\right\|_{L^2}$. To see this, apply the above bound for $F^{\le d}$ and absorb the term $F^{>d}$ into w. Theorem 6.1 is proved. ∎

7. Low-degree projections of empirical measures. Consider two probability measures ν and μ on $\{-1,1\}^p$, and let f and g denote their densities (or probability mass functions):

$$f(z) = \nu(\{z\}) \quad \text{and} \quad g(z) = \mu(\{z\}), \quad z \in \{-1,1\}^p.$$

The densities of the empirical probability measures ν_n and μ_m are

$$f_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_i} \quad \text{and} \quad g_m = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}_{y_i} \tag{7.1}$$

where $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ are i.i.d. points drawn from the cube according to the densities f and g, respectively. The functions f_n and g_m provide unbiased estimators of f and g:

$$\mathbb{E}\, f_n = f, \quad \mathbb{E}\, g_m = g.$$

Assume that $f(z) = 0$ whenever $g(z) = 0$. Consider the function

$$\tilde{g}_m := (f/g) g_m. \tag{7.2}$$

Although $\tilde{g}_m$ is supported on the sample drawn from density g, it provides an unbiased estimator of f:

$$\mathbb{E}\, \tilde{g}_m = (f/g)\, \mathbb{E}\, g_m = (f/g) g = f.$$

This property will be crucial in the proof of Theorem 3.1.

Let us look at the low-degree projections of f_n and $\tilde{g}_m$ and try to bound their mean magnitude and deviation from the mean. Toward this end, note that

$$\forall x \in \{-1,1\}^p, \quad \left\| (\mathbf{1}_x)^{\le d} \right\|_{L^2} = \binom{p}{\le d}^{1/2} \frac{1}{2^p}. \tag{7.3}$$

Indeed, to see this, use Parseval's identity

$$\left\| (\mathbf{1}_x)^{\le d} \right\|_{L^2}^2 = \sum_{|J| \le d} \langle \mathbf{1}_x, w_J \rangle_{L^2}^2 = \sum_{|J| \le d} \Big(\frac{1}{2^p} w_J(x) \Big)^2$$

and recall that the Walsh function w_J takes ± 1 values. Furthermore, by definition of f_n and the triangle inequality, (7.3) yields

$$\left\| (f_n)^{\le d} \right\|_{L^2} \le \binom{p}{\le d}^{1/2} \frac{1}{2^p} \quad \text{deterministically.} \tag{7.4}$$

Lemma 7.1 (Deviation). *We have*

$$\Big(\mathbb{E} \left\| (f_n - f)^{\le d} \right\|_{L^2}^2 \Big)^{1/2} \le \binom{p}{\le d}^{1/2} \frac{1}{\sqrt{n} 2^p}.$$

Moreover, if $\left\| f/g \right\|_{L^2} \le \kappa$ *then we have*

$$\Big(\mathbb{E} \left\| (\tilde{g}_m - f)^{\le d} \right\|_{L^2} \Big)^{1/2} \le \binom{p}{\le d}^{1/2} \frac{\kappa}{\sqrt{m} 2^p}.$$

Proof. By Parseval's identity,

$$\left\|(f_n - f)^{\le d}\right\|_{L^2}^2 = \sum_{|J|\le d} \langle f_n - f, w_J\rangle_{L^2}^2. \tag{7.5}$$

By definition (7.1) of f_n, each term of this sum can be expressed as

$$\langle f_n - f, w_J\rangle_{L^2} = \frac{1}{n}\sum_{i=1}^{n} \langle \mathbf{1}_{x_i} - f, w_J\rangle_{L^2}.$$

The terms on the right hand side are i.i.d. mean zero random variables, so

$$\begin{aligned}
\mathbb{E}\langle f_n - f, w_J\rangle_{L^2}^2 &= \frac{1}{n}\,\mathbb{E}\langle \mathbf{1}_{x_1} - f, w_J\rangle_{L^2}^2 \\
&\le \frac{1}{n}\,\mathbb{E}\langle \mathbf{1}_{x_1}, w_J\rangle_{L^2}^2 \quad \text{(the variance is bounded by the second moment)} \\
&= \frac{1}{n}\,\mathbb{E}\left(\frac{1}{2^p} w_J(x_1)\right)^2 = \frac{1}{n2^{2p}},
\end{aligned}$$

since the Walsh function w_J takes ± 1 values. Substitute this bound into Parseval's identity (7.5) to get

$$\mathbb{E}\left\|(f_n - f)^{\le d}\right\|_{L^2}^2 \le \binom{p}{\le d} \cdot \frac{1}{n2^{2p}}.$$

This proves the first part of the lemma.

The second part of the lemma can be derived similarly. Indeed,

$$\left\|(\tilde{g}_m - f)^{\le d}\right\|_{L^2}^2 = \sum_{|J|\le d} \langle \tilde{g}_m - f, w_J\rangle_{L^2}^2. \tag{7.6}$$

By definition (7.1) of g_m and (7.2) of $\tilde{g}_m$, each term of this sum can be expressed as

$$\langle \tilde{g}_m - f, w_J\rangle_{L^2} = \frac{1}{m}\sum_{i=1}^{m} \left\langle \frac{f(y_i)}{g(y_i)} \cdot \mathbf{1}_{y_i} - f, w_J \right\rangle_{L^2}.$$

The terms on the right hand side are i.i.d. mean zero random variables, so

$$\begin{aligned}
\mathbb{E}\langle \tilde{g}_m - f, w_J\rangle_{L^2}^2 &= \frac{1}{m}\,\mathbb{E}\left\langle \frac{f(y_1)}{g(y_1)} \cdot \mathbf{1}_{y_1} - f, w_J \right\rangle_{L^2}^2 \\
&\le \frac{1}{m}\,\mathbb{E}\left\langle \frac{f(y_1)}{g(y_1)} \cdot \mathbf{1}_{y_1}, w_J \right\rangle_{L^2}^2 \text{(the variance is bounded by the second moment)} \\
&= \frac{1}{m}\,\mathbb{E}\left(\frac{1}{2^p}\frac{f(y_1)}{g(y_1)} w_J(y_1)\right)^2 \\
&= \frac{1}{m2^{2p}}\left\|f/g\right\|_{L^2}^2 \le \frac{\kappa^2}{m2^{2p}},
\end{aligned}$$

where in the last line we used the fact that the Walsh function w_J takes ± 1 values and the assumption on f/g. Substitute this bound into Parseval's identity (7.6) to get

$$\mathbb{E}\left\|(\tilde{g}_m - f)^{\le d}\right\|_{L^2}^2 \le \binom{p}{\le d} \cdot \frac{\kappa^2}{m2^{2p}}.$$

This proves the second part of the lemma. ■

8. Proof of Theorem 3.1.

The following master theorem is a more general version of Theorem 3.1, as we will see shortly. Recall that $g_m, \mu_m, \tilde{g}_m$ are defined in (7.1).

Theorem 8.1. *Let f and g be densities on the cube $\{-1,1\}^p$, and let f_n and g_m be their empirical counterparts. Assume that $\|f/g\|_{L^2} \le \kappa$ for some $\kappa \ge 1$ and that g is bounded below by $\alpha/2^p$. If*

$$n \ge 16(\alpha\delta)^{-2}\gamma^{-1}e^{2d}\binom{p}{\le d} \qquad \text{and} \qquad m \ge 16(\alpha\delta)^{-2}\gamma^{-1}\kappa^2 e^{2d}\binom{p}{\le d}, \tag{8.1}$$

then the following holds with probability $1-2\gamma$. There exists $h : \{-1,1\}^p \to \mathbb{R}$ that satisfies

$$h^{\le d} = f_n^{\le d}, \quad \operatorname{supp}(h) \subset \operatorname{supp}(g_m), \quad \|h - (f/g)g_m\|_\infty \le \frac{\delta}{m}.$$

Proof. Let $\tilde{g}_m = (f/g)g_m$ and apply Theorem 6.1 for the function $F = f_n - \tilde{g}_m$. With probability $1-\gamma$, there exists $w \in W^{>d}$ such that

$$f_n - \tilde{g}_m - w \in S_{\mu_m} \quad \text{and} \quad \|f_n - \tilde{g}_m - w\|_\infty \le \frac{2e^d 2^p}{\alpha m}\left\|(f_n - \tilde{g}_m)^{\le d}\right\|_{L^2}. \tag{8.2}$$

Set

$$h = f_n - w.$$

Since $w \in W^{>d}$, we have $h^{\le d} = f_n^{\le d}$ as claimed. Since both $\tilde{g}_m$ and $h - \tilde{g}_m = f_n - \tilde{g}_m - w$ lie in S_{μ_m}, so does h, as claimed.

