Improving the Gaussian Mechanism for Differential Privacy: Analytical Calibration and Optimal Denoising

Borja Balle [1] **Yu-Xiang Wang** [2 3]

Abstract

The Gaussian mechanism is an essential building block used in multitude of differentially private data analysis algorithms. In this paper we revisit the Gaussian mechanism and show that the original analysis has several important limitations. Our analysis reveals that the variance formula for the original mechanism is far from tight in the high privacy regime ($\varepsilon \to 0$) and it cannot be extended to the low privacy regime ($\varepsilon \to \infty$). We address these limitations by developing an optimal Gaussian mechanism whose variance is calibrated directly using the Gaussian cumulative density function instead of a tail bound approximation. We also propose to equip the Gaussian mechanism with a post-processing step based on adaptive estimation techniques by leveraging that the distribution of the perturbation is known. Our experiments show that analytical calibration removes at least a third of the variance of the noise compared to the classical Gaussian mechanism, and that denoising dramatically improves the accuracy of the Gaussian mechanism in the high-dimensional regime.

1. Introduction

Output perturbation is a cornerstone of mechanism design in differential privacy (DP). Well-known mechanisms in this class are the Laplace and Gaussian mechanisms (Dwork et al., 2006; Dwork & Roth, 2014). More complex mechanisms are often obtained by composing multiple applications of these basic output perturbation mechanisms. For example, the Laplace mechanism is the basic building block of the sparse vector mechanism (Dwork et al., 2009), and the Gaussian mechanism is the building block of private empirical risk minimization algorithms based on stochastic gradient descent (Bassily et al., 2014). Analysing the privacy of such complex mechanisms turns out to be a delicate and error-prone task (Lyu et al., 2017). In particular, obtaining tight privacy analyses leading to optimal utility is one of the main challenges in the design of advanced DP mechanisms. An alternative to tight *a-priori* analyses is to equip complex mechanisms with *algorithmic* noise calibration and accounting methods. These methods use numerical computations to, e.g. calibrate perturbations and compute cumulative privacy losses at run time, without relying on hand-crafted worst-case bounds. For example, recent works have proposed methods to account for the privacy loss under compositions occurring in complex mechanisms (Rogers et al., 2016; Abadi et al., 2016).

In this work we revisit the Gaussian mechanism and develop two ideas to improve the utility of output perturbation DP mechanisms based on Gaussian noise. The first improvement is an algorithmic noise calibration strategy that uses numerical evaluations of the Gaussian cumulative density function (CDF) to obtain the optimal variance to achieve DP using Gaussian perturbation. The analysis and the resulting algorithm are provided in Section 3. In order to motivate the need for a numerical approach to calibrate the noise of a DP Gaussian perturbation mechanism, we start with an analysis of the main limitations of the classical Gaussian mechanism in Section 2. A numerical evaluation provided in Section 5.1 showcases the advantages of our optimal calibration procedure.

The second improvement equips the Gaussian perturbation mechanism with a post-processing step which denoises the output using adaptive estimation techniques from the statistics literature. Since DP is preserved by post-processing and the distribution of the perturbation added to the desired outcome is known, this allows a mechanism to achieve the desired privacy guarantee while increasing the accuracy of the released value. The relevant denoising estimators and their utility guarantees are discussed in Section 4. Results presented in this section are not new: they are the product of a century's worth of research in statistical estimation. Our contribution is to compile relevant results scattered

[1]Amazon Research, Cambridge, UK [2]Amazon Web Services, Palo Alto, USA [3]University of California, Santa Barbara, USA. Correspondence to: Borja Balle <pigem@amazon.co.uk>, Yu-Xiang Wang <yuxiangw@amazon.com>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

throughout the literature in a single place and showcase their practical impact in synthetic (Section 5.2) and real (Section 5.3) datasets, thus providing useful pointers and guidelines for practitioners.

2. Limitations of the Classical Gaussian Mechanism

Let $\mathbb{X}$ be an input space equipped with a symmetric neighbouring relation $x \simeq x'$. Let $\varepsilon \geq 0$ and $\delta \in [0,1]$ be two privacy parameters. A $\mathbb{Y}$-valued randomized algorithm $M : \mathbb{X} \to \mathbb{Y}$ is (ε, δ)-DP (Dwork et al., 2006) if for every pair of neighbouring inputs $x \simeq x'$ and every possible (measurable) output set $E \subseteq \mathbb{Y}$ the following inequality holds:

$$\mathbb{P}[M(x) \in E] \leq e^{\varepsilon}\mathbb{P}[M(x') \in E] + \delta \ . \tag{1}$$

The definition of DP captures the intuition that a computation on private data will not reveal sensitive information about individuals in a dataset if removing or replacing an individual in the dataset has a negligible effect in the output distribution.

In this paper we focus on the family of so-called output perturbation DP mechanisms. An output perturbation mechanism M for a deterministic vector-valued computation $f : \mathbb{X} \to \mathbb{R}^d$ is obtained by computing the function f on the input data x and then adding random noise sampled from a random variable Z to the output. The amount of noise required to ensure the mechanism $M(x) = f(x) + Z$ satisfies a given privacy guarantee typically depends on how sensitive the function f is to changes in the input and the specific distribution chosen for Z. The Gaussian mechanism gives a way to calibrate a zero mean isotropic Gaussian perturbation $Z \sim \mathcal{N}(0, \sigma^2 I)$ to the global L_2 sensitivity $\Delta = \sup_{x \simeq x'} \|f(x) - f(x')\|$ of f as follows.

Theorem 1 (Classical Gaussian Mechanism)**.** *For any $\varepsilon, \delta \in (0,1)$, the Gaussian output perturbation mechanism with $\sigma = \Delta\sqrt{2\log(1.25/\delta)}/\varepsilon$ is (ε, δ)-DP.*

A natural question one can ask about this result is whether this value of σ provides the minimal amount of noise required to obtain (ε, δ)-DP with Gaussian perturbations. Another natural question is what happens in the case $\varepsilon \geq 1$. This section addresses both these questions. First we show that the value of σ given in Theorem 1 is suboptimal in the high privacy regime $\varepsilon \to 0$. Then we show that this problem is in fact inherent to the usual proof strategy used to analyze the Gaussian mechanism. We conclude the section by showing that for large values of ε the standard deviation of a Gaussian perturbation that provides (ε, δ)-DP must scale like $\Omega(1/\sqrt{\varepsilon})$. This implies that the scaling $\Theta(1/\varepsilon)$ provided by the classical Gaussian mechanism in the range $\varepsilon \in (0,1)$ cannot be extended beyond any bounded interval.

