Supplementary Material to Paper ”Local Differential Privacy for Sampling”

Abstract

This is the Supplementary Material to Paper ”Local Differential Privacy for Sampling”. Notation “main file” indicates reference to the submitted draft.

Appendix: table of contents

1 Proofs and formal results

1.1 Proof of Lemma 2

For any $x \in \mathcal{X}$ and $P, P' \in \mathcal{D}(\mathcal{X})$, we have $\mathbf{Pr}[A(P) = x] \in \mathcal{M}$ and $\mathbf{Pr}[A(P') = x] \in \mathcal{M}$ by the fact that $A(P)$ samples from densities that lie in the mollifier $\mathcal{M}$. By definition of ε-mollifiers, the density ratio between any two densities in the ε-mollifiers is bounded by $\exp(\varepsilon)$, meaning we have

$$\frac{\mathbf{Pr}[A(P) = x]}{\mathbf{Pr}[A(P') = x]} \leq \exp(\varepsilon), \tag{1}$$

and thus A is an ε-private sampler.

1.2 Proof of Theorem 4

The proof follows from two Lemma which we state and prove.

Lemma 1 *For any $T \in \mathbb{N}_*$, we have that*

$$\sum_{t=1}^{T} \theta_t(\varepsilon) = \sum_{t=1}^{T} \left(\frac{\varepsilon}{\varepsilon + 4\log(2)} \right)^t < \frac{\varepsilon}{4\log(2)}. \tag{2}$$

Proof Since $(\varepsilon/(\varepsilon + 4\log(2)) < 1$ for any ε and noting that $\theta_t(\varepsilon) = (\varepsilon/(\varepsilon + 4\log(2))\theta_{t-1}(\varepsilon)$, we can conclude that $\theta_t(\varepsilon)$ is a geometric sequence. For any geometric series with ratio r, we have that

$$\sum_{t=1}^{T} r^t = r\left(\frac{1 - r^T}{1 - r}\right) \tag{3}$$

$$= \frac{r}{1 - r} - \frac{r^{T+1}}{1 - r} \tag{4}$$

$$< \frac{r}{1 - r} \tag{5}$$

Indeed, $\frac{r}{1-r}$ is the limit of the geometric series above when $T \to \infty$. In our case, we let $r = (\varepsilon/(\varepsilon + 4\log(2)))$ to show that

$$\frac{r}{1 - r} = \frac{\frac{\varepsilon}{\varepsilon + 4\log(2)}}{1 - \frac{\varepsilon}{\varepsilon + 4\log(2)}} = \frac{\frac{\varepsilon}{\varepsilon + 4\log(2)}}{\frac{4\log(2)}{\varepsilon + 4\log(2)}} = \frac{\varepsilon}{4\log(2)}, \tag{6}$$

which concludes the proof. ■

Lemma 2 *For any $\varepsilon > 0$ and $T \in \mathbb{N}_*$, let $\theta(\varepsilon) = (\theta_1(\varepsilon), \ldots, \theta_T(\varepsilon))$ denote the parameters and $c = (c_1, \ldots, c_t)$ denote the sufficient statistics returned by Algorithm 1, then we have*

$$-\frac{\varepsilon}{2} \leq \langle \theta(\varepsilon), c \rangle - \varphi(\theta(\varepsilon)) \leq \frac{\varepsilon}{2}. \tag{7}$$

Proof Since the algorithm returns classifiers such that $c_t(x) \in [-\log 2, \log 2]$ for all $1 \leq t \leq T$, we have from Lemma 1,

$$\sum_{t=1}^{T} \theta_t(\varepsilon) c_t \leq \log(2) \sum_{t=1}^{T} \theta_t(\varepsilon) < \log(2) \frac{\varepsilon}{4 \log(2)} = \frac{\varepsilon}{4}, \tag{8}$$

and similarly,

$$\sum_{t=1}^{T} \theta_t(\varepsilon) c_t \geq -\log(2) \sum_{t=1}^{T} \theta_t(\varepsilon) > -\log(2) \frac{\varepsilon}{4 \log(2)} = -\frac{\varepsilon}{4}. \tag{9}$$

Thus we have

$$-\frac{\varepsilon}{4} \leq \langle \theta(\varepsilon), c \rangle \leq \frac{\varepsilon}{4}. \tag{10}$$

