Differentially Private Distributed Convex Optimization via Objective Perturbation

Erfan Nozari Pavankumar Tallapragada Jorge Cortés

***Abstract*— This paper studies the problem of differentially private distributed convex unconstrained optimization for multi-agent systems. A group of agents seek to minimize the aggregate sum of their individual objective functions. Each agent only knows its own objective function and wants to keep it private from other agents or eavesdroppers listening to the network communications. Our design strategy consists of perturbing the objective functions with Laplace noise so that any query on the functions or their attributes is differentially private. This, together with the fact that differential privacy is immune to post-processing, allows us to employ any distributed algorithm that solves the unconstrained convex optimization problem on the perturbed objective functions. Our technical approach carefully describes how these perturbations can be selected so that the resulting functions retain the requirements on smoothness and convexity critical to many optimization algorithms. We quantify the magnitude of the expected deviation of the algorithm output from the true optimizer. The specific choice of distributed optimization algorithm determines the requirements on the network communication graph. Simulations illustrate the strengths of the proposed approach.**

I. Introduction

Privacy in cyberphysical systems such as power networks, industrial control, and smart traffic systems is an increasingly critical issue that plays a key role in preventing catastrophic physical accidents as well as easing the social adoption of new technology. In these scenarios, the problem of optimizing the operation of a group of networked resources is a very common and important task, where the individual objective functions associated to the entities, the estimates of the optimizer, or even the constraints on the optimization might reveal sensitive information. Our work here is motivated by the need for efficient distributed coordination algorithms that accurately solve networked optimization problems with privacy guarantees.

Literature review: Our work here has connections with several areas, most notably distributed convex optimization and differential privacy. In the context of multi-agent networked systems, there exists an increasing body of research on distributed convex optimization that designs provably correct coordination algorithms either in discrete or continuous time and for deterministic as well as stochastic problems, see [1]–[5] and references therein. The notion of differential privacy, first introduced to protect the privacy of databases subject to public queries [6]–[8], has since then been applied to numerous problems in computer science, machine learning, and cyberphysical systems. In machine learning, the problem of differentially private optimization is studied as an intermediate step to achieve other statistical tasks, see e.g. [9]–[14]. In these works, the sensitive information is a finite database (collection of records or training samples) which constitute the parameters of a parameterized objective function (in contrast to our present work, where the sensitive information are the infinite-dimensional objective functions themselves). Although the proposed methods preserve the privacy of the underlying finite-dimensional parameter set, they cannot keep functional data private. Other limitations include the lack of characterization of the effect of noise on the accuracy of the optimization or formal study of the problems arising from the potential lack of convexity or smoothness of the perturbed functions. The problem considered in [14] is different in the sense that the total space of parameters for optimization is finite, which makes it possible to use the exponential mechanism of [15] to achieve differential privacy. More closely related to our present work are efforts [16]–[18] on differentially private distributed optimization for networked systems, where the sensitive information are the objective functions in [16], the optimization constraints in [17], and the agents' states in [18]. In these methods, agents follow a standard distributed optimization algorithm and add Gaussian or Laplace noise to their messages, either to neighbors or a central aggregator. An advantage of these *message-perturbing* strategies is that the optimization algorithm operates on the original objective functions, which satisfy all the required smoothness and convexity requirements. A common drawback of strategies based on message perturbation is that, as the privacy requirement on the executions is weakened, exact convergence to the optimizer is not necessarily recovered. This observation, that we properly justify in the body of the paper, is a consequence of the design nature of these strategies and the fundamental tradeoffs between ensuring differential privacy and ensuring convergence via an underlying dynamics that is sufficiently robust to noise. Also of relevance to our paper is [19], which proposes the use of metrics instead of binary adjacency relations for differential privacy. The advantage of this approach is that one neither needs to calculate conservative sensitivity measures nor needs to limit the allowable distance between datasets a priori.

Statement of contributions: We consider the problem of a multi-agent network optimizing a sum of convex functions in a differentially private way. Our contributions can be divided in three main categories. Our first contribution is an impossibility result for distributed optimization algorithms based on message perturbation that shows that they cannot be differentially private if they are sufficiently robust to noise. This result explains why strategies designed

