A. Proofs

In this appendix we present supporting proofs for all the results mentioned in the main text.

A.1. Proofs from Section 2

Proof of Theorem 2. An simple way to see that $(0, \delta)$-DP is achievable with Gaussian noise is to recall that $(0, \delta)$-DP is equivalent to a bound of δ on the total variation (TV) distance between the output distributions of $M(x)$ and $M(x')$ for any neighbouring pair $x \simeq x'$. If $M(x)$ is an output perturbation mechanism for $f(x)$ with noise $Z \sim \mathcal{N}(0, \sigma^2 I)$, then using Pinsker's inequality we have

$$\begin{aligned}\mathsf{TV}(M(x), M(x')) &\leq \sqrt{\frac{\mathsf{KL}(M(x)|M(x'))}{2}} \\ &= \sqrt{\frac{\mathsf{KL}(\mathcal{N}(f(x), \sigma^2 I)|\mathcal{N}(f(x'), \sigma^2 I))}{2}} \\ &= \frac{\|f(x) - f(x')\|}{2\sigma} \leq \frac{\Delta}{2\sigma} \ .\end{aligned}$$

Thus, we see that a Gaussian perturbation with standard deviation $\sigma = \Delta/2\delta$ is enough to achieve $(0, \delta)$-DP. □

Proof of Theorem 4. Note that the proof of Theorem 9 shows that a Gaussian perturbation with $\sigma = \Delta/\sqrt{2\varepsilon}$ yields a $(\varepsilon, \delta_0(\varepsilon))$-DP mechanism, where $\delta_0(\varepsilon) = \Phi(0) - e^{\varepsilon}\Phi(-\sqrt{2\varepsilon})$. Thus, it is not possible to attain (ε, δ)-DP with $\delta < \delta_0(\varepsilon)$ without increasing the variance of the perturbation.

The result follows by showing that the upper bound for δ proposed in Theorem 4 is a lower bound for $\delta_0(\varepsilon)$. Since $\Phi(0) = 1/2$, all we need to show is $e^{\varepsilon}\Phi(-\sqrt{2\varepsilon}) < \frac{e^{-3\varepsilon}}{\sqrt{4\pi\varepsilon}}$.

Let $\Phi^c(t) = \mathbb{P}[\mathcal{N}(0, 1) \geq t] = 1 - \Phi(t)$ be the complementary of the standard Gaussian CDF. The Mill's ratio for the Gaussian distribution is the quantity $r(t) = \sqrt{2\pi}e^{t^2/2}\Phi^c(t)$. Bounding the Mill's ratio is a standard approach to approximate the tail of the Gaussian distribution. A well-known bound for the Mill's ratio is Gordon's inequality $r(t) < 1/t$ (Gordon, 1941). By using the symmetry $\Phi(-t) = \Phi^c(t)$ we obtain :

$$e^{\varepsilon}\Phi(-\sqrt{2\varepsilon}) = e^{\varepsilon}\Phi^c(\sqrt{2\varepsilon}) = \frac{e^{-3\varepsilon}}{\sqrt{2\pi}} r(\sqrt{2\varepsilon}) < \frac{e^{-3\varepsilon}}{\sqrt{4\pi\varepsilon}} \ .$$

□

Proof of Lemma 3. Recall that the density of the Gaussian output perturbation mechanism $M(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ is given by $p_{M(x)}(y) = \exp(-\|y - f(x)\|^2/2\sigma^2)/\sqrt{2\pi\sigma^2}$. Plugging this expression into the definition of the privacy loss function and performing a quick computation we get

$$\begin{aligned}\ell_{M,x,x'}(y) &= \frac{\|y - f(x')\|^2 - \|y - f(x)\|^2}{2\sigma^2} \\ &= \frac{\|f(x) - f(x')\|^2}{2\sigma^2} + \frac{\langle y - f(x), f(x) - f(x')\rangle}{\sigma^2} \ .\end{aligned}$$

