Privacy-Preserving Constrained Quadratic Optimization with Fisher Information

Farhad Farokhi

Abstract—**Noisy (stochastic) gradient descent is used to develop privacy-preserving algorithms for solving constrained quadratic optimization problems. The variance of the error of an adversary's estimate of the parameters of the quadratic cost function based on iterates of the algorithm is related to the Fisher information of the noise using the Cramér-Rao bound. This motivates using the Fisher information as a measure of privacy. Noting that the performance degradation in noisy gradient descent is proportional to the variance of the noise, a measure of utility is defined to be equal to the variance of the noise. Trade-off between privacy and utility is balanced by minimizing the Fisher information subject to a constraint on the variance of the noise. The optimal privacy-preserving noise is proved to be Gaussian, which implies that the developed privacy-preserving optimization algorithm also guarantees differential privacy.**

Index Terms—**Optimization algorithm; Constrained optimization; Quadratic cost; Noisy gradient descent; Fisher information; Privacy-preserving optimization.**

I. Introduction

Privacy has been increasingly under threat in the era of big data [1], [2]. This motivates development of methods for preserving privacy of individuals. GDPR[1] has undoubtedly further fuelled privacy research by industry so that corporations can recruit and maintain a larger set of customers willing to share their private data under proper guarantees.

An important area for privacy preservation is numerical optimization. The iterates of optimization algorithms, at the heart of machine learning and artificial intelligence, and their final solutions can leak significant amount of information about cost functions and the constraints of the problem often stemming from personal data or sensitive information [3]–[5].

A natural candidate for alleviating these privacy concerns is differential privacy [6]. Differential privacy, originally defined in the context of performing privacy-preserving analytic on databases, ensures that the output of a query on a dataset is systematically corrupted, for instance by an additive noise, so that its statistics do not significantly change by changing the private data of one individual. This makes reverse engineering the private dataset difficult [7]. Differential privacy is regularly used to ensure privacy in optimization [8]–[11] and machine learning [12]–[16]. These methods rely on the use of independently and identically distributed privacy-preserving additive noise at each iteration of the optimization/learning algorithm to ensure differential privacy. However, differential privacy can result in low utility [17]–[19] and is susceptible to adversarial attacks [20]. Further, there is no systematic way for setting the so-called privacy parameter, which can render it ineffective in practice [21], [22].

Another approach for privacy is to use metrics from information or estimation theory to measure private information leakage when responding to queries on private datasets [23]–[25]. These studies date back to wiretap channels [26] and their extensions [27]. Information-theoretic measures of privacy rely on mutual information for measuring private information leakage and cast the privacy problem as a generalized rate-distortion problem [23], [28]. Although possessing physical interpretations for privacy and offering strong guarantees, using mutual information as a measure of privacy forces the private dataset to be statistically distributed with known distributions, which makes the framework susceptible to adoption of wrong distribution for analysis and design. Hence, Fisher information has been suggested to be used for privacy [29]–[31]. Although successful in treating static database queries [30], control problems [32], smart meter privacy [29], Fisher information privacy have not been used for developing privacy-preserving optimization algorithms. This is the topic of this paper.

Motivated by these observations, in this paper, we focus on developing privacy-preserving optimization algorithms using Fisher information as a measure of privacy. We particularly develop privacy-preserving algorithms for solving constrained quadratic optimization problems based on noisy gradient descent algorithm. The choice of quadratic programs is relevant as trust-region optimization can be used to solve a larger class of optimization problems using a sequence of constrained quadratic programs recursively [33, Chapter 4]. Quadratic cost functions are also prevalent in conventional machine learning problems, such as regression [34]. Our focus on gradient descent algorithm is also motivated by favourable results on first-order methods for solving very large optimization problems [35]. We use the Cramér-Rao bound (see, e.g., [36]) to relate the performance of an adversary's estimate of the parameters of the cost function (in the sense of the variance of the estimation error) to the Fisher information of the noise in the noisy gradient descent algorithm. Therefore, we can use the Fisher information as a measure of privacy. The degradation caused in the performance of the noisy gradient descent algorithm is proportional to the variance of the noise. Therefore, we can use the variance of the noise as a measure of utility. We cast the problem of balancing between privacy and utility as an optimization problem minimizing the Fisher information of the noise subject to a constraint on the variance of the noise. We show that the optimal privacy-preserving noise is indeed Gaussian. This implies that the privacy-preserving noisy gradient descent algorithm in this paper also guarantees