The authors are with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, {enozari,ptallapragada,cortes}@ucsd.edu

within this approach have zero-input dynamics that are not asymptotically stable, leading to the drawback that, as the privacy requirement on the executions is weakened, exact convergence to the optimizer is not recovered. To overcome the limitations of message-perturbing strategies, our second contribution explores the general framework of functional differential privacy over Hilbert spaces. We propose a novel formulation of differential privacy for functional data using normed vector spaces instead of binary adjacency relations. These vector spaces are carefully designed so that differential privacy is achievable with bounded perturbations. Building on this result, our third contribution outlines an alternative design strategy to solve the distributed convex optimization problem based on the perturbation of the agents' objective functions. The basic idea is to have agents independently perturb their objective functions in a differentially private way and then have them participate in a distributed optimization algorithm with the perturbed objective functions instead of their original ones. We carefully examine how to ensure that the perturbed functions satisfy the smoothness and regularity requirements required for the convergence of distributed optimization algorithms. Finally, we formally characterize the accuracy (i.e., the deviation from the optimizer of the original problem) of the proposed class of algorithms. Our technical approach relies on a novel characterization of a Lipschitz property of the argmin map for the class of smooth strongly convex functions over compact domains. Our design methodology leads to differentially private, distributed algorithms that recover perfect accuracy in the absence of noise. For space reasons, all proofs are omitted and will appear elsewhere.

II. Preliminaries

In this section we introduce the notation and basic concepts on functional and Hilbert spaces and input-to-state stability of discrete-time systems.

A. Notation

The space of real-valued infinite sequences is denoted by $\mathbb{R}^{\mathbb{N}}$ whose elements are written in bold-faced letters. $\ell_2 \subset \mathbb{R}^{\mathbb{N}}$ denotes the space of square-summable infinite sequences. The pairs $|\cdot|$ and $\|\cdot\|$ are used for the 2-norm in finite and infinite-dimensional normed vector spaces, respectively. The d-sphere with center c and radius r is written as $B(c,r)$. The n-vector of all ones is denoted by $\mathbf{1}_n$.

For $X \subseteq \mathbb{R}^d$, X^o denotes its interior and $L_2(X)$ and $C^2(X)$ denote the set of square-integrable measurable functions and the set of twice continuously differentiable functions over X, respectively. Following [20], a map $M : X \to Y$ between two normed vector spaces is called L-Lipschitz of order $\alpha > 0$ if $\|M(x_1) - M(x_2)\|_Y \leq L\|x_1 - x_2\|_X^\alpha$ for all x_1 and x_2 in the domain of M. The Riemann zeta function $\zeta : (1,\infty) \to \mathbb{R}$ is defined as

$$\zeta(p) = \sum_{k=1}^{\infty} \frac{1}{k^p}.$$

The (zero-mean) Laplace distribution with scale b is a continuous distribution with probability density function

$$\mathcal{L}(x;b) = \frac{1}{2b} e^{-\frac{|x|}{b}}.$$

It is clear that $\frac{\mathcal{L}(x;b)}{\mathcal{L}(y;b)} \leq e^{\frac{|x-y|}{b}}$. We use $\eta \sim \mathrm{Lap}(b)$ to denote a random variable η with Laplace distribution. It is easy to see that if $\eta \sim \mathrm{Lap}(b)$, $|\eta|$ has an exponential distribution with rate $\lambda = \frac{1}{b}$. Given any random variable η and any convex function ϕ, Jensen's inequality states that $\mathbb{E}[\phi(\eta)] \geq \phi(\mathbb{E}[\eta])$. Clearly, the opposite holds if ϕ is concave.

B. Hilbert Spaces and Orthonormal Bases

We review some basic facts of Hilbert spaces in this section. The interested reader can consult [21] for a comprehensive study. Recall that a Hilbert space $\mathcal{H}$ is a complete inner-product space. A set $\{e_k\}_{k\in I} \subset \mathcal{H}$ is an orthonormal system if $\langle e_k, e_j\rangle = 0$ for $k \neq j$ and $\langle e_k, e_k\rangle = \|e_k\|^2 = 1$ for all $k \in I$. If, in addition, $\mathrm{span}(\{e_k\}_{k\in I})$ is dense in $\mathcal{H}$, then $\{e_k\}_{k\in I}$ is an orthonormal basis. Note that here I might be uncountable. However, if $\mathcal{H}$ is separable, i.e., it has a countable dense subset, then any orthonormal basis is countable. In this case, we assume $I = \mathbb{N}$ for simplicity. Consequently, for any $h \in \mathcal{H}$,

$$h = \sum_{k=1}^{\infty} \langle h, e_k\rangle e_k.$$

Define the coefficient sequence $\boldsymbol{\theta} \in \mathbb{R}^{\mathbb{N}}$ by $\theta_k = \langle h, e_k\rangle$ for $k \in \mathbb{N}$. Then, $\boldsymbol{\theta} \in \ell_2$ and, by Parseval's identity, $\|h\| = \|\boldsymbol{\theta}\|$. For ease of notation, we define $\Phi : \ell_2 \to \mathcal{H}$ to be the linear bijection that maps the coefficient sequence $\boldsymbol{\theta}$ to h.