To compute the privacy loss random variable $L_{M,x,x'}$ we need to plug $Y = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ in the above inner product. By observing that $\langle Z, f(x) - f(x')\rangle \sim \mathcal{N}(0, \sigma^2\|f(x) - f(x')\|^2)$ we obtain the distribution of the privacy loss random variable is given by

$$L_{M,x,x'} \sim \mathcal{N}\left(\frac{\|f(x) - f(x')\|^2}{2\sigma^2}, \frac{\|f(x) - f(x')\|^2}{\sigma^2}\right) \ .$$

Therefore, the privacy loss of the Gaussian mechanism has the form $\mathcal{N}(\eta, 2\eta)$ for $\eta = D^2/2\sigma^2$. □

A.2. Proofs from Section 3

Proof of Theorem 5. Given a pair of neighbouring datasets $x \simeq x'$ let $p = p_{M(x)}$ and $p' = p_{M(x')}$ be the densities of the output random variables $Y = M(x)$ and $Y' = M(x')$. Note that given an event $E \subseteq \mathbb{Y}$ one can rewrite (1) as follows:

$$\int_E (p(y) - e^{\varepsilon}p'(y)) \leq \delta \ . \tag{9}$$

Defining the event $E_* = \{y : p(y) \geq e^{\varepsilon}p'(y)\}$ and its complementary $\bar{E}_* = \mathbb{Y} \setminus E_*$, we can partition E into the sets $E_+ = E \cap E_*$ and $E_- = E \cap \bar{E}_*$. Therefore, by the definition of E_* we have

$$\begin{aligned}\int_E (p(y) - e^{\varepsilon}p'(y)) &= \int_{E_+} (p(y) - e^{\varepsilon}p'(y)) \\ &\quad + \int_{E_-} (p(y) - e^{\varepsilon}p'(y)) \\ &\leq \int_{E_+} (p(y) - e^{\varepsilon}p'(y)) \\ &\leq \int_{E_*} (p(y) - e^{\varepsilon}p'(y)) \ .\end{aligned}$$

Because (9) has to hold for any event E and the upper bound above holds for any event, we conclude that M is (ε, δ)-DP if and only if

$$\int_{E_*} (p(y) - e^{\varepsilon}p'(y)) \leq \delta \tag{10}$$

holds for any $x \simeq x'$. To complete the proof we need to show that (10) is equivalent to (3). Expanding the definition of $L_{M,x,x'}$ we get:

$$\begin{aligned}\mathbb{P}[L_{M,x,x'} \geq \varepsilon] &= \mathbb{P}[\log(p(Y)/p'(Y)) \geq \varepsilon] \\ &= \mathbb{P}[p(Y) \geq e^{\varepsilon} p'(Y)] \\ &= \int_{\mathbb{Y}} \mathbf{1}[p(y) \geq e^{\varepsilon} p'(y)] p(y) \\ &= \int_{E_*} p(y) \ .\end{aligned}$$

A similar argument with $L_{M,x',x}$ also shows:

$$\mathbb{P}[L_{M,x',x} \leq -\varepsilon] = \int_{E_*} p'(y) \ .$$

Putting the last two equations together we obtain see that the left hand side of (3) equals the left hand side of (10). □

Proof of Lemma 6. Note that Lemma 3 shows that the privacy loss random variables $L_{M,x,x'}$ and $L_{M,x'x}$ both follow the same distribution $\mathcal{N}(\eta, 2\eta)$ with $\eta = D^2/2\sigma^2$. This allows us to write the left hand side of (4) in terms of the Gaussian CDF Φ as follows:

$$\begin{aligned}\mathbb{P}[L_{M,x,x'} \geq \varepsilon] &= \mathbb{P}[\mathcal{N}(\eta, 2\eta) \geq \varepsilon] \\ &= \mathbb{P}\left[\mathcal{N}(0,1) \geq \frac{-\eta + \varepsilon}{\sqrt{2\eta}}\right] \\ &= \mathbb{P}\left[\mathcal{N}(0,1) \leq \frac{\eta - \varepsilon}{\sqrt{2\eta}}\right] \\ &= \Phi\left(\sqrt{\frac{\eta}{2}} - \frac{\varepsilon}{\sqrt{2\eta}}\right) \\ &= \Phi\left(\frac{D}{2\sigma} - \frac{\varepsilon\sigma}{D}\right) \ ,\end{aligned}$$

