# Differentially Private Policy Evaluation (Supplementary Material)

Borja Balle[1], Maziar Gomrokchi[2], and Doina Precup[2]

[1]*Department of Mathematics and Statistics, Lancaster University, UK*
[2]*School of Computer Science, McGill University, Canada*

## 1 Smoothed Gaussian Perturbation

A proof of Lemma 1 in the paper can be found in the pre-print [2]. For the sake of completeness, we provide here an elementary proof (albeit with slightly worse constants). In particular, we are going to prove the following.

**Lemma 1.** *Let $A$ be an algorithm that on input $X$ computes a vector $\mu_X \in \mathbb{R}^d$ deterministically and then outputs $Z_X \sim \mathcal{N}(\mu_X, \sigma_X^2 I)$, where $\sigma_X^2$ is a variance that depends on $X$. Let $\alpha = \alpha(\varepsilon, \delta) = 15\sqrt{2\ln(4/\delta)}/\varepsilon$ and $\beta = \beta(\varepsilon, \delta, d) = (2\ln 2)\varepsilon/5(\sqrt{d} + \sqrt{2\ln(4/\delta)})^2$. Suppose that $\varepsilon \leq 5$, $\delta$ and $d$ are such $\beta \leq \ln 2$, and the following are satisfied for every pair of neighbouring datasets $X \simeq X'$:*

1. $\sigma_X \geq \alpha \|\mu_X - \mu_{X'}\|_2$,
2. $|\ln(\sigma_X^2) - \ln(\sigma_{X'}^2)| \leq \beta$.

*Then $A$ is $(\varepsilon, \delta)$-differentially private.*

We start with a simple characterization of $(\varepsilon, \delta)$-differential privacy that will be useful for our proof.

**Lemma 2.** *Let $A(X) = \theta_X \in \mathbb{R}^d$ be the output of a randomized algorithm on input $X$. Write $f_{\theta_X}(\theta)$ for the probability density of the output of $A$ on input $X$. Suppose that for every pair of neighbouring datasets $X \simeq X'$ there exists a measurable set $\Theta_{X,X'} \subset \mathbb{R}^d$ such that the following are satisfied:*

1. $\mathbb{P}[\theta_X \notin \Theta_{X,X'}] \leq \delta$;
2. *for all $\theta \in \Theta_{X,X'}$ we have $f_{\theta_X}(\theta) \leq e^{\varepsilon} f_{\theta_{X'}}(\theta)$.*

*Then $A$ is $(\varepsilon, \delta)$-differentially private.*

*Proof.* Fix a pair of neighbouring datasets $X \simeq X'$ and let $E \subseteq \mathbb{R}^d$ be any measurable set. Let $\Theta_{X,X'}$ be as in the statement and write $\Theta_{X,X'}^{\mathsf{c}} = \mathbb{R}^d \setminus \Theta_{X,X'}$. Using the assumptions on $\Theta_{X,X'}$ we see that

$$\begin{aligned}\mathbb{P}[\theta_X \in E] &= \mathbb{P}[\theta_X \in E \cap \Theta_{X,X'}] + \mathbb{P}[\theta_X \in E \cap \Theta_{X,X'}^{\mathsf{c}}] \\ &\leq e^{\varepsilon}\mathbb{P}[\theta_{X'} \in E \cap \Theta_{X,X'}] + \delta \\ &\leq e^{\varepsilon}\mathbb{P}[\theta_{X'} \in E] + \delta \ .\end{aligned}$$

□

Now we proceed with the proof of Lemma 1. Let $X \simeq X'$ be two neighbouring datasets and let us write $Z_1 = Z_X$ and $Z_2 = Z_{X'}$ for simplicity. Thus, for $i = 1, 2$ we have that $Z_i \sim \mathcal{N}(\mu_i, \sigma_i^2 I)$ are $d$-dimensional independent Gaussian random variables whose means and variances satisfy the assumptions of Lemma 1 for some $\varepsilon, \delta > 0$. The density function of $Z_i$ is denoted by $f_{Z_i}(z)$. In order to be able to apply Lemma 2 we want to show that the privacy loss between $Z_1$ and $Z_2$ defined as

$$L(z) = \ln \frac{f_{Z_1}(z)}{f_{Z_2}(z)} \tag{1}$$

is bounded by $\varepsilon$ for all $z \in \Omega$, where $\Omega \subset \mathbb{R}^d$ is an event with probability at least $1-\delta$ under $Z_1$.

We can start by identifying a candidate $\Omega$. Since $\Omega$ has to have high probability w.r.t. $Z_1$, it should contain $\mu_1$ because a ball around the mean is the event with the highest probability under a spherical Gaussian distribution (among those with the same Lebesgue measure). For technical reasons, instead of a ball we will take a slightly more complicated region, which for now we will parametrize by two quantities $a, b > 0$. The definition of this region will depend on the difference of means $\Delta = \mu_2 - \mu_1$:

$$\Omega = \Omega_a \cap \Omega_b = \{z + \mu_1 \in \mathbb{R}^d \mid |\langle z, \Delta \rangle| \leq a\} \cap \{z + \mu_1 \in \mathbb{R}^d \mid \|z\| \leq b\} \ . \tag{2}$$

We need to choose $a$ and $b$ such that the probability $\mathbb{P}[Z_1 \notin \Omega] \leq \delta$, and for that we shall combine two different tail bounds. On the one hand, note that $Z = \langle Z_1 - \mu_1, \Delta \rangle / (\sigma_1 \|\Delta\|) \sim \mathcal{N}(0,1)$ is a one dimensional standard Gaussian random variable and recall that for any $t \geq 0$:

$$\mathbb{P}[|Z| > t] \leq 2e^{-t^2/2} \ . \tag{3}$$

On the other hand, $X = \|Z_1 - \mu_1\|^2 / \sigma_1^2 \sim \chi^2_d$ follows a chi-squared distribution with $d$ degrees of freedom, for which is known [1] that for all $t \geq 0$:

$$\mathbb{P}[X > d + 2\sqrt{dt} + 2t] \leq e^{-t} \ . \tag{4}$$

To make our choices for $a$ and $b$ we can take them such that $\mathbb{P}[Z_1 \notin \Omega_a], \mathbb{P}[Z_1 \notin \Omega_b] \leq \delta/2$, since then by a union bound we will get

$$\mathbb{P}[Z_1 \notin \Omega] \leq \mathbb{P}[Z_1 \notin \Omega_A] + \mathbb{P}[Z_1 \notin \Omega_B] \leq \delta \ . \tag{5}$$

Since $Z$ satisfies $|Z| \leq \sqrt{2\ln(4/\delta)}$ with probability at least $1 - \delta/2$, we can take

$$a = \sigma_1 \|\Delta\| \sqrt{2 \ln \frac{4}{\delta}} = \sigma_1 \|\Delta\| C_\delta \ . \tag{6}$$