The space $L_p(X)$ for an arbitrary $X \subseteq \mathbb{R}^d$ is a Hilbert space if and only if $p = 2$ and the inner product in this case is the integral of the product of functions. Moreover, $L_2(X)$ is separable, so every orthonormal basis is countable. In the remainder of the paper, we assume $\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis for $L_2(X)$ and $\Phi : \ell_2 \to L_2(X)$ is the corresponding linear bijection between coefficient sequences and functions.

C. Input-to-State Stability of Discrete-Time Systems

We briefly review the concept of ISS for discrete-time systems following [22]. Consider a dynamical system

$$x(k+1) = f(k, x(k), u(k)), \tag{1}$$

with input $u : \mathbb{Z}_{\geq 0} \to \mathbb{R}^m$ and state $x : \mathbb{Z}_{\geq 0} \to \mathbb{R}^n$ where $f : \mathbb{Z}_{\geq 0} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ satisfies $f(k,0,0) = 0$ for all $k \in \mathbb{Z}_{\geq 0}$. The system (1) is said to be globally input-to-state stable (ISS) if there exist $\beta \in \mathcal{KL}$ and $\gamma \in \mathcal{K}$ such that, for any bounded input u, any initial time $k_0 \in \mathbb{Z}_{\geq 0}$, any initial condition $x(k_0) \in \mathbb{R}^n$, and all $k \geq k_0$,

$$\|x(k)\| \leq \beta(\|x(k_0)\|, k - k_0) + \gamma(\|u_{[k_0]}(\cdot)\|_\infty),$$

where $\|u_{[k_0]}(\cdot)\|_\infty = \sup_{k\geq k_0} \|u(k)\|$. If (1) is ISS, then it has a $\mathcal{K}$-asymptotic gain, i.e., there exists $\gamma_a \in \mathcal{K}$ such that, for any initial condition $x_0 \in \mathbb{R}^n$,

$$\limsup_{k\to\infty} \|x(k)\| \leq \gamma_a\Big(\limsup_{k\to\infty} \|u(k)\|\Big).$$

III. Problem Statement

Consider a group of n agents whose communication topology is described by a digraph $\mathcal{G}$. Each agent $i \in \{1, \ldots, n\}$ has a local objective function $f_i : X \to \mathbb{R}$, where $X \subset \mathbb{R}^d$ is the domain of optimization known to all agents. Given $0 < \alpha < \beta$, let $\mathcal{S} \subseteq L_2(X)$ be the space of smooth functions whose Hessian is uniformly lower and upper bounded with αI_d and βI_d, respectively, and assume $f_i \in \mathcal{S}$ for all $i \in \{1, \ldots, n\}$. We also assume that X is compact and has nonempty interior but is sufficiently large so that the minimizers of $\{f_i\}_{i=1}^n$ are sufficiently far from the boundaries of X. This is a reasonable assumption as the optimization problem is unconstrained and the compactness of X is only required for $L_2(X)$ to include convex functions. The aggregate objective function is then given by

$$f(x) \triangleq \sum_{i=1}^{n} f_i(x),$$

for each $x \in X$. The objective of the agents is to optimize f over X in a distributed and differentially private fashion. By distributed, we mean that agents can only interact with their neighbors in the graph $\mathcal{G}$. To properly define differential privacy, we first need to introduce the notion of adjacency.

Definition 3.1 (Adjacency): Given any subset $\mathcal{V}$ of a normed vector space, two sets of functions $F^{(1)} = \{f_i^{(1)}\}_{i=1}^n$ and $F^{(2)} = \{f_i^{(2)}\}_{i=1}^n$ in $\mathcal{S}^n$ are $\mathcal{V}$-adjacent if there exists some $i_0 \in \{1, \ldots, n\}$ such that $f_i^{(1)} = f_i^{(2)}$ for all $i \neq i_0$ and $f_{i_0}^{(1)} - f_{i_0}^{(2)} \in \mathcal{V}$. •

The set $\mathcal{V}$ is a design choice that will be specified later in our developments. Next, we provide the formal definition of differential privacy. We represent the overall optimization algorithm as a map $\mathcal{M}$ that assigns to $\{f_i\}_{i=1}^n$ some output which is observable by an "adversary" who is trying to estimate the functions based on its observations.