where we used $\mathcal{N}(\eta, 2\eta) = \eta + \mathcal{N}(0,1)/\sqrt{2\eta}$ and the symmetry $\mathbb{P}[\mathcal{N}(0,1) \geq t] = \mathbb{P}[\mathcal{N}(0,1) \leq -t]$ of the distribution of the Gaussian distribution around its mean. A similar argument applied to the left hand side of (5) yields:

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

□

Proof of Lemma 7. We prove the result by using Leibniz's rule for differentiation under the integral sign to show that the function of interest has non-negative derivatives. First note that from the derivation of (4) we have

$$\mathbb{P}[\mathcal{N}(\eta, 2\eta) \geq \varepsilon] = \Phi(a(\eta)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a(\eta)} e^{-y^2/2} dy \ ,$$

where $a(\eta) = \sqrt{\eta/2} - \varepsilon/\sqrt{2\eta}$. Now we can use Leibniz's rule to write

$$\begin{aligned}\frac{d}{d\eta} \int_{-\infty}^{a(\eta)} e^{-y^2/2} dy &= e^{-a(\eta)^2/2} a'(\eta) \\ &= e^{-a(\eta)^2/2} \left(\frac{1}{\sqrt{8\eta}} + \frac{\varepsilon}{\sqrt{8\eta^3}}\right) \ .\end{aligned}$$

Similarly, for the second term in the function we have $\mathbb{P}[\mathcal{N}(\eta, 2\eta) \leq -\varepsilon] = \Phi(b(\eta))$ where $b(\eta) = -\sqrt{\eta/2} - \varepsilon/\sqrt{2\eta}$. Using Leibniz's rule again we get

$$\frac{d}{d\eta} \int_{-\infty}^{b(\eta)} e^{-y^2/2} dy = e^{-b(\eta)^2/2} \left(-\frac{1}{\sqrt{8\eta}} + \frac{\varepsilon}{\sqrt{8\eta^3}}\right) \ .$$

Therefore, we see that the derivative of h satisfies:

$$\begin{aligned}h'(\eta) &= \frac{1}{4\sqrt{\pi\eta}} \left(e^{-a(\eta)^2/2} + e^{\varepsilon} e^{-b(\eta)^2/2}\right) \\ &\quad + \frac{\varepsilon}{4\sqrt{\pi\eta^3}} \left(e^{-a(\eta)^2/2} - e^{\varepsilon} e^{-b(\eta)^2/2}\right) \\ &= \frac{1}{4\sqrt{\pi\eta}} \left(e^{-a(\eta)^2/2} + e^{\varepsilon} e^{-b(\eta)^2/2}\right) \geq 0 \ ,\end{aligned}$$

where we used that $a(\eta)^2 + 2\varepsilon = b(\eta)^2$. □

Proof of Theorem 9. Recall that the derivations in Section 3 establish that in order to calibrate a Gaussian perturbation to achieve (ε, δ)-DP all that is required is find the smallest σ such that

$$\Phi\left(\frac{\Delta}{2\sigma} - \frac{\varepsilon\sigma}{\Delta}\right) - e^{\varepsilon} \Phi\left(-\frac{\Delta}{2\sigma} - \frac{\varepsilon\sigma}{\Delta}\right) \leq \delta \ . \tag{11}$$

To establish the correctness of the analytic Gaussian mechanism we begin by observing that the argument in the first term of (11) changes sign at $\sigma = \Delta/\sqrt{2\varepsilon}$, while the argument for the second terms is always negative. Thus, we substitute $\sigma = \alpha\Delta/\sqrt{2\varepsilon}$ in the expression above and obtain:

$$B_{\varepsilon}(\alpha) = \Phi\left(\sqrt{\frac{\varepsilon}{2}}\left(\frac{1}{\alpha} - \alpha\right)\right) - e^{\varepsilon} \Phi\left(-\sqrt{\frac{\varepsilon}{2}}\left(\frac{1}{\alpha} + \alpha\right)\right) \ .$$

