Journal of Machine Learning **ISSN**: 2790-2048(e), 2790-203X(p)

Tangent Differential Privacy

Lexing Ying *

Department of Mathematics, Stanford University, Stanford, CA 94305, USA.

Abstract. Differential privacy is a framework for protecting the identity of individual data points in the decision-making process. In this note, we propose a new form of differential privacy, known as tangent differential privacy. Compared to the usual differential privacy, which is defined uniformly across data distributions, tangent differential privacy is tailored to a specific data distribution of interest. It also allows for general distribution distances such as total variation distance and Wasserstein distance. In the context of risk minimization, we demonstrate that entropic regularization ensures tangent differential privacy under relatively general conditions on the risk function.

Keywords:
Differential privacy,
Entropic regularization.

Article Info.:
Volume: X
Number: X
Pages: 1 - 9
Date: /2025
doi.org/10.4208/jml.240928

Article History:
Received: 28/09/2024
Accepted: 12/04/2025

Communicated by:
Song Mei

1 Introduction

Differential privacy is a framework for protecting the identity of individual data points in the machine learning process. The most commonly discussed differential privacy is ϵ-differential privacy. A randomized algorithm is called ϵ-differential private if, for any two input data distributions that differ by one element, the ratio of the probabilities at any outcome is bounded by at most $\exp(\epsilon)$. The definition clearly shows that differential privacy is a uniform concept across all data distributions. In many machine learning applications, one often cares about a specific data distribution and raises privacy concerns about the impact of deleting or adding a single or small number of data points to this specific data distribution.

To address such questions, we propose here tangent differential privacy, a privacy concept tailored to a specific data distribution. When applying the case of risk minimization (such as supervised learning), we demonstrate that entropic regularization guarantees tangent differential privacy under relatively general conditions.

Related work. The concepts of ϵ-differential privacy and (ϵ, δ)-differential privacy were first proposed in [9,10] and a wonderful resource for this vast literature is [11]. Several efforts have been devoted to relax or reformulate differential privacy, with examples including Renyi differential privacy [17], concentrated differential privacy [6,12], and Lipschitz privacy [15]. In a broader context, other related forms of privacy concepts have also been developed, such as local differential privacy [8,13,14] and the recently proposed metric privacy [4,5]. The concept of tangent differential privacy proposed here is closely

*Corresponding author. lexing@stanford.edu

related to Lipschitz privacy, though the latter is defined as a uniform concept across all data distributions.

Contents. The rest of the note is organized as follows. Section 2 introduces the concept of tangent differential privacy. Section 3 considers the risk minimization problem and proposes entropic regularization as a solution of tangent differential privacy for both total variation and Wasserstein distances. Section 4 concludes with some discussions.

2 Tangent differential privacy

Let X be the metric space of the data points, and $\mathcal{P}(X)$ be the space of distributions over X. Let W be the metric space of outputs, $\mathcal{P}(W)$ be the space of distributions over W, and $\mathcal{F}(W)$ be the space of bounded functions over W. Here, the output space W can be quite general, such as $\mathbb{R}^n$, the space of regression functions, or the space of neural network weights [1]. To discuss differential privacy, let A be a randomized algorithm that takes $p \in \mathcal{P}(X)$ and produces a randomized output w. Because A is random, we can regard it as a (typically nonlinear) map

$$A : \mathcal{P}(X) \;\; \to \;\; \mathcal{P}(W),$$

taking $p(x)$ to a distribution $q(w)$. When $q(w)$ has a bounded density, we can also consider its logarithm

$$\log \circ A : \mathcal{P}(X) \;\; \to \;\; \mathcal{F}(W),$$

taking $p(x)$ to a function $(\log q)(w)$.

Let us denote T_p and T_q as the tangent spaces of signed measures at p and q, respectively. The tangent map of A at p is $DA_p : T_p \to T_q$. Suppose that p is the data distribution of interest. For any $\tilde{p}$ close to p, the linear approximation suggests that

$$A\tilde{p} - Ap \approx DA_p \cdot (\tilde{p} - p). \tag{2.1}$$

In the usual setting, p can be an empirical distribution with N data samples $\{x_i\}$ and $\tilde{p}$ is obtained by removing a distinguished sample x_k

$$p(x) = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}(x), \quad \tilde{p}(x) = \frac{1}{N-1}\sum_{i \neq k} \delta_{x_i}(x).$$

This also extends naturally to the situation where $\tilde{p}$ is obtained from p by changing a small number of data points.