Definition 3.2: (Differential Privacy): Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and consider a random map

$$\mathcal{M} : \mathcal{S}^n \times \Omega \to \mathcal{X} \tag{2}$$

from the function space $\mathcal{S}^n$ to an arbitrary finite or infinite-dimensional space $\mathcal{X}$. Given $\epsilon \in \mathbb{R}^n_{>0}$, the map $\mathcal{M}$ is ϵ-differentially private if, for any two $\mathcal{V}$-adjacent sets of functions $F^{(1)} = \{f_i^{(1)}\}_{i=1}^n$ and $F^{(2)} = \{f_i^{(2)}\}_{i=1}^n$ that (at most) differ in their i_0'th element and any set $\mathcal{O} \subseteq \mathcal{X}$,

$$\begin{aligned} &\mathbb{P}\{\omega \in \Omega \mid \mathcal{M}(F^{(2)}, \omega) \in \mathcal{O}\} \\ &\qquad \leq e^{\epsilon_{i_0} \|f_{i_0}^{(1)} - f_{i_0}^{(2)}\|_{\mathcal{V}}} \mathbb{P}\{\omega \in \Omega \mid \mathcal{M}(F^{(1)}, \omega) \in \mathcal{O}\}. \end{aligned} \tag{3}$$ •

We refer to ϵ as the level of privacy of the map $\mathcal{M}$. Note that here, we require the statistics of the output of $\mathcal{M}$ to change only (relatively) slightly if the objective function of one agent changes and the change is in $\mathcal{V}$. In case $\mathcal{M}$ has internal initial conditions, (3) has to hold for any fixed value of them.

Problem 1: (Differentially Private Distributed Optimization): Design a distributed, differentially private optimization algorithm of the form (2) with guaranteed accuracy such that, in the absence of noise, its output is the exact optimizer of the aggregate objective function. •

Problem 1 captures our desire to have better algorithm performance as the level of privacy decreases (or equivalently, as the noise variance vanishes). This ensures that the cause of the performance loss in obtaining the exact optimizer is only due to the noise introduced to ensure privacy.

IV. Rationale for Design Strategy

In this section, we justify the rationale for the design of our objective-perturbing algorithm based on the proposed concept of functional differential privacy.

A. Limitations of Message-Perturbing Strategies

We use the term message-perturbing strategies to refer to algorithms designed to solve the distributed optimization problem stated in Section III in which agents follow a standard distributed optimization algorithm and add (Gaussian or Laplace) noise to the messages they send to either neighbors or a central aggregator. The following result shows that a message-perturbing strategy cannot achieve differential privacy if the underlying optimization algorithm is ISS.

Proposition 4.1: (Impossibility Result for Message-Perturbing Algorithms): Consider a general distributed message-perturbing optimization algorithm of the form

$$\begin{aligned} x(k+1) &= a_F(k, x(k), \xi(k)), \\ \xi(k) &= x(k) + \eta(k), \end{aligned} \tag{4}$$

where $\boldsymbol{\eta} \in \mathbb{R}^{\mathbb{N}}$ is a perturbation sequence and $a_F : \mathbb{Z}_{\geq 0} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ depends on the agents' function set $F = \{f_i\}_{i=1}^n$ with associated optimizer x_F^*. Let $\tilde{a}_F(k, x(k), \eta(k)) = a_F(k, x(k), x(k) + \eta(k))$ denote the input-to-state dynamics. Assume $\eta_i(k) \sim \text{Lap}(b_i(k))$ and $b_i(k) \to 0$ for all $i \in \{1, \ldots, n\}$ at least as fast as $\frac{1}{k^p}$ for some $p > 0$. Then, if $x(k+1) = \tilde{a}_F(k, x(k), \eta(k))$ is ISS relative to x_F^* for at least two function sets $F^{(1)}$ and $F^{(2)}$ such that $x_{F^{(1)}}^* \neq x_{F^{(2)}}^*$, then the algorithm does not preserve ϵ-differentially privacy with respect to the function set F for any ϵ. •

Remark 4.2: (Relaxing the Assumptions in Proposition 4.1): The requirement of input-to-state stability of (4) in Proposition 4.1 is only sufficient but not necessary. For instance, we can prove the same impossibility of differential privacy if (4) is only zero-perturbation globally asymptotically stable (0-GAS) and $b_i(k) \to 0$ as fast as $\frac{1}{k^p}$ with $p > 2$ using the equivalence of 0-GAS and integral ISS given in [23] and the bounded energy convergence state property of [24]. Both the 0-GAS of (4) and the minimum decay rate of $b_i(k)$ can be further relaxed using the local notions of input-to-state stability. Finally, similar results can be proved if the noise sequence is normally distributed. •

From Proposition 4.1 and Remark 4.2, it is clear that a GAS optimization algorithm cannot be made differentially private by perturbing the inter-agent messages with asymptotically-vanishing noise. In the solution proposed in [16], this problem is circumvented by choosing a finite-sum sequence of step-sizes that leads to a dynamical system

which is not 0-GAS. In [17], on the other hand, a constant-variance noise is used but the algorithm is terminated after a finite number of steps. In both cases, this forced lack of asymptotic stability of the zero-input dynamics causes an error which is present independently of the amount of noise required for privacy.

