Differentially Private Distributed Convex Optimization via Functional Perturbation

Erfan Nozari   Pavankumar Tallapragada   Jorge Cortés

***Abstract*—We study a class of distributed convex constrained optimization problems where a group of agents aim to minimize the sum of individual objective functions while each desires that any information about its objective function is kept private. We prove the impossibility of achieving differential privacy using strategies based on perturbing the inter-agent messages with noise when the underlying noise-free dynamics are asymptotically stable. This justifies our algorithmic solution based on the perturbation of individual functions with Laplace noise. To this end, we establish a general framework for differentially private handling of functional data. We further design post-processing steps that ensure the perturbed functions regain the smoothness and convexity properties of the original functions while preserving the differentially private guarantees of the functional perturbation step. This methodology allows us to use any distributed coordination algorithm to solve the optimization problem on the noisy functions. Finally, we explicitly bound the magnitude of the expected distance between the perturbed and true optimizers which leads to an upper bound on the privacy-accuracy trade-off curve. Simulations illustrate our results.**

***Index Terms*—Networks of Autonomous Agents, Optimization, Distributed Algorithms/Control, Differential Privacy, Cyber-Physical Systems**

I. Introduction

Privacy preservation is an increasingly critical issue that plays a key role in preventing catastrophic failures in physical infrastructure as well as easing the social adoption of new technology. Power networks, manufacturing systems, and smart transportation are just but a few examples of cyberphysical applications in need of privacy-aware design of control and coordination strategies. In these scenarios, the problem of optimizing the operation of a group of networked resources is a 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 goal of synthesizing distributed coordination algorithms that solve networked optimization problems with privacy guarantees *and* high accuracy.

Literature review: Our work builds upon the existing literature of distributed convex optimization and differential privacy. In the area of networked systems, an increasing body of research, e.g., [1]–[6] and references therein, designs and analyzes algorithms for distributed convex optimization both in discrete and continuous time as well as in deterministic and stochastic scenarios. While these works consider an ambitious suite of topics related to convergence and performance under various constraints imposed by real-world applications, privacy is an aspect generally absent in their treatment. The concept of differential privacy [7], [8] was originally proposed for databases of individual records subject to public queries and has been extended to several areas thereafter. The recent work [9] provides a comprehensive recent account of this area. In machine learning, the problem of differentially private optimization has received attention, see e.g. [10]–[16], as an intermediate, usually centralized, step for solving other learning or statistical tasks. The common paradigm is having the sensitive information correspond to the entries of a finite database of records or training data that usually constitute the parameters of an additive objective function. Threat models are varied, including releasing to the adversary the whole sequence of internal states of the optimization process or only the algorithm's final output. In [10], the authors design a differentially private classifier by perturbing the objective function with a linear finite-dimensional function (hyper-plane). It is shown in [11] that this method also works in the presence of constraints and non-differentiable regularizers. Although this is sufficient to preserve the privacy of the underlying finite-dimensional parameter set (learning samples), it cannot keep the whole objective functions private. The authors of [12] design a sensitivity-based differentially private algorithm for regression analysis which, instead of perturbing the optimal weight vector, perturbs the regression cost function by injecting noise into the coefficients of the quadratic truncation of its Taylor expansion. This truncation limits the functional space to the (finite-dimensional) space of quadratic functions. In [13], the authors propose the addition of a sample path of a Gaussian random process to the objective function, but do not explore the generalization to arbitrary dimensions or ensure the smoothness and convexity of the resulting function. In general, the proposed algorithms are not distributed and neither designed for nor capable of preserving the privacy of infinite-dimensional objective functions. Furthermore, the work in this area does not rigorously study the effect of added noise on the global optimizer or on the smoothness and convexity properties of the objective functions. In addition to addressing these issues, the present treatment is applicable to scenarios where the sensitive information consists of elements of any separable Hilbert space, including (objective) functions coming from the (infinite-dimensional) L_2 space.

Of more relevance to our paper are recent works [17]–[19] that are focused on differentially private distributed optimization problems for multi-agent systems. These papers consider as private information, respectively, the objective functions, the optimization constraints, and the agents' states. The underlying

A preliminary version of this paper has been submitted to the 2016 American Control Conference, Boston, MA

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

commonality is the algorithm design approach based on the idea of message perturbation. This idea consists of adopting a standard distributed optimization algorithm and modifying it by having agents perturb the messages to their neighbors or a central coordinator with Laplace or Gaussian noise. This approach has the advantage of working with the original objective functions and thus is easy to implement. However, for fixed design parameters, the algorithm's output does not correspond to the true optimizer in the absence of noise, suggesting the presence of a steady-state accuracy error. This problem is addressed in [18] by terminating the algorithm after a finite number of steps and optimizing this number offline as a function of the desired level of privacy. Nevertheless, for any fixed level of privacy, there exists an amount of bias in the algorithm's output which is not due to the added noise but to the lack of asymptotic stability of the underlying noiseless dynamics. To address this issue, our approach explores the use of functional perturbation to achieve differential privacy. The concept of functional differential privacy combines the benefits of metrics and adjacency relations. The authors of [20] also employ metrics instead of binary adjacency relations in the context of differential privacy. This approach has the advantage that the difference between the probabilities of events corresponding to any pair of data sets is bounded by a function of the distance between the data sets, eliminating the need for the computation of conservative sensitivity bounds.

Statement of contributions: We consider a group of agents that seek to minimize the sum of their individual objective functions over a communication network in a differentially private manner. Our first contribution is to show that coordination algorithms which rely on perturbing the agents' messages with noise cannot satisfy the requirements of differential privacy if the underlying noiseless dynamics are locally asymptotically stable. The presence of noise necessary to ensure differential privacy is known to affect the algorithm accuracy in solving the distributed convex optimization problem. However, this result explains why message-perturbing strategies incur additional inaccuracies that are present even if no noise is added. Our second contribution is motivated by the goal of guaranteeing that the algorithm accuracy is only affected by the presence of noise. We propose a general framework for functional differential privacy over Hilbert spaces and introduce a novel definition of adjacency using adjacency spaces. The latter notion is quite flexible and includes, as a special case, the conventional bounded-difference notion of adjacency. We carefully specify these adjacency spaces within the L_2 space such that the requirement of differential privacy can be satisfied with bounded perturbations. Our third contribution builds on these results on functional perturbation to design a class of distributed, differentially private coordination algorithms. We let each agent perturb its own objective function based on its desired level of privacy, and then the group uses any provably correct distributed coordination algorithm to optimize the sum of the individual perturbed functions. Two challenges arise to successfully apply this strategy: the fact that the perturbed functions might lose the smoothness and convexity properties of the original functions and the need to characterize the effect of the added noise on the minimizer of the resulting problem. We address the first challenge using a cascade of smoothening and projection steps that maintain the differential privacy of the functional perturbation step. We address the second challenge by explicitly bounding the absolute expected deviation from the original optimizer using a novel Lipschitz characterization of the argmin map. By construction, the resulting coordination algorithms satisfy the requirement of recovering perfect accuracy in the absence of noise. Various simulations illustrate our results.