Similarly, if $T_{\log q}$ is the tangent space at $\log q$, the tangent map of $\log \circ A$ at p is

$$D(\log \circ A)_p : T_p \;\; \to \;\; T_{\log q}.$$

For any $\tilde{p}$ close to p, we have

$$\log(A\tilde{p}) - \log(Ap) \approx D(\log \circ A)_p \cdot (\tilde{p} - p). \tag{2.2}$$

When $\mathcal{P}(X)$ and $\mathcal{P}(W)$ are endowed with distances, the distance functions induce corresponding norms on T_p, T_q, and $T_{\log q}$. For $\mathcal{P}(W)$, the most relevant distance function to differential privacy is the total variation distance

$$d_{\mathrm{TV}}(q, \tilde{q}) = 2 \cdot \max_{S \subset W} |q(S) - \tilde{q}(S)|.$$

Here, we introduce an additional factor of 2 to ensure consistency with the L^1 norm. The corresponding norm for T_q is the L^1 norm $\|\epsilon\|_{L^1} := \int_W |\epsilon(w)| dw$ and the one for $T_{\log q}$ is the L^∞ norm $\|\epsilon\|_{L^\infty} := \sup_{w \in W} |\epsilon(w)|$. For $\mathcal{P}(X)$, there are two common choices for the distance and the corresponding norm $\|\cdot\|_{T_p}$ at p.

Example 2.1. Consider the total variation distance on $\mathcal{P}(X)$, i.e.

$$d_{\mathrm{TV}}(p, \tilde{p}) = 2 \cdot \max_{S \subset X} |p(S) - \tilde{p}(S)|.$$

This setup results in the L^1 norm for T_p

$$\|\epsilon\|_{L^1} := \int_X |\epsilon(x)| dx.$$

Example 2.2. Consider the Wasserstein-2 distance on $\mathcal{P}(X)$ and still the total variation distance on $\mathcal{P}(W)$. This results in the following weighted Sobolev norm for T_p [18]:

$$\|\epsilon\|^2_{\dot{H}^{-1}(p)} := \min_{f : \nabla \cdot (fp) = \epsilon} \int_X |f(x)|^2 p(x) dx, \quad \epsilon \in T_p.$$

With these preparations, we are ready to introduce the following definitions.

Definition 2.1. *A is Lipschitz at p for the norm pair $(\|\cdot\|_{T_p}, L^1)$ with bound C_p if*

$$\|DA_p\|_{\|\cdot\|_{T_p} \to L^1} \leq C_p. \tag{2.3}$$

For $\tilde{p}$ with $\|\tilde{p} - p\|_{T_p}$ small, using (2.1) and (2.3) leads to

$$\|Ap - A\tilde{p}\|_{L^1} \leq \big(C_p + o(1)\big) \|\tilde{p} - p\|_{T_p},$$

i.e. for any set $S \subset W$,

$$|(Ap)(S) - (A\tilde{p})(S)| \leq \big(C_p + o(1)\big) \|\tilde{p} - p\|_{T_p}. \tag{2.4}$$

Definition 2.2. *A is log-Lipschitz or satisfies tangent differential privacy at p for the norm pair $(\|\cdot\|_{T_p}, L^\infty)$ with bound C_p if*

$$\|D(\log \circ A)_p\|_{\|\cdot\|_{T_p} \to L^\infty} \leq C_p. \tag{2.5}$$

For $\tilde{p}$ with $\|\tilde{p} - p\|_{T_p}$ small, using (2.2) and (2.5) leads to

$$\|\log(Ap) - \log(A\tilde{p})\|_{L^\infty} = \left\| \log\left(\frac{Ap}{A\tilde{p}}\right) \right\|_{L^\infty} \leq \big(C_p + o(1)\big) \|\tilde{p} - p\|_{T_p},$$

i.e. for any $w \in W$,

$$\exp\big(-\big(C_p+o(1)\big)\|\tilde{p}-p\|_{T_p}\big) \le \frac{(Ap)(w)}{(A\tilde{p})(w)} \le \exp\big(\big(C_p+o(1)\big)\|\tilde{p}-p\|_{T_p}\big).$$

Therefore, for any set $S \subset W$,

$$\exp\big(-\big(C_p+o(1)\big)\|\tilde{p}-p\|_{T_p}\big) \le \frac{\int_S (Ap)(w)dw}{\int_S (A\tilde{p})(w)dw} \le \exp\big(\big(C_p+o(1)\big)\|\tilde{p}-p\|_{T_p}\big). \quad (2.6)$$

This is a more quantitative version of differential privacy adapted to the data distribution p.