F. Farokhi is with CSIRO's Data61 and the Department of Electrical and Electronic Engineering at the University of Melbourne, Australia.

e-mails: farhad.farokhi@data61.csiro.au; farhad.farokhi@unimelb.edu.au

[1] The General Data Protection Regulation 2016/679 is a regulation in EU law on data protection and privacy. https://gdpr-info.eu

differential privacy due to results on Gaussian mechanisms; see, e.g., [37]. We finally demonstrate the results for training linear regression models.

II. Privacy-Preserving Quadratic Optimization

We start by an unconstrained quadratic optimization problem in the from of

$$\min_{x\in\mathbb{R}^n}\left\{f(x)=\frac{1}{2}x^\top Qx+q^\top x\right\},\tag{1}$$

where Q is a symmetric semi-positive definite matrix. We will add constraints to this optimization problem later. As stated earlier, focusing on quadratic programs is relevant due to prevalence of quadratic optimization problems in machine learning [34] and trust-region methods [33]. In addition, the quadratic utility functions are widely used in signal processing, e.g., weighted least squares, and machine learning, e.g. support vector machines. We can solve this optimization problem by gradient descent according to

$$\begin{aligned}x[k+1]&=x[k]-\alpha_k\nabla f(x[k])\\&=x[k]-\alpha_k(Qx[k]+q),\end{aligned}\tag{2}$$

where $\alpha_k>0$ is an appropriately selected step size to ensure convergence to the optimal solution (see Proposition 2 for more information). We assume that the adversary can access the iterates $x[k]$ of (2). This is a common assumption in privacy-preserving optimization literature [8]–[16]. Instead of following the updates in (2) that can potentially leak information about the private parameters of the cost function [3], we follow the perturbed (noisy) update rule in

$$x[k+1]=x[k]-\alpha_k(\nabla f(x[k])+w[k]),\tag{3}$$

where $w[k]\in\mathbb{R}^n$, $\forall k$, are independently and identically distributed random variables according to the probability density function γ with support set $\mathbb{R}^n$. The framework can be extended to admit quadratic optimisation problems with constraints as in

$$\min_{x\in\mathcal{C}}\left\{f(x)=\frac{1}{2}x^\top Qx+q^\top x\right\},\tag{4}$$

where $\mathcal{C}\subseteq\mathbb{R}^n$ is a convex set. To solve this problem in a privacy-preserving manner, we propose the use of projected gradient descent algorithm in

$$x[k+1]=\mathcal{P}_{\mathcal{C}}\left[x[k]-\alpha_k(\nabla f(x[k])+w[k])\right],\tag{5}$$

where $\mathcal{P}_{\mathcal{C}}[x]:=\arg\min_{x'\in\mathcal{C}}\|x-x'\|_2$ denotes projection into $\mathcal{C}$. In what follows, we use the notation $x[0:T]$ to denote the sequence $(x[0],\dots,x[T])$. Furthermore, I denotes the identity matrix of an appropriate size. For any matrix A, $\|A\|_F$ is its Frobenius norm.