By taking exponential, integrand (w.r.t Q_0) and logarithm of 10, we get

$$\log \int_{\mathcal{X}} \exp\left(-\frac{\varepsilon}{4}\right) dQ_0 \leq \log \int_{\mathcal{X}} \exp\left(\langle \theta(\varepsilon), c \rangle\right) dQ_0 \leq \log \int_{\mathcal{X}} \exp\left(\frac{\varepsilon}{4}\right) dQ_0 \tag{11}$$

$$-\frac{\varepsilon}{4} \leq \varphi(\theta(\varepsilon)) \leq \frac{\varepsilon}{4} \tag{12}$$

Since $\langle \theta(\varepsilon), c \rangle \in [-\varepsilon/4, \varepsilon/4]$ and $\varphi(\theta(\varepsilon)) \in [-\varepsilon/4, \varepsilon/4]$, the proof concludes by considering highest and lowest values. ■

The proof of Theorem 4 now follows from taking the $\exp$ of all quantities in (7), which makes appear Q_T in the middle and conditions for membership to $\mathcal{M}_\varepsilon$ in the bounds.

1.3 Proof of Theorem 5

We begin by first deriving the KL drop expression. At each iteration, we learn a classifier c_t, fix some step size $\theta > 0$ and multiply Q_{t-1} by $\exp(\theta \cdot c_t)$ and renormalize to get a new distribution which we will denote by $Q_t(\theta)$ to make the dependence of θ explicit.

Lemma 3 *For any $\theta > 0$, let $\varphi(\theta) = \log \int_{\mathcal{X}} \exp(\theta \cdot c_t) dQ_{t-1}$. The drop in KL is*

$$\textit{DROP}(\theta) := \textit{KL}(P, Q_{t-1}) - \textit{KL}(P, Q_t(\theta)) = \theta \cdot \int_{\mathcal{X}} c_t dP - \varphi(\theta) \tag{13}$$

Proof Note that $Q_t(\theta)$ is indeed a one dimensional exponential family with natural parameter θ, sufficient statistic c_t, log-partition function $\varphi(\theta)$ and base measure Q_{t-1}. We can write out the KL

divergence as

$$\mathrm{KL}(P, Q_{t-1}) - \mathrm{KL}(P, Q_t(\theta)) = \int_{\mathcal{X}} \log\left(\frac{P}{Q_{t-1}}\right) dP - \int_{\mathcal{X}} \log\left(\frac{P}{\exp(\theta \cdot c_t - \varphi(\theta))Q_{t-1}}\right) dP \tag{14}$$

$$= \int_{\mathcal{X}} \log\left(\frac{\exp(\theta \cdot c_t - \varphi(\theta))Q_{t-1}}{Q_{t-1}}\right) dP \tag{15}$$

$$= \int_{\mathcal{X}} \theta \cdot c_t - \varphi(\theta) dP \tag{16}$$

$$= \theta \cdot \int_{\mathcal{X}} c_t dP - \varphi(\theta) \tag{17}$$

■

It is not hard to see that the drop is indeed a concave function of θ, suggesting that there exists an optimal step size at each iteration. We split our analysis by considering two cases and begin when $\gamma_Q^t < 1/3$. Since $\theta > 0$, we can lowerbound the first term of the KL drop using WLA. The trickier part however, is bounding $\varphi(\theta)$ which we make use of Hoeffding's lemma.

Lemma 4 (Hoeffding's Lemma) *Let X be a random variable with distribution Q, with $a \leq X \leq b$ such that $\mathbb{E}_Q[X] = 0$, then for all $\lambda > 0$, we have*

$$\mathbb{E}_Q[\exp(\lambda \cdot X)] \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right) \tag{18}$$

Lemma 5 *For any classifier c_t satisfying Assumption 3 (WLA), we have*

$$\mathbb{E}_{Q_{t-1}}[\exp(\theta_t(\varepsilon) \cdot c_t)] \leq \exp\left(\theta_t^2(\varepsilon) \cdot \frac{(c_t^*)^2}{2} - \theta_t(\varepsilon) \cdot \gamma_Q^t \cdot c_t^*\right) \tag{19}$$