2.1. Limitations in the High Privacy Regime

To illustrate the sub-optimality of the classical Gaussian mechanism in the regime $\varepsilon \to 0$ we start by showing it is possible to achieve $(0, \delta)$-DP using Gaussian perturbations. This clearly falls outside the capabilities of the classical Gaussian mechanism, since the standard deviation $\sigma = \Theta(1/\varepsilon)$ provided by Theorem 1 grows to infinity as $\varepsilon \to 0$.

Theorem 2. *A Gaussian output perturbation mechanism with $\sigma = \Delta/2\delta$ is $(0, \delta)$-DP*[1].

Previous analyses of the Gaussian mechanism are based on a simple sufficient condition for DP in terms of the privacy loss random variable (Dwork & Roth, 2014). The next section explains why the usual analysis of the Gaussian mechanism cannot yield tight bounds for the regime $\varepsilon \to 0$. This shows that our example is not a corner case, but a fundamental limitation of trying to establish (ε, δ)-DP through said sufficient condition.

2.2. Limitations of Privacy Loss Analyses

Given a vector-valued mechanism M let $p_{M(x)}(y)$ denote the density of the random variable $Y = M(x)$. The privacy loss function of M on a pair of neighbouring inputs $x \simeq x'$ is defined as

$$\ell_{M,x,x'}(y) = \log\left(\frac{p_{M(x)}(y)}{p_{M(x')}(y)}\right) \ .$$

The privacy loss random variable $L_{M,x,x'} = \ell_{M,x,x'}(Y)$ is the transformation of the output random variable $Y = M(x)$ by the function $\ell_{M,x,x'}$. For the particular case of a Gaussian mechanism $M(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ it is well-known that the privacy loss random variable is also Gaussian (Dwork & Rothblum, 2016).

Lemma 3. *The privacy loss $L_{M,x,x'}$ of a Gaussian output perturbation mechanism follows a distribution $\mathcal{N}(\eta, 2\eta)$ with $\eta = D^2/2\sigma^2$, where $D = \|f(x) - f(x')\|$.*

The privacy analysis of the classical Gaussian mechanism relies on the following sufficient condition: a mechanism M is (ε, δ)-DP if the privacy loss $L_{M,x,x'}$ satisfies

$$\forall x \simeq x' : \mathbb{P}[L_{M,x,x'} \geq \varepsilon] \leq \delta \ . \tag{2}$$

Since Lemma 3 shows the privacy loss $L_{M,x,x'}$ of the Gaussian mechanism is a Gaussian random variable with mean $\|f(x) - f(x')\|^2/2\sigma^2$, we have $\mathbb{P}[L_{M,x,x'} > 0] \geq 1/2$ for any pair of datasets with $f(x) \neq f(x')$. This observation shows that in general it is not possible to use this sufficient condition for (ε, δ)-DP to prove that the Gaussian mechanism achieves $(0, \delta)$-DP for any $\delta < 1/2$. In other words,

[1] Proofs for all results given in the paper are presented in Appendix A.

the sufficient condition is not necessary in the regime $\varepsilon \to 0$. We conclude that an alternative analysis is required in order to improve the dependence on ε in the Gaussian mechanism.

2.3. Limitations in the Low Privacy Regime

The last question we address in this section is whether the order of magnitude $\sigma = \Theta(1/\varepsilon)$ given by Theorem 1 for $\varepsilon \leq 1$ can be extended to privacy parameters of the form $\varepsilon > 1$. We show this is not the case by providing the following lower bound.

Theorem 4. *Let $f : \mathbb{X} \to \mathbb{R}^d$ have global L_2 sensitivity Δ. Suppose $\varepsilon > 0$ and $0 < \delta < 1/2 - e^{-3\varepsilon}/\sqrt{4\pi\varepsilon}$. If the mechanism $M(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ is (ε, δ)-DP, then $\sigma \geq \Delta/\sqrt{2\varepsilon}$.*

Note that as $\varepsilon \to \infty$ the upper bound on δ in Theorem 4 converges to $1/2$. Thus, as ε increases the range of δ's requiring noise of the order $\Omega(1/\sqrt{\varepsilon})$ increases to include all parameters of practical interest. This shows that the rate $\sigma = \Theta(1/\varepsilon)$ provided by the classical Gaussian mechanism cannot be extended beyond the interval $\varepsilon \in (0, 1)$. Note this provides an interesting contrast with the Laplace mechanism, which can achieve ε-DP with standard deviation $\Theta(1/\varepsilon)$ in the low privacy regime.

3. The Analytic Gaussian Mechanism

The limitations of the classical Gaussian mechanism described in the previous section suggest there is room for improvement in the calibration of the variance of a Gaussian perturbation to the corresponding global L_2 sensitivity. Here we present a method for optimal noise calibration for Gaussian perturbations that we call *analytic Gaussian mechanism*. To do so we must address the two sources of slack in the classical analysis: the sufficient condition (2) used to reduce the analysis to finding an upper bound for $\mathbb{P}[\mathcal{N}(\eta, 2\eta) > \varepsilon]$, and the use of a Gaussian tail approximation to obtain such upper bound. We address the first source of slack by showing that the sufficient condition in terms of the privacy loss random variable comes from a relaxation of a necessary and sufficient condition involving two privacy loss random variables. When specialized to the Gaussian mechanism, this condition involves probabilities about Gaussian random variables, which instead of approximating by a tail bound we represent explicitly in terms of the CDF of the standard univariate Gaussian distribution:

$$\Phi(t) = \mathbb{P}[\mathcal{N}(0, 1) \leq t] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-y^2/2} dy \ .$$

Using this point of view, we introduce a calibration strategy for Gaussian perturbations that requires solving a simple optimization problem involving $\Phi(t)$. We discuss how to solve this optimization at the end of this section.