For $X$ we have that $d + 2\sqrt{d\ln(2/\delta)} + 2\ln(2/\delta) \leq d + 2\sqrt{2d\ln(2/\delta)} + 2\ln(2/\delta) = (\sqrt{d} + \sqrt{2\ln(2/\delta)})^2$. Hence, we choose

$$b = \sigma_1(\sqrt{d} + \sqrt{2\ln(2/\delta)}) = \sigma_1 D_\delta \ . \tag{7}$$

Fixing this choice of $\Omega$, we now proceed to see under what conditions on $\sigma_1$ and $\sigma_2$ we can get $L(z) \leq \varepsilon$ for all $z \in \Omega$. We start by expanding the definition of $L(z)$ to get

$$L(z) = \frac{d}{2} \ln \frac{\sigma_2^2}{\sigma_1^2} + \frac{\|\mu_2 - z\|^2}{2\sigma_2^2} - \frac{\|\mu_1 - z\|^2}{2\sigma_1^2} \ . \tag{8}$$

The easiest thing to do is to separate this quantity into several parts and insist on each part being at most a fraction of $\varepsilon$. To simplify calculations we will just require that each part is at most $\epsilon = \varepsilon/5$. This reasoning applied to the first term shows that we must satisfy

$$\frac{\sigma_2^2}{\sigma_1^2} \leq e^{2\epsilon/d} \ . \tag{9}$$

Note that this becomes more restrictive as $\epsilon \approx 0$ or $d \to \infty$, in which case we have $e^{\epsilon/d} \approx 1$.

Next we look at the second part and write $z = z' + \mu_1$ because this is the form of the vectors in $\Omega$. With some algebra we get:

$$\frac{\|\mu_2 - (z' + \mu_1)\|^2}{2\sigma_2^2} - \frac{\|\mu_1 - (z' + \mu_1)\|^2}{2\sigma_1^2} = \frac{\|\Delta\|^2 + \|z'\|^2 - 2\langle z', \Delta \rangle}{2\sigma_2^2} - \frac{\|z'\|^2}{2\sigma_1^2} \ . \tag{10}$$

To further decompose this quantity we write $z' \in \mathbb{R}^d$ as $z' = z_p + z_o$, where $z_p = \Delta \langle z', \Delta \rangle / \|\Delta\|^2$ is the orthogonal projection of $z$ onto the line spanned by the vector $\Delta$, and $z_o$ is the corresponding orthogonal complement. Pythagora's Theorem implies $\|z'\|^2 = \|z_p\|^2 + \|z_o\|^2$, and the RHS in the above expression is equal to

$$\frac{\|\Delta\|^2}{2\sigma_2^2} - \frac{\langle z', \Delta \rangle}{\sigma_2^2} + \frac{|\langle z', \Delta \rangle|^2}{2\|\Delta\|^2} \left( \frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \right) + \frac{\|z_o\|^2}{2} \left( \frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \right) \ . \tag{11}$$

Now note that the last two terms can be upper bounded by zero if $\sigma_1 \leq \sigma_2$, but need to be taken into account otherwise. Furthermore, if it were the case that $\sigma_1 \gg \sigma_2 \approx 0$, then these terms could grow unboundedly. Thus we shall require that a bound of the form

$$\frac{\sigma_1^2}{\sigma_2^2} \leq \gamma \ , \tag{12}$$

holds for some $\gamma \geq 1$ to be specified later. Nonetheless, we observe that under this assumption

$$\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \leq \frac{\gamma - 1}{\sigma_1^2} \ . \tag{13}$$

Furthermore, $z \in \Omega$ implies $\|z_o\|^2 \leq \|z'\|^2 = \|z - \mu_1\| \leq b^2$ and $|\langle z', \Delta \rangle|^2 = |\langle z - \mu_1, \Delta \rangle|^2 \leq a^2$. Thus we see that

$$\frac{|\langle z', \Delta \rangle|^2}{2\|\Delta\|^2} \left( \frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \right) \leq \frac{C_\delta^2 (\gamma - 1)}{2} \ , \tag{14}$$

and

$$\frac{\|z_o\|^2}{2} \left( \frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \right) \leq \frac{D_\delta^2 (\gamma - 1)}{2} \ . \tag{15}$$

By requiring that each of these bounds is at most $\epsilon$ we obtain the following constraint for $\gamma$:

$$\gamma \leq 1 + \frac{2\epsilon}{\max\{C_\delta^2, D_\delta^2\}} \ , \tag{16}$$

which can be satisfied by taking, for example:

$$\gamma = 1 + \frac{2\epsilon}{\left( \sqrt{d} + \sqrt{2 \ln(4/\delta)} \right)^2} \ . \tag{17}$$

Note that for fixed $\delta$, small $\epsilon$ and/or large $d$ this choice of $\gamma$ will make (12) behave much like the bound (9) we assumed above for $\sigma_2^2/\sigma_1^2$. In fact, using that $1 + x \geq e^{x \ln 2}$ for all $0 \leq x \leq 1$ we see that (12) can be satisfied if $2\epsilon/(\sqrt{d} + \sqrt{2\ln(4/\delta)})^2 \leq 1$ and

$$\frac{\sigma_1^2}{\sigma_2^2} \leq \exp\left( \frac{(2 \ln 2)\epsilon}{\left( \sqrt{d} + \sqrt{2 \ln(4/\delta)} \right)^2} \right) \ . \tag{18}$$

From here it is immediate to see that if the second condition $|\ln(\sigma_1^2) - \ln(\sigma_2^2)| \leq \beta$ in Lemma 1 is satisfied, then (9) and (18) are both satisfied.

The missing ingredient to show that $L(z) \leq \varepsilon$ for all $z \in \Omega$ is an absolute lower bound on $\sigma_1$. This will follow from bounding the remaining terms in $L(z)$ as follows:

$$\frac{\|\Delta\|^2}{2\sigma_2^2} - \frac{\langle z', \Delta \rangle}{\sigma_2^2} \leq \frac{\|\Delta\|^2 + 2\sigma_1 \|\Delta\| C_\delta}{2\sigma_2^2} \tag{19}$$

$$\leq \frac{\gamma}{2} \frac{\|\Delta\|^2 + 2\sigma_1 \|\Delta\| C_\delta}{\sigma_1^2} \tag{20}$$

$$\leq \frac{3}{2} \frac{\|\Delta\|^2 + 2\sigma_1 \|\Delta\| C_\delta}{\sigma_1^2} \tag{21}$$

$$= \frac{3\|\Delta\|^2}{2\sigma_1^2} + \frac{3\|\Delta\| C_\delta}{\sigma_1} \ , \tag{22}$$

where we used that $\epsilon \leq 1$ implies $\gamma \leq 3$. If we require each of these two terms to be at most $\epsilon$, we obtain the constraint:

$$\sigma_1 \geq \|\Delta\| \max\left\{ \sqrt{\frac{3}{2\epsilon}}, \frac{3C_\delta}{\epsilon} \right\} = \frac{3\|\Delta\| C_\delta}{\epsilon} \ . \tag{23}$$

To conclude the proof just note that the above bound can be rewritten as $\sigma_1 \geq \alpha\|\Delta\|$, which is precisely the first condition in Lemma 1.