Organization: We introduce our notation and basic preliminaries in Section II and formulate the private distributed optimization problem in Section III. Section IV presents the rationale for our design strategy and Section V describes a general framework for functional differential privacy. We formulate our solution to the private distributed optimization problem in Section VI. We present simulations in Section VII and collect our conclusions and ideas for future work in Section VIII. Appendix A gathers our results on the Lipschitzness of the argmin map under suitable assumptions.

II. Preliminaries

In this section, we introduce our notational conventions and some fundamental facts about Hilbert spaces and robust stability of discrete-time systems.

A. Notation

We use $\mathbb{R}$, $\mathbb{R}_{>0}$, $\mathbb{Z}_{\geq 0}$, and $\mathbb{N}$ to denote the set of reals, positive reals, nonnegative integers, and positive integers, respectively. The space of scalar- and n-vector-valued infinite sequences are denoted by $\mathbb{R}^{\mathbb{N}}$ and $(\mathbb{R}^n)^{\mathbb{N}}$, respectively. Given $K \in \mathbb{N}$ and an element $\boldsymbol{\eta} = \{\eta(k)\}_{k=0}^{\infty}$ of $\mathbb{R}^{\mathbb{N}}$ or $(\mathbb{R}^n)^{\mathbb{N}}$, we use the shorthand notation $\boldsymbol{\eta}_K = \{\eta(k)\}_{k=0}^{K}$. If the index of $\boldsymbol{\eta}$ starts at $k = 1$, with a slight abuse of notation we also denote $\{\eta(k)\}_{k=1}^{K}$ by $\boldsymbol{\eta}_K$. We denote by $\ell_2 \subset \mathbb{R}^{\mathbb{N}}$ the space of square-summable infinite sequences. We use $|\cdot|_p$ and $\|\cdot\|_p$ for the p-norm in finite and infinite-dimensional normed vector spaces, respectively (we drop the index p for $p = 2$). We let $B(c, r)$ denote the closed ball with center c and radius r in Euclidean space. For $D \subseteq \mathbb{R}^d$, D^o denotes its interior and $L_2(D)$ and $C^2(D)$ denote the set of square-integrable measurable functions and the set of twice continuously differentiable functions over D, respectively. Throughout the paper, $m(\cdot)$ denotes the Lebesgue measure. If $\{E_k\}_{k=1}^{\infty}$ is a sequence of subsets of Ω such that $E_k \subseteq E_{k+1}$ and $E = \bigcup_k E_k$, then we write $E_k \uparrow E$ as $k \to \infty$. We say $E_k \downarrow E$ as $k \to \infty$ if $E_k^c \uparrow E^c$ as $k \to \infty$, where $E^c = \Omega \setminus E$ is the complement of E. Given any closed and convex subset $\mathcal{S} \subseteq \mathcal{H}$ of a Hilbert space, we denote by $\operatorname{proj}_{\mathcal{S}}$ the orthogonal projection onto $\mathcal{S}$.

We denote by $\mathcal{K}$ the set of strictly increasing continuous functions $\alpha : [0, \infty) \to [0, \infty)$ such that $\alpha(0) = 0$. A function α belongs to $\mathcal{K}_\infty$ if $\alpha \in \mathcal{K}$ and $\lim_{r\to\infty} \alpha(r) = \infty$. We denote by $\mathcal{KL}$ the set of functions $\beta : [0, \infty) \times [0, \infty) \to [0, \infty)$ such that, for each $s \in [0, \infty)$, $r \mapsto \beta(r, s)$ is nondecreasing and continuous and $\beta(0, s) = 0$ and, for each $r \in [0, \infty)$, $s \mapsto \beta(r, s)$ is monotonically decreasing with $\beta(r, s) \to 0$ as $s \to \infty$. A map $M : X \to Y$ between two normed vector spaces is $\mathcal{K}$-Lipschitz if there exists $\kappa \in \mathcal{K}_\infty$ such that $\|M(x_1) - M(x_2)\|_Y \leq \kappa(\|x_1 - x_2\|_X)$ for all $x_1, x_2 \in X$.

The (zero-mean) Laplace distribution with scale $b \in \mathbb{R}_{>0}$ 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 $\lambda = \frac{1}{b}$. Similarly, we use the notation $\eta \sim \mathcal{N}(\mu, \sigma^2)$ when η is normally distributed with mean μ and variance σ^2. The error function $\mathrm{erf} : \mathbb{R} \to \mathbb{R}$ is defined as

$$\mathrm{erf}(x) \triangleq \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt \geq 1 - e^{-x^2}.$$

Therefore, $\mathbb{P}\{|\eta| \leq r\} = \mathrm{erf}(r/\sqrt{2}\sigma)$ if $\eta \sim \mathcal{N}(0, \sigma^2)$. Given any random variable η and any convex function ϕ, Jensen's inequality states that $\mathbb{E}[\phi(\eta)] \geq \phi(\mathbb{E}[\eta])$. The opposite inequality holds if ϕ is concave.

B. Hilbert Spaces and Orthonormal Bases

We review some basic facts on Hilbert spaces and refer the reader to [21] for a comprehensive treatment. 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, the set of linear combinations of $\{e_k\}_{k \in I}$ is dense in $\mathcal{H}$, then $\{e_k\}_{k \in I}$ is an orthonormal basis. 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 have

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

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

C. Robust Stability of Discrete-Time Systems

We briefly present some definitions and results on robust stability of discrete-time systems following [22]. Given the vector field $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, consider the system

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

with state $\mathbf{x} : \mathbb{Z}_{\geq 0} \to \mathbb{R}^n$ and input $\boldsymbol{\eta} : \mathbb{Z}_{\geq 0} \to \mathbb{R}^m$. Given an equilibrium point $x^* \in \mathbb{R}^n$ of the unforced system, we say that (1) is

(i) *0-input locally asymptotically stable (0-LAS) relative to* x^* if, by setting $\boldsymbol{\eta} = 0$, there exists $\rho > 0$ and $\gamma \in \mathcal{KL}$ such that, for every initial condition $x(0) \in B(x^*, \rho)$, we have for all $k \in \mathbb{Z}_{\geq 0}$,

$$|x(k) - x^*| \leq \gamma(|x(0) - x^*|, k).$$

(ii) *locally input-to-state stable (LISS) relative to* x^* if there exist $\rho > 0$, $\gamma \in \mathcal{KL}$, and $\kappa \in \mathcal{K}$ such that, for every initial condition $x(0) \in B(x^*, \rho)$ and every input satisfying $\|\boldsymbol{\eta}\|_\infty \leq \rho$, we have

$$|x(k) - x^*| \leq \max\{\gamma(|x(0) - x^*|, k), \kappa(|\boldsymbol{\eta}_{k-1}|_\infty)\}, \tag{2}$$

for all $k \in \mathbb{N}$. In this case, we refer to ρ as the *robust stability radius* of (1) relative to x^*.