Remark 2.1. (a) By working directly with the space of distributions $\mathcal{P}(X)$, the concept of tangent differential privacy is defined without direct reference to the number of data samples in the distribution. Therefore, it allows for changing either a single data sample or a small fraction of samples.

(b) Working with different distances on $\mathcal{P}(X)$ leads to different types of privacy considerations. For example, the total variation distance on $\mathcal{P}(X)$ corresponds to the Hamming distance case of the ϵ-differential privacy. The Wasserstein distance case is related to the metric privacy setup.

(c) One key property of the standard differential privacy is the postprocessing property, i.e. any operation of the output w does not invalidate the differential privacy property. Here, if A is Lipschitz or satisfies tangent differential privacy (i.e. log-Lipschitz), the postprocessing property still holds. This can be seen from the fact that any postprocessing does not change the L^1 norm (2.4) and the L^∞ norm (2.6).

(d) Another important feature of standard differential privacy is the composition property. It is not difficult to demonstrate that this also holds here. Let $A : \mathcal{P}(X) \to \mathcal{P}(W)$ and $A' : P(X) \to \mathcal{P}(W')$ be Lipschitz with constant C_p and C'_p, respectively. The composition $A \times A' : \mathcal{P}(X) \to \mathcal{P}(W \times W')$ maps to the product measure, i.e. $(A \times A')(p) = (Ap) \otimes (A'p)$. Then, for $\tilde{p}$ close to p

$$\begin{aligned} &\|(A \times A')(p) - (A \times A')(\tilde{p})\|_{L^1} \\ &= \|(Ap) \otimes (A'p) - (A\tilde{p}) \otimes (A'\tilde{p})\|_{L^1} \\ &\le \big(C_p + C'_p + o(1)\big)\|p-\tilde{p}\|_{T_p} \end{aligned}$$

shows that $A \times A'$ is also Lipschitz with constant $C_p + C'_p$.

Let $A : \mathcal{P}(X) \to \mathcal{P}(W)$ and $A' : P(X) \to \mathcal{P}(W')$ be log-Lipschitz (i.e. satisfy tangent differential privacy) with constant C_p and C'_p, respectively, then for $\tilde{p}$ close to p,

$$\begin{aligned} &\|\log(A \times A')(p) - \log(A \times A')(\tilde{p})\|_{L^\infty} \\ &= \|\log(Ap) + \log(A'p) - \log(A\tilde{p}) - \log(A'\tilde{p})\|_{L^\infty} \\ &\le \big(C_p + C'_p + o(1)\big)\|p-\tilde{p}\|_{T_p} \end{aligned}$$

showing that $A \times A'$ is also log-Lipschitz (i.e. satisfies tangent differential privacy) with constant $C_p + C'_p$.

3 Risk minimization

We consider the case where the output in W is obtained via an optimization procedure, such as empirical risk minimization. Given the data distribution $p(x)$ and risk function $r(w, x) \geq 0$, the goal is

$$\min_w \int r(w, x)p(x)dx.$$

The solution w of this minimization problem depends deterministically on $p(x)$. To discuss differential privacy, one needs to consider a randomized algorithm with its output distributed over W. Here, we propose to adopt entropic regularization following [16] and seek $q(w) \in \mathcal{P}(W)$:

$$q = \operatorname*{argmin}_{q \in \mathcal{P}(W)} \int q(w) \Big(\int r(w, x)p(x)dx \Big) dw + \beta^{-1} \int q(w) \ln q(w) dw.$$

The solution is the Gibbs distribution

$$q(w) = \frac{\exp\Big(-\beta \int r(w, x)p(x)dx \Big)}{\int_W \exp\Big(-\beta \int r(w', x)p(x)dx \Big) dw'}, \tag{3.1}$$

simply written as $q(w) \propto \exp(-\beta \int r(w, x)p(x)dx)$. Then, the map $A : \mathcal{P}(X) \to \mathcal{P}(W)$ takes from $p(x)$ to $q(w)$.