Proposition 1: Let $\hat{Q}(x[0:T])$ and $\hat{q}(x[0:T])$ be an unbiased estimate of Q and q based on the iterates $x[0:T]$ in (3) or (5). Then, there exists a constant $c>0$, such that

$$\mathbb{E}\{\|Q-\hat{Q}(x[0:T])\|_F^2+\|q-\hat{q}(x[0:T])\|_2^2\}\geq\frac{c}{\mathrm{trace}(\mathcal{I})},$$

where $\mathcal{I}$ is the Fisher information matrix associated with the probability density function γ defined as

$$\mathcal{I}:=\mathbb{E}\{\nabla\log(\gamma(w))\nabla\log(\gamma(w))^\top\}.$$

Proof: We first present the proof for the iterates of (3). Note that we can rewrite (3) as

$$\begin{bmatrix}-(x[k]^\top\otimes I) & -I\end{bmatrix}\begin{bmatrix}\mathrm{vec}(Q)\\ q\end{bmatrix}+w[k]=\frac{x[k+1]-x[k]}{\alpha_k}.$$

We can define

$$y=\begin{bmatrix}\dfrac{x[T]-x[T-1]}{\alpha_k}\\ \vdots\\ \dfrac{x[1]-x[0]}{\alpha_k}\end{bmatrix},\quad w=\begin{bmatrix}w[T-1]\\ \vdots\\ w[0]\end{bmatrix},$$

and

$$A=\begin{bmatrix}-x[T-1]^\top\otimes I) & -I\\ \vdots & \\ -(x[0]^\top\otimes I) & -I\end{bmatrix}.$$

This implies that the adversary has access to noisy measurements of Q and q of the form of

$$y=A\begin{bmatrix}\mathrm{vec}(Q)\\ q\end{bmatrix}+w.\tag{6}$$

Therefore, according to the Cramér-Rao bound, we have

$$\begin{aligned}&\mathbb{E}\{\|Q-\hat{Q}(x[0:T])\|_F^2+\|q-\hat{q}(x[0:T])\|_2^2\}\\&=\mathbb{E}\left\{\left\|\begin{bmatrix}\mathrm{vec}(Q-\hat{Q}(x[0:T]))\\ q-\hat{q}(x[0:T])\end{bmatrix}\right\|_2^2\right\}\\&\geq\mathrm{trace}(\mathbb{E}\{\nabla\log(p(x[0:T]|\,\mathrm{vec}(Q),q))\\&\quad\times\nabla\log(p(x[0:T]|\,\mathrm{vec}(Q),q))^\top\}^{-1}),\end{aligned}$$

where

$$p(x[0:T]|\,\mathrm{vec}(Q),q)=\prod_{k=0}^{T-1}\gamma\left(A_k\begin{bmatrix}\mathrm{vec}(Q)\\ q\end{bmatrix}-y_k\right),$$

where A_k and y_k are, respectively, the k-th row of A and the k-th entry of y. Note that

$$\begin{aligned}&\nabla\log(p(x[0:T]|\,\mathrm{vec}(Q),q))\\&=\nabla\sum_{k=0}^{T-1}\log\left(\gamma\left(A_k\begin{bmatrix}\mathrm{vec}(Q)\\ q\end{bmatrix}-y_k\right)\right)\\&=\left.\frac{1}{\gamma\left(A_k\begin{bmatrix}\mathrm{vec}(Q)\\ q\end{bmatrix}-y_k\right)}A_k\nabla\gamma(w)\right|_{w=A_k\begin{bmatrix}\mathrm{vec}(Q)\\ q\end{bmatrix}-y_k}.\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\mathrm{trace}(\mathbb{E}\{\nabla \log(p(x[0:T]|\,\mathrm{vec}(Q),q))\\
&\qquad\times \nabla \log(p(x[0:T]|\,\mathrm{vec}(Q),q))^\top\}^{-1})\\
&\qquad\geq (n^2+n)^2\mathrm{trace}(\mathbb{E}\{\nabla \log(p(x[0:T]|\,\mathrm{vec}(Q),q))\\
&\qquad\qquad\times \nabla \log(p(x[0:T]|\,\mathrm{vec}(Q),q))^\top\})^{-1}\\
&\qquad= (n^2+n)^2\mathrm{trace}\left(\sum_{k=0}^{T-1} A_k\mathcal{I}A_k^\top\right)^{-1}\\
&\qquad= (n^2+n)^2\mathrm{trace}\left(\left(\sum_{k=0}^{T-1} A_k^\top A_k\right)\mathcal{I}\right)^{-1}\\
&\qquad\geq (n^2+n)^2\lambda_{\max}\left(\sum_{k=0}^{T-1} A_k^\top A_k\right)^{-1}\mathrm{trace}(\mathcal{I})^{-1},
\end{aligned}
$$