By definition, if the system (1) is LISS, then it is also 0-LAS. The converse is also true, cf. [22, Theorem 1]. The following result is a local version of [23, Lemma 3.8] and states an important asymptotic behavior of LISS systems.

Proposition II.1. *(Asymptotic gain of LISS systems): Assume system* (1) *is LISS relative to* x^* *with associated robust stability radius* ρ. *If* $x(0) \in B(x^*, \rho)$ *and* $\|\boldsymbol{\eta}\|_\infty \leq \min\{\kappa^{-1}(\rho), \rho\}$ *(where* $\kappa^{-1}(\rho) = \infty$ *if* ρ *is not in the range of* κ*), then*

$$\limsup_{k\to\infty} |x(k) - x^*| \leq \kappa(\limsup_{k\to\infty} |\eta(k)|).$$

In particular, $x(k) \to x^*$ *if* $\eta(k) \to 0$ *as* $k \to \infty$.

Proof: From (2), we have

$$\begin{aligned} |x(k) - x^*| &\leq \max\{\gamma(|x(0) - x^*|, k), \kappa(\|\boldsymbol{\eta}\|_\infty)\} \\ &\leq \max\{\gamma(\rho, k), \kappa(\|\boldsymbol{\eta}\|_\infty)\}, \end{aligned} \tag{3}$$

where we have used $x(0) \in B(x^*, \rho)$. Now, for each $k \in \mathbb{N}$, let $\boldsymbol{\eta}_{[k]} \in (\mathbb{R}^n)^{\mathbb{N}}$ be defined by $\eta_{[k]}(\ell) = \eta(k+\ell)$ for all $\ell \in \mathbb{Z}_{\geq 0}$. If there exists k_0 such that $\|\boldsymbol{\eta}_{[k_0]}\|_\infty = 0$, then we need to show that $\lim_{k\to\infty} |x(k) - x^*| = 0$. Since $\gamma \in \mathcal{KL}$, there exists $K \in \mathbb{Z}_{\geq 0}$ such that $\gamma(\rho, k) \leq \rho$ for all $k \geq K$, and since $\kappa(\|\boldsymbol{\eta}\|_\infty) \leq \rho$ as well, it follows from (3) that $x(k) \in B(x^*, \rho)$ for all $k \geq K$. Let $\bar{k} = \max\{k_0, K\}$. Using (2), we get

$$|x(k) - x^*| \leq \gamma(|x(\bar{k}) - x^*|, k - \bar{k}), \quad \forall k > \bar{k},$$

and the result follows. Assume then that no k_0 exists such that $\|\boldsymbol{\eta}_{[k_0]}\|_\infty = 0$. Let $K_0 = 0$ and, for each $j \in \mathbb{N}$, let K_j be such that $\gamma(\rho, k - K_{j-1}) \leq \kappa(\|\boldsymbol{\eta}_{[K_{j-1}]}\|_\infty)$ for all $k \geq K_j$ (this sequence is well-defined because $\gamma \in \mathcal{KL}$). Since $\kappa(\|\boldsymbol{\eta}_{[K_{j-1}]}\|_\infty) \leq \kappa(\|\boldsymbol{\eta}\|_\infty) \leq \rho$, (2) holds if we set the "initial" state to $x(K_{j-1})$ which implies that $|x(k) - x^*| \leq \kappa(\|\boldsymbol{\eta}_{[K_{j-1}]}\|_\infty)$ for all $k \geq K_j$. Therefore,

$$\limsup_{k\to\infty} |x(k) - x^*| \leq \kappa(\|\boldsymbol{\eta}_{[K_j]}\|_\infty), \quad \forall j \in \mathbb{Z}_{\geq 0}.$$

The result follows by taking limit of both sides as $j \to \infty$. ■

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 $f_i : D \to \mathbb{R}$, where $D \subset \mathbb{R}^d$ is convex and compact and has nonempty interior. We assume that each $f_i, i \in \{1, \ldots, n\}$ is convex and twice continuously differentiable, and use the shorthand notation $F = \{f_i\}_{i=1}^n$. Consider the following convex optimization problem

$$\begin{aligned} \underset{x \in D}{\text{minimize}} \quad & f(x) \triangleq \sum_{i=1}^{n} f_i(x) \\ \text{subject to} \quad & G(x) \leq 0, \\ & Ax = b, \end{aligned}$$

where the component functions of $G : D \to \mathbb{R}^m$ are convex, $A \in \mathbb{R}^{s\times d}$, and $b \in \mathbb{R}^s$. Denote by $X \subseteq D$ the feasibility set. The optimization problem can be equivalently written as,

$$\underset{x\in X}{\text{minimize}} \;\; f(x). \tag{4}$$

We assume that X is a global piece of information known to all agents.

The group objective is to solve the convex optimization problem (4) in a distributed and private way. By distributed, we mean that each agent can only interact with its neighbors in $\mathcal{G}$. For privacy, we consider the case where the function f_i (or some of its attributes) constitute the local and sensitive information known to agent $i \in \{1, \dots, n\}$ that has to be kept confidential. Each agent assumes (the worst-case where) the adversary has access to all the "external" information (including all the network communications and all other objective functions). This setting is sometimes called local (differential) privacy in the literature, see e.g., [24]. In order to define privacy, we first introduce the notion of adjacency. Given any normed vector space $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ with $\mathcal{V} \subseteq L_2(D)$, two sets of functions $F, F' \subset L_2(D)$ are $\mathcal{V}$-adjacent if there exists $i_0 \in \{1, \dots, n\}$ such that

$$f_i = f'_i, \; i \neq i_0 \quad \text{and} \quad f_{i_0} - f'_{i_0} \in \mathcal{V}.$$

The set $\mathcal{V}$ is a design choice that we specify later in Section V-B. Moreover, this definition can be readily extended to the case where $\mathcal{V}$ is any subset of another normed vector space $\mathcal{W} \subseteq L_2(D)$. With this generalization, the conventional bounded-difference notion of adjacency becomes a special case of the definition above, where $\mathcal{V}$ is a closed ball around the origin. We provide next a general definition of differential privacy for a map.