Remark 3.1. The distribution (3.1) can, in principle, be sampled using Monte Carlo methods, such as Langevin dynamics. One popular differentially private algorithm is noisy-SGD [2, 3], and there is a close connection between noisy-SGD and Langevin dynamics [7, 19].

Fixing $p(x)$, let us compute the differential $DA_p : T_p \to T_q$. Its kernel as a function of (w, x) is given by

$$-\beta \int \big(q(w)\delta(w - w') - q(w)q(w') \big) r(w', x)dw'.$$

When X and W are finite sets, this can be written in the matrix form as

$$-\beta \big(\operatorname{diag}(q) - qq^\top \big) r,$$

where here r denotes a matrix with value $r(w, x)$ at entry (w, x).

The differential of $D(\log \circ A)_p : T_p \to T_{\log q}$ can also be computed easily with the chain rule. Its kernel as a function of (w, x) is

$$-\beta \int \big(\delta(w - w') - q(w') \big) r(w', x)dw'.$$

Again, when X and W are finite sets, the matrix form is $-\beta(I - \mathbf{1}q^\top)r$, where $\mathbf{1}$ stands for the all-one column vector. Below, we show that the entropic regularization guarantees tangent differential privacy under rather general conditions for both the TV distance and the Wasserstein distance on $\mathcal{P}(X)$.

3.1 Total variation distance on $\mathcal{P}(X)$

Recall from Example 2.1 that one has L^1 norm for T_p, L^1 norm for T_q, and L^∞ norm for $T_{\log q}$.

Theorem 3.1. *If $\max_x \int_W q(w)r(w,x)dw \le R$, then A is differentiable at p for the norm pair (L^1, L^1) with bound $2\beta R$.*

Proof. Pick any signed measure $\epsilon(x) \in T_p$. Up to the $-\beta$ factor, $(DA_p\epsilon)(w)$ is equal to

$$\begin{aligned} &q(w) \iint \left(\delta(w - w') - q(w')\right) r(w', x)\epsilon(x)dxdw' \\ &= q(w) \int r(w, x)\epsilon(x)dx - q(w) \iint q(w')r(w', x)\epsilon(x)dxdw'. \end{aligned}$$

Among the two terms, the TV norm of the first term is bounded by

$$\int_W \left| \int_X q(w)r(w,x)\epsilon(x)dx \right| dw \le \left(\max_x \int_W |q(w)r(w,x)|dw \right) \cdot \|\epsilon\|_{L^1} \le R\|\epsilon\|_{L^1},$$

where we use the non-negativity of $q(w)r(w,x)$. The same estimate applies to the second term as well. Putting together shows that $\|DA_p\|_{L^1 \to L^1} \le 2\beta R$. □

Remark 3.2. (a) The product βR controls the sensitivity of A. One can achieve this by adopting either a small R (a safer risk function) or a small β (stronger entropic regularization).

(b) One way to ensure $\max_x \int_W q(w)r(w,x)dw \le R$ is $\max_{w,x} |r(w,x)| \le R$. But this can be strict as it does not take into consideration the distributions $p(x)$ and $q(w)$.

(c) The quantity $\max_x \int_W q(w)r(w,x)dw$ can be estimated. Suppose we have an algorithm that can sample w from $q(w)$. First, for each w, iterate over the data point x and accumulate $r(w,x)$ for each x. Second, for each data point x, dividing the accumulated value by the number of w gives the estimate of $\int_W q(w)r(w,x)dw$ for x. Finally, taking the maximum of these estimates over x gives the approximation to $\max_x \int_W q(w)r(w,x)dw$.

Theorem 3.2. *If $\max_{x,w} |r(w,x)| \le R$, then A satisfies tangent differential privacy at p for the norm pair (L^1, L^∞) with bound $2\beta R$.*

Proof. Pick any $\epsilon \in T_p$. Up to the $-\beta$ factor, $(D(\log \circ A)_p\epsilon)(w)$ at each w is

$$\begin{aligned} &\iint \left(\delta(w - w') - q(w')\right) r(w', x)\epsilon(x)dxdw' \\ &= \int r(w, x)\epsilon(x)dx - \iint q(w')r(w', x)\epsilon(x)dxdw'. \end{aligned}$$

Among the two terms, the first one is bounded at w with

$$\left| \int r(w,x)\epsilon(x)dx \right| \leq R\|\epsilon\|_{L^1}.$$

The second one can be bounded in the same way. Therefore,

$$\|D(\log \circ A)_p\|_{L^1 \to L^\infty} \leq 2\beta R.$$

The proof is complete. □

Remark 3.3. Examples of bounded $r(w,x)$ include the Savage loss, the tangent loss, and the $0/1$ loss. Using any of these losses automatically guarantees tangent differential privacy for the pair (L^1, L^∞).