where the first inequality follows from the first inequality in the proof of Proposition 5 in [30].

Now, we present the proof for the iterates of (5). We can rewrite (5) as

$$
\begin{aligned}
\frac{1}{\alpha_k}\mathcal{P}_{\mathcal{P}}&\left[x[k]+\alpha_k\begin{bmatrix}-(x[k]^\top\otimes I) & -I\end{bmatrix}\begin{bmatrix}\mathrm{vec}(Q)\\ q\end{bmatrix}+\alpha_k w[k]\right]\\
&-\frac{1}{\alpha_k}x[k]=\frac{x[k+1]-x[k]}{\alpha_k}.
\end{aligned}
$$

Note that in this case, the noisy measurements of Q and q take the form of

$$
y_k=\phi_k\left(A_k\begin{bmatrix}\mathrm{vec}(Q)\\ q\end{bmatrix}+w[k]\right),
$$

where $\phi_k(z):=(1/\alpha_k)\mathcal{P}_{\mathcal{P}}[x[k]+\alpha_k z]-(1/\alpha_k)x[k]$. The data processing inequality for the Fisher information [38] implies that post-processing the measurements in (6) by any function, including ϕ_k, only reduces their information content and increase the expected estimation error of the adversary. Therefore, the lower bound for the estimation error of the adversary cannot be improved. ■

The tightness of the bound in Proposition 1 hinges on two factors. The first factor is the tightness of the Cramér-Rao bound. It is known that the Cramér-Rao bound is in fact tight for the linear-Gaussian setup and asymptotically in more general cases [39]. The other factor is tightness of the bound relating trace($\mathcal{I}^{-1}$) and trace($\mathcal{I}$)$^{-1}$; see Proposition 5 in [30]. This bounds is tight for $n=1$; however, it gets looser with increasing n unless the Fisher information matrix is diagonal. If we derive a lower bound for the adversary's performance based on trace($\mathcal{I}^{-1}$), the bound gets tighter while the problem of finding the optimal privacy-preserving noise becomes non-convex and requires solving nonlinear partial differential equations [30].

Propositions 1 illustrates that, by reducing the Fisher information, we can increase the variance of the estimation error of the adversary when estimating the private information encoded in Q and q. Therefore, we can select trace($\mathcal{I}$) as a measure of privacy. Note that, instead of maximizing the lower bounds Propositions 1 through minimizing the trace of the Fisher information matrix, we can directly minimize $\mathbb{E}\{\|Q-\hat{Q}(x[0:T])\|_F^2+\|q-\hat{q}(x[0:T])\|_2^2\}$. However, doing so, the problem formulation becomes a function of the adversary's policy $\hat{Q}(x[0:T])$ and $\hat{q}(x[0:T])$, which might not be known in advance. In security and privacy, it is often desired to have the least assumptions on the adversary [28]. Now, we can introduce a measure of quality.