To solve the optimization $\inf\{\alpha > 0 : B_{\varepsilon}(\alpha) \leq \delta\}$ using numerical evaluations of Φ it is convenient to consider the cases $\alpha \geq 1$ and $\alpha < 1$ separately. In the case $\alpha \geq 1$ we define $u = (\alpha - 1/\alpha)^2/2$ and substitute the corresponding α in B_{ε} to obtain

$$B_{\varepsilon}^{-}(u) = \Phi(-\sqrt{\varepsilon u}) - e^{\varepsilon} \Phi(-\sqrt{\varepsilon(u+2)}) \ .$$

Similarly, by taking $v = (1/\alpha - \alpha)^2/2$ in the case $\alpha < 1$ we obtain

$$B_{\varepsilon}^{+}(v) = \Phi(\sqrt{\varepsilon v}) - e^{\varepsilon} \Phi(-\sqrt{\varepsilon(v+2)}) \ .$$

Note that, as expected, these definitions satisfy $\lim_{v\to\infty} B_\varepsilon^+(v) = 1$ and $\lim_{u\to\infty} B_\varepsilon^-(u) = 0$, since the limits correspond to $\lim_{\alpha\to 0} B(\alpha) = 1$ and $\lim_{\alpha\to\infty} B(\alpha) = 0$, respectively. Furthermore, we have

$$B_\varepsilon^+(0) = B_\varepsilon^-(0) = \Phi(0) - e^\varepsilon \Phi(-\sqrt{2\varepsilon}) = \delta_0(\varepsilon) \ ,$$

which corresponds to the privacy guarantee $(\varepsilon, \delta_0(\varepsilon))$-DP obtained by taking $\sigma = \Delta/\sqrt{2\varepsilon}$; i.e. $\alpha = 1$.

These observations motivate the mechanism described in Algorithm 1. In particular, for $\delta \geq \delta_0(\varepsilon)$ we can achieve (ε, δ)-DP with $\alpha < 1$, and the smallest $\alpha < 1$ such that $B_\varepsilon(\alpha) \leq \delta$ corresponds to the largest $v \geq 0$ such that $B_\varepsilon^+(v) \leq \delta$. Similarly, for $\delta < \delta_0(\varepsilon)$ we require $\alpha \geq 1$, and the smallest $\alpha \geq 1$ such that $B_\varepsilon(\alpha) \leq \delta$ corresponds to the smallest $u \geq 0$ such that $B_\varepsilon^-(u) \leq \delta$. □

A.3. Proofs from Section 4

The proofs in this section are well-known and not part of the contribution of the current paper. We include these proofs because they are short and revealing and we hope to be self-contained as much as possible.

Proof of Theorem 10. Let P be the distribution of $f(x)$ induced by $x \sim \pi$. Let $\theta \in \mathbb{R}^d$, define its posterior error

$$r(\theta|\hat{y}) = \int \|\theta - f(x)\|^2 dP(f(x)|\hat{y}).$$

Take the gradient with respect to θ on both sides and apply Fubini's theorem

$$\frac{\partial}{\partial\theta} r(\theta|\hat{y}) = \frac{\partial}{\partial\theta} \int \|\theta - f(x)\|^2 dP(f(x)|\hat{y})$$
$$= 2\theta - 2\int f(x) dP(f(x)|\hat{y}) = 2(\theta - \mathbb{E}[f(x)|\hat{y}]).$$

Assign the gradient to 0 we get that the minimizer is $\mathbb{E}[f(x)|\hat{y}]$.