The first step in our analysis is to provide a necessary and sufficient condition for differential privacy in terms of privacy loss random variables. This is captured by the following result.

Theorem 5. *A mechanism $M : \mathbb{X} \to \mathbb{Y}$ is (ε, δ)-DP if and only if the following holds for every $x \simeq x'$:*

$$\mathbb{P}[L_{M,x,x'} \geq \varepsilon] - e^{\varepsilon} \mathbb{P}[L_{M,x',x} \leq -\varepsilon] \leq \delta \ . \tag{3}$$

Note that Theorem 5 immediately implies the sufficient condition given in (2) through the inequality

$$\mathbb{P}[L_{M,x,x'} \geq \varepsilon] - e^{\varepsilon} \mathbb{P}[L_{M,x',x} \leq -\varepsilon] \leq \mathbb{P}[L_{M,x,x'} \geq \varepsilon] \ .$$

Now we can use Lemma 3 to specialize (3) for a Gaussian output perturbation mechanism. The relevant computations are packaged in the following result, where we express the probabilities in (3) in terms of the Gaussian CDF Φ.

Lemma 6. *Suppose $M(x) = f(x) + Z$ is a Gaussian output perturbation mechanism with $Z \sim \mathcal{N}(0, \sigma^2 I)$. For any $x \simeq x'$ let $D = \|f(x) - f(x')\|$. Then the following hold for any $\varepsilon \geq 0$:*

$$\mathbb{P}[L_{M,x,x'} \geq \varepsilon] = \Phi\left(\frac{D}{2\sigma} - \frac{\varepsilon\sigma}{D}\right) \ , \tag{4}$$

$$\mathbb{P}[L_{M,x',x} \leq -\varepsilon] = \Phi\left(-\frac{D}{2\sigma} - \frac{\varepsilon\sigma}{D}\right) \ . \tag{5}$$

This result specializes the left hand side of (3) in terms of the distance $D = \|f(x) - f(x')\|$ between the output means on a pair of neighbouring datasets. To complete the derivation of our analytic Gaussian mechanism we need to ensure that (3) is satisfied for *every* pair $x \simeq x'$. The next lemma shows that this reduces to plugging the global L_2 sensitivity Δ in the place of D in (4) and (5).

Lemma 7. *For any $\varepsilon \geq 0$, the function $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$ defined as follows is monotonically increasing:*

$$h(\eta) = \mathbb{P}[\mathcal{N}(\eta, 2\eta) \geq \varepsilon] - e^{\varepsilon} \mathbb{P}[\mathcal{N}(\eta, 2\eta) \leq -\varepsilon] \ .$$

Now we are ready to state our main result, whose proof follows directly from Theorem 5, Lemma 7, and equations (4) and (5).

Theorem 8 (Analytic Gaussian Mechanism)**.** *Let $f : \mathbb{X} \to \mathbb{R}^d$ be a function with global L_2 sensitivity Δ. For any $\varepsilon \geq 0$ and $\delta \in [0, 1]$, the Gaussian output perturbation mechanism $M(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ is (ε, δ)-DP if and only if*

$$\Phi\left(\frac{\Delta}{2\sigma} - \frac{\varepsilon\sigma}{\Delta}\right) - e^{\varepsilon} \Phi\left(-\frac{\Delta}{2\sigma} - \frac{\varepsilon\sigma}{\Delta}\right) \leq \delta \ . \tag{6}$$

This result shows that in order to obtain an (ε, δ)-DP Gaussian output perturbation mechanism for a function f with

Algorithm 1: Analytic Gaussian Mechanism

Public Inputs: f, Δ, ε, δ
Private Inputs: x
Let $\delta_0 = \Phi(0) - e^{\varepsilon}\Phi(-\sqrt{2\varepsilon})$
if $\delta \geq \delta_0$ **then**
  Define $B_{\varepsilon}^{+}(v) = \Phi(\sqrt{\varepsilon v}) - e^{\varepsilon}\Phi(-\sqrt{\varepsilon(v+2)})$
  Compute $v^* = \sup\{v \in \mathbb{R}_{\geq 0} : B_{\varepsilon}^{+}(v) \leq \delta\}$
  Let $\alpha = \sqrt{1 + v^*/2} - \sqrt{v^*/2}$
else
  Define $B_{\varepsilon}^{-}(u) = \Phi(-\sqrt{\varepsilon u}) - e^{\varepsilon}\Phi(-\sqrt{\varepsilon(u+2)})$
  Compute $u^* = \inf\{u \in \mathbb{R}_{\geq 0} : B_{\varepsilon}^{-}(u) \leq \delta\}$
  Let $\alpha = \sqrt{1 + u^*/2} + \sqrt{u^*/2}$
Let $\sigma = \alpha\Delta/\sqrt{2\varepsilon}$
Return $f(x) + \mathcal{N}(0, \sigma^2 I)$

global L_2 sensitivity Δ it is enough to find a noise variance σ^2 satisfying (6). One could now use upper and lower bounds for the tail of the Gaussian CDF to derive an analytic expression for a parameter σ satisfying this constraint. However, this again leads to a suboptimal result due to the slack in these tail bounds in the non-asymptotic regime. Instead, we propose to find σ using a numerical algorithm by leveraging the fact that the Gaussian CDF can be written as $\Phi(t) = (1 + \mathsf{erf}(t/\sqrt{2}))/2$, where erf is the standard error function. Efficient implementations of this function to very high accuracies are provided by most statistical and numerical software packages. However, this strategy requires some care in order to avoid numerical stability issues around the point where the expression $\Delta/2\sigma - \varepsilon\sigma/\Delta$ in (6) changes sign. Thus, we further massage the left hand side (6) we obtain the implementation of the analytic Gaussian mechanism given in Algorithm 1. The correctness of this implementation is provided by the following result.