Proposition 2: Assume $\lambda_{\min}(Q)\geq\lambda$ and $\alpha_k=1/(\lambda k)$. Then,

$$
\mathbb{E}\Big\{f(x[T])-\min_{x\in\mathcal{C}}f(x)\Big\}\leq\frac{17(1+\log(T))}{\lambda T}\mathrm{trace}(\mathbb{E}\{ww^\top\}).
$$

Proof: The proof follows from the application of Theorem 1 in [40]. ■

Proposition 2 motivates the choice of trace($\mathbb{E}\{ww^\top\}$) as a measure of utility. This is because we can improve the performance of the noisy gradient descent algorithm in (3) and the projected stochastic gradient descent in (5) by reducing the variance of the additive noise trace($\mathbb{E}\{ww^\top\}$). Therefore, we can balance between privacy and utility of privacy-preserving optimization algorithms in (3) and (5) by solving the optimization problem in

$$
\min_{\gamma\in\Gamma}\quad \mathrm{trace}(\mathcal{I}), \tag{7a}
$$

$$
\text{s.t.}\quad \mathrm{trace}(\mathbb{E}\{ww^\top\})\leq\varepsilon. \tag{7b}
$$

The solution of this optimization problem, capturing the utility-privacy trade-off, is provided in the next theorem.

Theorem 1: The solution of (7a) is equal to

$$
\gamma(w)=\frac{1}{\sqrt{(2\pi\varepsilon/n)^n}}\exp\left(-\frac{n}{2\varepsilon}w^\top w\right).
$$

Proof: Following [30], we know that the optimal density function γ is Gaussian. The constraint in (7a) implies that the covariance matrix of the additive noise $\mathbb{E}\{ww^\top\}$ must be such that trace($\mathbb{E}\{ww^\top\}$) $\leq\varepsilon$. Note that, for Gaussian γ, we have trace($\mathcal{I}$) = trace($\mathbb{E}\{ww^\top\}^{-1}$). Hence, trace($\mathcal{I}$) is a decreasing function of each eigenvalue of $\mathbb{E}\{ww^\top\}$. Therefore, we must have trace($\mathbb{E}\{ww^\top\}$) = ε. Therefore, $\mathbb{E}\{ww^\top\}=(\varepsilon/n)I$. This concludes the proof. ■

III. Numerical Example

We use the Adult dataset containing nearly 49,000 records from the 1994 Census database [41]. The dataset $\{(u_i,z_i)\}_{i=1}^N$ contains features z_i, such as age, education, and work, and labels z_i indicating whether an individual earns more than $50,000. We transform all the categorical entries to reals. We are interested in training a linear regression model of the form $z=x^\top[u^\top\ 1]^\top$. The cost function for training the regression model is

$$
\begin{aligned}
f(x)=&\frac{1}{N}\sum_{i=1}^N\left(x^\top\begin{bmatrix}u_i\\1\end{bmatrix}-z_i\right)^2+\frac{1}{2}x^\top x\\
=&\frac{1}{2}x^\top\left(I+\frac{2}{N}\sum_{i=1}^N\begin{bmatrix}u_i\\1\end{bmatrix}^\top\begin{bmatrix}u_i\\1\end{bmatrix}\right)x\\
&+\left(-\frac{2}{N}\sum_{i=1}^N\begin{bmatrix}u_i\\1\end{bmatrix}^\top z_i\right)^\top x+\left(\frac{1}{N}\sum_{i=1}^N z_i^2\right).
\end{aligned}
$$