B. Algorithm Design via Objective Perturbation

To overcome the limitations of message-perturbing strategies, here we outline an alternative design strategy to solve Problem 1 based on the perturbation of the agents' objective functions. The basic idea is to have agents independently perturb their objective functions in a differentially private way and then have them participate in a distributed optimization algorithm with the perturbed objective functions instead of their original ones. The following result, which is a special case of [25, Theorem 1], ensures that the combination with the distributed optimization algorithm does not deteriorate the differential privacy at the functional level.

Proposition 4.3: (Resilience to Post-processing): Let $\mathcal{M} : \mathcal{S}^n \times \Omega \to \mathcal{S}^n$ be an ϵ-differentially private map and $\mathcal{F} : \mathcal{S}^n \to \mathcal{X}$ be any deterministic map. Then, the map $\mathcal{M}' = \mathcal{F} \circ \mathcal{M} : \mathcal{S}^n \times \Omega \to \mathcal{X}$ is also ϵ-differentially private. •

This design strategy based on the perturbation of the individual objective functions requires solving the following challenges to be successful:

(i) establish a differentially private procedure to perturb the individual objective functions;
(ii) ensure that the resulting perturbed functions enjoy the smoothness and regularity properties required by distributed optimization algorithms to converge;
(iii) with (i) and (ii) in place, characterize the accuracy of the resulting differentially private, distributed coordination algorithm.

Section V addresses challenge (i) and Section VI deals with challenges (ii) and (iii).

V. Functional Differential Privacy with Laplace Noise

Motivated by the discussion of Section IV, here we explore the concept of functional differential privacy. The generality of the notion of functional differential privacy allows it to be used for any problem where the sensitive information is a whole function or some of its attributes (e.g., sample points, optimizers, derivatives and integrals). For simplicity of exposition and without loss of generality, we only consider in this section the privacy of a single function.

A. Functional Perturbation via Laplace Noise

Let $f \in L_2(X)$ be a function whose differential privacy has to be preserved. With the notations of Section II, we decompose f into its coefficients $\Phi^{-1}(f)$ and perturb this sequence by adding noise to all of its elements. Specifically,

$$\mathcal{M}(f, \boldsymbol{\eta}) = \Phi\left(\Phi^{-1}(f) + \boldsymbol{\eta}\right) = f + \Phi(\boldsymbol{\eta}), \tag{5}$$

where

$$\eta_k \sim \mathrm{Lap}(b_k), \tag{6}$$

for all $k \in \mathbb{N}$. Clearly, for $\boldsymbol{\eta}$ to belong to ℓ_2 and for $\Phi(\boldsymbol{\eta})$ (as a series) to converge, the scales $\{b_k\}_{k=1}^{\infty}$ cannot be arbitrary.

Lemma 5.1: (Sufficient condition for boundedness of perturbed functions): If there exists $K \in \mathbb{N}$ such that, for some $p > \frac{1}{2}$ and $s > 1$,

$$b_k \leq \frac{1}{k^p \log k^s}, \quad \forall k \geq K, \tag{7}$$

then $\boldsymbol{\eta}$ defined by (6) belongs to ℓ_2 with probability one. In particular, if for some $p > \frac{1}{2}$ and $\gamma > 0$,

$$b_k \leq \frac{\gamma}{k^p}, \quad \forall k \in \mathbb{N}, \tag{8}$$

then $\boldsymbol{\eta}$ defined by (6) belongs to ℓ_2 with probability one. •

B. Differential Privacy of Functional Perturbation

Here we establish the differential privacy of the map (5) using Lemma 5.1. Given any $\rho > 1$, consider the weight sequence $\{w_k = k^\rho\}_{k=1}^{\infty}$ and let the adjacency vector space be the image of the $\boldsymbol{w}$-weighted ℓ_2 space under Φ,

$$\mathcal{V}_\rho = \Phi\Big(\Big\{\boldsymbol{\delta} \in \mathbb{R}^{\mathbb{N}} \mid \sum_{k=1}^{\infty} w_k^2 \delta_k^2 < \infty\Big\}\Big).$$

It is not difficult to see that $\mathcal{V}_\rho$ is a vector space. Moreover,

$$\|f\|_{\mathcal{V}_\rho} \triangleq \Big(\sum_{k=1}^{\infty} w_k^2 \delta_k^2\Big)^{\frac{1}{2}}, \quad \text{with } \boldsymbol{\delta} = \Phi^{-1}(f),$$

is a norm on $\mathcal{V}_\rho$. The next result establishes the differential privacy of the map (5) for an appropriately chosen noise scale sequence $\boldsymbol{b}$.