3.2 Wasserstein distance on $\mathcal{P}(X)$

Recall from Example 2.2 that we have the $\dot{H}^{-1}(p)$ norm for T_p, L^1 norm for T_q, and L^∞ norm for $T_{\log q}$. Recall that the $\dot{H}^{-1}(p)$ norm and its dual norm are given by

$$\|\epsilon\|^2_{\dot{H}^{-1}(p)} = \min_{f:\nabla\cdot(fp)=\epsilon} \int |f(x)|^2 p(x)dx, \quad \|g\|^2_{\dot{H}^1(p)} = \int |\nabla g(x)|^2 p(x)dx.$$

Theorem 3.3. *If*

$$\left\| \int_W q(w)r(w,\cdot)dw \right\|_{\dot{H}^1(p)} \leq R,$$

then A is differentiable at p for the norm pair $(\dot{H}^{-1}(p), L^1)$ with bound $2\beta R$.

Proof. Pick any $\epsilon(x) \in T_p$. Up to the $-\beta$ factor, $(DA_p\epsilon)(w)$ is equal to

$$q(w)\int r(w,x)\epsilon(x)dx - q(w)\iint q(w')r(w',x)\epsilon(x)dxdw'.$$

The L^1 norm of the first term can be bounded by

$$\int_W \left| \int_X q(w)r(w,x)\epsilon(x)dx \right| dw \leq \left\| \int |q(w)r(w,\cdot)|dw \right\|_{\dot{H}^1(p)} \|\epsilon\|_{\dot{H}^{-1}(p)} \leq R\|\epsilon\|_{\dot{H}^{-1}(p)},$$

where we use the non-negativity of $q(w)r(w,\cdot)$. The same estimate can bound the second term. Therefore, $\|DA_p\|_{\dot{H}^{-1}(p)\to L^1} \leq 2\beta R$. □

Remark 3.4. (a) One way to ensure

$$\left\| \int_W q(w)r(w,\cdot)dw \right\|_{\dot{H}^1(p)} \leq R$$

is

$$\max_w \|r(w,\cdot)\|_{\dot{H}^1(p)} \leq R.$$

However, this can be too strict as it does not take into consideration the distributions $p(x)$ and $q(w)$.

(b) The quantity

$$\left\| \int_W q(w) r(w,\cdot) dw \right\|_{\dot{H}^1(p)} = \left(\left| \int_W q(w) \nabla_x r(w,x) dw \right|^2 p(x) dx \right)^{\frac{1}{2}}$$

can be estimated instead. Suppose that we have an algorithm that samples $w \sim q(w)$. First, for each w, iterate over x and accumulate $\nabla_x r(w,x)$ for each x. Second, for each x, divide the accumulated value by the number of w to get an estimate of $|\int_W q(w)\nabla_x r(w,x)dw|$. Its square is an estimate for $|\int_W q(w)\nabla_x r(w,x)dw|^2$. Finally, averaging the squares over x and taking the square root gives an approximation to $\|\int_W q(w) r(w,\cdot)dw\|_{\dot{H}^1(p)}$.

Theorem 3.4. *If* $\max_w \|r(w,\cdot)\|_{\dot{H}^1(p)} \le R$, *then A satisfies tangent differential privacy at p for the norm pair* $(\dot{H}^{-1}(p), L^\infty)$ *with bound* $2\beta R$.

Proof. Pick any $\epsilon \in T_p$. Up to the $-\beta$ factor, $(D(\log \circ A)_p \epsilon)(w)$ at each w is

$$\begin{aligned} &\iint \left(\delta(w-w') - q(w')\right) r(w',x)\epsilon(x) dx dw' \\ &= \int r(w,x)\epsilon(x)dx - \iint q(w') r(w',x)\epsilon(x) dx dw'. \end{aligned}$$

The first term is bounded at w by

$$\int r(w,x)\epsilon(x)dx \le \|r(w,\cdot)\|_{\dot{H}^1(p)} \|\epsilon\|_{\dot{H}^{-1}(p)} \le R\|\epsilon\|_{\dot{H}^{-1}(p)}.$$