Definition III.1. *(Differential Privacy): Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and consider a random map*

$$\mathcal{M} : L_2(D)^n \times \Omega \to \mathcal{X}$$

from the function space $L_2(D)^n$ to an arbitrary set $\mathcal{X}$. 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 and F' that (at most) differ in their i_0'th element and any set $\mathcal{O} \subseteq \mathcal{X}$, one has

$$\begin{aligned}\mathbb{P}\{\omega \in & \Omega \mid \mathcal{M}(F', \omega) \in \mathcal{O}\} \\ & \leq e^{\epsilon_{i_0}\|f_{i_0} - f'_{i_0}\|_{\mathcal{V}}} \mathbb{P}\{\omega \in \Omega \mid \mathcal{M}(F, \omega) \in \mathcal{O}\}. \quad \square\end{aligned} \tag{5}$$

Essentially, this notion requires 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 $\mathcal{V}$), making it hard for an adversary who observes the output of $\mathcal{M}$ to determine the change. In the case of an iterative asymptotic distributed optimization algorithm, $\mathcal{M}$ represents the action (observed by the adversary) of the algorithm on the set of local functions F. In other words, $\mathcal{M}$ is the map (parameterized by the initial network condition) that assigns to F the whole sequence of messages transmitted over the network. In this case, (5) has to hold for all allowable initial conditions. We are ready to formally state the network objective.

Problem 1. *(Differentially private distributed optimization): Design a distributed and differentially private optimization algorithm whose guarantee on accuracy improves as the level of privacy decreases, leading to the exact optimizer of the aggregate objective function in the absence of privacy.* $\square$

The reason for the requirement of recovering the exact optimizer in the absence of privacy in Problem 1 is the following. It is well-known in the literature of differential privacy that there always exists a cost for an algorithm to be differentially private, i.e., the algorithm inevitably suffers a performance loss that increases as the level of privacy increases. This phenomenon is a result of the noise added in the map $\mathcal{M}$, whose variance increases as ϵ decreases. With the above requirement on the noise-free behavior of the algorithm, we aim to make sure that the cause of this performance loss is *only* due to the added noise and not to any other factor.

Example III.2. *(Linear Classification with Logistic Loss Function):* We introduce here a supervised classification problem that will serve to illustrate the discussion along the paper. Consider a database of training records composed by the labeled samples $\{(a_i, b_i)\}_{i=1}^N$, where each $a_i \in \mathbb{R}^d$ (containing the features of a corresponding object) may belong to one of two possible classes and $b_i \in \{-1, 1\}$ determines to which class it belongs. The goal is to train a classifier with the samples so that it can automatically classify future unlabeled samples. For simplicity, we let $d = 2$ and assume $a_i \in [0, 1]^2$ and $b_i \in \{-1, 1\}$ are independently and uniformly randomly selected. The aim is to find the best hyperplane $x^T a$ that can separate the two classes. The parameters x defining the hyperplane can be found by solving the convex problem,

$$x^* = \underset{x\in X}{\operatorname{argmin}} \sum_{i=1}^{N} \Big(\ell(x; a_i, b_i) + \frac{\lambda}{2}|x|^2\Big), \tag{6}$$

where $\ell : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}_{>0}$ is the loss function and $(\lambda/2)|x|^2$ is the regularizing term. Since the objective function is strongly convex, we choose X large enough so that x^* is the same as the unique unconstrained minimizer. Popular choices of ℓ are the logistic loss $\ell(x; a_i, b_i) = \ln(1 + e^{-b_i a_i^T x})$ and the hinge loss $\ell(x; a_i, b_i) = \max\{0, 1 - b_i a_i^T x\}$. We focus on the logistic loss here due to its smoothness.

Consider a group of n agents, each one owning a portion $N_d = N/n$ of the training samples, who seek to collectively solve (6) in a distributed fashion, i.e., only by communicating with their neighbors (without a central aggregator). Various iterative algorithms have been proposed in the literature, cf. [2]–[6], to address this problem. As an example, [2] proposes that each agent $i \in \{1, \dots, n\}$ starts with an initial estimate $x_i(0)$ of x^* and, at each iteration k, update its estimate as

$$x_i(k+1) = \operatorname{proj}_X(z_i(k) - \alpha_k \nabla f_i(z_i(k))), \tag{7a}$$

$$z_i(k) = \sum_{j=1}^{n} a_{ij} x_j(k), \tag{7b}$$

where $\{a_{ij}\}_{j=1}^n$ are the edge weights of the communication graph at node i and α_k is the stepsize. From (7b), one can see that agents only need to share their estimates with their neighbors to run the algorithm. Under reasonable connectivity assumptions, one can show [2] that $x_i(k)$ converges to x^* asymptotically if the sequence of stepsizes is square-summable

($\sum_k \alpha_k^2 < \infty$) but not summable ($\sum_k \alpha_k = \infty$). In this paper, we are interested in endowing distributed coordination algorithms such as this with privacy guarantees so that their execution does not reveal information about the local objective functions to the adversary. □

IV. Rationale for Design Strategy

In this section, we discuss two algorithm design strategies to solve Problem 1 based on the perturbation of either inter-agent messages or the local objective functions. We point out an important limitation of the former, and this provides justification for the ensuing design of our objective-perturbing algorithm based on functional differential privacy.

A. Limitations of Message-Perturbing Strategies

We use the term *message-perturbing strategy* to refer to the result of modifying any of the distributed optimization algorithms available in the literature by adding (Gaussian or Laplace) noise to the messages agents send to either neighbors or a central aggregator in order to preserve privacy. A generic message-perturbing distributed algorithm takes the form

$$\begin{aligned} x(k+1) &= a_{\mathcal{I}}(x(k), \xi(k)), \\ \xi(k) &= x(k) + \eta(k), \end{aligned} \tag{8}$$

where $\boldsymbol{\xi}$, $\boldsymbol{\eta} : \mathbb{Z}_{\geq 0} \to \mathbb{R}^n$ are the sequences of messages and perturbations, respectively, and $a_{\mathcal{I}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ depends on the agents' sensitive information set $\mathcal{I}$ with associated optimizer $x_{\mathcal{I}}^*$. This formulation is quite general and can also encode algorithmic solutions for optimization problems other than the one in Section III, such as the ones studied in [18], [19]. In the problem of interest here, $\mathcal{I} = F = \{f_i\}_{i=1}^n$.

The following result provides conditions on the noise variance that ensure that the noise vanishes asymptotically almost surely and remains bounded with nonzero probability.

Lemma IV.1. *(Convergence and boundedness of Laplace and normal random sequences with decaying variance): Let $\boldsymbol{\eta}$ be a sequence of independent random variables defined over the sample space $\Omega = \mathbb{R}^{\mathbb{N}}$, with $\eta(k) \sim \mathrm{Lap}(b(k))$ or $\eta(k) \sim \mathcal{N}(0, b(k))$ for all $k \in \mathbb{N}$. Given $r > 0$, consider the events*

$$\begin{aligned} E &= \{\boldsymbol{\eta} \in \Omega \mid \lim_{k\to\infty} \eta(k) = 0\}, \\ F_r &= \{\boldsymbol{\eta} \in \Omega \mid \forall k \in \mathbb{N} \quad |\eta(k)| \leq r\}. \end{aligned}$$

If $b(k)$ is $O(\frac{1}{k^p})$ for some $p > 0$, then $\mathbb{P}(E) = 1$ and $\mathbb{P}(F_r) = \mathbb{P}(F_r \cap E) > 0$ for all $r > 0$.