# 2 Privacy Analysis of DP-LSW

**Lemma 3.** *Let $X \simeq X'$ be two neighbouring datasets of $m$ trajectories with $X = (x_1, \ldots, x_{m-1}, x)$ and $X' = (x_1, \ldots, x_{m-1}, x')$. Let $X^\circ = (x_1, \ldots, x_{m-1})$. Let $\mathcal{S}_x$ (resp. $\mathcal{S}_{x'}$) denote the set of states visited by $x$ (resp. $x'$). Then we have*

$$\|F_X - F_{X'}\|_{2,\Gamma} \leq \frac{R_{\max}}{1-\gamma}\sqrt{\sum_{s\in\mathcal{S}_x\cup\mathcal{S}_{x'}} \frac{w_s}{(|X_s^\circ|+1)^2}} \ .$$

*Proof.* We start by noting that if $s \in \mathcal{S} \setminus (\mathcal{S}_x \cup \mathcal{S}_{x'})$, then $F_{X,s} = F_{X',s}$. In the case $s \in \mathcal{S}_x \cup \mathcal{S}_{x'}$ we can write $F_{X,s} = (|X_s^\circ|F_{X^\circ,s} + F_{x,s})/(|X_s^\circ|+1)$. Using a symmetric expression for $F_{X',s}$ we see that in this case

$$|F_{X,s} - F_{X',s}| = \frac{1}{|X_s^\circ|+1}|F_{x,s} - F_{x',s}| \leq \frac{1}{|X_s^\circ|+1}\max\{F_{x,s}, F_{x',s}\} \leq \frac{1}{|X_s^\circ|+1}\frac{R_{\max}}{1-\gamma} \ ,$$

where we used that $0 \leq F_{x,s} \leq R_{\max}/(1-\gamma)$ for all $s$ and $x$. When $s \in \mathcal{S}_x \setminus \mathcal{S}_{x'}$ we can use the same expression as before for $F_{X,s}$ and write $F_{X',s} = F_{X^\circ,s}$. A similar argument as in the previous case then yields

$$|F_{X,s} - F_{X',s}| = \frac{1}{|X_s^\circ|+1}|F_{x,s} - F_{X^\circ,s}| \leq \frac{1}{|X_s^\circ|+1}\frac{R_{\max}}{1-\gamma} \ .$$

Note the same bound also holds for the case $s \in \mathcal{S}_{x'} \setminus \mathcal{S}_x$. Finally, since we have seen that the same bound holds for all $s \in \mathcal{S}_x \cup \mathcal{S}_{x'}$, we obtain

$$\sum_{s\in\mathcal{S}} w_s(F_{X,s} - F_{X',s})^2 \leq \frac{R_{\max}^2}{(1-\gamma)^2}\sum_{s\in\mathcal{S}_x\cup\mathcal{S}_{x'}} \frac{w_s}{(|X_s^\circ|+1)^2} \ ,$$

which yields the desired bound. □

**Corollary 4.** *If $X$ is a dataset of trajectories, then the following holds for every neighbouring dataset $X' \simeq X$:*

$$\|F_X - F_{X'}\|_{2,\Gamma} \leq \frac{R_{\max}}{1-\gamma}\sqrt{\sum_{s\in\mathcal{S}} \frac{w_s}{\max\{|X_s|,1\}^2}} \ .$$

*Proof.* Using the notation from Lemma 3 we observe that $|X_s| = |X_s^\circ| + 1$ if $s \in \mathcal{S}_x$, and $|X_s| = |X_s^\circ|$ if $s \notin \mathcal{S}_x$. Therefore, the following holds for any trajectories $x, x'$:

$$\sum_{s\in\mathcal{S}_x\cup\mathcal{S}_{x'}} \frac{w_s}{(|X_s^\circ|+1)^2} \leq \sum_{s\in\mathcal{S}} \frac{w_s}{(|X_s^\circ|+1)^2} = \sum_{s\in\mathcal{S}_x} \frac{w_s}{|X_s|^2} + \sum_{s\in\mathcal{S}\setminus\mathcal{S}_x} \frac{w_s}{(|X_s|+1)^2} \leq \sum_{s\in\mathcal{S}_X} \frac{w_s}{|X_s|^2} + \sum_{s\in\mathcal{S}\setminus\mathcal{S}_X} w_s \ ,$$

where $\mathcal{S}_X$ denotes the set of states visited by at least one trajectory from $X$. Since $s \notin \mathcal{S}_X$ implies $|X_s| = 0$, we can plug this bound into the result of Lemma 3 as follows:

$$\|F_X - F_{X'}\|_{2,\Gamma} \leq \frac{R_{\max}}{1-\gamma}\sqrt{\sum_{s\in\mathcal{S}_X} \frac{w_s}{|X_s|^2} + \sum_{s\in\mathcal{S}\setminus\mathcal{S}_X} w_s} = \frac{R_{\max}}{1-\gamma}\sqrt{\sum_{s\in\mathcal{S}} \frac{w_s}{\max\{|X_s|,1\}^2}} \ .$$ □

**Lemma 5.** *The following holds for every $v \in \mathbb{N}^{\mathcal{S}}$:*

$$\varphi_k^w(v) = \sum_{s\in\mathcal{S}} \frac{w_s}{\max\{v_s - k, 1\}^2} \ .$$

*Furthermore, for every $k \geq \|v\|_\infty - 1$ we have $\varphi_k^w(v) = \sum_s w_s$.*

*Proof.* Recall that $\varphi_k^w(v) = \max_{\|v'-v\|_\infty \le k} \varphi^w(v')$ with $\varphi^w(v) = \sum_s w_s / \max\{v_s, 1\}^2$ and observe the result follows immediately because

$$\varphi_k^w(v) = \sum_{s\in\mathcal{S}} \frac{w_s}{\min_{-k\le l\le k} \max\{v_s + l, 1\}^2} = \sum_{s\in\mathcal{S}} \frac{w_s}{\max\{v_s - k, 1\}^2} \ . \qquad \square$$

# 3 Privacy Analysis of DP-LSL

**Lemma 6.** *Let $X \simeq X'$ be two neighbouring datasets of $m$ trajectories with $X = (x_1, \dots, x_{m-1}, x)$ and $X' = (x_1, \dots, x_{m-1}, x')$. Let $F_x \in \mathbb{R}^{\mathcal{S}}$ (resp. $F_{x'} \in \mathbb{R}^{\mathcal{S}}$) be the vector given by $F_x(s) = F_{x,s}$ (resp. $F_{x'}(s) = F_{x',s}$). Define the diagonal matrices $\Gamma_\rho, \Delta_{x,x'} \in \mathbb{R}^{\mathcal{S}\times\mathcal{S}}$ given by $\Gamma_\rho(s,s) = \rho_s$ and $\Delta_{x,x'}(s,s) = \mathbb{I}_{s\in x} - \mathbb{I}_{s\in x'}$. If the regularization parameter satisfies $\lambda > \|\Phi^\top \Delta_{x,x'} \Gamma_\rho \Phi\|$, then the following holds:*

$$\frac{\|\theta_X^\lambda - \theta_{X'}^\lambda\|_2}{2} \le \frac{\left\| \left(\Delta_{x,x'}\Phi\theta_X^\lambda - F_x + F_{x'}\right)^\top \Gamma_\rho \Phi \right\|_2}{\lambda - \|\Phi^\top \Delta_{x,x'} \Gamma_\rho \Phi\|} \ . \tag{24}$$