Proof Let $X = c_t - \cdot\mathbb{E}_{Q_{t-1}}[c_t]$, $b = c_t^*$, $a = -c_t^*$ and $\lambda = \theta_t(\varepsilon)$ and noticing that

$$\mathbb{E}_{Q_{t-1}}[\lambda \cdot X] = \mathbb{E}_{Q_{t-1}}[c_t - \mathbb{E}_{Q_{t-1}}[c_t]] = \mathbb{E}_{Q_{t-1}}[c_t] - \mathbb{E}_{Q_{t-1}}[c_t] = 0, \tag{20}$$

allows us to apply Lemma 4. By first realizing that

$$\exp(\lambda \cdot X) = \exp(\theta_t(\varepsilon) \cdot c_t) \cdot \exp(\theta_t(\varepsilon) \cdot \mathbb{E}_{Q_{t-1}}[-c_t]), \tag{21}$$

We get that

$$\mathbb{E}_{Q_{t-1}}[\exp(\theta_t(\varepsilon) \cdot c_t)] \cdot \exp\left(\theta_t(\varepsilon) \cdot \mathbb{E}_{Q_{t-1}}[-c_t]\right) \leq \exp\left(\theta_t^2(\varepsilon) \cdot \frac{(c_t^*)^2}{2}\right). \tag{22}$$

Re-arranging and using the WLA inequality yields

$$\mathbb{E}_{Q_{t-1}}[\exp(\theta_t(\varepsilon) \cdot c_t)] \leq \exp\left(\theta_t^2(\varepsilon) \cdot \frac{(c_t^*)^2}{2} - \theta_t(\varepsilon) \cdot \mathbb{E}_{Q_{t-1}}[-c_t]\right) \tag{23}$$

$$\leq \exp\left(\theta_t^2(\varepsilon) \cdot \frac{(c_t^*)^2}{2} - \theta_t(\varepsilon) \cdot \gamma_Q^t \cdot c_t^*\right) \tag{24}$$

■

Applying Lemma 5 and Lemma 3 (writing $Q_t = Q_t(\varepsilon)$) together gives us

$$\begin{aligned}
\mathrm{KL}(P, Q_t) &= \mathrm{KL}(P, Q_{t-1}) - \mathrm{DROP}(\theta_t(\varepsilon)) && (25)\\
&= \mathrm{KL}(P, Q_{t-1}) - \theta_t(\varepsilon) \cdot \int_{\mathcal{X}} c_t dP + \log \mathbb{E}_{Q_{t-1}}[\exp(\theta_t(\varepsilon) \cdot c_t)] && (26)\\
&\leq \mathrm{KL}(P, Q_{t-1}) - c_t^* \cdot \theta_t(\varepsilon) \cdot \left(\frac{1}{c_t^*} \int_{\mathcal{X}} c_t dP \right) + \left(\theta_t^2(\varepsilon) \cdot \frac{(c_t^*)^2}{2} - \theta_t(\varepsilon) \cdot \gamma_Q^t \cdot c_t^* \right) && (27)\\
&\leq \mathrm{KL}(P, Q_{t-1}) - c_t^* \theta_t(\varepsilon) \left(\gamma_P^t + \gamma_Q^t - \frac{c_t^* \cdot \theta_t(\varepsilon)}{2} \right) && (28)
\end{aligned}$$

Now we move to the case of $\gamma_Q^t \geq 1/3$.

Lemma 6 *For any classifier c_t returned by Algorithm 1, we have that*

$$\mathbb{E}_{Q_{t-1}}[\exp(c_t)] \leq \exp\left(-\Gamma(\gamma_Q^t)\right) \tag{29}$$

where $\Gamma(z) = \log(4/(5 - 3z))$.

Proof Consider the straight line between $(-\log 2, 1/2)$ and $(\log 2, 2)$ given by $y = 5/4 + (3/(4 \cdot \log 2))x$, which by convexity is greater then $y = \exp(x)$ on the interval $[-\log 2, \log 2]$. To this end, we define the function

$$f(x) = \begin{cases} \frac{5}{4} + \frac{3}{4 \cdot \log 2} \cdot x, & \text{if } x \in [-\log 2, \log 2] \\ 0, & \text{otherwise} \end{cases} \tag{30}$$