Furthermore, combining both bounds of Lemma 7.1 via the Minkowski inequality, we get

$$\left(\mathbb{E}\left\|(f_n - \tilde{g}_m)^{\le d}\right\|_{L^2}^2\right)^{1/2} \le \binom{p}{\le d}^{1/2}\left(\frac{1}{\sqrt{n}} + \frac{\kappa}{\sqrt{m}}\right)\frac{1}{2^p}.$$

By Chebyshev's inequality, with probability at least $1-\gamma$ we have

$$\left\|(f_n - \tilde{g}_m)^{\le d}\right\|_{L^2} \le \gamma^{-1/2}\binom{p}{\le d}^{1/2}\left(\frac{1}{\sqrt{n}} + \frac{\kappa}{\sqrt{m}}\right)\frac{1}{2^p}.$$

We substitute this into (8.2) and get

$$\|h - \tilde{g}_m\|_{L^\infty(\nu_m)} \le \frac{2e^d 2^p}{\alpha m} \cdot \gamma^{-1/2}\binom{p}{\le d}^{1/2}\left(\frac{1}{\sqrt{n}} + \frac{\kappa}{\sqrt{m}}\right)\frac{1}{2^p} \le \frac{\delta}{m},$$

where we used the assumption on n and m in the last bound. ■

8.1. Proof of Theorem 3.1.

Let us explain how Theorem 8.1 is a more general form of Theorem 3.1. Let f and g be the densities of the two distributions in the statement of Theorem 3.1, $X = (x_1, \dots, x_n)$ and $S = (y_1, \dots, y_m)$ be the samples drawn according to these densities, and $f_n = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{x_i}$ and $g_m = \frac{1}{m}\sum_{i=1}^m \mathbf{1}_{y_i}$ be the empirical densities. The regularity assumption implies that

$$f/g \asymp 1 \quad \text{pointwise,} \tag{8.3}$$

and in particular the requirement $\left\|f/g\right\|_{L^2} = O(1)$ holds in Theorem 8.1. The function h we obtain from that result is supported on $S = \mathrm{supp}(g_m)$ and satisfies

$$h \geq (f/g)g_m - \frac{\delta}{m} \gtrsim \frac{1}{m} \quad \text{everywhere on } S.$$

(In the last step we used (8.3) that $g_m = \frac{1}{m}\sum_{i=1}^m \mathbf{1}_{y_i}$ is lower bounded by $1/m$ on S.) In particular, h is positive on S. The condition $h^{\leq d} = f_n^{\leq d}$ means that h has exactly the same marginals up to dimension d as f_n, the uniform probability distribution on X. Since f_n is a density, the sum of all of its values equals 1. The same must be true for h, since the sum of the values can be expressed as the zero-dimensional marginal, which must be the same for h and f_n. In other words, h must be a density, too. Theorem 3.1 is proved.

8.2. A one-sample version. Here is a one-sample version of Theorem 8.1. It is a rigorous version of Theorem 3.6 we stated informally in the introduction.

Theorem 8.2. *Let f be a density on the cube $\{-1,1\}^p$ that is bounded below by $\alpha/2^p$, and let f_m be its empirical counterpart. If $m \geq 16(\alpha\delta)^{-2}\gamma^{-1}e^{2d}\binom{p}{\leq d}$ then the following holds with probability $1-2\gamma$. There exists a density h on $\mathrm{supp}(f_m)$ that satisfies*

$$h^{\leq d} = f^{\leq d}, \quad \|h - f_m\|_\infty \leq \frac{\delta}{m}.$$

Proof. The proof is similar to that of Theorem 3.1 above. Choose $g = f$, $n = m$, hence $\tilde{g}_m = (f/g)g_m = f_m$, and use $F = f - \tilde{g}_m$. Apply only the first bound in Lemma 7.1.

Note that the bound in the conclusion and the fact that $f_m = 1/m$ on its support implies that $h \geq 1/m - \delta/m > 0$ on $\mathrm{supp}(f_m)$, and thus h is a density.

We leave the details to the reader. ■

9. Solution space. Our next focus is on proving Theorem 3.5, which gives guarantees for privacy and accuracy of the synthetic data created by Algorithm 3.1.

Let us formally introduce the solution space – the space of all functions on the reduced sample space S that have the same marginals as a given function u.

Definition 9.1 (Solution space). *Let μ be a probability measure on the cube $\{-1,1\}^p$, and μ_m be its empirical counterpart. For any function $u : \{-1,1\}^p \to \mathbb{R}$, consider the affine subspace $H(u)$ of all functions supported on $\mathrm{supp}(\mu_m)$ and that have the same marginals up to dimension d as the function u, i.e.*

$$H(u) := \left\{h \in S_{\mu_m} : \; h^{\leq d} = u^{\leq d}\right\} = \left(u - W^{>d}\right) \cap S_{\mu_m},$$

where S_{μ_m}, as before, denotes the linear space of all functions supported on the reduced space $S = \mathrm{supp}(\mu_m)$.

9.1. Success with high probability. The Algorithm 3.1 succeeds, i.e. does not return "Failure", when the reduced space $S = \{\theta_1, \ldots, \theta_m\}$ is well conditioned. By definition, this happens if

$$s_{\min}(M) \geq \frac{\sqrt{m}}{2e^d} \tag{9.1}$$

where $s_{\min}$ denotes the smallest singular value, and M is the $m \times \binom{p}{\leq d}$ matrix whose entries are $w_J(\theta_i)$ for $|J| \leq d$, i.e. the matrix whose rows are indexed by the points $\theta_i \in S$, and whose columns are indexed by Walsh functions w_J of degree at most d.

Let us reformulate the condition (9.1) in the dual form, and then deduce from Theorem 6.1 that that it holds with high probability.

Lemma 9.2 (Well conditioned reduced space). *The reduced space S is well conditioned if and only if any function $F : \{-1,1\}^p \to \mathbb{R}$ satisfies*

$$\inf\left\{\|F-w\|_{L^2(\mu_m)} : \ w \in W^{>d},\ F-w \in S_{\mu_m}\right\} \leq \frac{2e^d 2^p}{m}\left\|F^{\leq d}\right\|_{L^2}. \tag{9.2}$$

Proof. Decomposing $F = F^{\leq d} + F^{>d}$ we see that $F^{\leq d}$ in the right hand side of (9.2) may be replaced by F without loss of generality. Furthermore, since $\|f\|_{L^2(\mu_m)} = \sqrt{2^p/m}\|f\|_{L^2}$ for any $f \in S_{\mu_m}$, we can rewrite condition (9.2) equivalently as

$$\inf\left\{\|F-w\|_{L^2} : \ w \in W^{>d},\ F-w \in S_{\mu_m}\right\} \leq B\|F\|_{L^2} \tag{9.3}$$

where

$$B = 2e^d\sqrt{\frac{2^p}{m}}.$$

We will employ a duality argument similar to the one we used in the proof of Theorem 6.1. Given $\delta > 0$, consider the weighted Hilbert space L^2_δ where the norm is defined by

$$\|F\|^2_{L^2_\delta} := \|F\mathbf{1}_S\|^2_{L^2} + \delta\|F\mathbf{1}_{S^c}\|^2_{L^2}.$$

where $\mathbf{1}_S$ denotes the indicator function of S. Then (9.3) is equivalent to

$$\inf\left\{\|F-w\|_{L^2_{1/\delta}} : \ w \in W^{>d}\right\} \leq B\|F\|_{L^2} \quad \forall \delta > 0.$$

(To see this, note that taking $\delta \to 0_+$ enforces $F - w\mathbf{1}_{S^c} = 0$, or equivalently $F - w \in S_{\mu_m}$.) This can be interpreted as a bound on the norm of the quotient map Q:

$$\left\|Q : \ L^2 \to L^2_{1/\delta}/W^{>d}\right\| \leq B \quad \forall \delta > 0.$$

Let us dualize this bound. The adjoint operator has the same norm, so

$$\left\|Q^* : \ \left(L^2\right)^* \to \left(L^2_{1/\delta}/W^{>d}\right)^*\right\| \leq B \quad \forall \delta > 0.$$

The adjoint of the quotient map is the canonical (identity) embedding; the Hilbert space L^2 is self-dual, and the dual of a quotient space is a subspace of the dual, i.e.

$$\left(L^2_{1/\delta}/W^{>d}\right)^* = ((W^{>d})^\perp, \|\,\|_{(L^2_{1/\delta})^*}) = (W^{\leq d}, \|\,\|_{L^2_\delta}).$$

Thus, the bound is equivalent to

$$\left\|\mathrm{Id} : \ (W^{\leq d}, \|\,\|_{L^2_\delta}) \to L^2\right\| \leq B \quad \forall \delta > 0.$$