*Proof.* In order to simplify our notation we write $\bar\theta = \theta_X^\lambda$ and $\bar\theta' = \theta_{X'}^\lambda$ for the rest of the proof. Given a trajectory $x$ and a vector $\theta \in \mathbb{R}^d$ we shall also write $\ell(x,\theta) = \sum_{s\in\mathcal{S}_x} \rho_s (F_{x,s} - \phi_s^\top \theta)^2$ so that $J_X(\theta) = \frac{1}{m}\sum_{i=1}^m \ell(x_i,\theta)$. Now we proceed with the proof.

Let us start by noting that because $J_X^\lambda(\theta)$ is $\lambda/m$-strongly convex, we have $J_X^\lambda(\theta_1) - J_X^\lambda(\theta_2) \ge \langle \nabla J_X^\lambda(\theta_2), \theta_1 - \theta_2\rangle + \frac{\lambda}{2m}\|\theta_1 - \theta_2\|_2^2$ for any $\theta_1, \theta_2 \in \mathbb{R}^d$. Thus, using that optimality implies $\nabla J_X^\lambda(\bar\theta) = \nabla J_{X'}^\lambda(\bar\theta') = 0$, we get

$$\begin{aligned}
\frac{\lambda}{m}\|\bar\theta - \bar\theta'\|_2^2 &\le J_X^\lambda(\bar\theta') - J_X^\lambda(\bar\theta) + J_{X'}^\lambda(\bar\theta) - J_{X'}^\lambda(\bar\theta') \\
&= J_X(\bar\theta') - J_X(\bar\theta) + J_{X'}(\bar\theta) - J_{X'}(\bar\theta') \\
&= \frac{1}{m}\left(\ell(x,\bar\theta') - \ell(x,\bar\theta) + \ell(x',\bar\theta) - \ell(x',\bar\theta')\right) \ ,
\end{aligned}$$

where the equalities follows from definitions of $X$, $X'$, $J_X^\lambda$ and $J_X$. If we now expand the definition of $\ell(x,\theta)$ we see that

$$\begin{aligned}
\ell(x,\bar\theta') - \ell(x,\bar\theta) &= \sum_{s\in\mathcal{S}_x} \rho_s \left((\phi_s^\top\bar\theta')^2 - (\phi_s^\top\bar\theta)^2 - 2F_{x,s}\phi_s^\top(\bar\theta' - \bar\theta)\right) \ , \\
\ell(x',\bar\theta) - \ell(x',\bar\theta') &= \sum_{s\in\mathcal{S}_{x'}} \rho_s \left((\phi_s^\top\bar\theta)^2 - (\phi_s^\top\bar\theta')^2 - 2F_{x',s}\phi_s^\top(\bar\theta - \bar\theta')\right) \ .
\end{aligned}$$

Using the identity $(\phi_s^\top\bar\theta')^2 - (\phi_s^\top\bar\theta)^2 = (\bar\theta' + \bar\theta)^\top \phi_s\phi_s^\top(\bar\theta' - \bar\theta)$, we rewrite $\ell(x,\bar\theta') - \ell(x,\bar\theta) + \ell(x',\bar\theta) - \ell(x',\bar\theta')$ as

$$\sum_{s\in\mathcal{S}} \rho_s \left[(\mathbb{I}_{s\in x} - \mathbb{I}_{s\in x'})(\bar\theta' + \bar\theta)^\top \phi_s\phi_s^\top - 2(F_{x,s} - F_{x',s})\phi_s^\top\right](\bar\theta' - \bar\theta) \ , \tag{25}$$

where we implicitly used that $F_{x,s} = 0$ whenever $s \notin x$. Finally, using the definitions in the statement we can rearrange the above expression to show that

$$\begin{aligned}
\frac{\lambda}{m}\|\bar\theta - \bar\theta'\|_2^2 &\le \frac{1}{m}\left((\bar\theta' + \bar\theta)^\top\Phi^\top\Delta_{x,x'} - 2(F_x - F_{x'})^\top\right)\Gamma_\rho\Phi(\bar\theta' - \bar\theta) \\
&= \frac{2}{m}\left(\bar\theta^\top\Phi^\top\Delta_{x,x'} - (F_x - F_{x'})^\top\right)\Gamma_\rho\Phi(\bar\theta' - \bar\theta) + \frac{1}{m}(\bar\theta' - \bar\theta)^\top\Phi^\top\Delta_{x,x'}\Gamma_\rho\Phi(\bar\theta' - \bar\theta) \\
&\le \frac{2}{m}\|\left(\bar\theta^\top\Phi^\top\Delta_{x,x'} - (F_x - F_{x'})^\top\right)\Gamma_\rho\Phi\|_2\|\bar\theta' - \bar\theta\|_2 + \frac{1}{m}\|\Phi^\top\Delta_{x,x'}\Gamma_\rho\Phi\|\|\bar\theta' - \bar\theta\|_2^2 \ ,
\end{aligned}$$

where we used the Cauchy–Schwartz inequality and the definition of operator norm. The result now follows by solving for $\|\bar\theta - \bar\theta'\|_2$ in the above inequality. $\square$

**Corollary 7.** *Let $X$ be a dataset of trajectories and suppose $\lambda > \|\Phi\|^2\|\rho\|_\infty$. Then the following holds for any neighbouring dataset $X' \simeq X$:*

$$\|\theta_X^\lambda - \theta_{X'}^\lambda\|_2 \leq \frac{2R_{\max}\|\Phi\|}{(1-\gamma)(\lambda - \|\Phi\|^2\|\rho\|_\infty)}\sqrt{\varphi_X^\lambda} \ ,$$

*where*

$$\varphi_X^\lambda = \left(\frac{\|\Phi\|\|\rho\|_\infty}{\sqrt{2\lambda}}\sqrt{\sum_{s\in\mathcal{S}}\rho_s|X_s|} + \|\rho\|_2\right)^2 \ .$$

*Proof.* We start by noting that $\|\Delta_{x,x'}\| \leq 1$ and $\|\Gamma_\rho\| = \|\rho\|_\infty$, hence submultiplicativity of matrix operator norms yields $\|\Phi^\top\Delta_{x,x'}\Gamma_\rho\Phi\| \leq \|\Phi\|^2\|\rho\|_\infty$. On the other hand, for the numerator in (24) we have

$$\left\|\left(\Delta_{x,x'}\Phi\theta_X^\lambda - F_x + F_{x'}\right)^\top\Gamma_\rho\Phi\right\|_2 \leq \left(\|\theta_X^\lambda\|_2\|\Phi\|\|\rho\|_\infty + \|(F_x - F_{x'})^\top\Gamma_\rho\|_2\right)\|\Phi\| \ . \tag{26}$$