Theorem 9. *Let f be a function with global L_2 sensitivity Δ. For any $\varepsilon > 0$ and $\delta \in (0, 1)$, the mechanism described in Algorithm 1 is (ε, δ)-DP.*

Given a numerical oracle for computing $\Phi(t)$ based on the error function it is relatively straightforward to implement a solver for finding the values v^* and u^* needed in Algorithm 1. For example, using the fact that $B_{\varepsilon}^{+}(v)$ is monotonically increasing we see that computing v^* is a root finding problem for which one can use Newton's method since the derivative of $\Phi(t)$ can be computed in closed form using Leibniz's rule. In practice we find that a simple scheme based on binary search initiated from an interval obtained by finding the smallest $k \in \mathbb{N}$ such that $B_{\varepsilon}^{+}(2^k) > \delta$ provides a very efficient and robust way to find v^* up to arbitrary accuracies (the same applies to u^*).

4. Optimal Denoising

Can we improve the performance of analytical Gaussian mechanism even further? The answer is "yes" and "no". We can't because Algorithm 1 is already the exact calibration of the Gaussian noise level to the given privacy budget. But if we consider the problem of designing the best differentially private procedure $M(x)$ that approximates $f(x)$, then there could still be room for improvement.

In this section, we consider a specific class of mechanisms that *denoise* the output of a Gaussian mechanism. Let $\hat{y} \sim \mathcal{N}(f(x), \sigma^2 I)$, we are interested in designing a post-processing function g such that $\tilde{y} = g(\hat{y})$ is closer to $f(x)$ than $\hat{y}$. This class of mechanisms are of particular interest for differential privacy because (1) since differential privacy is preserved by post-processing, releasing a function $\tilde{y} = g(\hat{y})$ of a differentially private output is again differentially private; (2) since information about f and the distribution of the noise are publicly known, this information can be leveraged to design denoising functions.

This is a statistical estimation problem, where $f(x)$ is the underlying parameter and $\hat{y}$ is the data. Since in this case we are adding the noise ourselves, it is possible to use the classical statistical theory on Gaussian models *as is* because the Gaussian assumption is now true by construction. This is however an unusual estimation problem where all we observe is a single data point. Since $\hat{y}$ is the maximum likelihood estimator, if there is no additional information about $f(x)$, we cannot hope to improve the estimation error *uniformly* over all $f(x) \in \mathbb{R}^d$. But there is still something we can do when we consider either of the following assumptions: (A.1) x is drawn from some underlying distribution, thus inducing some distribution on $f(x)$; or, (A.2) $\|f(x)\|_p \leq B$ for some $p, B > 0$, where $\|\cdot\|_p$ is the L_p-norm (or pseudo-norm when $p < 1$).

Optimal Bayesian denoising. Assumption A.1 translates the problem of optimal denoising into a Bayesian estimation problem, where the underlying parameter $f(x)$ has a prior distribution, and the task is to find an estimator that attains the Bayes risk — the minimum of the average estimation error integrated over a prior π, defined as

$$R(\pi) = \min_{g:\mathbb{R}^d \to \mathbb{R}^d} \mathbb{E}\big[\mathbb{E}[\|g(\hat{y}) - f(x)\|^2 | f(x)]\big] \ .$$

For square loss, the Bayes estimator is simply the posterior mean estimator, as the following theorem shows:

Theorem 10. *Let $x \sim \pi$ and assume the induced distribution of $f(x)$ is square integrable. Then the Bayes estimator $\tilde{y}_{\mathrm{Bayes}}$ is given by*

$$\tilde{y}_{\mathrm{Bayes}} = \operatorname*{argmin}_{g:\mathbb{R}^d \to \mathbb{R}^d} \mathbb{E}\left[\|g(\hat{y}) - f(x)\|^2\right] = \mathbb{E}[f(x)|\hat{y}] \ .$$

The proof can be found in any standard statistics textbook (see, e.g., Lehmann & Casella, 2006). One may ask what the corresponding MSE is and how much it improves over the version without post-processing. The answer depends on the prior and the amount of noise added for differential privacy. When $f(x) \sim \mathcal{N}(0, w^2 I)$, the posterior mean estimator can be written analytically into $\tilde{y}_{\text{Bayes}} = (w^2/(w^2+\sigma^2))\hat{y}$, and the corresponding Bayes risk is $\mathbb{E}[\|\tilde{y}_{\text{Bayes}} - f(x)\|^2] = dw^2\sigma^2/(\sigma^2+w^2)$. In other word, we get a factor of $w^2/(w^2+\sigma^2)$ improvement over simply using $\hat{y}$.

In general, there is no analytical form for the posterior mean, but if we can evaluate the density of $f(x)$ or sample from the distribution of x, then we can obtain an arbitrarily good approximation of $\tilde{y}_{\text{Bayes}}$ using Markov Chain Monte Carlo techniques.

Optimal frequentist denoising. Assumption A.2 spells out a minimax estimation problem, where the underlying parameter $f(x)$ is assumed to be within a set $S \subset \mathbb{R}^d$. In particular, we are interested in finding $\tilde{y}_{\text{minimax}}$ that attains the minimax risk

$$R(S) = \min_{g:\mathbb{R}^d \to \mathbb{R}^d} \max_{f(x)\in S} \mathbb{E}\left[\|g(\hat{y}) - f(x)\|^2\right],$$

on L_p balls $S = \mathcal{B}(p, B) = \{y \in \mathbb{R}^d \mid \|y\|_p \le B\}$ of radius B.