By definition of the operator norm and the norm in L^2_δ, this bound is equivalent to saying that

$$\|F\|_{L^2}^2 \le B^2 \left(\|F\mathbf{1}_S\|_{L^2}^2 + \delta\|F\mathbf{1}_{S^c}\|_{L^2}^2\right) \quad \forall F \in W^{\le d}, \forall \delta > 0.$$

Taking $\delta \to 0_+$, we see that this is equivalent to

$$\|F\|_{L^2}^2 \le B^2\|F\mathbf{1}_S\|_{L^2}^2 = \frac{B^2}{2^p}\|F\mathbf{1}_S\|_{\ell^2}^2 = \frac{4e^{2d}}{m}\|F\mathbf{1}_S\|_{\ell^2}^2 \quad \forall F \in W^{\le d}.$$

Expressing F through its orthogonal decomposition $F = \sum_{|J|\le d} a_J w_J$, we can rewrite the latter condition as

$$\sum_{|J|\le d} a_J^2 \le \frac{4e^{2d}}{m}\left\|\sum_{|J|\le d} a_J w_J \mathbf{1}_S\right\|_{\ell^2}^2 = \frac{4e^{2d}}{m}\sum_{i=1}^m \Big(\sum_{|J|\le d} a_J w_J(\theta_i)\Big)^2 \quad \forall \text{ choice of coefficients } a_J.$$

This in turn is equivalent to

$$\|a\|_{\ell^2}^2 \le \frac{4e^{2d}}{m}\|Ma\|_{\ell^2}^2\,,$$

which is finally equivalent to (9.1). ∎

Proposition 9.3 (Success with high probability). *If $m \ge 16\gamma^{-2}e^{2d}\binom{p}{\le d}$, then Algorithm 3.1 succeeds (i.e. does not return "Failure") with probability at least $1-\gamma$.*

Proof. By definition, Algorithm 3.1 succeeds if the reduced space S is well conditioned. Then the conclusion immediately follows from Theorem 6.1 for the uniform density μ, Lemma 9.2 and the fact that the $L^2(\mu_m)$ norm is bounded by the sup-norm. ∎

9.2. All solution spaces are translates of each other. First let us show that with high probability in μ_m, all solution spaces $H(u)$ are nonempty and are translates of each other. The following elementary lemma will help us.

Proposition 9.4. *If the reduced space S is well conditioned, the solution spaces $H(u)$ for all $u : \{-1,1\}^p \to \mathbb{R}$ are nonempty and are translates of each other.*

Proof. Let $F : \{-1,1\}^p \to \mathbb{R}$ be an arbitrary function. If S is well conditioned, Lemma 9.2 for $F = u$ yields the existence of $w \in W^{>d}$ and $s \in S_{\mu_m}$ such that $u = s + w$. This implies that $u - W^{>d} = s - W^{>d}$. Hence

$$H(u) = \left(u - W^{>d}\right) \cap S_{\mu_m} = \left(s - W^{>d}\right) \cap S_{\mu_m} = s - \left(W^{>d} \cap S_{\mu_m}\right).$$

The linear subspace $W^{>d} \cap S_{\mu_m}$ is nonempty as it contains the origin. Therefore, all solution spaces $H(u)$ are translates of this linear space, and thus of each other. ∎

9.3. Sensitivity of the solution space. Next, we will check that the map $u \mapsto H(u)$ is Lipschitz in the Hausdorff metric. Recall that the Hausdorff distance between two subsets A and B of a normed space X is defined as

$$d_X(A,B) = \max\left\{\sup_{a\in A}\inf_{b\in B}\|a-b\|_X\,,\ \sup_{b\in B}\inf_{a\in A}\|a-b\|_X\right\}.$$

When A and B are affine subspaces that are translates of each other, we have

$$d_X(A, B) = \inf_{b \in B} \|a - b\|_X = \operatorname{dist}_X(a, B) \quad \text{for any } a \in A.$$

When the norm is clear from the context, we skip the subscript X. When $X = L^q$ we simply write $d_q(A, B)$.

Lemma 9.5 (Sensitivity of the solution space). *If the reduced space S is well conditioned, then any pair of functions $u_1, u_2 : \{-1, 1\}^p \to \mathbb{R}$ satisfies*

$$d_\infty \left(H(u_1),\, H(u_2)\right) \le \frac{2e^d 2^p}{\sqrt{m}} \left\|(u_1 - u_2)^{\le d}\right\|_{L^2}. \tag{9.4}$$

Proof. Since, by Proposition 9.4, the affine subspaces $H(u_1)$ and $H(u_2)$ are translates of each other, it suffices to bound $\inf_{s_2 \in H(u_2)} \|s_1 - s_2\|_\infty$ for any $s_1 \in H(u_1)$.

Pick any $s_1 \in H(u_1)$. Since $H(u_1) = (u_1 - W^{>d}) \cap S_{\mu_m}$, there exists $w_1 \in W^{>d}$ such that $s_1 = u_1 - w_1 \in S_{\mu_m}$. Apply the bound in Lemma 9.2 for $F = s_1 - u_2$. There exists $w_2 \in W^{>d}$ such that $s_1 - u_2 - w_2 \in S_{\mu_m}$ and

$$\|s_1 - u_2 - w_2\|_\infty \le \sqrt{m}\|s_1 - u_2 - w_2\|_{L^2(\mu_m)} \le \frac{2e^d 2^p}{\sqrt{m}} \left\|(s_1 - u_2 - w_2)^{\le d}\right\|_{L^2}. \tag{9.5}$$

Since both s_1 and $s_1 - u_2 - w_2$ lie in the linear subspace S_{μ_m}, it must be that $s_2 := u_2 + w_2 \in S_{\mu_m}$ as well. Since $w_2 \in W^{>d}$, it follows that $s_2 \in (u_2 + W^{>d}) \cap S_{\mu_m} = H(u_2)$.

Furthermore,

$$(s_1 - u_2 - w_2)^{\le d} = (u_1 - w_1 - u_2 - w_2)^{\le d} = (u_1 - u_2)^{\le d}.$$

(In the last step, we used that w_1 and w_2 are in $W^{>d}$ and so $(w_1)^{\le d} = (w_2)^{\le d} = 0$.)

Therefore, we can rewrite (9.5) as

$$\|s_1 - s_2\|_\infty \le \frac{2e^d 2^p}{\sqrt{m}} \left\|(u_1 - u_2)^{\le d}\right\|_{L^2}.$$

The proof is complete. ■

9.4. Changing a single data point. The Sensitivity Lemma 9.5 will be applied in the situation where u_1 and u_2 are the uniform densities on the two datasets X_1 and X_2 that are different by a single element. Let us specialize the bound (9.4) to this case.

Suppose $X_1 = (x_1, \ldots, x_n)$ and $X_2 = (x_1, \ldots, x_n, x_{n+1})$. Here, in our discussion of privacy, we allow x_i be arbitrary points drawn from $\{-1, 1\}^p$; they do not need to be random. The corresponding densities are

$$f_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{x_i} \quad \text{and} \quad f_{n+1} = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbf{1}_{x_i}.$$

A direct calculation yields

$$f_{n+1} - f_n = \frac{1}{n+1} \left(\mathbf{1}_{x_{n+1}} - f_n\right).$$

Using triangle inequality and then (7.3) and (7.4), we get

$$\left\|(f_{n+1}-f_n)^{\le d}\right\|_{L^2} \le \frac{1}{n+1}\left(\left\|(\mathbf{1}_{x_{n+1}})^{\le d}\right\|_{L^2} + \left\|(f_n)^{\le d}\right\|_{L^2}\right) \le \frac{2}{n}\binom{p}{\le d}^{1/2}\frac{1}{2^p}. \tag{9.6}$$

10. Selection rule. Next, we want to extend sensitivity to the selection rule. Can we pick one point from a solution space in such a way that a small change in the solution space always leads to a small change in the selected point?

10.1. L^2 sensitivity. We do not know the best selection rule in the L^∞ metric. The problem is simpler for the L^2 metric: the proximal point (to a given reference point) is a good selection rule.

Lemma 10.1 (Sensitivity of the closest point in the Hilbert space). *Consider a Hilbert space X and a reference point $r \in X$. Let $x(K)$ denote a point in a nonempty closed set $K \subset X$ that is closest to r, i.e.*

$$x_r(K) = \operatorname{argmin}\left\{\|x-r\| : x \in K\right\}.$$

Then, for any two nonempty closed convex sets $K_1, K_2 \subset X$, we have

$$\left\|x_r(K_1) - x_r(K_2)\right\|^2 \le 4\max\left(\operatorname{dist}(r,K_1), \operatorname{dist}(r,K_2)\right)\cdot d(K_1,K_2).$$

In order to prove this lemma, we first observe:

Lemma 10.2. *Suppose that K is a nonempty closed convex subset of a Hilbert space X. Let $r \in X$. Let $x_0 = \operatorname{argmin}\left\{\|x-r\| : x \in K\right\}$. Then*

$$\|x_0 - y\|^2 \le 2\left(\|y-r\|^2 - \|x_0-r\|^2\right)$$

for all $y \in K$.