Now, assume $\tilde{y}_{\text{Bayes}}$ is suboptimal, there exists $\tilde{y}^* \neq \tilde{y}_{\text{Bayes}}$ such that

$$\mathbb{E}\left[\|f(x) - \tilde{y}_{\text{Bayes}}\|^2\right] > \mathbb{E}\left[\|f(x) - \tilde{y}^*\|^2\right]$$
$$= \mathbb{E}[r(\tilde{y}^*|\hat{y})] \geq \mathbb{E}[r(\tilde{y}_{\text{Bayes}})] = \mathbb{E}\left[\|f(x) - \tilde{y}_{\text{Bayes}}\|^2\right].$$

which is a contradiction. □

Proof of Theorem 11. Note that $\frac{\|\hat{y}\|^2}{w^2+\sigma^2}$ follows a χ^2 distribution with degree of freedom d. The likelihood function

$$p(\|\hat{y}\|^2|w^2) \propto (\frac{\|\hat{y}\|^2}{w^2+\sigma^2})^{d/2-1} e^{-\frac{\|\hat{y}\|^2}{2(w^2+\sigma^2)}}.$$

The gradient w.r.t. w^2 of the log-likelihood, we get

$$-\frac{d/2-1}{w^2+\sigma^2} + \frac{\|y\|^2}{2(w^2+\sigma^2)^2}.$$

Assigning it to 0, we get the maximum likelihood estimate $w^2 = \frac{\|y\|^2}{k-2} - \sigma^2$. Substituting it into $\tilde{y}_{\text{Bayes}} = (w^2/(w^2 + \sigma^2))\hat{y}$ produces $\tilde{y}_{\text{JS}}$ as stated and the calculation of its MSE is straightforward. □

B. Additional Experiments

Here we present additional experimental results. Figure 3 provides more plots for the setups explored in Sections 5.1 and 5.2. The next two sections present further experiments on a sparse histogram denoising task and on the New York City taxi dataset.

B.1. Denoising for Histogram Release

We evaluate the accuracy of our new Gaussian perturbation mechanisms on a second task involving private histogram release. In this problem the dataset $x = (x_1, \ldots, x_n)$ contains elements $x_i \in [d]$ from a finite set with d. The deterministic functionality is the empirical histogram $y = f(x) \in \mathbb{R}^d$ where $y_j = (1/n)\sum_{i=1}^n \mathbb{I}[x_i = j]$. In this case the global L_2 sensitivity $\Delta_2 = \sqrt{2}/n$ and a global L_1 sensitivity $\Delta_1 = 2/n$ (with respect to replacing one individual in a dataset by another arbitrary individual).

For this task, each dataset is sampled from a multinomial distribution with parameters sampled from a symmetric Dirichlet with $\alpha = 1/d$. The parameters are resampled for each individual experiment. The choice of α guarantees that the resulting histograms are highly sparse (Telgarsky, 2013). Our setup follows the same structure as the one for the experiments from previous section. The results are presented in Figure 4. We observe that in this problem the Laplace mechanism is better than the classical Gaussian mechanism, and in the setting $\varepsilon = 1$ it is even better than the analytic Gaussian mechanism, with and without denoising. However, as we decrease ε the utility of the analytic Gaussian mechanism becomes better than that of the Laplace mechanism, and denoising provides a significant advantage over mechanisms without denoising. Finally, we note that due to the sparsity of the underlying datapoint, denoising via soft thresholding provides better utility in this case than denoising via shrinking.

B.2. New York City Taxi Heat Maps

Here we present a second qualitative experiment with the New York City taxi dataset. The difference with the previous experiment is that we use data for a different time of the same day, leading to a different structure in the activities around the city; see Figure 2. This illustrates that the