Theorem 5.2 (Differential Privacy): Given $\rho > 1$, $\gamma > 0$ and $p \in \left(\frac{1}{2}, \rho - \frac{1}{2}\right)$, let

$$b_k = \frac{\gamma}{k^p}, \quad k \in \mathbb{N}. \tag{9}$$

Then, the map (5) is ϵ-differentially private with

$$\epsilon = \frac{1}{\gamma}\sqrt{\zeta(2\rho - 2p)}, \tag{10}$$

where ζ is the Riemann zeta function. •

VI. Differentially Private Distributed Optimization

In this section, we show how one can use the framework of functional differential privacy to efficiently and accurately address the problem of private distributed optimization formulated in Section III for a group of $n \in \mathbb{N}$ agents.

A. Smoothness and Regularity of the Perturbed Functions

To exploit the framework of functional differential privacy for optimization, the first issue is that of ensuring that the perturbed functions provided to the distributed coordination algorithm have the required smoothness and regularity properties. In general, the output (5) might neither be smooth nor convex. We detail next how to address these problems by defining appropriate maps that, when composed with $\mathcal{M}$, yield functions with the desired properties. Proposition 4.3 ensures that differential privacy is retained throughout this procedure.

1) Ensuring Smoothness: To ensure smoothness, we rely on the fact that $\mathcal{S}_0 = C^2(X)$ is dense in $L_2(X)$, and therefore, given any function g in $L_2(X)$, there exists a smooth function arbitrarily close to it, i.e.,

$$\forall \varepsilon > 0,\ \exists \hat{g}^s \in \mathcal{S}_0 \quad \text{such that} \quad \|g - \hat{g}^s\| < \varepsilon.$$

Here, ε is a design parameter and can be chosen sufficiently small (later, we show how to do this so that the accuracy of the coordination algorithm is not affected).

2) Ensuring Strong Convexity and Bounded Hessian: As given in Section III,

$$\mathcal{S} = \{h \in \mathcal{S}_0 \mid \alpha I_d \leq \nabla^2 h(x) \leq \beta I_d,\ \forall x \in X^o\}, \quad (11)$$

is the space of original objective functions. The next result ensures that the orthogonal projection from $\mathcal{S}_0$ onto $\mathcal{S}$ is well defined, and can therefore be used to ensure strong convexity and bounded Hessian of the perturbed functions.

Proposition 6.1: (Convexity of $\mathcal{S}$ and closedness relative to $\mathcal{S}_0$): The set $\mathcal{S}$ is convex and closed as a subset of $\mathcal{S}_0$ under the 2-norm. •

Given the result in Proposition 6.1, the best approximation in $\mathcal{S}$ of a function $h \in \mathcal{S}_0$ is its unique projection onto $\mathcal{S}$,

$$\tilde{h} = \text{proj}_{\mathcal{S}}(h).$$

By definition, the projected function has bounded gradient and Hessian.

B. Lipschitz Property of the argmin Map

Here we establish, under appropriate conditions, the Lipschitzness of the argmin map. This is a strong result of independent interest given that argmin is not even continuous for arbitrary C^2 functions. We rely on this result later to characterize the accuracy of our proposed coordination algorithm to solve the distributed optimization problem.

Lemma 6.2: (Lipschitzness of argmin*):* For any two functions $f, g \in \mathcal{S}$,

$$\left| \operatorname*{argmin}_{x \in X} f - \operatorname*{argmin}_{x \in X} g \right| \leq L \|f - g\|^{\frac{2}{d+4}}, \quad (12)$$

where

$$L = \frac{d(d+2)(d+4)(\beta - \alpha)^{d+2}\Gamma(d/2)}{4(\alpha\beta)^{d/2+2}\pi^{d/2}}. \quad •$$