Proof. Without loss of generality, assume that $r = 0$. Let $y \in K$. Since $\frac{x_0+y}{2} \in K$, we have $\left\|\frac{x_0+y}{2}\right\| \ge \|x_0\|$, so

$$\left\|\frac{x_0-y}{2}\right\|^2 + \|x_0\|^2 \le \left\|\frac{x_0-y}{2}\right\|^2 + \left\|\frac{x_0+y}{2}\right\|^2 = \frac{1}{2}(\|x_0\|^2 + \|y\|^2).$$

Thus, $\|x_0-y\|^2 \le 2(\|y\|^2 - \|x_0\|^2)$. ■

Proof of Lemma 10.1. If $d(K_1,K_2) \ge d(r,K_1) + d(r,K_2)$, then we are done, since

$$\begin{aligned}\|x_r(K_1) - x_r(K_2)\| &\le \|x_r(K_1) - r\| + \|x_r(K_2) - r\| \\ &= d(r,K_1) + d(r,K_2) \le \sqrt{(d(r,K_1)+d(r,K_2))d(K_1,K_2)}.\end{aligned}$$

Thus, we may assume that $d(K_1,K_2) \le d(r,K_1) + d(r,K_2)$. Without loss of generality, we may also assume that $d(r,K_2) \le d(r,K_1)$. By Lemma 10.2,

$$\|x_r(K_1) - y\|^2 \le 2(\|y-r\|^2 - d(r,K_1)^2),$$

for all $y \in K_1$. Note that we can write $x_r(K_2) = y + d(K_1, K_2)z$ for some $y \in K_1$ and $z \in X$ with $\|z\| \leq 1$. Since

$$\|y - r\| \leq \|x_r(K_2) - r\| + d(K_1, K_2) = d(r, K_2) + d(K_1, K_2),$$

it follows that

$$\begin{aligned}
&\|x_r(K_1) - y\|^2 \\
\leq& 2[(d(r, K_2) + d(K_1, K_2))^2 - d(r, K_1)^2] \\
=& 2[d(r, K_2) + d(K_1, K_2) + d(r, K_1)][d(r, K_2) + d(K_1, K_2) - d(r, K_1)] \\
\leq& 2[d(r, K_2) + d(K_1, K_2) + d(r, K_1)]d(K_1, K_2) \\
\leq& 4(d(r, K_1) + d(r, K_2))d(K_1, K_2),
\end{aligned}$$

where the second inequality follows from the assumption that $d(r, K_2) \leq d(r, K_1)$ and the last inequality follows from the assumption that $d(K_1, K_2) \leq d(r, K_1) + d(r, K_2)$. ■

10.2. Restriction onto the cube. Functions that comprise the solution space $H(u)$ may take negative values, hence not all of $H(u)$ consists of densities. So, our next goal is to restrict the affine space $H(u)$ to the positive orthant $[0, \infty)^m$ and check that sensitivity still holds. Our Algorithm 3.1 makes a more aggressive restriction onto the cube $[2\delta/m, (\Delta - \delta)/m]^m$. This is what we will analyze now.

Lemma 10.3 (Restriction onto a cube). *Let H_1 and H_2 be a pair of parallel affine subspaces of $\mathbb{R}^m$ with equal dimensions. Assume that for some scalars $a < b$, we have*

$$H_i \cap [a, b]^m \neq \emptyset, \quad i = 1, 2.$$

Fix any $\lambda > 0$ and consider the cube $Q = [a - \lambda, b + \lambda]^m$. Then

$$d_\infty (H_1 \cap Q,\, H_2 \cap Q) \leq \left(\frac{b - a}{\lambda} + 2 \right) d_\infty (H_1, H_2).$$

Proof. Due to symmetry, it is enough to bound the quantity

$$\sup_{h_1 \in H_1 \cap Q} \inf_{h_2 \in H_2 \cap Q} \|h_1 - h_2\|_\infty .$$

So let us fix any $h_1 \in H_1 \cap Q$ and find $h_2 \in H_2 \cap Q$ for which $\|h_1 - h_2\|_\infty$ is small. To this end, fix a vector

$$x_1 \in H_1 \cap [a, b]^m, \tag{10.1}$$

which exists by assumption. Due to the definition of Hausdorff distance, we can find $x_2 \in H_2$ such that

$$\|x_2 - x_1\|_\infty \leq d_\infty(H_1, H_2) =: \delta. \tag{10.2}$$

Consider the vector

$$y := x_1 + \frac{\lambda}{\delta}(x_2 - x_1)$$

and set h_2 to be the following convex combination of h_1 and y:

$$h_2 := \Big(1 - \frac{\delta}{\lambda}\Big) h_1 + \frac{\delta}{\lambda} y.$$

(Here we assume that $\delta \le \lambda$. Otherwise, the result follows immediately, since the diameter of Q in L^∞-norm is $b - a + 2\lambda$.) Figure 1 might help to visualize our construction.

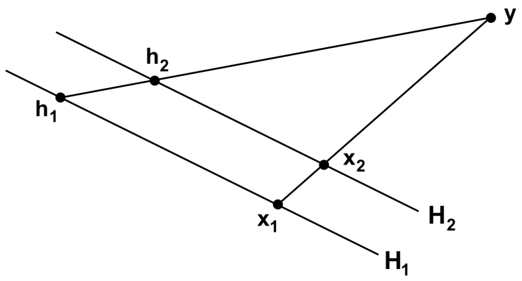

Figure 1: Construction in the proof of Lemma 10.3.

Let us check that the vector h_2 constructed this way satisfies all the required properties. First, we claim that

$$y \in Q.$$

Indeed, the definition of y combined with (10.1) and (10.2) yields

$$y \in [a, b]^m + \frac{\lambda}{\delta}[-\delta, \delta]^m = [a - \lambda, b + \lambda]^m = Q.$$

We claim that

$$h_2 \in H_2.$$

Indeed, substituting the definition of y into the expression for h_2, we get

$$h_2 = \Big(1 - \frac{\delta}{\lambda}\Big)\Big(h_1 - x_1\Big) + x_2 \qquad (10.3)$$

By the assumption, H_1 and H_2 are translates of the same linear subspace. This linear subspace can be expressed as $H_1 - x_1$ or, equivalently, as $H_2 - x_2$ since $x_1 \in H_1$ and $x_2 \in H_2$. In particular, we have $t(H_1 - x_1) = H_2 - x_2$ for any $t \in \mathbb{R}$, or equivalently $H_2 = t(H_1 - x_1) + x_2$. Since $h_1 \in H_1$, it follows from (10.3) that $h_2 \in H_2$ as claimed.

Next, since both h_1 and y lie in Q, their convex combination must lie there, too, so

$$h_2 \in Q.$$

Finally, using the definition of h_2 and recalling that h_1 and y lie in Q, we get

$$h_1 - h_2 = \frac{\delta}{\lambda}(h_1 - y) \in \frac{\delta}{\lambda}(Q - Q) = \frac{\delta}{\lambda}\left[-(b - a + 2\lambda), b - a + 2\lambda\right]^m.$$

Thus

$$\|h_1 - h_2\|_\infty \le \frac{\delta}{\lambda}(b - a + 2\lambda) = \Big(\frac{b - a}{\lambda} + 2\Big)\,\delta.$$

The proof if complete. ■

10.3. L^∞ sensitivity of the selection rule. We are ready to analyze the sensitivity of the L^2-proximal selection rule:

Lemma 10.4 (L^∞ sensitivity of the selection rule). *Let $0 < a < c < (a+b)/2$. Let H_1 and H_2 be a pair of parallel affine subspaces of $\mathbb{R}^m$ with equal dimensions. Assume that*

$$H_i \cap [a,b]^m \neq \emptyset, \quad i = 1,2.$$

Let

$$h_i = \operatorname{argmin}\left\{\|x - c \cdot \mathbf{1}_m\|_2 : \ x \in H_i \cap [a-\lambda, b+\lambda]^m\right\}, \quad i = 1,2.$$

Then

$$\|h_1 - h_2\|_\infty^2 \le 4m(b-c)\Big(\frac{b-a}{\lambda} + 2\Big)\, d_\infty\left(H_1, H_2\right).$$