A complete characterization of this minimax risk (up to a constant) is given by Birgé & Massart (2001, Proposition 5), who show that in the non-trivial region[2] of the signal to noise ratio B/σ, the ball $S = \mathcal{B}(p, B)$ satisfies

$$R(S) = \Theta\left(B^p \sigma^{2-p}\left(1 + \log\left(\frac{d\sigma^p}{B^p}\right)\right)^{1-p/2}\right) \quad (7)$$

for $0 < p < 2$ and when $p \ge 2$, Donoho et al. (1990) show that

$$R(S) = \Theta\left(\frac{B^2\sigma^2}{\sigma^2 + B^2/d}\right).$$

Deriving exact minimax estimators is challenging and most analyses assume certain asymptotic regimes (see the case for $p = 2$ by Bickel et al. (1981)). Nonetheless, some techniques have been shown to match $R(\mathcal{B}(p, B))$ up to a small constant factor in the finite sample regime (see, e.g., Donoho et al., 1990; Donoho & Johnstone, 1994). This means that we can often improve the square error from $d\sigma^2$ to $R(\mathcal{B}(p, B))$ when we have the additional information that $f(x)$ is in some L_p ball. This could be especially helpful in the high-dimensional case for $p < 2$. For instance if $p = 1$ and $B = \sigma$, then we obtain a risk $\sigma B\sqrt{1 + \log(d\sigma/B)}$, which improves exponentially in d over the $d\sigma^2$ risk of $\hat{y}$. More practically, if $f(x)$ is a sparse histogram with s non-zero elements, then taking $p \to 0$ will result in an error bound on the order of $s\sigma^2(1 + \log(d))$, which is linear in the sparsity s rather than the dimension d.

[2] When $\sqrt{\log d} \le B/\sigma \le c_p d^{1/p}$ for a constant c_p that depends only on p.

Adaptive estimation. What if we do not know the prior parameter w^2, or a right choice of B and p? Can we still come up with estimators that take advantage of these structures? It turns out that this is the problem of designing *adaptive* estimators which sits at the heart of statistical research. An *adaptive* estimator in our case, is one that does not need to know w^2 or a pair of B and p, yet behave nearly as well as Bayes estimator that knows w^2 or the minimax estimator that knows B and p for each parameter regime.

We first give an example of an adaptive Bayes estimator that does not require us to specify a prior, yet can perform almost as well as the optimal Bayes estimator for all isotropic Gaussian prior simultaneously.

Theorem 11 (James-Stein estimator and its adaptivity)**.** *When $d \ge 3$, substituting w^2 in $\tilde{y}_{\text{Bayes}}$ with its maximum likelihood estimate under*

$$f(x) \sim \mathcal{N}(0, w^2 I) \ , \ \hat{y}|f(x) \sim \mathcal{N}(f(x), \sigma^2 I)$$

produces the James-Stein estimator

$$\tilde{y}_{\text{JS}} = \left(1 - \frac{(d-2)\sigma^2}{\|\hat{y}\|^2}\right)\hat{y}.$$

Moreover, it has an MSE

$$\mathbb{E}\left[\|\tilde{y}_{\text{JS}} - f(x)\|^2\right] = d\sigma^2\left(1 - \frac{(d-2)^2}{d^2}\frac{\sigma^2}{w^2+\sigma^2}\right).$$

The James-Stein estimator has the property that it always improves the MLE $\hat{y}$ when $d \ge 3$ (Stein, 1956) and it always achieves a risk that is within a $d^2/(d-2)^2$ multiplicative factor of the Bayes risk of $\tilde{y}_{\text{Bayes}}$ for any w^2.

We now move on to describe a method that is adaptive to B and p in minimax estimation. Quite remarkably, Donoho (1995) shows that choosing $\lambda = \sigma\sqrt{2\log d}$ in the soft-thresholding estimator

$$\tilde{y}_{\text{TH}} = \text{sign}(\hat{y})\max\{0, |\hat{y}| - \lambda\} \quad (8)$$

yields a nearly optimal estimator for every L_p ball.

Theorem 12 (The adaptivity of soft-thresholding, Theorem 4.2 of (Donoho, 1995))**.** *Let $S = \mathcal{B}(p, B)$ for some $p, B > 0$. The soft-thresholding estimator with $\lambda = \sigma\sqrt{2\log d}$ obeys that*

$$\sup_{f(x)\in S} \mathbb{E}\left[\|\tilde{y}_{\text{TH}} - f(x)\|^2\right] \le (2\log d + 1)(\sigma^2 + 2.22R(S)) \ .$$

The result implies that the soft-thresholding estimator is nearly optimal for all balls up to a multiplicative factor of $4.44\log(d)$.

Thanks to the fact that we know the parameter σ exactly, both $\tilde{y}_{\mathrm{JS}}$ and $\tilde{y}_{\mathrm{TH}}$ are now completely free of tuning parameters. Yet, they can achieve remarkable adaptivity that covers a large class of model assumptions and function classes. A relatively small price to pay for such adaptivity is that we might lose a constant (or a $\log(d)$) factor. Whether such adaptivity is worth will vary on a case-by-case basis.

Take the problem of private releasing a histogram of n items in d bins. Theorem 12 and Equation (7) with $p \leq 1$ imply that the soft-thresholding estimator obeys

$$\mathbb{E}\left[\|\tilde{y}_{\mathrm{TH}} - f(x)\|^2\right] = \tilde{O}\left(\min\{s\sigma^2, n^{1/k}\sigma^{2-1/k}\}\right).$$

where s denotes the number of nonzero elements in $f(x)$ and k is the largest power-law exponent greater than 1 such that order statistics $f(x)^{(d-i+1)} \leq ni^{-k}$ for all $i = 1, ..., d$ and $\tilde{O}$ hides logarithmic factors in $d, d\sigma/n$. The fact that $s \leq d$ implies that the soft-thresholding estimator improves over the naive private release for all d, n, s and the $n^{1/k}$ factor suggests that we can take advantage of an unknown power law distribution even if the histogram is not really sparse. This makes our technique effective in the many data mining problems where power law distributions occur naturally (Faloutsos et al., 1999).

Related work. Denoising as a post-processing step in the context of differentially privacy is not a new idea. Notably, Barak et al. (2007); Hay et al. (2009) show that a post-processing step enforcing consistency of contingency table releases and graph degree sequences leads to more accurate estimations. Williams & McSherry (2010) sets up the general statistical (Bayesian) inference problem of DP releases by integrating auxiliary information (a prior). Karwa et al. (2016) exploits knowledge of the noise distribution use to achieve DP in the inference procedure of a network model and shows that it helps to preserve asymptotic efficiency. Nikolov et al. (2013) demonstrates that projecting linear regression solutions to a known ℓ_1-ball improves the estimation error from $O(\mathrm{poly}(d))$ to $O(\mathrm{polylog}(d))$ when the underlying ground truth is sparse. Bernstein et al. (2017) uses Expectation–Maximization to denoise the parameters of a class of graphical models starting from noisy empirical moments obtained using the Laplace mechanism.