Proof: First, consider the case where $\eta(k) \sim \mathrm{Lap}(b(k))$. By the independence of the random variables and the fact that $|\eta(k)|$ is exponentially distributed with rate $\frac{1}{b(k)}$,

$$\mathbb{P}(F_r) = \prod_{k=1}^{\infty} \left(1 - e^{-\frac{r}{b(k)}}\right).$$

By assumption, $b(k) \leq \frac{c}{k^p}$ for all $k \in \mathbb{N}$ and some $p, c > 0$. Thus, given that the series $\sum_{k=1}^{\infty} e^{-\frac{r}{c}k^p}$ converges [25, §1.14],

$$\mathbb{P}(F_r) \geq \prod_{k=1}^{\infty} \left(1 - e^{-\frac{r}{c}k^p}\right) > 0.$$

Next, let $E_{\ell,K} = \{\boldsymbol{\eta} \in \Omega \mid \forall k \geq K \quad |\eta(k)| < \upsilon_\ell\}$ where $\{\upsilon_\ell\}_{\ell=1}^{\infty}$ is a monotonically decreasing sequence that converges to zero as $\ell \to \infty$ (e.g., $\upsilon_\ell = \frac{1}{\ell}$). Note that

$$E = \bigcap_{\ell=1}^{\infty} \bigcup_{K=1}^{\infty} E_{\ell,K}.$$

Since $E_{\ell,K} \uparrow \bigcup_{K=1}^{\infty} E_{\ell,K}$ for all $\ell \in \mathbb{N}$ as $K \to \infty$, and $\bigcup_{K=1}^{\infty} E_{\ell,K} \downarrow E$ as $\ell \to \infty$, we have

$$\begin{aligned} \mathbb{P}(E) &= \lim_{\ell\to\infty} \lim_{K\to\infty} \mathbb{P}(E_{\ell,K}) = \lim_{\ell\to\infty} \lim_{K\to\infty} \prod_{k=K}^{\infty} \left(1 - e^{-\frac{\upsilon_\ell}{b(k)}}\right) \\ &\geq \lim_{\ell\to\infty} \lim_{K\to\infty} \prod_{k=K}^{\infty} \left(1 - e^{-\frac{\upsilon_\ell}{c}k^p}\right) = 1. \end{aligned}$$

Then, $\mathbb{P}(F_r \cap E) = \mathbb{P}(F_r) - \mathbb{P}(F_r \cap E^c) = \mathbb{P}(F_r) > 0$. For the case of normal distribution of random variables,

$$\mathbb{P}\{|\eta(k)| \leq r\} = \mathrm{erf}\left(\frac{r}{\sqrt{2b(k)}}\right) \geq 1 - e^{-\frac{r^2}{2b(k)}},$$

and the results follows from the arguments above. ■

Note that Lemma IV.1 also ensures that the probability that the noise simultaneously converges to zero and remains bounded is nonzero. One might expect that Lemma IV.1 would hold if $b(k) \to 0$ at any rate. However, this is not true. For instance, if $b(k) = \frac{1}{\log k}$, one can show that the probability that $\eta(k)$ eventually remains bounded is zero for any bound $r \leq 1$, so the probability that $\eta(k) \to 0$ is zero as well.

The following result shows that a message-perturbing algorithm of the form (8) cannot achieve differential privacy if the underlying (noise-free) dynamics are asymptotically stable. For convenience, we employ the short-hand notation $\tilde{a}_{\mathcal{I}}(x(k), \eta(k)) = a_{\mathcal{I}}(x(k), x(k) + \eta(k))$ to refer to (8).

Proposition IV.2. *(Impossibility result for 0-LAS message-perturbing algorithms): Consider* any *algorithm of the form* (8) *with either $\eta_i(k) \sim \mathrm{Lap}(b_i(k))$ or $\eta_i(k) \sim \mathcal{N}(0, b_i(k))$. If $\tilde{a}_{\mathcal{I}}$ is 0-LAS relative to $x_{\mathcal{I}}^*$ for two information sets $\mathcal{I}$ and $\mathcal{I}'$ with different optimizers $x_{\mathcal{I}}^* \neq x_{\mathcal{I}'}^*$ and associated robust stability radii ρ and ρ', respectively, $b_i(k)$ is $O(\frac{1}{k^p})$ for all $i \in \{1, \dots, n\}$ and some $p > 0$, and at least one of the following holds,*

- (i) *$x_{\mathcal{I}}^*$ is not an equilibrium point of $x(k+1) = \tilde{a}_{\mathcal{I}'}(x(k), 0)$ and $\tilde{a}_{\mathcal{I}'}$ is continuous,*
- (ii) *$x_{\mathcal{I}}^*$ belongs to the interior of $B(x_{\mathcal{I}'}^*, \rho')$,*

then, the algorithm cannot preserve the ϵ-differentially privacy of the information set $\mathcal{I}$ for any $\epsilon > 0$.

Proof: Our proof strategy consists of establishing that, if the initial state is close to the equilibrium of the system for one information set, the state trajectory converges to that equilibrium with positive probability but to the equilibrium of the system with the other information set with probability zero. We then use this fact to rule out differential privacy. For any fixed initial state x_0, if either of $\boldsymbol{\xi}$ or $\boldsymbol{\eta}$ is known, the other one can be uniquely determined from (8). Therefore, the mapping $\Xi_{\mathcal{I},x_0} : (\mathbb{R}^n)^{\mathbb{N}} \to (\mathbb{R}^n)^{\mathbb{N}}$ such that

$$\Xi_{\mathcal{I},x_0}(\boldsymbol{\eta}) = \boldsymbol{\xi}$$

is well-defined and bijective. Let $\kappa, \kappa' \in \mathcal{K}$ be as in (2) corresponding to $\tilde{a}_{\mathcal{I}}$ and $\tilde{a}_{\mathcal{I}'}$, respectively. Consider as initial condition $x_0 = x^*_{\mathcal{I}}$ and define

$$R = \big\{\boldsymbol{\eta} \in \Omega \mid \forall i \in \{1, \ldots, n\}, \lim_{k\to\infty} \eta_i(k) = 0 \text{ and } |\eta_i(k)| \leq \min\big\{\kappa^{-1}(\rho), \rho\big\}, \forall k \in \mathbb{N}\big\}.$$