Proof. Lemma 10.3 gives

$$d_\infty(K_1, K_2) \le \Big(\frac{b-a}{\lambda} + 2\Big)\, d_\infty\left(H_1, H_2\right) \tag{10.4}$$

where $K_i = H_i \cap [a-\lambda, b+\lambda]^m$. Let us apply Lemma 10.1 for $r = c \cdot \mathbf{1}_m$ and the L^2 norm on $\mathbb{R}^m$. Note that

$$\operatorname{dist}_{L^2}(r, K_i) \le \max_{h \in [a,b]^m} \|r - h\|_{L^2} \le \max_{h \in [a,b]^m} \|r - h\|_\infty = \max\left\{|a-c|, |c-b|\right\} = b - c.$$

Thus, Lemma 10.1 yields

$$\|h_1 - h_2\|_{L^2}^2 \le 4(b-c) \cdot d_{L^2}(K_1, K_2) \le 4(b-c) \cdot d_\infty(K_1, K_2).$$

To complete the proof, use (10.4) and note that $\|h_1 - h_2\|_\infty^2 \le m\|h_1 - h_2\|_{L^2}^2$. ∎

11. Shrinkage.

Another step of Algorithm 3.1 we need to control is shrinkage. We will check here that shrinkage onto a cube is Lipschitz in the L^∞-Hausdorff metric. Let us start with a general observation:

Lemma 11.1 (Shrinkage). *Let X be a normed space and $z \in X$ be a point such that $\|z\| \le 1 - \beta$ for some $\beta \in (0,1)$. Let $r : X \to X$ be the retraction map onto the unit ball of X toward z, i.e.*

$$r(x) = (1-\lambda)x + \lambda z$$

where $\lambda = \lambda(x)$ is the minimal number in $[0,1]$ such that $\big\|r(x)\big\| \le 1$. Then the Lipschitz norm of the map $\lambda(\cdot)$ is at most $1/\beta$, and the Lipschitz norm of the map $r(\cdot)$ is at most $2/\beta$.

Proof. Fix any pair of vectors $x_1, x_2 \in X$ and denote

$$\lambda_1 = \lambda(x_1), \quad \lambda_2 = \lambda(x_2), \quad \mu = \|x_1 - x_2\| / \beta.$$

The claim about the Lipschitz norm of $\lambda(\cdot)$ can be stated as $|\lambda_1 - \lambda_2| \le \mu$. By symmetry, it suffices to show that

$$\lambda_1 \le \lambda_2 + \mu. \tag{11.1}$$

This bound is trivial if $\lambda_2 + \mu > 1$ since we always have $\lambda_1 \le 1$. So we can assume from now on that $\lambda_2 + \mu \in [0, 1]$.

Due to the minimality property in the definition of $\lambda_1 = \lambda(x_1)$, in order to prove (11.1) it suffices to show that

$$\big\|(1 - \lambda_2 - \mu)x_1 + (\lambda_2 + \mu)z\big\| \le 1. \tag{11.2}$$

By triangle inequality, the left hand side is bounded by $\|A\| + \|B\|$ where

$$A = (1 - \lambda_2 - \mu)x_2 + (\lambda_2 + \mu)z, \quad B = (1 - \lambda_2 - \mu)(x_1 - x_2).$$

Rearranging the terms, we can rewrite

$$A = (1 - a)\big[(1 - \lambda_2)x_2 + \lambda_2 z\big] + az \quad \text{where} \quad a = \frac{\mu}{1 - \lambda_2}.$$

By assumption, $a \in [0, 1]$. Then A is a convex combination of the vector $(1 - \lambda_2)x_2 + \lambda_2 z$ whose norm is bounded by 1 by definition of $\lambda_2 = \lambda(x_2)$ and the vector z whose norm is bounded by $1 - \beta$ by assumption. Hence, by triangle inequality and definition of a and μ, we have

$$\|A\| \le (1 - a) \cdot 1 + a \cdot (1 - \beta) = 1 - a\beta \le 1 - \mu\beta = 1 - \|x_1 - x_2\|.$$

Furthermore, the assumption $1 - \lambda_2 - \mu \in [0, 1]$ yields

$$\|B\| \le \|x_1 - x_2\|.$$

Hence we showed that $\|A\| + \|B\| \le 1$, establishing (11.2) and completing the first part of the proof (about the Lipschitz norm of λ).

To prove the second part of the lemma, we need to show that

$$\big\|r(x_1) - r(x_2)\big\| \le (2/\beta)\|x_1 - x_2\|. \tag{11.3}$$

Let us first prove this inequality assuming that $\|x_1\| \le 1$ or $\|x_2\| \le 1$. Without loss of generality, assume $\|x_1\| \le 1$. Denoting $\mu_1 = 1 - \lambda_1$ and $\mu_2 = 1 - \lambda_2$ and using triangle inequality, we obtain

$$\big\|r(x_1) - r(x_2)\big\| = \|\mu_1 x_1 + \lambda_1 z - \mu_2 x_2 - \lambda_2 z\| \le \|\mu_1 x_1 - \mu_2 x_2\| + |\lambda_1 - \lambda_2|\|z\| \tag{11.4}$$

By the first part of the lemma and since $\|z\| \le 1 - \beta$, we have

$$|\lambda_1 - \lambda_2|\|z\| \le \frac{1}{\beta}\|x_1 - x_2\|(1 - \beta) = (1/\beta - 1)\|x_1 - x_2\|. \tag{11.5}$$

Furthermore, adding and subtracting the cross term $\mu_2 x_1$ and using triangle inequality, we get

$$\|\mu_1 x_1 - \mu_2 x_2\| \le |\mu_1 - \mu_2|\,\|x_1\| + \mu_2\,\|x_1 - x_2\|.$$

Now, $|\mu_1 - \mu_2| = |\lambda_1 - \lambda_2| \le \|x_1 - x_2\|/\beta$ by the first part of the lemma; $\|x_1\| \le 1$ by the standing assumption, and $\mu_2 \le 1$. Hence

$$\|\mu_1 x_1 - \mu_2 x_2\| \le (1/\beta + 1)\|x_1 - x_2\|. \tag{11.6}$$

Substitute (11.5) and (11.6) into (11.4), we conclude the claim (11.3).

Finally, consider the remaining case where both $\|x_1\| \geq 1$ and $\|x_2\| \geq 1$. Without loss of generality, $\lambda_1 \leq \lambda_2$, so the vectors

$$\tilde{x}_1 \coloneqq (1-\lambda_1)x_1 + \lambda_1 z \quad \text{and} \quad \tilde{x}_2 \coloneqq (1-\lambda_1)x_2 + \lambda_1 z$$

satisfy

$$\|\tilde{x}_1\| = 1 \quad \text{and} \quad \|\tilde{x}_2\| \geq 1.$$

Definition of retraction yields $r(\tilde{x}_1) = r(x_1)$ and $r(\tilde{x}_2) = r(x_2)$. Thus, applying (11.3) for $\tilde{x}_1$ and $\tilde{x}_2$, we get

$$\begin{aligned} \big\|r(x_1) - r(x_2)\big\| &= \big\|r(\tilde{x}_1) - r(\tilde{x}_2)\big\| \leq (2/\beta)\|\tilde{x}_1 - \tilde{x}_2\| \\ &= (2/\beta)(1-\lambda_1)\|x_1 - x_2\| \leq (2/\beta)\|x_1 - x_2\|. \end{aligned}$$

The lemma is proved. ■

Now we extend our analysis of shrinkage for affine subspaces:

Lemma 11.2 (Shrinkage for subspaces). *Let K be the unit ball of a finite dimensional normed space X. Let $z, z_0 \in X$ be points such that $z \in z_0 + (1-\beta)K$ for some $\beta \in (0,1)$. Given an affine subspace H in X, define the affine subspace $\tilde{H}$ by moving H toward z until it intersects the ball $z_0 + K$, i.e.*

$$\tilde{H} = (1-\lambda)H + \lambda z$$

where $\lambda = \lambda(H)$ is the minimal number in $[0,1]$ such that $\tilde{H} \cap (z_0 + K) \neq \emptyset$. Then for any two affine subspaces H_1 and H_2 that are translates of each other, the Hausdorff distance satisfies

$$d_X(\tilde{H}_1, \tilde{H}_2) \leq \frac{2}{\beta} d_X(H_1, H_2).$$

Proof. By translation, we can assume without loss of generality that $z_0 = 0$. The affine subspaces H_1 and H_2 are translates of some common linear subspace H_0. Apply Lemma 11.1 for the quotient space X/H_0 instead of X and for $H_z \coloneqq z + H_0$ instead of z.