C. Algorithm Design and Analysis

Here, we put together the discussion above to propose a class of differentially private, distributed optimization algorithms that solve Problem 1. Since the modifications in Section VI-A do not depend directly on the original function, Proposition 4.3 ensures that they do not deteriorate the level of differential privacy. Therefore, we require each agent $i \in \{1, \ldots, n\}$ to first compute

$$\hat{f}_i = \mathcal{M}(f_i, \boldsymbol{\eta}_i) = f_i + \Phi(\boldsymbol{\eta}_i),$$

where $\boldsymbol{\eta}_i$ is a sequence of Laplace noise generated according to (6) with the choice (9), then select $\hat{f}_i^s \in \mathcal{S}_0$ such that

$$\|\hat{f}_i - \hat{f}_i^s\| < \varepsilon_i,$$

and finally compute

$$\tilde{f}_i = \text{proj}_{\mathcal{S}}(\hat{f}_i^s). \quad (13)$$

After this process, agents participate in *any* distributed optimization algorithm with the modified objective functions $\{\tilde{f}_i\}_{i=1}^n$. Let

$$\tilde{x} = \operatorname*{argmin}_{x \in X} \sum_{i=1}^n \tilde{f}_i \quad \text{and} \quad x^* = \operatorname*{argmin}_{x \in X} \sum_{i=1}^n f_i,$$

denote, respectively, the output of the distributed algorithm and the optimizer for the original optimization problem with objective functions $\{f_i\}_{i=1}^n$. We measure the algorithm's accuracy by the expected value of the absolute deviation in the aggregate minimizer, i.e.,

$$\Delta = \mathbb{E}|\tilde{x} - x^*|.$$

The following result establishes the connection between Δ and the design parameters.

Theorem 6.3 (Accuracy): For a group of n agents who perturb their objective functions according to (13) and then participate in any distributed coordination algorithm that asymptotically converges to $\tilde{x}$, the differential privacy of the agents' objective functions is preserved and

$$\Delta \leq \frac{L}{n^2} \sum_{i=1}^n \left(\gamma_i^{\frac{2}{d+4}} \zeta(2p_i)^{\frac{1}{d+4}} + \varepsilon_i^{\frac{2}{d+4}} \right). \quad (14)$$

•

By looking at (14) and (10), one can see that there is a meaningful relationship between accuracy and privacy, as stated next. The proof is by direct substitution.

Corollary 6.4: (Privacy-Accuracy Relationship): Let $p_i = \frac{\rho_i}{2}$ in (9) for all $i \in \{1, \ldots, n\}$. Then,

$$\Delta \leq \frac{L}{n^2} \sum_{i=1}^n \left[\left(\frac{\zeta(\rho_i)}{\epsilon_i} \right)^{\frac{2}{d+4}} + \varepsilon_i^{\frac{2}{d+4}} \right]. \quad •$$

From this equation, it is clear that in order for the accuracy of the coordination algorithm not to be affected by the smoothening step, each agent has to take the value of ε_i sufficiently small so that it is negligible relative to $\zeta(\rho_i)/\epsilon_i$. Then, one can see that $\Delta \to 0$ and perfect accuracy is recovered if agents relax their privacy requirements, i.e., $\{\epsilon_i\}_{i=1}^n \to \infty$, satisfying our intuitive expectation in Problem 1.

VII. Simulations

In this section, we report simulation results of the algorithm proposed in Section VI-C over a group of $n = 10$ agents. The individual objective functions are two-dimensional quadratic functions, defined on $[-50, 50]^2$, with minimizers chosen uniformly randomly on the unit circle. The orthonormal basis is constructed from the Gram-Schmidt orthogonalization of the Taylor functions and the series is truncated to the fourth order, resulting in a 15-dimensional coefficient space. This truncation also acts as the smoothening step described in Section VI-A.1. The projection operator in (13) is performed by numerically solving the convex optimization problem $\min_{\tilde{f}_i} \|\tilde{f}_i - \hat{f}_i^s\|$, where $\tilde{f}_i^s$ is the result

of the truncation, subject to the inequality constraint $\tilde{f}_i \in \mathcal{S}$. We find the perturbed and original optimizers $\tilde{x}$ and x^* by applying an iterative interior-point algorithm to $\tilde{f}$ and f, respectively.

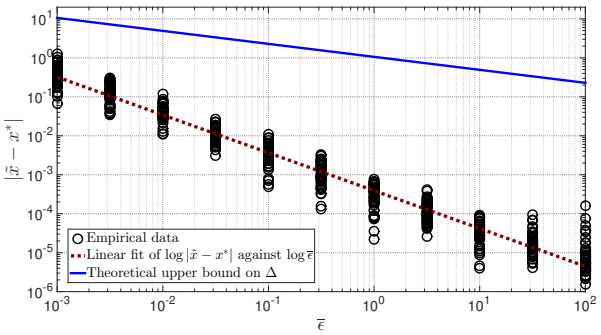

Fig. 1. Privacy-accuracy relationship. The circles, dotted line, and solid line illustrate the simulation results, the best linear fit of the simulation results in logarithmic scale, and the upper bound of Corollary 6.4, respectively

The privacy levels are taken the same for all agents, i.e., $\epsilon = \bar{\epsilon}\mathbf{1}_n$, and $\bar{\epsilon}$ is swept logarithmically over $[10^{-3}, 10^2]$. For each value of $\bar{\epsilon}$, the simulations are repeated 50 times to capture the stochasticity of the solutions. Figure 1 illustrates the error $|\tilde{x} - x^*|$ as a function of $\bar{\epsilon}$, together with the best linear fit of $\log|\tilde{x} - x^*|$ against $\log\bar{\epsilon}$, and the upper bound obtained in Corollary 6.4. The conservative nature of this upper bound can be explained by noting the approximations leading to the computation of L in Lemma 6.2, suggesting there is room for refining this result.