By Lemma IV.1, we have $\mathbb{P}(R) > 0$. By Proposition II.1, since $|x_0 - x^*_{\mathcal{I}}| = 0 \leq \rho$ and $\|\boldsymbol{\eta}\|_\infty \leq \min\{\kappa^{-1}(\rho), \rho\}$ for all $\boldsymbol{\eta} \in R$, the sequence $\Xi_{\mathcal{I},x_0}(\boldsymbol{\eta})$ converges to $x^*_{\mathcal{I}}$. Let $\mathcal{O} = \Xi_{\mathcal{I},x_0}(R)$ and $R' = \Xi^{-1}_{\mathcal{I}',x_0}(\mathcal{O})$ (where we are using the forward and inverse images of sets, respectively). Next, we show that no $\boldsymbol{\eta}' \in R'$ converges to 0 under either hypothesis (i) or (ii) of the statement. Under (i), there exists a neighborhood of $(x^*_{\mathcal{I}}, 0) \in \mathbb{R}^{2n}$ in which the infimum of the absolute value of at least one of the components of $\tilde{a}_{\mathcal{I}'}(x, \eta)$ is positive, so whenever (x, η) enters this neighborhood, it exits it in finite time. Therefore, given that any $\mathbf{x} \in \mathcal{O}$ converges to $x^*_{\mathcal{I}}$, no $\boldsymbol{\eta}' \in R'$ can converge to zero. Under (ii), there exists a neighborhood of $x^*_{\mathcal{I}}$ included in $B(x^*_{\mathcal{I}'}, \rho')$. Since $\Xi_{\mathcal{I}',x_0}(\boldsymbol{\eta}') \to x^*_{\mathcal{I}}$, there exists $K \in \mathbb{N}$ such that $\Xi_{\mathcal{I}',x_0}(\boldsymbol{\eta}')(k)$ belongs to $B(x^*_{\mathcal{I}'}, \rho')$ for all $k \geq K$. Therefore, if $|\eta'(k)| \leq \min\{(\kappa')^{-1}(\rho'), \rho'\}$ indefinitely after any point of time, $\Xi_{\mathcal{I}',x_0}(\boldsymbol{\eta}') \to x^*_{\mathcal{I}'}$ by Proposition II.1 which is a contradiction, so $\boldsymbol{\eta}'$ cannot converge to zero. In both cases, by Lemma IV.1, $\mathbb{P}(R') = 0$, which, together with $\mathbb{P}(R) > 0$ and the definition of ϵ-differential privacy, cf. (5), implies the result. ∎

Note that the hypotheses of Proposition IV.2 are mild and easily satisfied in most cases. In particular, the result holds if the dynamics are continuous and globally asymptotically stable relative to $x^*_{\mathcal{I}}$ for two information sets. The main take-away message of this result is that a globally asymptotically stable distributed optimization algorithm cannot be made differentially private by perturbing the inter-agent messages with asymptotically vanishing noise. This observation is at the core of the design choices made in the literature regarding the use of stepsizes with finite sum to make the zero-input dynamics not asymptotically stable, thereby causing a steady-state error in accuracy which is present independently of the amount of noise injected for privacy. For instance, the algorithmic solution proposed in [17] replaces (7b) by $z_i(k) = \sum_{j=1}^n a_{ij}\xi_j(k)$, where $\xi_j(k) = x_j(k) + \eta_j(k)$ is the perturbed message received from agent j, and chooses a finite-sum sequence of stepsizes $\{\alpha_k\}$ in the computation (7a), leading to a dynamical system which is not 0-GAS, see Figure 1. Similar observations can be made in the scenario considered in [18], where the agents' local constraints are the sensitive information (instead of the objective function). This algorithmic solution uses a constant-variance noise, which would make the dynamics unstable if executed over an infinite time horizon. This problem is circumvented by having the algorithm terminate after a finite number of steps, and optimizing this number offline as a function of the desired level of privacy ϵ.

B. Algorithm Design via Objective Perturbation

To overcome the limitations of message-perturbing strategies, here we outline an alternative design strategy to solve

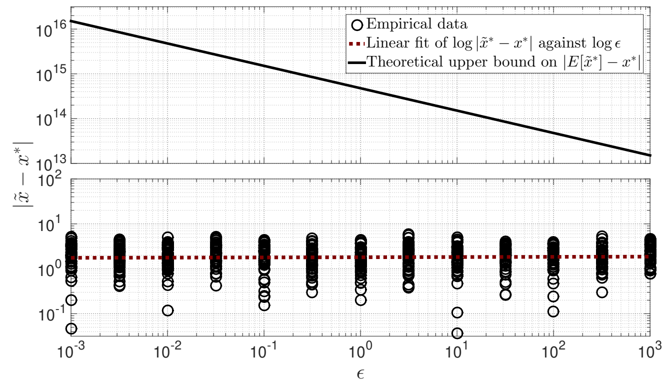


Fig. 1. Privacy-accuracy trade-off for the algorithm proposed in [17] applied to Example III.2 with $D = X = [-5, 5]^2$, $n = 10$, $N_d = 100$, and $\lambda = 0.01$. With that paper's notation, we set $q = 0.1$, $p = 0.11$, $c = 0.5$. The stepsize $\alpha_k = cq^{k-1}$ has finite sum. The circles, dotted line, and solid line illustrate simulation results for 50 executions, their best linear fit in logarithmic scale, and the upper bound on accuracy provided in [17], respectively. We have broken the vertical axis to better display the scale of the algorithm output.

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. In the context of Example III.2, this would correspond to leave (7b) and the sequence of stepsizes unchanged, and instead use perturbed functions in the computation (7a). The latter in turn automatically adds noise to the estimates shared with neighbors. The following result, which is a special case of [26, Theorem 1], ensures that the combination with the distributed optimization algorithm does not affect the differential privacy at the functional level.

Proposition IV.3. *(Resilience to post-processing): Let $\mathcal{M} : L_2(D)^n \times \Omega \to L_2(D)^n$ be ϵ-differentially private (cf. Definition III.1) and $\mathcal{F} : L_2(D)^n \to \mathcal{X}$, where $(\mathcal{X}, \Sigma_{\mathcal{X}})$ is an arbitrary measurable space. Then, $\mathcal{F} \circ \mathcal{M} : L_2(D)^n \times \Omega \to \mathcal{X}$ is ϵ-differentially private.*

Proof: First, note that although a slightly different definition of differential privacy is used in [26], the exact same proof of [26, Theorem 1] works with Definition III.1. Consider the σ-algebra $\mathcal{P}(L_2(D)^n)$ on $L_2(D)^n$ where $\mathcal{P}$ denotes the power set. With the notation of [26, Theorem 1], $M_2 = \mathcal{F} \circ \mathcal{M}$ is a deterministic function of the output of $M_1 = \mathcal{M}$. Then, it is easy to verify that, for any $S \in \Sigma_{\mathcal{X}}$,

$$\mathbb{P}(M_2(F) \in S \mid M_1(F)) = \chi_S(\mathcal{F}(M_1(F))),$$

(with $\chi_\cdot$ being the indicator function) is measurable as a function of $M_1(F)$ (because $\mathcal{F}$ and S are trivially measurable) and defines a probability measure on $(\mathcal{X}, \Sigma_{\mathcal{X}})$ (associated to a singleton), so it is a probability kernel. Hence, the conditions of [26, Theorem 1] are satisfied and $\mathcal{F} \circ \mathcal{M}$ is ϵ-differentially private. ∎

Our design strategy based on the perturbation of individual objective functions requires solving the following challenges:

(i) establishing a differentially private procedure to perturb the individual objective functions;

(ii) ensuring 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, characterizing the accuracy of the resulting differentially private, distributed coordination algorithm.