Since $c_t(x) \in [-\log 2, \log 2]$ for all $x \in \mathcal{X}$, we have that $f(c_t(x)) \geq \exp(c_t(x))$ for all $x \in \mathcal{X}$. Taking $\mathbb{E}_{Q_{t-1}}[\cdot]$ over both sides and using linearity of expectation gives

$$\begin{aligned}
\mathbb{E}_{Q_{t-1}}[\exp(c_t(x))] &\leq \mathbb{E}_{Q_{t-1}}[f(c_t(x))] && (31)\\
&= \frac{5}{4} + \frac{3}{4 \log 2} \left(\mathbb{E}_{Q_{t-1}}[c_t(x)] \right) && (32)\\
&= \frac{5}{4} - \frac{3}{4} \left(\frac{1}{\log 2} \mathbb{E}_{Q_{t-1}}[-c_t(x)] \right) && (33)\\
&< \frac{5}{4} - \frac{3}{4} \gamma_Q^t && (34)\\
&= \exp\left(-\log\left(\frac{5 - 3\gamma_Q^t}{4} \right)^{-1} \right) && (35)\\
&= \exp\left(-\log\left(\frac{4}{5 - 3\gamma_Q^t} \right) \right) && (36)\\
&= \exp\left(-\Gamma(\gamma_Q^t)\right), && (37)
\end{aligned}$$

as claimed. ■

Now we use Lemma 3 and Jensen's inequality since $\theta_t(\varepsilon) < 1$ so that

$$\begin{aligned}
\mathrm{KL}(P, Q_t) &= \mathrm{KL}(P, Q_{t-1}) - \mathrm{DROP}(\theta) && (38)\\
&= \mathrm{KL}(P, Q_{t-1}) - \theta_t(\varepsilon) \cdot \int_{\mathcal{X}} c_t dP + \log \mathbb{E}_{Q_{t-1}}[\exp(\theta_t \cdot c_t)] && (39)\\
&\leq \mathrm{KL}(P, Q_{t-1}) - \theta_t(\varepsilon) \cdot \mathbb{E}_P[c_t] + \theta_t \cdot \log \mathbb{E}_{Q_{t-1}}[\exp(c_t)] && (40)\\
&\leq \mathrm{KL}(P, Q_{t-1}) - \theta_t(\varepsilon) \left(\mathbb{E}_P[c_t] - \log \mathbb{E}_{Q_{t-1}}[\exp(c_t)]\right) && (41)\\
&= \mathrm{KL}(P, Q_{t-1}) - \theta_t(\varepsilon) \left(c_t^* \left(\frac{1}{c_t^*}\mathbb{E}_P[c_t]\right) - \log \mathbb{E}_{Q_{t-1}}[\exp(c_t)]\right) && (42)\\
&< \mathrm{KL}(P, Q_{t-1}) - \theta_t(\varepsilon) \left(c_t^* \gamma_P^t - \log\left(\exp\left(-\Gamma(\gamma_Q^t)\right)\right)\right) && (43)\\
&= \mathrm{KL}(P, Q_{t-1}) - \theta_t(\varepsilon) \left(c_t^* \gamma_P^t + \Gamma(\gamma_Q^t)\right). && (44)
\end{aligned}$$

1.4 Proof of Theorem 6

We first note that for any $Q \in \mathcal{M}_\varepsilon$,

$$\begin{aligned}
\mathrm{KL}(P, Q) &= \int_{\mathcal{X}} \log\left(\frac{P}{Q}\right) dP && (45)\\
&= \int_{\mathcal{X}} \log\left(\frac{P}{Q_0}\frac{Q_0}{Q}\right) dP && (46)\\
&= \int_{\mathcal{X}} \log\left(\frac{P}{Q_0}\right) dP - \int_{\mathcal{X}} \log\left(\frac{Q_0}{Q}\right) dP && (47)\\
&\geq \mathrm{KL}(P, Q_0) - \int_{\mathcal{X}} \frac{\varepsilon}{2} dP && (48)\\
&\geq \mathrm{KL}(P, Q_0) - \frac{\varepsilon}{2}, && (49)
\end{aligned}$$