The requirement of that lemma is satisfied since

$$\|H_z\|_{X/H_0} = \inf_{h \in H_z} \|h\|_X \leq \|z\|_X \leq 1 - \beta. \tag{11.7}$$

Indeed, the equality here is the definition of the norm in the quotient space, the first inequality holds since $z \in H_z$, and the last inequality is an equivalent form of the assumption $z \in (1-\beta)K$.

We claim that the retraction map $r(\cdot)$ in Lemma 11.1 satisfies

$$r(H) = \tilde{H} \quad \text{for any translate } H \text{ of } H_0.$$

Indeed, by definition we have

$$r(H) = (1-\lambda)H + \lambda H_z$$

where λ is the minimal number in $[0,1]$ such that $\left\|r(H)\right\|_{X/H_0} \le 1$. Since $\|H_z\|_{X/H_0} < 1$ by (11.7), continuity shows that $\lambda < 1$ and hence

$$r(H) = (1-\lambda)H + \lambda z.$$

Moreover, the condition that $\left\|r(H)\right\|_{X/H_0} \le 1$ is equivalent to $r(H) \cap K \ne \emptyset$. Hence the definitions of $r(H)$ and $\tilde{H}$ are equivalent as we claimed.

Lemma 11.1 yields

$$\left\|\tilde{H}_1 - \tilde{H}_2\right\|_{X/H_0} \le \frac{2}{\beta} \left\|H_1 - H_2\right\|_{X/H_0}.$$

It remains to note that, by definition,

$$\|H_1 - H_2\|_{X/H_0} = \inf_{h_1 \in H_1,\, h_2 \in H_2} \|h_1 - h_2\|_X = d_X(H_1, H_2),$$

and similarly for the distance between $\tilde{H}_1$ and $\tilde{H}_2$. The proof is complete. ■

Finally, we specialize our analysis to the shrinkage onto the cube:

Lemma 11.3 (Shrinkage onto a cube). *Let $0 < a < c < (a+b)/2$. Given an affine subspace H in $\mathbb{R}^m$, define the affine subspace $\tilde{H}$ by moving H toward $d\mathbf{1}_m$ until it intersects the cube $[a,b]^m$, i.e.*

$$\tilde{H} = (1-\lambda)H + \lambda \cdot c\mathbf{1}_m$$

where $\lambda = \lambda(H)$ is the minimal number in $[0,1]$ such that $\tilde{H} \cap [a,b]^m \ne \emptyset$. Then for any two affine subspaces H_1 and H_2 that are translates of each other, the Hausdorff distance in the L^∞ norm satisfies

$$d_\infty(\tilde{H}_1, \tilde{H}_2) \le \frac{b-a}{c-a}\, d_\infty(H_1, H_2).$$

Proof. Apply Lemma 11.2 for

$$z = c\mathbf{1}_m, \quad z_0 = \frac{a+b}{2}\,\mathbf{1}_m, \quad K = \Big[-\frac{b-a}{2}, \frac{b-a}{2}\Big]^m.$$

so that z_0 is the center of the cube $[a,b]^m$, K is the centered cube, and $z_0 + K = [a,b]^m$.

Now,

$$z - z_0 = \Big(c - \frac{a+b}{2}\Big)\,\mathbf{1}_m$$

and

$$0 \le \frac{a+b}{2} - c = (1-\beta)\frac{b-a}{2} \quad \text{for } \beta = \frac{2(c-a)}{b-a},$$

so $z - z_0 \in (1-\beta)K$ as required in Lemma 11.2. The conclusion of this lemma is that

$$d_X(\tilde{H}_1, \tilde{H}_2) \le \frac{2}{\beta}\, d_X(H_1, H_2).$$

Since the unit ball K of X is the cube $[-1,1]^m$ scaled by the factor $(b-a)/2$, the norm in X is the L^∞-norm scaled by that factor. Therefore, the conclusion holds for the L^∞ norm as well. ■

12. Privacy and accuracy of the algorithm.

We are ready to analyze the privacy and accuracy of Algorithm 3.1.

12.1. Algorithm.

For convenience we rewrite Algorithm 3.1, see Algorithm 12.1 below. Note that in Step 5 of Algorithm 12.1, the $L^2(S)$-norm is defined as $\|h\|_{L^2(S)}^2 = \frac{1}{m}\sum_{i=1}^m h(s_i)^2$.

Algorithm 12.1 Private sampling synthetic data algorithm

Input: a sequence X of n points in $\{-1,1\}^p$ (true data); m: cardinality of S; d: the degree of the marginals to be matched; parameters δ, Δ with $\Delta > \delta > 0$.

1. Draw a sequence $S = (\theta_1, \dots, \theta_m)$ of m points in the cube independently and uniformly (reduced space).
2. Form the $m \times \binom{p}{\le d}$ matrix M with entries $w_J(\theta_i)$, i.e. the matrix whose rows are indexed by the points of the reduced space S and whose columns are indexed by the Walsh functions of degree at most d. If the smallest singular value of M is bounded below by $\sqrt{m}/2e^d$, call S well conditioned and proceed. Otherwise return "Failure" and stop.
3. Let f_n be the uniform density on true data: $f_n = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{x_i}$. Consider the solution space
$$H = H(f_n) = \left\{ h : \{-1,1\}^p \to \mathbb{R} :\ \operatorname{supp}(h) \subset S,\ h^{\le d} = (f_n)^{\le d} \right\},$$
4. Shrink H toward the uniform density $u_m = \frac{1}{m}\sum_{i=1}^m \mathbf{1}_{s_i}$ on S: let
$$\tilde{H} = (1-\lambda)H + \lambda u_m$$
where $\lambda \in [0,1]$ is the minimal number such that $\tilde{H} \cap [2\delta/m, (\Delta-\delta)/m]^S \neq \emptyset$.
5. Pick a proximal point
$$h^* = \operatorname{argmin}\left\{ \|\tilde{h} - u_m\|_{L^2(S)} :\ \tilde{h} \in \tilde{H} \cap [\delta/m, \Delta/m]^S \right\}.$$

Output: a sequence $Y = (y_1, \dots, y_k)$ of k independent points drawn from S according to density h^*.

Remark 12.1. *The computational complexity of Algorithm 12.1 is governed by the linear program in Step 3 to compute the density h on S that is guaranteed by Theorem 8.1, which dominates the cost of the simple "line-search" optimization in Step 4 and the linear least squares problem in Step 5. The associated linear program has $|S| \le m$ variables (the values of the density on S), $\binom{p}{\le d}$ linear equations (to match the marginals to those of X), and $|S| \le m$ linear inequalities (to ensure the density is nonnegative on S), where $m \ge e^{2d}\binom{p}{\le d}$. The complexity of solving general linear programs is polynomial in the number of variables, see e.g. [31]. Hence (for fixed d) the complexity of Algorithm 12.1 is polynomial in p.*

As already discussed in Section 3.3, if S fails the well-condionedness condition in Step 2, we can simply resample S until it is well conditioned. Since the expected number or trials until success is ≤ 2 (under some mild conditions), Algorithm 3.1 succeeds deterministically, but its running time becomes random (with expected overhead time ≤ 2).

The standing assumption in this section is that the reduced space $S = (s_1, \dots, s_m)$ is random, and consists of points s_i drawn independently and uniformly from the cube. We would like to show that with high probability over S, the algorithm is differentially private.

12.2. Sensitivity of density. The privacy guarantee will be achieved via Private Sampling Lemma 3.3. To apply it, we need to bound the sensitivity of the density h^* computed by the algorithm.

Lemma 12.2. *Suppose the reduced space S is well conditioned. Then, for any pair of input datasets X_1 and X_2 that consist of at least n elements each and differ from each other by a single element, the densities h_1^* and h_2^* computed by the algorithm satisfy*

$$\left\|h_1^* - h_2^*\right\|_\infty \le \frac{4\sqrt{2}\Delta^{3/2}e^{d/2}}{\sqrt{\delta n}\, m^{1/4}} \binom{p}{\le d}^{1/4}.$$

Proof. By Proposition 9.4, the solution subspaces

$$H_1 = H(f_n) \quad \text{and} \quad H_2 = H(f_{n+1})$$

are translates of each other. The ambient space consists of all functions supported on an m-element set S, and thus can be identified with $\mathbb{R}^m$. Let $\tilde{H}_i$ be the result of shrinkage of the subspaces H_i toward the uniform distribution as specified in the algorithm, i.e. the shrinkage onto the cube $[\delta/m, \Delta/m]^m$ and toward the uniform distribution u_m. The selection rule for h^* specified in the algorithm is stable in the L^∞ metric. Indeed, Lemma 10.4 applied for the subspaces $\tilde{H}_i$ and for

$$a = \frac{2\delta}{m}, \quad b = \frac{\Delta - \delta}{m}, \quad c = \frac{1}{m}, \quad \lambda = \frac{\delta}{m}$$

yields

$$\left\|h_1^* - h_2^*\right\|_\infty^2 \le \frac{4\Delta^2}{\delta} \cdot d_\infty(\tilde{H}_1, \tilde{H}_2).$$