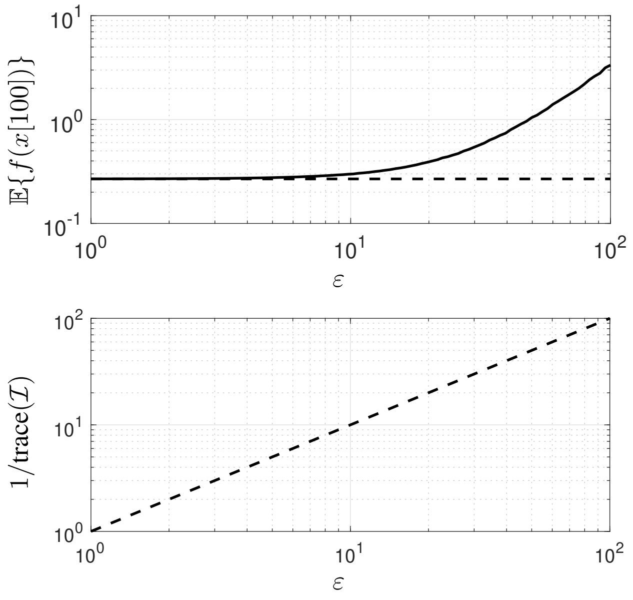

Fig. 1. [top] $\mathbb{E}\{f(x[100])\}$ versus ε after 100 iterations of (3) (solid) and $\min_{x\in\mathcal{C}} f(x)$ (dashed). [bottom] $1/\text{trace}(\mathcal{I})$ versus ε.

Forgetting about the last (constant) term, this cost function is in the form of (1). Therefore, we can use (3) to train the regression model in a privacy-preserving manner. Figure 1 [top] shows $\mathbb{E}\{f(x[100])\}$ versus ε after 100 iterations of (3) when using the privacy-preserving noise in Theorem 1 (solid) and $\min_{x\in\mathcal{C}} f(x)$ (dashed). The vertical axis in Figure 1 [top] is an indicator for the utility of the algorithm. An intended user is interested in computing $\min_{x\in\mathcal{C}} f(x)$, illustrated by the dashed horizontal line, which is achievable with high accuracy when ε is small. Figure 1 [bottom] shows $1/\text{trace}(\mathcal{I})$ versus ε. The vertical axis in Figure 1 [bottom] is an indicator of privacy. We can clearly observe the privacy-utility trade-off in this optimization problem.

IV. Conclusions

In this paper, we used the Cramér-Rao bound to relate the performance of an adversary's estimate of the parameters of the private cost function from the iterates of noisy gradient descent algorithm to the Fisher information of the noise. This motivated the use of Fisher information as a measure of privacy. We prove that the optimal privacy-preserving noise is Gaussian. Future work can focus on more general optimization problems.

References

[1] I. Rubinstein, "Big data: The end of privacy or a new beginning?," *International Data Privacy Law*, 2013. January 25, 2013.

[2] S. K. McNeil, "Privacy and the modern grid," *Harvard Journal of Law & Technology*, vol. 25, pp. 199–224, 2011.

[3] F. Farokhi, I. Shames, M. G. Rabbat, and M. Johansson, "On reconstructability of quadratic utility functions from the iterations in gradient methods," *Automatica*, vol. 66, pp. 254–261, 2016.

[4] S. Gentry, V. Saligrama, and E. Feron, "Dynamic inverse optimization," in *Proceedings of the 2001 American Control Conference*, vol. 6, pp. 4722–4727, IEEE, 2001.

[5] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *2017 IEEE Symposium on Security and Privacy (SP)*, pp. 3–18, IEEE, 2017.

[6] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography* (S. Halevi and T. Rabin, eds.), (Berlin, Heidelberg), pp. 265–284, Springer Berlin Heidelberg, 2006.

[7] W. Wang, L. Ying, and J. Zhang, "On the relation between identifiability, differential privacy, and mutual-information privacy," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5018–5029, 2016.

[8] S. Han, U. Topcu, and G. J. Pappas, "Differentially private distributed constrained optimization," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 50–64, 2016.

[9] S. Han, U. Topcu, and G. J. Pappas, "Differentially private convex optimization with piecewise affine objectives," in *53rd IEEE Conference on Decision and Control*, pp. 2160–2166, IEEE, 2014.

[10] M. T. Hale and M. Egerstedt, "Cloud-enabled differentially private multiagent optimization with constraints," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 4, pp. 1693–1706, 2017.

[11] R. Hoogervorst, Y. Zhang, G. Tillem, Z. Erkin, and S. Verwer, "Solving bin-packing problems under privacy preservation: possibilities and trade-offs," *Information Sciences*, vol. 500, pp. 203–216, 2019.

[12] 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*, pp. 308–318, 2016.