Bounding the individual entries in $F_x$ and $F_{x'}$ by $R_{\max}/(1-\gamma)$ we get $\|(F_x - F_{x'})^\top\Gamma_\rho\|_2 \leq R_{\max}\|\rho\|_2/(1-\gamma)$. The last step is to bound the norm $\|\theta_X^\lambda\|_2$, for which we use the closed-form solution to $\operatorname{argmin}_\theta J_X^\lambda(\theta)$ given in the paper and write:

$$\|\theta_X^\lambda\|_2 \leq \left\|(\Phi^\top\Gamma_X\Phi + \frac{\lambda}{2m}I)^{-1}\Phi^\top\Gamma_X^{1/2}\right\|\|F_X\|_{2,\Gamma_X} \leq \left\|(\Phi^\top\Gamma_X\Phi + \frac{\lambda}{2m}I)^{-1}\Phi^\top\Gamma_X^{1/2}\right\|\left(\frac{R_{\max}}{1-\gamma}\sqrt{\sum_{s\in\mathcal{S}}\frac{\rho_s|X_s|}{m}}\right) \ .$$

To bound the last remaining norm let use write $U\Sigma V^\top$ for the SVD of $\Gamma_X^{1/2}\Phi$, where $V \in \mathbb{R}^{d\times d}$ with $V^\top V = VV^\top = I$. With this we can write:

$$(\Phi^\top\Gamma_X\Phi + \frac{\lambda}{2m}I)^{-1}\Phi^\top\Gamma_X^{1/2} = V\left(\Sigma^2 + \frac{\lambda}{2m}I\right)^{-1}\Sigma U^\top \ . \tag{27}$$

Now we use that $\|U\| = \|V\| = 1$ and $x/(x^2+a) \leq 1/(2\sqrt{a})$ for any $x \geq 0$ to get $\|V(\Sigma^2 + (\lambda/2m)I)^{-1}\Sigma U^\top\| \leq \sqrt{m/2\lambda}$. Thus we get a bound for $\|\theta_X^\lambda\|_2$ that when plugged into (26) yields the desired result. □

**Lemma 8.** *The following holds for every $v \in \mathbb{N}^{\mathcal{S}}$:*

$$\varphi_k^\lambda(v) = \left(\frac{\|\Phi\|\|\rho\|_\infty}{\sqrt{2\lambda}}\sqrt{\sum_{s\in\mathcal{S}}\rho_s\max\{v_s + k, m\}} + \|\rho\|_2\right)^2 \ .$$

*Furthermore, for every $k \geq m - \min_s v_s$ we have $\varphi_k^\lambda(v) = \left(\frac{\|\Phi\|\|\rho\|_\infty\sqrt{m}}{\sqrt{2\lambda}}\sqrt{\sum_{s\in\mathcal{S}}\rho_s} + \|\rho\|_2\right)^2$.*

*Proof.* The proof is similar to that of Lemma 5 and is omitted. □

# 4 Utility Analysis of DP-LSW

The goal of this section is to show that as the size $m$ of the dataset $X$ grows, the differentially private solution $\theta_X^w$ provided by algorithm DP-LSW is not much worse than the one obtained by directly minimizing $J_X^w(\theta)$. In other words, for large datasets the noise introduced by the privacy constraint is negligible. We do so by proving a $O(1/m^2)$ bound for the expected empirical excess risk given by $\mathbb{E}_{X,\eta}[J_X^w(\hat{\theta}_X^w) - J_X^w(\theta_X^w)]$. Our analysis starts with a lemma that leverages the law of total expectation in order to reduce the bound to a quantity that only depends on $\mathbb{E}_X[\sigma_X^2]$.

**Lemma 9.**

$$\mathbb{E}_{X,\eta}[J_X^w(\hat{\theta}_X^w) - J_X^w(\theta_X^w)] = \|\Gamma^{1/2}\Phi\|_F^2\mathbb{E}_X[\sigma_X^2] \ . \tag{28}$$

*Proof.* By the law of total expectation it is enough to show that

$$\mathbb{E}_\eta[J_X^w(\hat{\theta}_X^w) - J_X^w(\theta_X^w)|X] = \sigma_X^2 \|\Gamma^{1/2}\Phi\|_F^2 \ . \tag{29}$$

Let $X$ be an arbitrary dataset. Expanding the definition of $J_X^w(\theta)$ we have that for any $\theta \in \mathbb{R}^d$

$$J_X^w(\theta) = F_X^\top \Gamma F_X + \theta^\top \Phi^\top \Gamma \Phi \theta - 2F_X^\top \Gamma \Phi \theta \ . \tag{30}$$

On the other hand, since $\nabla_\theta J_X^w(\theta_X^w) = 0$, we have ${\theta_X^w}^\top \Phi^\top \Gamma \Phi = F_X^\top \Gamma \Phi$. Thus, using the definition $\hat{\theta}_X^w = \theta_X^w + \eta$, a simple algebraic calculation yields

$$J_X^w(\hat{\theta}_X^w) - J_X^w(\theta_X^w) = \eta^\top \Phi^\top \Gamma \Phi \eta - F_X^\top \Gamma \Phi \eta - \eta^\top \Phi^\top \Gamma \Phi \theta_X^w \ . \tag{31}$$

Finally, taking the expectation over $\eta \sim \mathcal{N}(0, \sigma_X^2 I)$ of the above expression we get

$$\mathbb{E}_\eta[J_X^w(\hat{\theta}_X^w) - J_X^w(\theta_X^w)] = \mathbb{E}_\eta[\eta^\top \Phi^\top \Gamma \Phi \eta] = \sigma_X^2 \operatorname{Tr}(\Phi^\top \Gamma \Phi) = \sigma_X^2 \|\Gamma^{1/2}\Phi\|_F^2 \ . \tag{32}$$

□

In order to bound $\mathbb{E}_X[\sigma_X^2]$ we recall the variance has the form $\sigma_X^2 = C^2 \psi_X^w$, where $C$ is a constant independent of $X$ and

$$\psi_X^w = \max_{k \geq 0} e^{-k\beta} \sum_{s \in \mathcal{S}} \frac{w_s}{\max\{|X_s| - k, 1\}^2} \leq \sum_s w_s \left( \max_{k \geq 0} \frac{e^{-k\beta}}{\max\{|X_s| - k, 1\}^2} \right) \ . \tag{33}$$

Thus, we can bound $\mathbb{E}_X[\sigma_X^2] = C^2 \mathbb{E}_X[\psi_X^w]$ by providing a bound for the expectation of each individual maximum in (33). The two following technical lemmas will prove useful.