Next, recall that the shrinkage map is stable. Indeed, Lemma 11.3 applied for the same a, b, c yields

$$d_\infty(\tilde{H}_1, \tilde{H}_2) \le 2\Delta \cdot d_\infty(H_1, H_2).$$

Furthermore, the solution space is stable. Indeed, Lemma 9.5 for the uniform density μ on the cube yields

$$d_\infty(H_1, H_2) \le \frac{2e^d 2^p}{\sqrt{m}} \left\|(f_n - f_{n+1})^{\le d}\right\|_{L^2}.$$

Finally, recall from (9.6) that

$$\left\|(f_{n+1} - f_n)^{\le d}\right\|_{L^2} \le \frac{2}{n}\binom{p}{\le d}^{1/2} \frac{1}{2^p}.$$

Combining all these bounds, we conclude that

$$\left\|h_1^* - h_2^*\right\|_\infty^2 \le \frac{4\Delta^2}{\delta} \cdot 2\Delta \cdot \frac{2e^d 2^p}{\sqrt{m}} \cdot \frac{2}{n}\binom{p}{\le d}^{1/2} \frac{1}{2^p} \le \frac{32\Delta^3 e^d}{\delta n \sqrt{m}} \binom{p}{\le d}^{1/2}.$$

The proof is complete. ■

12.3. Privacy guarantee. Finally, we are ready to give the privacy guarantee of our algorithm:

Theorem 12.3 (Privacy). *If $k \le \frac{1}{4\sqrt{2}}\varepsilon\big(\frac{\delta}{\Delta}\big)^{3/2}e^{-d/2}\binom{p}{\le d}^{-1/4}\sqrt{n}/m^{3/4}$, then Algorithm 12.1 is ε-differentially private.*

Proof. Since the reduced space S is drawn independently of the input data X, we can condition on S. If S is ill conditioned, the algorithm returns "Failure" regardless of the input data, so the privacy holds trivially. Suppose S is well conditioned.

Let X_1 and X_2 be a pair of datasets that consist of at least n elements each and differ from each other by a single element. By the choice made in the algorithm and by sensitivity of density (Lemma 12.2), we have

$$h_2^* \ge \frac{\delta}{m} \quad \text{and} \quad \big|h_1^* - h_2^*\big| \le \frac{4\sqrt{2}\Delta^{3/2}e^{d/2}}{\sqrt{\delta n}\, m^{1/4}}\binom{p}{\le d}^{1/4} =: \eta$$

pointwise. Therefore

$$\big|h_1^*/h_2^*\big| \le 1 + \frac{\eta m}{\delta} \le \exp\Big(\frac{\eta m}{\delta}\Big) \le \exp\Big(\frac{\varepsilon}{k}\Big)$$

pointwise, where the last inequality indeed holds due to our assumption on k. Private Sampling Lemma 3.3 completes the proof. ■

Remark 12.4. *Suppose we chose the size m of the reduced space S so that $m \asymp e^{2d}\binom{p}{\le d}$. Simplifying the condition in Theorem 12.3, we conclude that if $k \ll \sqrt{n}/m$, then Algorithm 12.1 is $o(1)$-differentially private.*

12.4. Accuracy guarantee. The following is the accuracy guarantee of our algorithm:

Theorem 12.5 (Accuracy). *Assume the true data $X = (x_1, \ldots, x_n)$ is drawn independently from the cube according to some density f, which satisfies $\|f\|_\infty \le \Delta/2^p$. Assume that $n \ge 16\delta^{-2}\gamma^{-1}e^{2d}\binom{p}{\le d}$, $16\delta^{-2}\gamma^{-1}\Delta^2 e^{2d}\binom{p}{\le d} \le m \le 2^{p/4}$, and $k \ge 4\delta^{-2}(\log(2/\gamma) + \log\binom{p}{\le d})$. Then, with probability at least $1 - 4\gamma - \frac{1}{\sqrt{2^p}}$, the algorithm succeeds, and all marginals of the synthetic data Y up to dimension d are within 4δ from the corresponding marginals of the true data X.*

Proof. Proposition 9.3 and the choice of m guarantee that the algorithm succeeds with probability at least $1 - \gamma$.

Furthermore, the uniform density on the cube $g = 2^{-p}$ satisfies $\|f/g\|_{L^2} \le \|f/g\|_\infty = \|f\|_\infty \cdot 2^p \le \Delta$. Therefore, Theorem 8.1 implies that with probability at least $1 - 2\gamma$, there exists $h \in H = H(f_n)$ such that

$$\big\|h - (f/g)g_m\big\|_\infty \le \frac{\delta}{m}. \tag{12.1}$$

Since $(f/g)g_m$ is a nonnegative function, it follows that

$$h \ge -\frac{\delta}{m} \quad \text{pointwise.}$$

The assumption $m \le 2^{p/4}$ implies that with probability $1 - \frac{1}{\sqrt{2^p}}$ there are no repetitions in $y_1, \ldots, y_m$, which in turn implies that with probability $1 - \frac{1}{\sqrt{2^p}}$ we have $\|g_m\|_\infty \le 1/m$ (otherwise $\|g_m\|_\infty$ would scale with the number of repetitions in $y_1, \ldots, y_m$).

In the following we condition on the event that there are no repetitions in $y_1, \dots, y_m$. Since $\big\|f/g\big\|_\infty \le \Delta$ by above and $\|g_m\|_\infty \le 1/m$, we have $\big\|(f/g)g_m\big\|_\infty \le \Delta/m$, so

$$h \le \frac{\Delta + \delta}{m} \quad \text{pointwise.}$$

A combination of these two bounds on h implies that

$$\frac{2\delta}{m} \le (1 - 3\delta)h + \frac{3\delta}{m} \le \frac{\Delta - \delta}{m} \quad \text{pointwise,}$$

as long as $\Delta \ge 5/3$. Since $h \in H$, it follows that the affine subspace $(1 - 3\delta)H + 3\delta u_m$ has a nonempty intersection with $[2\delta/m, (\Delta - \delta)/m]^m$. The minimality property of λ in the algorithm yields

$$\lambda \le 3\delta. \tag{12.2}$$

Recall that a marginal of a function $f : \{-1, 1\}^p \to \mathbb{R}$ that corresponds to a subset $J \subset [p]$ of parameters and values $\theta_j \in \{-1, 1\}$ for $j \in J$, is defined as

$$P(f) = \sum_{x \in \{-1,1\}^p} f(x)v(x)$$

where $v(x) = \mathbf{1}_{\{x(j) = \theta_j \ \forall j \in J\}}$.

Recall that the solution h^* of the algorithm satisfies

$$h^* \in \tilde{H} = (1 - \lambda)H + \lambda u_m$$

and, by definition of H, all members of H have the same marginals up to dimension d as f_n. This and linearity implies that for any marginal up to dimension d,

$$P(h^*) = (1 - \lambda)P(f_n) + \lambda P(u_m)$$

Hence

$$\big|P(h^*) - P(f_n)\big| \le \lambda \big|P(u_m) - P(f_n)\big|$$

Since u_m and f_n are densities, all of their marginals must be within $[0, 1]$, so $\big|P(u_m) - P(f_n)\big| \le 1$. Combining this with (12.2), we get

$$\big|P(h^*) - P(f_n)\big| \le 3\delta, \tag{12.3}$$

for all marginals up to dimension d, with probability at least $1 - 2\gamma$.

Now we compare the marginals of the density h^* and its empirical counterpart h_k^*. We can express

$$P(h_k^*) - P(h^*) = \frac{1}{k} \sum_{i=1}^{k} \big(v(Y_i) - \mathbb{E}\, v(Y_i)\big)$$

where Y_i are i.i.d. random variables drawn according to the density h^*. Thus, we have a normalized and centered sum of i.i.d. Bernoulli random variables, so Bernstein's inequality (see e.g. [42, Theorem 2.8.4]) yields

$$\mathbb{P}\left\{\left|P(h_k^*) - P(h^*)\right| > \delta\right\} \leq 2\exp(-\delta^2 k/4) \leq \gamma \binom{p}{\leq d}^{-1}$$

if $k \geq 4\delta^{-2}(\log(2/\gamma) + \log \binom{p}{\leq d})$. Thus, by a union bound, we have

$$\left|P(h_k^*) - P(h^*)\right| \leq \delta,$$

simultaneously for all marginals up to dimension d, with probability at least $1-\gamma$.