which completes the proof of the upperbound To show (13), we have that

$$\begin{aligned}
\mathrm{KL}(P, Q_t) &\leq \mathrm{KL}(P, Q_{T-1}) - \theta_t(\varepsilon) \cdot \Lambda_t && (50)\\
&\leq \mathrm{KL}(P, Q_0) - \sum_{t=1}^{T-1} \theta_t(\varepsilon) \cdot \Lambda_t && (51)\\
&= \mathrm{KL}(P, Q_0) - \sum_{t=1}^{T-1} \theta_t(\varepsilon) \cdot \left(c_t^* \gamma_P^t + \Gamma(\gamma_Q^t)\right) && (52)\\
&\leq \mathrm{KL}(P, Q_0) - \sum_{t=1}^{T-1} \theta_t(\varepsilon) \cdot \left(\log 2 \cdot \gamma_P + \Gamma(\gamma_Q)\right) && (53)\\
&\leq \mathrm{KL}(P, Q_0) - \left(\log 2 \cdot \gamma_P + \log 2 \cdot \gamma_Q\right) \cdot \sum_{t=1}^{T-1} \theta_t(\varepsilon) && (54)\\
&\leq \mathrm{KL}(P, Q_0) - \left(\log 2 \cdot \gamma_P + \log 2 \cdot \gamma_Q\right) \cdot \sum_{t=1}^{T-1} \theta_t(\varepsilon) && (55)\\
&= \mathrm{KL}(P, Q_0) - \log 2 \cdot (\gamma_P + \gamma_Q) \cdot \theta_1(\varepsilon) \cdot \left(\frac{1 - \theta_t(\varepsilon)}{1 - \theta_1(\varepsilon)}\right) && (56)\\
&= \mathrm{KL}(P, Q_0) - \varepsilon \cdot \left(\frac{\gamma_P + \gamma_Q}{4}\right) \cdot \left(1 - \theta_t(\varepsilon)\right), && (57)
\end{aligned}$$

where we used the fact that $\Gamma(x) \geq \log 2 \cdot x$ and explicit geometric summation expression.

1.5 Proof of Theorems 7

We start by a general Lemma.

Lemma 7 *For any region of the support B, we have that*

$$\int_B dQ_t \geq \int_B dP - \int_B \log\left(\frac{P}{Q_t}\right) dP \tag{58}$$

Proof By first noting that for any region B,

$$\int_B (dP - dQ_t) = \int_B \left(1 - \frac{dQ_t}{dP}\right) dP \tag{59}$$

we then use the inequality $1 - x \leq \log(1/x)$ to get

$$\int_B (dP - dQ_t) = \int_B \left(1 - \frac{dP}{dQ_t}\right) dP \leq \int_B \log\left(\frac{dP}{dQ_t}\right) dP = \int_B \log\left(\frac{P}{Q_t}\right) dP \tag{60}$$

Re-arranging the above inequality gives us the bound. ■

Lemma 7 allows us to understand the relationship between two distributions P and Q_t in terms regions they capture. The general goal is to show that for a given region B (which includes the

highly dense mode regions), the amount of mass captured by the model $\int_B dQ_t$, is lower bounded by the target mass $\int_B dP$, and some small quantity. The inequality in Lemma 7 comments on this precisely with the small difference being a term that looks familiar to the KL-divergence - rather one that is bound to the specific region B. Though, this term can be understood to be small since by Theorem 5, we know that the global KL decreases, we give further refinements to show the importance of privacy parameters ε. We show that the term $\int_B \log(P/Q_t)dP$ can be decomposed in different ways, leading to our two Theorems to prove.

Lemma 8

$$\int_B \log\left(\frac{P}{Q_t}\right) dP \leq \int_B \log\left(\frac{P}{Q_0}\right) dP - \Delta + \frac{\varepsilon}{2}\left(1 - \int_B dP\right). \tag{61}$$

where $\Delta = KL(P, Q_0) - KL(P, Q_t)$

Proof We decompose the space $\mathcal{X}$ into B and the complement B^c to get

$$\begin{aligned}
\int_B \log\left(\frac{P}{Q_t}\right) dP &= \int_{\mathcal{X}} \log\left(\frac{P}{Q_t}\right) dP - \int_{B^c} \log\left(\frac{P}{Q_t}\right) dP && (62)\\
&= \mathrm{KL}(P, Q_t) - \int_{B^c} \log\left(\frac{P}{Q_t}\right) dP && (63)\\
&\leq \mathrm{KL}(P, Q_0) - \Delta - \int_{B^c} \log\left(\frac{P}{Q_t}\right) dP, && (64)
\end{aligned}$$

where we used Theorem 5, and letting $\theta = \theta(\varepsilon)$ for brevity, we also have

$$\begin{aligned}
\int_{B^c} \log\left(\frac{P}{Q_t}\right) dP &= \int_{B^c} \log\left(\frac{P}{Q_0 \exp\left(\langle \theta, c\rangle - \varphi(\theta)\right)}\right) dP && (65)\\
&= \int_{B^c} \log\left(\frac{P}{Q_0}\right) dP - \int_{B^c} \exp\left(\langle \theta, c\rangle - \varphi(\theta)\right) dP && (66)\\
&\geq \int_{B^c} \log\left(\frac{P}{Q_0}\right) dP - \int_{B^c} \frac{\varepsilon}{2} dP && (67)\\
&= \int_{B^c} \log\left(\frac{P}{Q_0}\right) dP - \frac{\varepsilon}{2}\left(1 - \int_B dP\right) && (68)
\end{aligned}$$