**Lemma 10.** *Let $b > 0$ and $a \geq 1$. Then the following holds:*

$$\max_{0 \leq x \leq a-1} \frac{e^{-bx}}{(a-x)^2} = \begin{cases} \frac{1}{a^2} & b < 2/a \\ e^{1-ab} & b > 2 \\ \frac{e^2}{4} b^2 e^{-ab} & \textit{otherwise} \end{cases} \tag{34}$$

*Proof.* The result follows from a simple calculation. □

**Lemma 11.** *Suppose $B_{m,p}$ is a binomial random variable with $m$ trials and success probability $p$. Then the following hold:*

$$\mathbb{E}\left[\frac{1}{B_{m,p}+1}\right] = \frac{1-(1-p)^{m+1}}{p(m+1)} \ ,$$
$$\mathbb{E}\left[\frac{1}{B_{m,p}^2}\mathbb{I}_{B_{m,p}\geq 1}\right] \leq \frac{6}{p(m+1)} \left( \frac{1-(1-p)^{m+2}}{p(m+2)} - (1-p)^{m+1} - \frac{p(m+1)}{2}(1-p)^m \right) \ .$$

*Proof.* The first expectation is a classical exercise in probability textbooks. The second one can be proved as follows:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{B_{m,p}^2}\mathbb{I}_{B_{m,p}\geq 1}\right] &= \sum_{k=1}^{m}\frac{1}{k^2}\binom{m}{k}p^k(1-p)^{m-k} \\
&\leq 6\sum_{k=1}^{m}\frac{1}{(k+1)(k+2)}\binom{m}{k}p^k(1-p)^{m-k} \\
&= \frac{6}{p(m+1)}\sum_{k=1}^{m}\frac{1}{k+2}\frac{(m+1)!}{(k+1)!(m-k)!}p^{k+1}(1-p)^{m-k} \\
&= \frac{6}{p(m+1)}\sum_{k=1}^{m}\frac{1}{k+2}\mathbb{P}[B_{m+1,p}=k+1] \\
&= \frac{6}{p(m+1)}\sum_{j=2}^{m+1}\frac{1}{j+1}\mathbb{P}[B_{m+1,p}=j] \\
&= \frac{6}{p(m+1)}\left(\mathbb{E}\left[\frac{1}{B_{m+1,p}+1}\right]-\mathbb{P}[B_{m+1,p}=0]-\frac{1}{2}\mathbb{P}[B_{m+1,p}=1]\right) \\
&= \frac{6}{p(m+1)}\left(\frac{1-(1-p)^{m+2}}{p(m+2)}-(1-p)^{m+1}-\frac{p(m+1)}{2}(1-p)^m\right) \ ,
\end{aligned}$$

where we used the first equation in the last step, and the bound $(k+1)(k+2)/k^2 \leq 6$ for $k \geq 1$ in the first inequality. □

Recall that $p_s$ denotes the probability that a trajectory from $X$ visits states $s$. Because these trajectories are i.i.d. we have that $|X_s| = B_{m,p_s}$ is a binomial random variable. Therefore, we can combine the last two lemmas to prove the following.

**Lemma 12.** *Suppose $\beta \leq 2$. Then we have:*

$$\mathbb{E}_X\left[\max_{k\geq 0}\frac{e^{-k\beta}}{\max\{|X_s|-k,1\}^2}\right] \leq \begin{cases} \frac{6}{p_s^2(m+1)(m+2)}+\frac{e^2\beta^2}{4}(1-(1-e^{-\beta})p_s)^m & p_s>0 \ , \\ 1 & p_s=0 \ . \end{cases} \tag{35}$$

*Proof.* Note in the first place that Lemma 10 implies

$$\max_{k\geq 0}\frac{e^{-k\beta}}{\max\{|X_s|-k,1\}^2} = \mathbb{I}_{|X_s|=0}+\mathbb{I}_{1\leq|X_s|<2/\beta}\frac{1}{|X_s|^2}+\mathbb{I}_{|X_s|\geq 2/\beta}\frac{e^2}{4}\beta^2e^{-\beta|X_s|} \ , \tag{36}$$

where we used that in the case $|X_s| = 0$ the maximum is 1. If $p_s = 0$, then obviously $|X_s| = 0$ almost surely and the expectation of (36) equals 1. On the other hand, when $p_s > 0$ we use the linearity of expectation and bound each term separately. Clearly, $\mathbb{E}_X[\mathbb{I}_{|X_s|=0}] = \mathbb{P}_X[B_{m,p_s} = 0] = (1-p_s)^m$. On the other hand, by looking up the moment generating function of a binomial distribution we have

$$\mathbb{E}_X[\mathbb{I}_{|X_s|\geq 2/\beta}\frac{e^2}{4}\beta^2e^{-\beta|X_s|}] \leq \frac{e^2}{4}\beta^2\mathbb{E}_X[e^{-\beta|X_s|}] = \frac{e^2}{4}\beta^2(1-(1-e^{-\beta})p_s)^m \ . \tag{37}$$

The remaining term is bounded by

$$\mathbb{E}_X\left[\mathbb{I}_{1\leq|X_s|<2/\beta}\frac{1}{|X_s|^2}\right] \leq \mathbb{E}_X\left[\mathbb{I}_{1\leq|X_s|}\frac{1}{|X_s|^2}\right] \ . \tag{38}$$

Therefore, applying Lemma 11 and upper bounding some negative terms by zero, we get

$$\mathbb{E}_X\left[\max_{k\geq 0}\frac{e^{-k\beta}}{\max\{|X_s|-k,1\}^2}\right] \leq \frac{6}{p_s^2(m+1)(m+2)}+\frac{e^2\beta^2}{4}(1-(1-e^{-\beta})p_s)^m \ . \tag{39}$$

□

Now we can combine Lemmas 16 and 12 using Equation 33 to get our final result.

**Theorem 13.** *Let $\mathcal{S}_0 = \{s \in \mathcal{S} | p_s = 0\}$ and $S_+ = \mathcal{S} \setminus \mathcal{S}_0$. Let $C = \alpha R_{\max} \|(\Gamma^{1/2}\Phi)^\dagger\| \|\Gamma^{1/2}\Phi\|_F/(1-\gamma)$. Suppose $\beta \leq 2$. Then we have the following:*

$$\mathbb{E}_{X,\eta}[J_X^w(\hat{\theta}_X^w) - J_X^w(\theta_X^w)] \leq C^2 \left( \sum_{s \in \mathcal{S}_0} w_s + \sum_{s \in \mathcal{S}_+} w_s \left( \frac{6}{p_s^2 (m+1)(m+2)} + \frac{e^2 \beta^2}{4} (1 - (1 - e^{-\beta}) p_s)^m \right) \right) \quad .$$

The following version is the one given in the paper for reasons of space. It is easily obtained by noting that $e^2/4 \leq 6$, $m^2 \leq (m+1)(m+2)$, and when $\beta \leq 1/2$ then $1 - (1 - e^{-\beta}) p_s \leq 1 - \beta p_s/2$.