Combining this with (12.3) via the triangle inequality, we conclude that

$$\left|P(h_k^*) - P(f_n)\right| \leq 4\delta,$$

for all marginals up to dimension d, with probability at least $1 - 3\gamma$. Recalling that we conditioned on an event with probability $1 - 1/\sqrt{p}$ and applying the union bound completes the proof. ■

Remark 12.6 (No shrinkage for regular densities). *If the density f from which the true data X is drawn is regular, specifically if $3\delta/2^p \leq f \leq (\Delta - 2\delta)/2^p$ pointwise for some positive numbers δ and Δ,* the algorithm does not apply any shrinkage. *Indeed, in this case we have $3\delta/m \leq (f/g)g_m \leq (\Delta - 2\delta)m$, so it follows from (12.1) that $2\delta/m \leq h \leq (\Delta - \delta)m$, and thus H has a nonempty intersection with $[2\delta/m, (\Delta - \delta)m]^S$, hence $\lambda = 0$.*

REFERENCES

[1] Nazmiye Ceren Abay, Yan Zhou, Murat Kantarcioglu, Bhavani Thuraisingham, and Latanya Sweeney. Privacy preserving synthetic data release using deep learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 510–526. Springer, 2018.

[2] John M Abowd and Simon D Woodcock. Disclosure limitation in longitudinal linked data. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*, 215277, 2001.

[3] Sergul Aydore, William Brown, Michael Kearns, Krishnaram Kenthapadi, Luca Melis, Aaron Roth, and Ankit Siva. Differentially private query release through adaptive projection, 2021.

[4] Boaz Barak, Kamalika Chaudhuri, Cynthia Dwork, Satyen Kale, Frank McSherry, and Kunal Talwar. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 273–282, 2007.

[5] Roberto J Bayardo and Rakesh Agrawal. Data privacy through optimal k-anonymization. In *21st International conference on data engineering (ICDE'05)*, pages 217–228. IEEE, 2005.

[6] Brett K Beaulieu-Jones, Zhiwei Steven Wu, Chris Williams, Ran Lee, Sanjeev P Bhavnani, James Brian Byrd, and Casey S Greene. Privacy-preserving generative deep neural networks support clinical data sharing. *Circulation: Cardiovascular Quality and Outcomes*, 12(7):e005122, 2019.

[7] Steven M Bellovin, Preetam K Dutta, and Nathan Reitinger. Privacy and synthetic datasets. *Stan. Tech. L. Rev.*, 22:1, 2019.

[8] Anat Reiner Benaim, Ronit Almog, Yuri Gorelik, Irit Hochberg, Laila Nassar, Tanya Mashiach, Mogher Khamaisi, Yael Lurie, Zaher S Azzam, Johad Khoury, et al. Analyzing medical research results based on synthetic data and their relation to real data results: systematic comparison from five observational studies. *JMIR medical informatics*, 8(2):e16492, 2020.

[9] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to noninteractive database privacy. *Journal of the ACM (JACM)*, 60(2):1–25, 2013.
[10] March Boedihardjo, Thomas Strohmer, and Roman Vershyin. Covariance's Loss is Privacy's Gain: Computationally Efficient, Private and Accurate Synthetic Data. *arXiv preprint arXiv:2107.05824*, 2021.
[11] March Boedihardjo, Thomas Strohmer, and Roman Vershyin. Privacy of synthetic data in the statistical framework, 2021. Manuscript.
[12] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.
[13] Jim Burridge. Information preserving statistical obfuscation. *Statistics and Computing*, 13(4):321–327, 2003.
[14] T Tony Cai, Yichen Wang, and Linjun Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics*, 2020, to appear.
[15] Jessamyn Dahmen and Diane Cook. Synsys: A synthetic data generation system for healthcare applications. *Sensors*, 19(5):1181, 2019.
[16] Laurent Jacques de Montjoye and Rémi Gribonval. Compressive learning with privacy guarantees. *Information and Inference*, to appear, 2021.
[17] Anne Marie Delaney, Eoin Brophy, and Tomas E Ward. Synthesis of realistic ECG using generative adversarial networks. *arXiv preprint arXiv:1909.09150*, 2019.
[18] John C Duchi, Michael I Jordan, and Martin J Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.
[19] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.
[20] Cynthia Dwork, Aleksandar Nikolov, and Kunal Talwar. Efficient algorithms for privately releasing marginals via convex relaxations. *Discrete & Computational Geometry*, 53(3):650–673, 2015.
[21] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.
[22] Ferdinando Fioretto, Cuong Tran, and Pascal Van Hentenryck. Decision making with differential privacy under a fairness lens. *arXiv preprint arXiv:2105.07513*, 2021.
[23] Moritz Hardt, Katrina Ligett, and Frank McSherry. A simple and practical algorithm for differentially private data release. *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*, 2012.
[24] Moritz Hardt and Guy N Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 61–70. IEEE, 2010.
[25] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the 42nd ACM symposium on Theory of computing, STOC '10*, pages 705–714, New York, NY, USA, 2010.
[26] James Jordon, Jinsung Yoon, and Mihaela Van Der Schaar. PATE-GAN: Generating synthetic data with differential privacy guarantees. In *International Conference on Learning Representations*, 2018.
[27] Michael Kearns and Aaron Roth. How much still needs to be done to make algorithms more ethical. *URL: https://www.shine.cn/opinion/2008214615/*, 2020.
[28] Haoran Li, Li Xiong, and Xiaoqian Jiang. Differentially private synthesization of multi-dimensional data using copula functions. In *Advances in database technology: proceedings. International conference on extending database technology*, volume 2014, page 475. NIH Public Access, 2014.
[29] Terrance Liu, Giuseppe Vietri, Thomas Steinke, Jonathan Ullman, and Zhiwei Steven Wu. Leveraging public data for practical private query release. *Preprint, arXiv:2102.08598*, 2021.
[30] Pei-Hsuan Lu and Chia-Mu Yu. Poster: A unified framework of differentially private synthetic data release with generative adversarial network. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 2547–2549, 2017.
[31] Nimrod Megiddo. *Progress in Mathematical Programming: Interior-Point and Related Methods*. Springer Science & Business Media, 2012.
[32] Ofer Mendelevitch and Michael D Lesh. Fidelity and privacy of synthetic medical data. *arXiv preprint arXiv:2101.08658*, 2021.
[33] Elaine M Newton, Latanya Sweeney, and Bradley Malin. Preserving privacy by de-identifying face images. *IEEE Transactions on Knowledge and Data Engineering*, 17(2):232–243, 2005.

[34] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: the sparse and approximate cases. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 351–360, 2013.

[35] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[36] Haoyue Ping, Julia Stoyanovich, and Bill Howe. Datasynthesizer: Privacy-preserving synthetic datasets. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*, pages 1–5, 2017.

[37] David Pujol, Ryan McKenna, Satya Kuppam, Michael Hay, Ashwin Machanavajjhala, and Gerome Miklau. Fair decision making using privacy-protected data. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 189–199, 2020.

[38] Zhongzheng Ren, Yong Jae Lee, and Michael S Ryoo. Learning to anonymize faces for privacy preserving action detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 620–636, 2018.

[39] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[40] Justin Thaler, Jonathan Ullman, and Salil Vadhan. Faster algorithms for privately releasing marginals. In *International Colloquium on Automata, Languages, and Programming*, pages 810–821. Springer, 2012.

[41] Jonathan Ullman and Salil Vadhan. PCPs and the hardness of generating private synthetic data. In *Theory of Cryptography Conference*, pages 400–416. Springer, 2011.

[42] Roman Vershynin. *High-dimensional probability. An introduction with applications in data science*. Cambridge University Press, 2018.

[43] Gus Wezerek and David Van Riper. Changes to the Census could make small towns disappear. *New York Times*, Feb. 6, 2020.

[44] Liyang Xie, Kaixiang Lin, Shu Wang, Fei Wang, and Jiayu Zhou. Differentially private generative adversarial network. *arXiv preprint arXiv:1802.06739*, 2018.

[45] Kaiyu Yang, Jacqueline Yau, Li Fei-Fei, Jia Deng, and Olga Russakovsky. A study of face obfuscation in ImageNet. *Preprint, arXiv:2103.06191*, 2021.

[46] Bin Zhou, Jian Pei, and WoShun Luk. A brief survey on anonymization techniques for privacy preserving publishing of social network data. *ACM Sigkdd Explorations Newsletter*, 10(2):12–22, 2008.

[47] Fei Zhu, Fei Ye, Yuchen Fu, Quan Liu, and Bairong Shen. Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. *Scientific reports*, 9(1):1–11, 2019.

[48] Keyu Zhu, Pascal Van Hentenryck, and Ferdinando Fioretto. Bias and variance of post-processing in differential privacy. *arXiv preprint arXiv:2010.04327*, 2020.