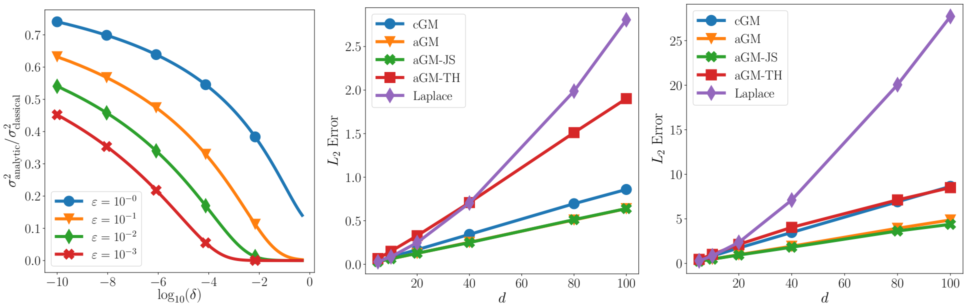


Figure 3. Left plot: Comparing the classical Gaussian mechanism (cGM) and the analytic Gaussian mechanism (aGM) in terms of gain in variance as a function of δ. Two rightmost plots: Mean estimation experiments showing L_2 error between the private mean estimate and the non-private empirical mean as a function of the dimension d with $\varepsilon = 1$ and $\varepsilon = 0.1$. Dataset size is fixed to $n = 500$ and privacy parameter is set to $\delta = 10^{-4}$.

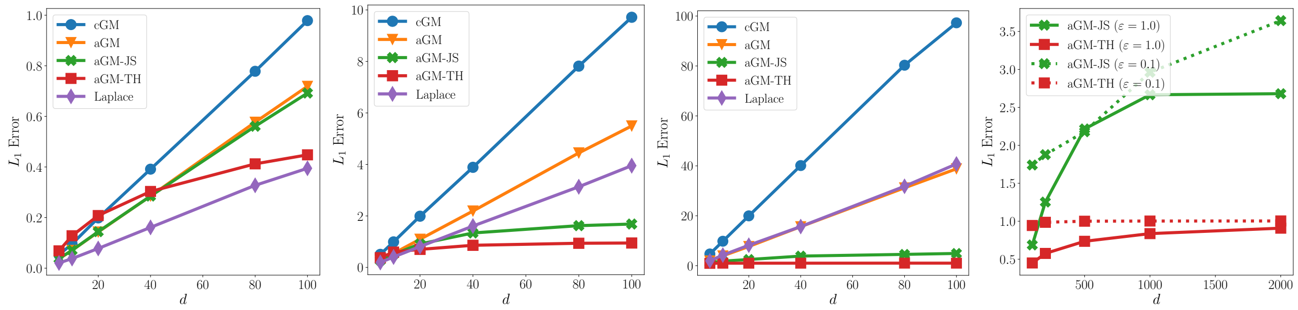
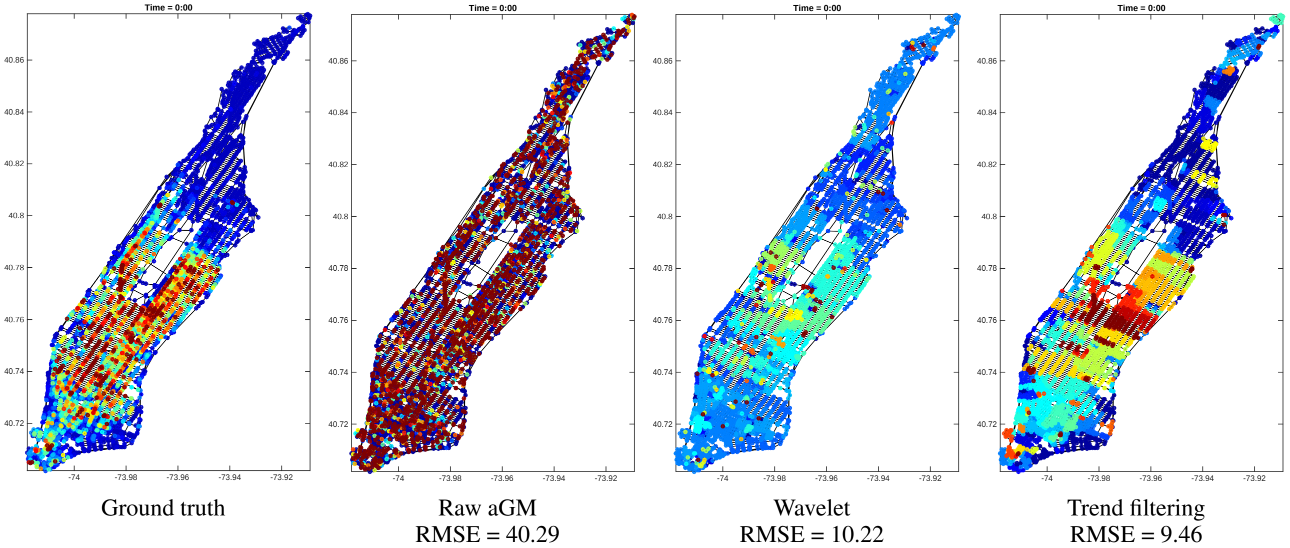