**Corollary 14.** *Let $\mathcal{S}_0 = \{s \in \mathcal{S} | p_s = 0\}$ and $S_+ = \mathcal{S} \setminus \mathcal{S}_0$. Let $C = \alpha R_{\max} \|(\Gamma^{1/2}\Phi)^\dagger\| \|\Gamma^{1/2}\Phi\|_F/(1-\gamma)$. Suppose $\beta \leq 1/2$. Then $\mathbb{E}_{X,\eta}[J_X^w(\hat{\theta}_X^w) - J_X^w(\theta_X^w)]$ is upper bounded by:*

$$C^2 \left( \sum_{s \in \mathcal{S}_0} w_s + 6 \sum_{s \in \mathcal{S}_+} w_s \left( \frac{1}{p_s^2 m^2} + \beta^2 \left( 1 - \frac{\beta p_s}{2} \right)^m \right) \right) \quad .$$

The following is an immediate consequence of these results.

**Corollary 15.** *If $w_s = 0$ for all $s \in \mathcal{S}_0$, then $\mathbb{E}_{X,\eta}[J_X^w(\hat{\theta}_X^w) - J_X^w(\theta_X^w)] = O(1/m^2)$.*

# 5 Utility Analysis of DP-LSL

The analysis in this section follows a scheme similar to the previous one. We start by taking the expectation of the excess empirical risk with respect to the Gaussian perturbation $\eta$.

**Lemma 16.**

$$\mathbb{E}_{X,\eta}[J_X^\lambda(\hat{\theta}_X^\lambda) - J_X^\lambda(\theta_X^\lambda)] = \mathbb{E}_X \left[ \left( \frac{\lambda d}{2m} + \frac{1}{m} \sum_{s \in \mathcal{S}} \rho_s \|\phi_s\|_2^2 |X_s| \right) \sigma_X^2 \right] \quad . \tag{40}$$

*Proof.* Let $X$ be an arbitrary dataset with $m$ trajectories. Recalling that $\hat{\theta}_X^\lambda = \theta_X^\lambda + \eta$ we get:

$$\begin{aligned} J_X^\lambda(\hat{\theta}_X^\lambda) - J_X^\lambda(\theta_X^\lambda) &= \frac{1}{m} \sum_{i=1}^m \sum_{s \in \mathcal{S}_{x_i}} \rho_s \left( (\phi_s^\top \hat{\theta}_X^\lambda)^2 - (\phi_s^\top \theta_X^\lambda)^2 - 2 F_{x_i,s} \phi_s^\top \eta \right) + \frac{\lambda}{2m} \left( \|\hat{\theta}_X^\lambda\|_2^2 - \|\theta_X^\lambda\|_2^2 \right) \\ &= \frac{1}{m} \sum_{i=1}^m \sum_{s \in \mathcal{S}_{x_i}} \rho_s \left( \eta^\top \phi_s \phi_s^\top \eta + 2 \eta^\top \phi_s \phi_s^\top \theta_X^\lambda - 2 F_{x_i,s} \phi_s^\top \eta \right) + \frac{\lambda}{2m} \left( \|\eta\|_2^2 + 2 \eta^\top \theta_X^\lambda \right) \quad . \end{aligned}$$

Taking the expectation over $\eta \sim \mathcal{N}(0, \sigma_X^2 I)$ in the above expression we get

$$\mathbb{E}_\eta[J_X^\lambda(\hat{\theta}_X^\lambda) - J_X^\lambda(\theta_X^\lambda)] = \frac{1}{m} \sum_{i=1}^m \sum_{s \in \mathcal{S}_{x_i}} \rho_s \operatorname{Tr}(\phi_s \phi_s^\top) \sigma_X^2 + \frac{\lambda}{2m} d \sigma_X^2 \quad .$$

The result now follows from noting that $\sum_{i=1}^m \sum_{s \in \mathcal{S}_{x_i}} \rho_s \operatorname{Tr}(\phi_s \phi_s^\top) = \sum_{s \in \mathcal{S}} \rho_s \|\phi_s\|_2^2 |X_s|$. □

In order to bound the expression given by previous lemma we will expand the definition of $\sigma_X = C_\lambda \sqrt{\psi_X^\lambda}$, with $C_\lambda = 2R_{\max}\|\Phi\|/(1-\gamma)(\lambda - \|\Phi\|^2 \|\rho\|_\infty)$, and note that using the straightforward bound $(a+b)^2 \leq 2a^2 + 2b^2$ we

have:

$$\psi_X^\lambda = \max_{k\geq 0} e^{-k\beta}\left(\|\rho\|_2 + \frac{\|\Phi\|\|\rho\|_\infty}{\sqrt{2\lambda}}\sqrt{\sum_{s\in\mathcal{S}}\rho_s \min\{|X_s|+k,m\}}\right)^2$$
$$\leq 2\|\rho\|_2^2 + \frac{\|\Phi\|^2\|\rho\|_\infty^2}{\lambda}\sum_{s\in\mathcal{S}}\rho_s \max_{k\geq 0}\left(e^{-2k\beta}\min\{|X_s|+k,m\}\right) \quad .$$

The following lemma can be used to bound the maximums inside this sum.

**Lemma 17.** *Suppose $a \geq 0$ and $b > 0$. Then the following holds:*

$$\max_{0\leq x\leq m-a} e^{-2bx}(a+x) = \begin{cases} a & b < a/2 \\ me^{-2b(m-a)} & b > m/2 \\ \frac{1}{2eb}e^{2ab} & \textit{otherwise} \end{cases} \tag{41}$$

Assuming we have $2\beta < 1 \leq m$, previous lemma yields:

$$\max_{k\geq 0}\left(e^{-2k\beta}\min\{|X_s|+k,m\}\right) = |X_s|\mathbb{I}_{|X_s|>2\beta} + \frac{1}{2e\beta}e^{2\beta|X_s|}\mathbb{I}_{|X_s|\leq 2\beta} \leq |X_s| + \frac{1}{2e\beta}\mathbb{I}_{|X_s|=0} \quad . \tag{42}$$

When taking the expectation of the upper bound for (40) obtained by plugging in (42), several quantities involving products of correlated binomial random variables will appear. Next lemma gives expressions for all these expectations.

**Lemma 18.** *Recall that $p_s = \mathbb{P}[s \in x]$ and $|X_s|$ is a binomial random variable with $m$ trials and success probability $p_s$. Define $p_{s,s'} = \mathbb{P}[s \in x \wedge s' \in x]$ and $\bar{p}_{s,s'} = \mathbb{P}[s \in x \wedge s' \notin x]$ for any $s, s' \in \mathcal{S}$. Then we have the following:*

1. $\mathbb{E}[|X_s|] = mp_s$,
2. $\mathbb{E}[\mathbb{I}_{|X_s|=0}] = (1-p_s)^m$,
3. $\mathbb{E}[|X_s|^2] = m^2p_s^2 + m(p_s - p_s^2)$,
4. $\mathbb{E}[|X_s||X_{s'}|] = m(m-1)p_sp_{s'} + mp_{s,s'}$,
5. $\mathbb{E}[|X_s|\mathbb{I}_{|X_{s'}|=0}] = m\bar{p}_{s,s'}(1-p_{s'})^{m-1}$.