In all the above references there is some prior knowledge (constraint sets, sparsity or Bayesian prior) that is exploited to improve the utility of DP releases. To the best of our knowledge, we are the first to consider "adaptive estimation" and demonstrate how classical techniques can be helpful even without such prior knowledge. These estimators are not new; they have been known in the statistics literature for decades. Our purpose is to compile facts that are relevant to the practice of DP and initiate a systematic study of how these ideas affect the utility of DP mechanisms, which we complement with the experimental evaluation presented in the next section.

5. Numerical Experiments

This section provides an experimental evaluation of the improvements in utility provided by optimal calibration and adaptive denoising. First we numerically compare the variance of the analytic Gaussian mechanism and the classical mechanism for a variety of privacy parameters. Then we evaluate the contributions of denoising and analytic calibration against a series of baselines for the task of private mean estimation using synthetic data. We also evaluate several denoising strategies on the task of releasing heat maps based on the New York City taxi dataset under differential privacy. Further experiments are presented in Appendix B, including an evaluation of denoising strategies for the task of private histogram release.

5.1. Analytic Gaussian Mechanism

We implemented Algorithm 1 in Python[3] and ran experiments to compare the variance of the perturbation obtained with the analytic Gaussian mechanism versus the variance required by the classical Gaussian mechanism. In all our experiments the values of v^* and u^* were solved up to an accuracy of 10^{-12} using binary search and the implementation of the erf function provided by SciPy (Jones et al., 2001).

The results are presented in the two leftmost panels in Figure 1. The plots show that as $\varepsilon \to 0$ the optimally calibrated perturbation outperforms the classical mechanism by several orders of magnitude. Furthermore, we see that even for values of ε close to 1 our mechanism reduces the variance by a factor of 1.4 or more, with higher improvements for larger values of δ.

5.2. Denoising for Mean Estimation

Our next experiment evaluates the utility of denoising combined with the analytical Gaussian mechanism for the task of private mean estimation. The input to the mechanism is a dataset $x = (x_1, \ldots, x_n)$ containing n vectors $x_i \in \mathbb{R}^d$ and the underlying deterministic functionality is $f(x) = (1/n)\sum_{i=1}^n x_i$. This relatively simple task is a classic example from the family of *linear queries* which are frequently considered in the differential privacy literature. We compare the accuracy of several mechanisms M for releasing a private version of $f(x)$ in terms of the Euclidean

[3] See https://github.com/BorjaBalle/analytic-gaussian-mechanism.

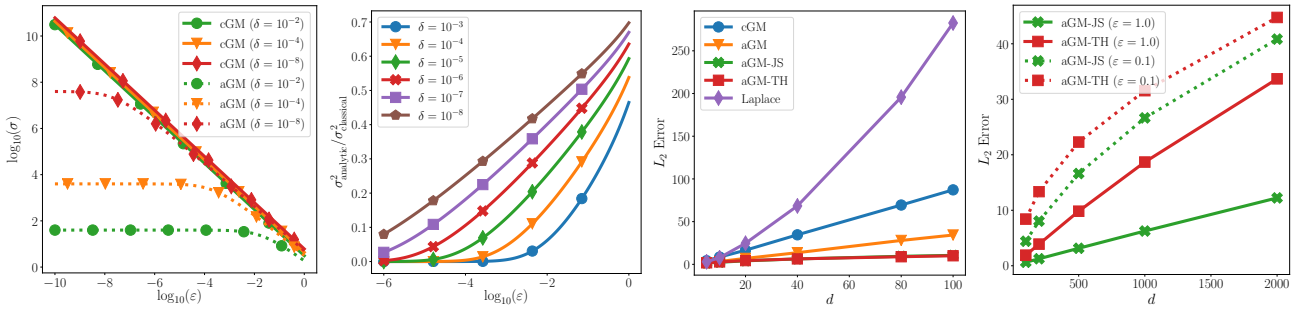


Figure 1. Two leftmost plots: Experiments comparing the classical Gaussian mechanism (cGM) and the analytic Gaussian mechanism (aGM), in terms of their absolute standard deviations as $\varepsilon \to 0$, and in terms of the gain in variance as a function of ε. Two rightmost plots: Mean estimation experiments showing L_2 error between the private mean estimate and the non-private empirical mean as a function of the dimension d.

distance $\|M(x) - f(x)\|_2$. In particular, we test the analytical Gaussian mechanism with either James-Stein denoising cf. Theorem 11 (aGM-JS) or optimal thresholding denoising cf. Theorem 12 (aGM-TH), as well as several baselines including: the classical Gaussian mechanism (cGM), the analytical Gaussian mechanism without denoising (aGM), and the Laplace mechanism (Lap) using the same ε parameter as the Gaussian mechanisms.

To provide a thorough comparison we explore of the different parameters of the problem on the final utility. The key parameters of the problem are the dimension d and the DP parameters ε and δ. The dimension affects the utility through the bounds provided in Theorem 11 and Theorem 12. The DP parameters affect the utility through the variance σ^2 of the mechanism, which is also affected by the sample size n via the global sensitivity. Thus, we can characterize the effect of σ^2 by keeping n fixed and changing the DP parameters. In our experiments we consider a fixed sample size $n = 500$ and privacy parameter $\delta = 10^{-4}$ while trying several values for ε.

The other parameter that affects the utility is the "size" of $f(x)$, controlled either through the variance w^2 or the norm ball S. Since the denoising estimators we use are adaptive to these parameters and do not need to know them in advance, we sample the dataset x repeatedly to obtain a diversity of values for $f(x)$. Each dataset x is sampled as follows: first sample a center $x_0 \sim \mathcal{N}(0, I)$ and then build $x = (x_1, \ldots, x_n)$ with $x_i = x_0 + \xi_i$, where each ξ_i is i.i.d. with independent coordinates sampled uniformly from the interval $[-1/2, 1/2]$. Thus, in each dataset the points x_i all lie in an L_∞-ball of radius 1, leading to a global L_2 sensitivity $\Delta_2 = \sqrt{d}/n$ and a global L_1 sensitivity $\Delta_1 = d/n$. These are used to calibrate the Gaussian and Laplace perturbations, respectively.