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

V. Functional Differential Privacy

We explore here the concept of functional differential privacy to address the challenge (i) laid out in Section IV-B. The generality of this notion makes it amenable for problems where the sensitive information is a function or some of its attributes (e.g., sample points, optimizers, derivatives and integrals). For simplicity of exposition and without loss of generality, we limit our discussion in this section to the privacy of a single function.

A. Functional Perturbation via Laplace Noise

Let $f \in L_2(D)$ be a function whose differential privacy has to be preserved. With the notation 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, we set

$$\mathcal{M}(f, \boldsymbol{\eta}) = \Phi\left(\Phi^{-1}(f) + \boldsymbol{\eta}\right) = f + \Phi(\boldsymbol{\eta}), \tag{9}$$

where

$$\eta_k \sim \mathrm{Lap}(b_k), \tag{10}$$

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

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

$$b_k \leq \frac{1}{k^p \log k^s}, \quad \forall k \geq K, \tag{11}$$

then $\boldsymbol{\eta}$ defined by (10) belongs to ℓ_2 with probability one. In particular, if for some $p > \frac{1}{2}$ and $\gamma > 0$,

$$b_k \leq \frac{\gamma}{k^p}, \quad \forall k \in \mathbb{N}, \tag{12}$$

then $\boldsymbol{\eta}$ defined by (10) belongs to ℓ_2 with probability one.

Proof: Equation (11) can be equivalently written as $e^{-\frac{1}{k^p b_k}} \leq \frac{1}{k^s}$, for $k \geq K$. In particular, this implies that $\sum_{k=1}^{\infty} e^{-\frac{1}{k^p b_k}}$ is convergent. Therefore [25, §1.14], $\prod_{k=1}^{\infty}\big(1-e^{-\frac{1}{k^p b_k}}\big)$ converges (i.e., the limit exists and is nonzero), so

$$1 = \lim_{K\to\infty} \prod_{k=K}^{\infty}\left(1 - e^{-\frac{1}{k^p b_k}}\right) = \lim_{K\to\infty} \mathbb{P}(E_K),$$

where $E_K = \left\{\boldsymbol{\eta} \in \mathbb{R}^{\mathbb{N}} \mid \forall k \geq K,\ |\eta_k| \leq \frac{1}{k^p}\right\}$ and we have used the fact that $|\eta_k|$ is exponentially distributed with rate $\frac{1}{b_k}$. Since $E_K \uparrow \bigcup_{K=1}^{\infty} E_K$ as $K \to \infty$, we have

$$1 = \mathbb{P}\Big(\bigcup_{K=1}^{\infty} E_K\Big) = \mathbb{P}\Big\{\boldsymbol{\eta} \in \mathbb{R}^{\mathbb{N}} \,|\, \exists K \in \mathbb{N} \text{ s.t. } \forall k \geq K{:}\, |\eta_k| \leq \frac{1}{k^p}\Big\} \leq \mathbb{P}\{\boldsymbol{\eta} \in \ell_2\},$$

as stated. If equation (12) holds, we define $\bar{p} = \frac{1}{2}(p + \frac{1}{2})$ and equivalently write (12) as

$$b_k \leq \frac{1}{k^{\bar{p}}} \frac{\gamma}{k^{p-\bar{p}}}, \quad \forall k \in \mathbb{N}.$$

Since $p - \bar{p} > 0$, for any $s > 1$ there exists $K \in \mathbb{N}$ such that $k^{p-\bar{p}} \geq \gamma \log k^s$ for all $k \geq K$, and the result follows. ■

Having established conditions on the noise variance under which the map (9) is well defined, we next turn our attention to establish its differential privacy.

B. Differential Privacy of Functional Perturbation

Here, we establish the differential privacy of the map (9). In order to do so, we first specify our choice of adjacency space $\mathcal{V}$. Given $q > 1$, consider the weight sequence $\{k^q\}_{k=1}^{\infty}$ and define the adjacency vector space to be the image of the resulting weighted ℓ_2 space under Φ, i.e.,

$$\mathcal{V}_q = \Phi\Big(\Big\{\boldsymbol{\delta} \in \mathbb{R}^{\mathbb{N}} \mid \sum_{k=1}^{\infty} (k^q \delta_k)^2 < \infty\Big\}\Big). \tag{13}$$

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

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

is a norm on $\mathcal{V}_q$. The next result establishes the differential privacy of (9) for a properly chosen noise scale sequence $\mathbf{b}$.

Theorem V.2. *(Differential privacy of functional perturbation): Given $q > 1$, $\gamma > 0$ and $p \in \left(\frac{1}{2}, q - \frac{1}{2}\right)$, let*

$$b_k = \frac{\gamma}{k^p}, \quad k \in \mathbb{N}. \tag{14}$$

Then, the map (9) is ϵ-differentially private with

$$\epsilon = \frac{1}{\gamma}\sqrt{\zeta(2(q-p))}, \tag{15}$$

where ζ is the Riemann zeta function.

Proof: Note that the map $\mathcal{M}$ defined by (9) is well defined because (14) ensures, by Lemma V.1, that $\boldsymbol{\eta}$ belongs to ℓ_2 almost surely. Our proof consists of showing that $\mathcal{M}$ satisfies the definition of differential privacy, cf. Definition III.1. To this effect, consider two functions f and f', with $f - f' \in \mathcal{V}_q$, and an arbitrary set $\mathcal{O} \subseteq L_2(D)$. Let $\Phi_K^{-1} : L_2(D) \to \mathbb{R}^K$ be the map that returns the first K coefficients of $\Phi^{-1}(\cdot)$ and

$$\mathcal{L}^K(\boldsymbol{\eta}_K; \mathbf{b}_K) \triangleq \prod_{k=1}^{K} \mathcal{L}(\eta_k; b_k).$$