Combining these inequalities together gives us:

$$\begin{aligned}
\int_B \log\left(\frac{P}{Q_t}\right) dP &\leq \mathrm{KL}(P, Q_0) - \Delta - \left(\int_{B^c} \log\left(\frac{P}{Q_0}\right) dP - \frac{\varepsilon}{2}\left(1 - \int_B dP\right)\right) && (69)\\
&= \int_{\mathcal{X}} \log\left(\frac{P}{Q_0}\right) dP - \int_{B^c} \log\left(\frac{P}{Q_0}\right) dP - \Delta + \frac{\varepsilon}{2}\left(1 - \int_B dP\right) && (70)\\
&= \int_B \log\left(\frac{P}{Q_0}\right) dP - \Delta + \frac{\varepsilon}{2}\left(1 - \int_B dP\right) && (71)
\end{aligned}$$

■

We are now in a position to prove Theorem 7. Using Lemma 8 into the inequality in Lemma 7 yields

$$\int_B dQ_t \geq \int_B dP - \left(\int_B \log\left(\frac{P}{Q_0}\right) dP - \Delta + \frac{\varepsilon}{2}\left(1 - \int_B dP\right)\right) \tag{72}$$

$$= \left(1 + \frac{\varepsilon}{2}\right)\int_B dP - \frac{\varepsilon}{2} - \int_B \log\left(\frac{P}{Q_0}\right) + \Delta. \tag{73}$$

Reorganising and using the Theorem's notations, we get

$$\textsc{m}(B, Q) \;\; \geq \;\; \textsc{m}(B, P) - KL(P, Q_0; B) + \frac{\varepsilon}{2} \cdot J(P, Q; B, \varepsilon), \tag{74}$$

where we recall that $J(P, Q; B, \varepsilon) \doteq \textsc{m}(B, P) + \frac{2\Delta(Q)}{\varepsilon} - 1$. Theorem 6 says that we have in the high boosting regime $2\Delta(Q_T)/\varepsilon \geq (\gamma_P + \gamma_Q)/2 - \theta_T(\varepsilon) \cdot (\gamma_P + \gamma_Q)/2$. Letting $\overline{\gamma} \doteq (\gamma_P + \gamma_Q)/2$ and $K \doteq 4 \log 2$, we have from MBDE in the high boosting regime:

$$\begin{aligned}
\frac{2\Delta(Q)}{\varepsilon} \;\; &\geq \;\; \overline{\gamma} \cdot \left(1 - \left(\frac{1}{1 + \frac{K}{\varepsilon}}\right)^T\right) \\
&\geq \;\; \overline{\gamma} \cdot \left(1 - \frac{1}{1 + \frac{TK}{\varepsilon}}\right) \\
&= \overline{\gamma} \cdot \frac{TK}{TK + \varepsilon}.
\end{aligned} \tag{75}$$

To have $J(P, Q; B, \varepsilon) \geq -(2/\varepsilon) \cdot \alpha \textsc{m}(B, P)$, it is thus sufficient that

$$\begin{aligned}
\textsc{m}(B, P) \;\; &\geq \;\; \frac{1}{1 + \frac{2\alpha}{\varepsilon}} \cdot \left(1 - \overline{\gamma} \cdot \frac{TK}{TK + \varepsilon}\right) \\
&= \varepsilon \cdot \frac{\varepsilon + (1 - \overline{\gamma})TK}{(\varepsilon + 2\alpha)(\varepsilon + TK)}.
\end{aligned} \tag{76}$$

In this case, we check that we have from (74)

$$\textsc{m}(B, Q) \;\; \geq \;\; (1 - \alpha)\textsc{m}(B, P) - KL(P, Q_0; B), \tag{77}$$

as claimed.