Figure 4. Histogram release experiments showing L_1 error between the private histogram and the non-private empirical histogram as a function of the dimension d. Dataset size is fixed to $n = 500$ and privacy parameter is set to $\delta = 10^{-4}$. The first three panels correspond to $\varepsilon = 1, 0.1, 0.01$ (left to right). The rightmost panel displays the two denoised mechanisms (aGM-JS and aGM-TH) in the high-dimensional case.

selected denoising methods are adaptive to the structure of the underlying data.

Furthermore, Figure 6 presents quantitative results where we compare the mean square error (MSE) of cGM, aGM as well as the aforementioned denoising techniques. As we can see, on the real datasets, aGM always improves over cGM by a constant factor and denoising techniques are able to leverage bias-variance trade-off and improve the recovery in MSE further. The benefits of denoising range from orders of magnitude (in the case when ε is tiny) to a small constant factor (when ε is moderate). In the low-privacy regime (e.g., $\varepsilon > 5$), soft-thresholding performs a little worse than not using it at all. This is the expected cost of adaptivity and it does appear in its error bound.



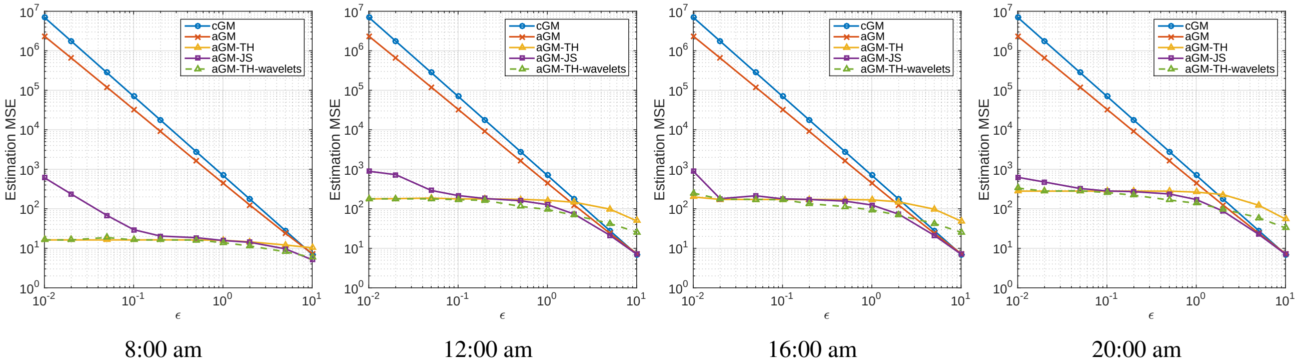
Ground truth

Raw aGM
RMSE = 40.29

Wavelet
RMSE = 10.22

Trend filtering
RMSE = 9.46

Figure 5. Illustration of the denoising in differentially private release of NYC taxi density during 12:00 - 13:00 pm Sept 24, 2014. Comparing to the figure in the midnight of Figure 2, the figures look structurally different. More activities center around the midtown and upper west sides.



8:00 am

12:00 am

16:00 am

20:00 am

Figure 6. Experiments for releasing NYC taxi heat maps. The plots compare the MSE of the reeased heat map as a function of the privacy loss parameter ε. We take $\Delta = 5$ and $\delta = 10^{-6}$ for all experiments. The wavelet basis is generated using Sharpnack et al. (2013) and the soft-thresholding's hyperparameter is chosen as $\sigma\sqrt{2\log d}$.