The second term can be bounded in the same way. Therefore,

$$\|D(\log \circ A)_p\|_{\dot{H}^{-1}(p) \to L^\infty} \le 2\beta R.$$

The proof is complete. □

Remark 3.5. Note that if W is finite, then $\max_w \|r(w,\cdot)\|_{\dot{H}^1(p)}$ can be estimated by evaluating $\|r(w,\cdot)\|_{\dot{H}^1(p)}$ for each $w \in W$. Otherwise, $\max_{w,x} |\nabla_x r(w,x)|$ provides an upper bound for $\max_w \|r(w,\cdot)\|_{\dot{H}^1(p)}$.

4 Discussion

In this note, we propose tangent differential privacy as a new form of differential privacy. Compared to the usual differential privacy, which is defined uniformly across data distributions, tangent differential privacy is tailored to a specific data distribution of interest. For empirical risk minimization of supervised learning, entropic regularization guarantees tangent differential privacy under rather general conditions on the risk function. Some directions for future work include

- Extend the framework to unsupervised learning and online learning problems.
- Explore alternatives or approximations to (3.1) since sampling the Gibbs distribution $q(w)$ can be challenging when it exhibits meta-stability.

Acknowledgments

The author thanks Yiping Lu and the reviewers for constructive discussions.
This work is partially supported by the NSF (Grant Nos. DMS-2011699, DMS-2208163).

References

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

[2] R. Bassily, V. Feldman, K. Talwar, and A. Guha Thakurta, Private stochastic convex optimization with optimal rates, in: *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 32:11250–11259, 2019.

[3] R. Bassily, A. Smith, and A. Thakurta, Private empirical risk minimization: Efficient algorithms and tight error bounds, in: *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, IEEE, 464–473, 2014.

[4] M. Boedihardjo, T. Strohmer, and R. Vershynin, Metric geometry of the privacy-utility tradeoff, *arXiv:2405.00329*, 2024.

[5] M. Boedihardjo, T. Strohmer, and R. Vershynin, Private measures, random walks, and synthetic data, *Probab. Theory Relat. Fields*, 189(1-2):569–611, 2024.

[6] M. Bun and T. Steinke, Concentrated differential privacy: Simplifications, extensions, and lower bounds, in: *Lecture Notes in Computer Science*, Vol. 9985, Springer, 635–658, 2016.

[7] X. Cheng, D. Yin, P. Bartlett, and M. Jordan, Stochastic gradient and Langevin processes, in: *Proceedings of the 37th International Conference on Machine Learning*, JMLR.org, 1810–1819, 2020.

[8] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, Local privacy and statistical minimax rates, in: *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, IEEE, 429–438, 2013.

[9] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, Our data, ourselves: Privacy via distributed noise generation, in: *Advances in Cryptology – EUROCRYPT 2006. Lecture Notes in Computer Science*, Vol. 4004, Springer, 486–503, 2006.

[10] C. Dwork, F. McSherry, K. Nissim, and A. Smith, Calibrating noise to sensitivity in private data analysis, in: *Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science*, Vol. 3876, Springer, 265–284, 2006.

[11] C. Dwork and A. Roth, The algorithmic foundations of differential privacy, *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

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

[13] A. Evfimievski, J. Gehrke, and R. Srikant, Limiting privacy breaches in privacy preserving data mining, in: *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, ACM, 211–222, 2003.

[14] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, What can we learn privately?, *SIAM J. Comput.*, 40(3):793–826, 2011.

[15] F. Koufogiannis, S. Han, and G. J. Pappas, Optimality of the Laplace mechanism in differential privacy, *arXiv:1504.00065*, 2015.

[16] F. McSherry and K. Talwar, Mechanism design via differential privacy, in: *48th Annual IEEE Symposium on Foundations of Computer Science*, IEEE, 94–103, 2007.

[17] I. Mironov, Rényi differential privacy, in: *2017 IEEE 30th Computer Security Foundations Symposium*, IEEE, 263–275, 2017.

[18] R. Peyre, Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance, *Control Optim. Calc. Var.*, 24(4):1489–1501, 2018.

[19] M. Welling and Y. W. Teh, Bayesian learning via stochastic gradient Langevin dynamics, in: *Proceedings of the 28th International Conference on Machine Learning*, Omnipress, 681–688, 2011.