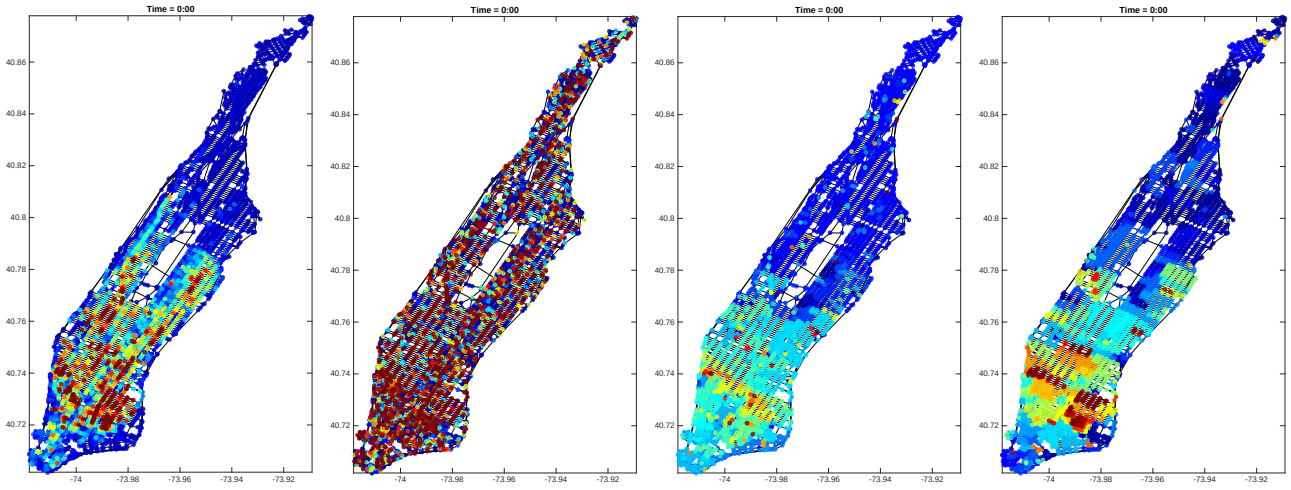
The results are presented in two rightmost panels of Figure 1. Each point in every plot is the result of averaging the error over 100 repetitions with different datasets. The first plot uses $\varepsilon = 0.01$ and shows how denoised methods improve the accuracy over all the other methods, sometimes by orders of magnitude. The second plot shows that for this problem the James-Stein estimator provides better accuracy in the high-dimensional setting.

5.3. New York City Taxi Heat Maps

In this section, we apply our method to New York City taxi data. The dataset is a collection of time-stamped pick-ups and drop-offs of taxi drivers and we are interested in sharing a density map of such pick-ups and drop-offs in Manhattan at a specific time of a specific day under differential privacy.

This is a problem of significant practical interest. Ever since the NYC Taxi & Limousine Commission released this dataset, there has been multiple independent reports concerning the security and privacy risks this dataset poses for taxi drivers and their passengers (see, e.g., Pandurangan; Douriez et al., 2016). The techniques presented in this paper allow us to provably prevent individuals (on both the per-trip level and per-cab level) in the dataset from being identified, while remarkably, permitting the release of rich information about the data with fine-grained spatial and temporal resolution.

Specifically, we apply the analytical Gaussian mechanism to release the number of picks-ups and drop-offs at every traffic junction in Manhattan. There are a total of 3,784 such traffic junctions and they are connected by 7,070 sections of roads. We will treat them as nodes and edges on a graph. In the post-processing phase, we apply graph smoothing techniques to reveal the underlying signal despite the noise due to aGM. Specifically, we compare the JS-estimator and the soft-thresholding estimator we described in Section 4, as well as the same soft-thresholding estimator applied to the coefficients of a graph wavelet transform due to Sharpnack et al. (2013). The basis transformation is important be-



Ground truth | Raw aGM RMSE = 40.29 | Wavelets RMSE = 12.28 | Trend filtering RMSE = 10.07

Figure 2. Illustration of the denoising in differentially private release of NYC taxi density during 00:00 - 01:00 am Sept 24, 2014. From left to right, the figures represent the true data, the output of the analytical Gaussian mechanism, the reconstructed signal from soft-thresholded wavelet coefficients with spanning-tree wavelet transform (Sharpnack et al., 2013), and the results of trend filtering on graphs (Wang et al., 2016). We observe that adding appropriate post-processing significantly reduces the estimation error and also makes the underlying structures visible.

cause the data might be sparser in the transformed domain. For reference, we also include the state-of-the-art graph smoothing techniques called graph trend filtering (Wang et al., 2016), which has one additional tuning parameter but has been shown to perform significantly better than wavelet smoothing in practice.

Our experiments provide cab-level differential privacy by assuming that every driver does a maximum of 5 trips within an hour so that we have a global L_2-sensitivity of $\Delta = 5$. This is a conservative but reasonable estimate and can be enforced by preprocessing the data. Data within each hour is gathered and distributed to each traffic junction using a kernel density estimator; further details are documented in Doraiswamy et al. (2014).

We present some qualitative comparisons in Figure 2, where we visualize the privately released heat map with and without post-processing. Relatively speaking, trend filtering performs better than wavelet smoothing, but both approaches significantly improves the RMSE over the DP release without post-processing. The results in Appendix B provide quantitative results by comparing the mean square error of cGM, aGM as well as the aforementioned denoising techniques for data corresponding to different time intervals.

6. Conclusion and Discussion

In this paper, we embark on a journey of pushing the utility limit of Gaussian mechanism for (ε, δ)-differential privacy. We propose a novel method to obtain the optimal calibration of Gaussian perturbations required to attain a given DP guarantee. We also review decades of research in statistical estimation theory and show that combining these techniques with differential privacy one obtains powerful *adaptivity* that denoises differentially private outputs nearly optimally without additional hyperparameters. On synthetic data and on the New York City Taxi dataset we illustrate a significant gain in estimation error and fine-grained spatial-temporal resolution.