*Proof.* All equations follow from straightforward calculations. □

**Theorem 19.** *Suppose $\beta < 1/2$ and $\lambda > \|\Phi\|^2\|\rho\|_\infty$. Let $C_\lambda = 2\alpha R_{\max}\|\Phi\|/(1-\gamma)(\lambda - \|\Phi\|^2\|\rho\|_\infty)$. Then we have*

$$\mathbb{E}_{X,\eta}[J_X^\lambda(\hat{\theta}_X^\lambda) - J_X^\lambda(\theta_X^\lambda)] \leq C_\lambda^2\left\{\sum_{s\in\mathcal{S}}\rho_s p_s\left(\frac{d\|\Phi\|^2\|\rho\|_\infty^2}{2} + 2\|\rho\|_2^2\|\phi_s\|_2^2\right)\right.$$
$$+ \frac{\lambda}{m}d\|\rho\|_2^2 + \frac{1}{m}\frac{d\|\Phi\|^2\|\rho\|_\infty^2}{4e\beta}\sum_{s\in\mathcal{S}}\rho_s(1-p_s)^m + \frac{m}{\lambda}\|\Phi\|^2\|\rho\|_\infty^2\sum_{s,s'\in\mathcal{S}}\rho_s\rho_{s'}p_sp_{s'}\|\phi_s\|_2^2$$
$$\left. + \frac{1}{\lambda}\|\Phi\|^2\|\rho\|_\infty^2\left(\sum_{s\in\mathcal{S}}\rho_s^2\|\phi_s\|_2^2(p_s-p_s^2) + \sum_{\substack{s,s'\in\mathcal{S}\\ s\neq s'}}\rho_s\rho_{s'}\|\phi_s\|_2^2\left(p_{s,s'} - p_sp_{s'} + \frac{1}{2e\beta}\bar{p}_{s,s'}(1-p_{s'})^{m-1}\right)\right)\right\} \quad .$$

*Proof.* Combining Lemma 16 with (42) and the definition of $\sigma_X^2$ yields the following upper bound for $\mathbb{E}_{X,\eta}[J_X^\lambda(\hat{\theta}_X^\lambda) - J_X^\lambda(\theta_X^\lambda)]$:

$$C_\lambda^2 \mathbb{E}_X \left[ \left( \frac{\lambda d}{2m} + \frac{1}{m} \sum_{s \in \mathcal{S}} \rho_s \|\phi_s\|_2^2 |X_s| \right) \left( 2\|\rho\|_2^2 + \frac{\|\Phi\|^2 \|\rho\|_\infty^2}{\lambda} \sum_{s \in \mathcal{S}} \rho_s \left( |X_s| + \frac{1}{2e\beta} \mathbb{I}_{|X_s|=0} \right) \right) \right] \quad .$$

Terms that do not involve products of the form $|X_s||X_{s'}|$ or $|X_s|\mathbb{I}_{|X_{s'}|=0}$ can be straightforwardly reduced to linear combinations of expectations in Lemma 18. The remaining term yields the following:

$$\begin{aligned}
&\mathbb{E}_X \left[ \sum_{s,s' \in \mathcal{S}} \rho_s \rho_{s'} \|\phi_s\|_2^2 |X_s| \left( |X_{s'}| + \frac{1}{2e\beta} \mathbb{I}_{|X_{s'}|=0} \right) \right] \\
&= \sum_{s \in \mathcal{S}} \rho_s^2 \|\phi_s\|_2^2 \mathbb{E}_X \left[ |X_s| \left( |X_s| + \frac{1}{2e\beta} \mathbb{I}_{|X_s|=0} \right) \right] \\
&\quad + \sum_{\substack{s,s' \in \mathcal{S} \\ s \neq s'}} \rho_s \rho_{s'} \|\phi_s\|_2^2 \mathbb{E}_X \left[ |X_s| \left( |X_{s'}| + \frac{1}{2e\beta} \mathbb{I}_{|X_{s'}|=0} \right) \right] \\
&= \sum_{s \in \mathcal{S}} \rho_s^2 \|\phi_s\|_2^2 \left( m^2 p_s^2 + m(p_s - p_s^2) \right) \\
&\quad + \sum_{\substack{s,s' \in \mathcal{S} \\ s \neq s'}} \rho_s \rho_{s'} \|\phi_s\|_2^2 \left( m(m-1) p_s p_{s'} + m p_{s,s'} + \frac{1}{2e\beta} m \bar{p}_{s,s'} (1 - p_{s'})^{m-1} \right) \quad ,
\end{aligned}$$

where we used Lemma 18 again. Thus we get:

$$\begin{aligned}
\mathbb{E}_{X,\eta}[J_X^\lambda(\hat{\theta}_X^\lambda) - J_X^\lambda(\theta_X^\lambda)] \leq C_\lambda^2 \Bigg\{ & \frac{\lambda d \|\rho\|_2^2}{m} + \frac{d \|\Phi\|^2 \|\rho\|_\infty^2}{2m} \sum_{s \in \mathcal{S}} \rho_s \left( m p_s + \frac{1}{2e\beta} (1 - p_s)^m \right) \\
& + \frac{2\|\rho\|_2^2}{m} \sum_{s \in \mathcal{S}} \rho_s p_s \|\phi_s\|_2^2 m + \frac{\|\Phi\|^2 \|\rho\|_\infty^2}{\lambda m} \sum_{s \in \mathcal{S}} \rho_s^2 \|\phi_s\|_2^2 \left( m^2 p_s^2 + m(p_s - p_s^2) \right) \\
& + \frac{\|\Phi\|^2 \|\rho\|_\infty^2}{\lambda m} \sum_{\substack{s,s' \in \mathcal{S} \\ s \neq s'}} \rho_s \rho_{s'} \|\phi_s\|_2^2 \left( m(m-1) p_s p_{s'} + m p_{s,s'} + \frac{1}{2e\beta} m \bar{p}_{s,s'} (1 - p_{s'})^{m-1} \right) \Bigg\}
\end{aligned}$$

The final result is obtained by grouping the terms in this expression by their dependence in $\lambda$ and $m$. □

Note that if we take $\lambda = \omega(1)$ with respect to $m$ in the above theorem, then $C_\lambda = O(1/\lambda)$ and we get the following corollary.

**Corollary 20.** *Suppose $\lambda = \omega(1)$ with respect to $m$. Then we have*

$$\mathbb{E}_{X,\eta}[J_X^\lambda(\hat{\theta}_X^\lambda) - J_X^\lambda(\theta_X^\lambda)] = O\left( \frac{1}{\lambda m} + \frac{1}{\lambda^2} + \frac{m}{\lambda^3} \right) \quad . \tag{43}$$

# References


[1] Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, pages 1302–1338, 2000.

[2] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis, 2011.