1.6 Additional formal results

One might ask what such a strong model of privacy allows to keep from the accuracy standpoint in general. Perhaps paradoxically at first sight, it is not hard to show that privacy can bring approximation guarantees on learning: *if* we learn Q_ε within an ε-mollifier $\mathcal{M}$ (hence, we get ε-privacy for sampling from Q_ε), *then* each time *some* Q_ε in $\mathcal{M}$ accurately fits P, we are guaranteed that the one *we learn* also accurately fits P — albeit eventually more moderately —. We let $Q_\varepsilon(;.)$ denote the density learned, where . is the dataset argument.

Lemma 9 *Suppose $\exists$ ε-mollifier $\mathcal{M}$ s.t. $Q_\varepsilon \in \mathcal{M}$, then $(\exists P, D', \delta : \textbf{\textit{KL}}(P, Q_\varepsilon(;P')) \leq \delta) \Rightarrow (\forall D, \textbf{\textit{KL}}(P, Q_\varepsilon(;P)) \leq \delta + \varepsilon)$.*

Proof The proof is straightforward; we give it for completeness: for any dataset D, we have

$$\begin{aligned}
\mathrm{KL}(P, Q_\varepsilon(;P)) &= \int_{\mathcal{X}} \log\left(\frac{P}{Q_\varepsilon(;P)}\right) dP && (78)\\
&= \int_{\mathcal{X}} \log\left(\frac{P}{Q_\varepsilon(;P')}\right) dP + \int_{\mathcal{X}} \log\left(\frac{Q_\varepsilon(;P)}{Q_\varepsilon(;P')}\right) dP && (79)\\
&\leq \int_{\mathcal{X}} \log\left(\frac{P}{Q_\varepsilon(;P')}\right) dP + \varepsilon \cdot \int_{\mathcal{X}} dP && (80)\\
&= \mathrm{KL}(P, Q_\varepsilon(;P')) + \varepsilon && (81)\\
&\leq \delta + \varepsilon, && (82)
\end{aligned}$$

from which we derive the statement of Lemma 9 assuming $\mathcal{A}$ is ε-IP (the inequalities follow from the Lemma's assumption). ■

In the jargon of (computational) information geometry [1], we can summarize Lemma 9 as saying that if there exists an eligible[1] density in a small KL-ball relatively to P, we are guaranteed to find a density also in a small KL-ball relatively to P. This result is obviously good when the premises hold true, but it does not tell the full story when they do not. In fact, when there exists an eligible density outside a big KL-ball relatively to P, it is not hard to show using the same arguments as for the Lemma that we *cannot* find a good one, and this is not a feature of MBDE: this would hold regardless of the algorithm. This limitation is intrinsic to the likelihood ratio constraint of differential privacy, as the following Lemma shows. In the context of ε-DP, we assume that all input datasets have the same size, say m.

Lemma 10 *Let $\mathcal{A}$ denote an algorithm learning an ε-differentially private density. Denote $D \sim P$ an input of the algorithm and $\mathcal{Q}_\varepsilon(D)$ the set of all densities that can be the output of $\mathcal{A}$ on input D, taking in considerations all internal randomisations of $\mathcal{A}$. Suppose there exists an input D' for which one of these densities is far from the target: $\exists D', \exists Q \in \mathcal{Q}_\varepsilon(D') : \textbf{\textit{KL}}(P, Q(;D')) \geq \Delta$ for some "big" $\Delta > 0$. Then the output Q of $\mathcal{A}$ obtained from **any** input $D \sim P$ satisfies: $\textbf{\textit{KL}}(P, Q(;D)) \geq \Delta - m\varepsilon$.*

Proof Denote D the actual input of $\mathcal{A}$. There exists a sequence $\mathcal{D}$ of datasets of the same size, whose length is at most m, which transforms D into D' by repeatedly changing one observation in the current dataset: call it $\mathcal{D} = \{D, D_1, D_2, ..., D_k, D'\}$, with $k \leq m - 1$. Denote $Q(;D'')$ any

[1] Within the chosen ε-mollifier.