There are a number of theoretical problems of interest for future work. First, on the problem of differentially private estimation. Our post-processing approach effectively restricts our choice of algorithms to the composition of privacy release and post-processing. While we now know that we are optimal in both components, it is unclear whether we lose anything relative to the best differentially private algorithms. Secondly, the analytical calibration proposed in this paper is optimal for achieving (ε, δ)-DP with Gaussian noise. But when building complex mechanisms we are stuck in the dilemma of choosing between (a) using the aGM with the advanced composition (Kairouz et al., 2015); or, (b) using Rényi DP (Mironov, 2017) or zCDP (Bun & Steinke, 2016) for tighter composition and calculate the (ε, δ) from moment bounds. While (a) is tighter in the calculation the privacy parameters of each intermediate value, (b) is tighter in the composition but cannot take advantage of aGM. It would be interesting if we could get the best of both worlds.

Acknowledgments

We thank Doraiswamy et al. (2014) for sharing their pre-processed NYC taxi dataset, the anonymous reviewers for helpful comments that led to improvements of the paper and Stephen E. Fienberg for discussions that inspired the authors to think about optimal post-processing.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318. ACM, 2016.

Barak, B., Chaudhuri, K., Dwork, C., Kale, S., McSherry, F., and Talwar, K. 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*, pp. 273–282. ACM, 2007.

Bassily, R., Smith, A., and Thakurta, A. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pp. 464–473. IEEE, 2014.

Bernstein, G., McKenna, R., Sun, T., Sheldon, D., Hay, M., and Miklau, G. Differentially private learning of undirected graphical models using collective graphical models. In *International Conference on Machine Learning (ICML)*, 2017.

Bickel, P. et al. Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *The Annals of Statistics*, 9(6):1301–1309, 1981.

Birgé, L. and Massart, P. Gaussian model selection. *Journal of the European Mathematical Society*, 3(3):203–268, 2001.

Bun, M. and Steinke, T. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pp. 635–658. Springer, 2016.

Donoho, D. L. De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3):613–627, 1995.

Donoho, D. L. and Johnstone, I. M. Minimax risk over p-balls for p-error. *Probability Theory and Related Fields*, 99(2):277–303, 1994.

Donoho, D. L., Liu, R. C., and MacGibbon, B. Minimax risk over hyperrectangles, and implications. *The Annals of Statistics*, pp. 1416–1437, 1990.

Doraiswamy, H., Ferreira, N., Damoulas, T., Freire, J., and Silva, C. T. Using topological analysis to support event-guided exploration in urban data. *IEEE transactions on visualization and computer graphics*, 20(12):2634–2643, 2014.

Douriez, M., Doraiswamy, H., Freire, J., and Silva, C. T. Anonymizing nyc taxi data: Does it matter? In *Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference on*, pp. 140–148. IEEE, 2016.

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

Dwork, C. and Rothblum, G. N. Concentrated differential privacy. *arXiv preprint arXiv:1603.01887*, 2016.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography*, pp. 265–284. Springer, 2006.

Dwork, C., Naor, M., Reingold, O., Rothblum, G. N., and Vadhan, S. On the complexity of differentially private data release: efficient algorithms and hardness results. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pp. 381–390. ACM, 2009.

Faloutsos, M., Faloutsos, P., and Faloutsos, C. On power-law relationships of the internet topology. In *ACM SIGCOMM computer communication review*, volume 29, pp. 251–262. ACM, 1999.

Gordon, R. D. Values of mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *The Annals of Mathematical Statistics*, 12(3):364–366, 1941.

Hay, M., Li, C., Miklau, G., and Jensen, D. Accurate estimation of the degree distribution of private networks. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*, pp. 169–178. IEEE, 2009.

Jones, E., Oliphant, T., Peterson, P., et al. SciPy: Open source scientific tools for Python, 2001. URL `http://www.scipy.org/`.

Kairouz, P., Oh, S., and Viswanath, P. The composition theorem for differential privacy. In *International Conference on Machine Learning (ICML)*, 2015.

Karwa, V., Slavković, A., et al. Inference using noisy degrees: Differentially private β-model and synthetic graphs. *The Annals of Statistics*, 44(1):87–112, 2016.

Lehmann, E. L. and Casella, G. *Theory of point estimation*. Springer Science & Business Media, 2006.

Lyu, M., Su, D., and Li, N. Understanding the sparse vector technique for differential privacy. *Proceedings of the VLDB Endowment*, 10(6):637–648, 2017.

Mironov, I. Renyi differential privacy. In *Computer Security Foundations Symposium (CSF), 2017 IEEE 30th*, pp. 263–275. IEEE, 2017.

Nikolov, A., Talwar, K., and Zhang, L. The geometry of differential privacy: the sparse and approximate cases. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pp. 351–360. ACM, 2013.

Pandurangan, V. On taxi and rainbows. https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1. Accessed: 2014-06-21.

Rogers, R. M., Roth, A., Ullman, J., and Vadhan, S. Privacy odometers and filters: Pay-as-you-go composition. In *Advances in Neural Information Processing Systems*, pp. 1921–1929, 2016.

Sharpnack, J., Singh, A., and Krishnamurthy, A. Detecting activations over graphs using spanning tree wavelet bases. In *Artificial Intelligence and Statistics*, pp. 536–544, 2013.

Stein, C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*, volume 1, pp. 197–206, 1956.

Telgarsky, M. Dirichlet draws are sparse with high probability. *CoRR*, abs/1301.4917, 2013.

Wang, Y.-X., Sharpnack, J., Smola, A. J., and Tibshirani, R. J. Trend filtering on graphs. *Journal of Machine Learning Research*, 17(105):1–41, 2016.

Williams, O. and McSherry, F. Probabilistic inference and differential privacy. In *Advances in Neural Information Processing Systems*, pp. 2451–2459, 2010.