VIII. Conclusions and Future Work

We have studied the problem of a multi-agent network optimizing a sum of strongly convex objective functions over a compact domain in a differentially private fashion. We first noted the limitations of distributed strategies based on message perturbation due to the incompatibility, in the context of distributed convex optimization, between the strong stability properties required to guarantee convergence in the absence of noise and differential privacy. This led us to develop the framework of functional differential privacy for functions belonging to arbitrary L_2 spaces, paying particular attention to ensuring that the perturbed functions obtained after adding Laplace noise have the smoothness and regularity properties required by the distributed coordination algorithm to guarantee convergence. Future work will include the study of the effect of different orthonormal bases on the accuracy of the overall optimization, the tuning of the injected noise to optimize the trade-off between privacy and accuracy, and applications to machine learning, online optimization, and signal processing.

Acknowledgments

Research supported by NSF Award CMMI-1300272.

References

[1] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[2] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[3] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.

[4] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2014.

[5] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: convergence analysis and network scaling," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 592–606, 2012.

[6] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proceedings of the 3rd Theory of Cryptography Conference*, New York, NY, Mar. 2006, pp. 265–284.

[7] C. Dwork, "Differential privacy," in *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP)*, Venice, Italy, July 2006, pp. 1–12.

[8] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3-4, pp. 211–407, Aug. 2014.

[9] J. Zhang, Z. Zhang, X. Xiao, Y. Yang, and M. Winslett, "Functional mechanism: Regression analysis under differential privacy," *Proceedings of the VLDB Endowment*, vol. 5, no. 11, pp. 1364–1375, July 2012.

[10] R. Hall, A. Rinaldo, and L. Wasserman, "Differential privacy for functions and functional data," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 703–727, Jan. 2013.

[11] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate, "Differentially private empirical risk minimization," *The Journal of Machine Learning Research*, vol. 12, pp. 1069–1109, 2011.

[12] A. Rajkumar and S. Agarwal, "A differentially private stochastic gradient descent algorithm for multiparty classification," in *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics*, vol. 22, 2012, pp. 933–941.

[13] S. Song, K. Chaudhuri, and A. D. Sarwate, "Stochastic gradient descent with differentially private updates," in *Proceedings of the Global Conference on Signal and Information Processing*. IEEE, Dec. 2013, pp. 245–248.

[14] K. Chaudhuri, D. J. Hsu, and S. Song, "The large margin mechanism for differentially private maximization," in *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc., 2014, pp. 1287–1295.

[15] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 94–103.

[16] Z. Huang, S. Mitra, and N. Vaidya, "Differentially private distributed optimization," in *Proceedings of the 2015 International Conference on Distributed Computing and Networking*, Pilani, India, Jan. 2015.

[17] S. Han, U. Topcu, and G. J. Pappas, "Differentially private distributed constrained optimization," *IEEE Transactions on Automatic Control*, 2014, submitted. [Online]. Available: http://arxiv.org/abs/1411.4105

[18] M. T. Hale and M. Egerstedt, "Differentially private cloud-based multi-agent optimization with constraints," in *American Control Conference*. Chicago, IL: IEEE, July 2015, pp. 1235–1240.

[19] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi, "Broadening the scope of differential privacy using metrics," in *Privacy Enhancing Technologies*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, vol. 7981, pp. 82–102.

[20] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, 3rd ed. Cambridge University Press, 1999.

[21] E. Kreyszig, *Introductory Functional Analysis with Applications*. John Wiley & Sons, 1989.

[22] Z.-P. Jiang and Y. Wang, "Input-to-state stability for discrete-time nonlinear systems," *Automatica*, vol. 37, no. 6, pp. 857–869, 2001.

[23] D. Angeli, "Intrinsic robustness of global asymptotic stability," *Systems & Control Letters*, vol. 38, no. 4-5, pp. 297–307, 1999.

[24] E. D. Sontag, "Comments on integral variants of ISS," *Systems & Control Letters*, vol. 34, no. 1-2, pp. 93–100, 1998.

[25] J. L. Ny and G. J. Pappas, "Differentially private filtering," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 341–354, 2014.