element of $\mathcal{Q}_\varepsilon(D'')$ for $D'' \in \mathcal{D}$. Since $\mathcal{A}$ is ε-differentially private, we have:

$$\Delta \leq \mathrm{KL}(P, Q(;D')) \tag{83}$$

$$= \int_{\mathcal{X}} \log\left(\frac{P}{Q(;D')}\right) dP \tag{84}$$

$$= \int_{\mathcal{X}} \log\left(\frac{P}{Q(;D)}\right) dP + \int_{\mathcal{X}} \log\left(\frac{Q(;D)}{Q(;D_1)}\right) dP + \sum_{j=1}^{k-1} \int_{\mathcal{X}} \log\left(\frac{Q(;D_j)}{Q(;D_{j+1})}\right) dP + \int_{\mathcal{X}} \log\left(\frac{Q(;D_k)}{Q(;D')}\right) dP \tag{85}$$

$$= \mathrm{KL}(P, Q(;D)) + \int_{\mathcal{X}} \log\left(\frac{Q(;D)}{Q(;D_1)}\right) dP + \sum_{j=1}^{k-1} \int_{\mathcal{X}} \log\left(\frac{Q(;D_j)}{Q(;D_{j+1})}\right) dP + \int_{\mathcal{X}} \log\left(\frac{Q(;D_k)}{Q(;D')}\right) dP \tag{86}$$

$$\leq \mathrm{KL}(P, Q(;D)) + m\varepsilon, \tag{87}$$

from which we derive the statement of Lemma 10. ■

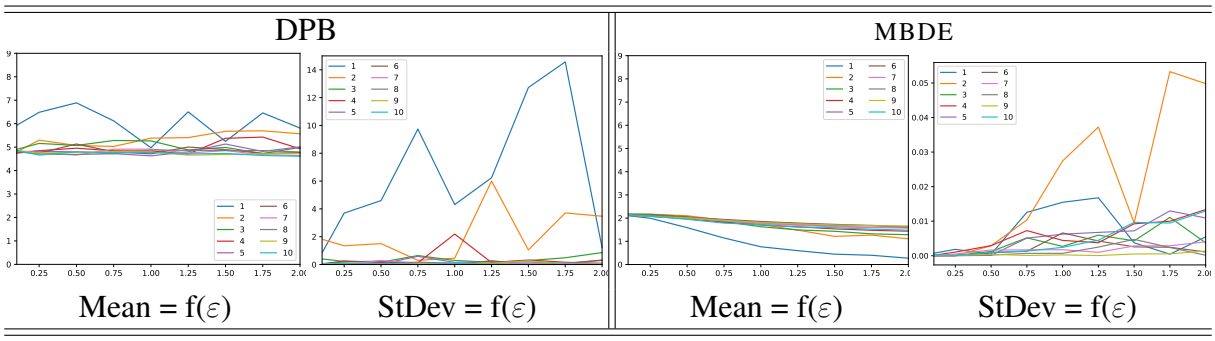

Figure 1: NLL metrics (mean and standard deviation) on the 1D random Gaussian problem for DPB (left pane) and MBDE (right pane), for a varying number of $m = 1, \ldots, 10$ random Gaussians. The lower the better on each metric. Remark the different scales for StDev (see text).

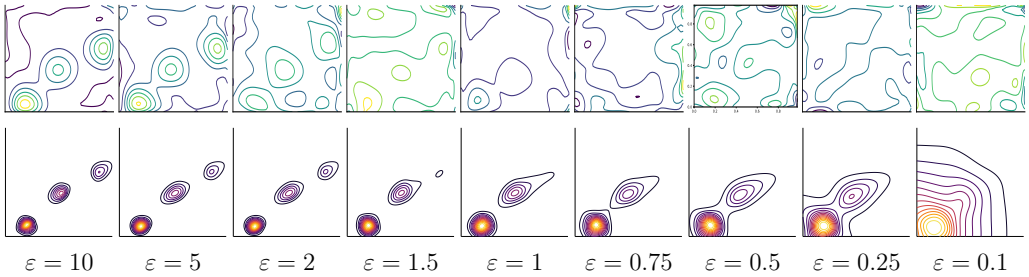

Figure 2: Randomly placed Gaussian convergence comparison for DPB (upper) against MBDE (lower).

2 Additional experiments

We provide here additional results to the main file. Figure 1 provides NLL values for the random 1D Gaussian problem. Figure 2 displays that picking Q_0 a standard Gaussian does not prevent to obtain good results — and beat DPB — when sampling random Gaussians.

References

[1] J.-D. Boissonnat, F. Nielsen, and R. Nock. Bregman voronoi diagrams. *DCG*, 44(2):281–307, 2010.