[13] K. Chaudhuri and C. Monteleoni, "Privacy-preserving logistic regression," in *Advances in neural information processing systems*, pp. 289–296, 2009.

[14] T. Zhang and Q. Zhu, "Dynamic differential privacy for admm-based distributed classification learning," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 172–187, 2016.

[15] 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*, pp. 429–438, IEEE, 2013.

[16] N. Wu, F. Farokhi, D. Smith, and M. A. Kaafar, "The value of collaboration in convex machine learning with differential privacy," in *2020 IEEE Symposium on Security and Privacy (SP)*, 2020.

[17] J. Bambauer, K. Muralidhar, and R. Sarathy, "Fool's gold: an illustrated critique of differential privacy," *Vand. J. Ent. & Tech. L.*, vol. 16, p. 701, 2013.

[18] K. Muralidhar and R. Sarathy, "Does differential privacy protect terry gross' privacy?," in *Privacy in Statistical Databases* (J. Domingo-Ferrer and E. Magkos, eds.), (Berlin, Heidelberg), pp. 200–209, Springer Berlin Heidelberg, 2010.

[19] S. L. Garfinkel, J. M. Abowd, and S. Powazek, "Issues encountered deploying differential privacy," in *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, pp. 133–137, 2018.

[20] A. Haeberlen, B. C. Pierce, and A. Narayan, "Differential privacy under fire," in *USENIX Security Symposium*, 2011.

[21] A. Greenberg, "How one of Apple's key privacy safeguards falls short," 2017. https://www.wired.com/story/apple-differential-privacy-shortcomings/.

[22] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, "Privacy loss in Apple's implementation of differential privacy on macOS 10.12," *arXiv preprint arXiv:1709.02753*, 2017.

[23] Y. Liang, H. V. Poor, and S. Shamai, "Information theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4–5, pp. 355–580, 2009.

[24] P. Braca, R. Lazzeretti, S. Marano, and V. Matta, "Learning with privacy in consensus + obfuscation," *IEEE Signal Processing Letters*, vol. 23, no. 9, pp. 1174–1178, 2016.

[25] Y. H. Liu, S.-H. Lee, and A. Khisti, "Information-theoretic privacy in smart metering systems using cascaded rechargeable batteries," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 314–318, 2017.

[26] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal, The*, vol. 54, no. 8, pp. 1355–1387, 1975.

[27] H. Yamamoto, "A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers," *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 918–923, 1983.

[28] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Transactions on Information Theory*, 2019.

[29] F. Farokhi and H. Sandberg, "Fisher information as a measure of privacy: Preserving privacy of households with smart meters using batteries," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4726–4734, 2018.

[30] F. Farokhi and H. Sandberg, "Ensuring privacy with constrained additive noise by minimizing fisher information," *Automatica*, vol. 99, pp. 275–288, 2019.

[31] F. Farokhi and H. Sandberg, "Fisher information privacy with application to smart meter privacy using HVAC units," in *Privacy in Dynamical Systems*, pp. 3–17, Springer, 2020.

[32] I. M. Ziemann and H. Sandberg, "Parameter privacy versus control performance: Fisher information regularized control," in *2020 American Control Conference (ACC)*, 2020.

[33] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer, 1999.

[34] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, Springer New York, 2013.

[35] V. Cevher, S. Becker, and M. Schmidt, "Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 32–43, 2014.

[36] J. Shao, *Mathematical statistics*. Springer Texts in Statistics, New York: Springer-Verlag, 2003.

[37] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[38] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1246–1250, 1998.

[39] Y. C. Eldar, "Minimum variance in biased estimation: Bounds and asymptotically optimal estimators," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1915–1930, 2004.

[40] O. Shamir and T. Zhang, "Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes," in *International Conference on Machine Learning*, pp. 71–79, 2013.

[41] D. Dheeru and E. Karra Taniskidou, "UCI machine learning repository," 2017. University of California, Irvine, http://archive.ics.uci.edu/ml.