We have

$$\mathbb{P}\{f + \Phi(\boldsymbol{\eta}) \in \mathcal{O}\} = \mathbb{P}\{\boldsymbol{\eta} \in \Phi^{-1}(\mathcal{O} - f)\} = \lim_{K\to\infty} \int_{\Phi_K^{-1}(\mathcal{O}-f)} \mathcal{L}^K(\boldsymbol{\eta}_K; \mathbf{b}_K) d\boldsymbol{\eta}_K,$$

where $\Phi_K^{-1}(\mathcal{O}-f)$ denotes the inverse image of the set $\mathcal{O}-f = \{g \in L_2(D) \mid g+f \in \mathcal{O}\}$ and the second equality follows from the continuity of probability [27, Theorem 1.1.1.iv] (since $\Phi_K^{-1}(\mathcal{O}-f) \times \mathbb{R}^{\mathbb{N}} \downarrow \Phi^{-1}(\mathcal{O}-f)$ as $K \to \infty$). Similarly,

$$\mathbb{P}\{f' + \Phi(\boldsymbol{\eta}') \in \mathcal{O}\} = \lim_{K\to\infty} \int_{\Phi_K^{-1}(\mathcal{O}-f')} \mathcal{L}^K(\boldsymbol{\eta}'_K; \mathbf{b}_K) d\boldsymbol{\eta}'_K.$$

By linearity of Φ_K, we have $\Phi_K^{-1}(\mathcal{O}-f') = \Phi_K^{-1}(\mathcal{O}-f) + \boldsymbol{\delta}_K$ where $\boldsymbol{\delta} = \Phi^{-1}(f - f')$. Therefore,

$$\mathbb{P}\{f' + \Phi(\boldsymbol{\eta}') \in \mathcal{O}\} = \lim_{K\to\infty} \int_{\Phi_K^{-1}(\mathcal{O}-f)} \mathcal{L}^K(\boldsymbol{\eta}_K + \boldsymbol{\delta}_K; \mathbf{b}_K) d\boldsymbol{\eta}_K.$$

Note that

$$\frac{\mathcal{L}^K(\boldsymbol{\eta}_K + \boldsymbol{\delta}_K; \mathbf{b}_K)}{\mathcal{L}^K(\boldsymbol{\eta}_K; \mathbf{b}_K)} = \prod_{k=1}^{K} \frac{\mathcal{L}(\eta_k + \delta_k; b_k)}{\mathcal{L}(\eta_k; b_k)} \leq e^{\sum_{k=1}^{K} \frac{|\delta_k|}{b_k}}.$$

After multiplying both sides by $\mathcal{L}^K(\boldsymbol{\eta}_K; \mathbf{b}_K)$, integrating over $\Phi_K^{-1}(\mathcal{O}-f)$, and letting $K \to \infty$, we have

$$\mathbb{P}\{f' + \Phi(\boldsymbol{\eta}') \in \mathcal{O}\} \leq e^{\sum_{k=1}^{\infty} \frac{|\delta_k|}{b_k}} \mathbb{P}\{f + \Phi(\boldsymbol{\eta}) \in \mathcal{O}\}.$$

Finally, the coefficient of the exponential can be upper bounded using Holder's inequality with $p = q = 2$ as

$$\sum_{k=1}^{\infty} \frac{|\delta_k|}{b_k} = \sum_{k=1}^{\infty} \frac{k^q |\delta_k|}{k^q b_k} \leq \left(\sum_{k=1}^{\infty} \frac{1}{(k^q b_k)^2} \right)^{\frac{1}{2}} \left(\sum_{k=1}^{\infty} (k^q \delta_k)^2 \right)^{\frac{1}{2}}$$
$$= \left(\sum_{k=1}^{\infty} \frac{1}{(\gamma k^{q-p})^2} \right)^{\frac{1}{2}} \|f - f'\|_{\mathcal{V}_q} = \frac{1}{\gamma} \sqrt{\zeta(2(q-p))}\, \|f - f'\|_{\mathcal{V}_q},$$

which completes the proof. ■

Remark V.3. *(Choice of q):* The choice of parameter q affects the trade-off between the size of the adjacency space $\mathcal{V}_q$ and the noise required for privacy. From (13), we see that decreasing q makes $\mathcal{V}_q$ larger, allowing for the privacy preservation of a larger collection of functions. However, as expected, preserving privacy in a larger space requires more noise. From (12), γ will be larger for a fixed ϵ, (since p cannot be decreased by the same amount as q and ζ is monotonically decreasing), resulting in larger b_k and larger noise. We show later in Theorem VI.2 that the guaranteed upper bound on the expected minimizer deviation also increases as $\{q_i\}_{i=1}^n$ decrease. □

VI. Differentially Private Distributed Optimization

In this section, we employ functional differential privacy to solve the differentially private distributed optimization problem formulated in Section III for a group of $n \in \mathbb{N}$ agents. For convenience, we introduce the shorthand notation $\mathcal{S}_0 = C^2(D) \subset L_2(D)$ and, for given $\overline{u} > 0$, $0 < \alpha < \beta$,

$$\mathcal{S} = \{h \in \mathcal{S}_0 \mid |\nabla h(x)| \leq \overline{u},\ \forall x \in D \text{ and } \alpha I_d \leq \nabla^2 h(x) \leq \beta I_d,\ \forall x \in D^o\},$$

for twice continuously differentiable functions with bounded gradients and Hessians. In the rest of the paper, we assume that the agents' local objective functions $f_1, \dots, f_n$ belong to $\mathcal{S}$.

A. Smoothness and Regularity of the Perturbed Functions

We address here the challenge (ii) laid out in Section IV-B. To exploit the framework of functional differential privacy for optimization, we need to ensure that the perturbed functions have the smoothness and regularity properties required by the distributed coordination algorithm. In general, the output of (9) might neither be smooth nor convex. We detail next how to address these problems by defining appropriate maps that, when composed with $\mathcal{M}$ in (9), yield functions with the desired properties. Proposition IV.3 ensures that differential privacy is retained throughout this procedure.

1) Ensuring Smoothness: To ensure smoothness, we rely on the fact that $\mathcal{S}_0$ is dense in $L_2(D)$ and, therefore, given any function g in $L_2(D)$, 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).

Remark VI.1. *(Smoothening and Truncation):* A natural choice for the smoothening step, if the basis functions are smooth (i.e., $\{e_k\}_{k=1}^\infty \subset \mathcal{S}_0$), is truncating the infinite expansion of g. Such truncation is also inevitable in practical implementations due to the impossibility of handling infinite series. The appropriate truncation order depends on the specific function, the basis set, and the noise decay rate (p in (14)). □

2) Ensuring Strong Convexity and Bounded Hessian: 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 VI.2. *(Convexity of $\mathcal{S}$ and its 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.*

Proof: The set $\mathcal{S}$ is clearly convex because, if $h_1, h_2 \in \mathcal{S}$ and $\lambda \in [0, 1]$, then for all $x \in D^o$,

$$\begin{aligned}\nabla^2((1-\lambda)h_1(x) + \lambda h_2(x)) &= (1-\lambda)\nabla^2 h_1(x) + \lambda \nabla^2 h_2(x) \\ &\geq (1-\lambda)\alpha I_d + \lambda \alpha I_d = \alpha I_d.\end{aligned}$$

Similarly, $\nabla^2((1-\lambda)h_1(x) + \lambda h_2(x)) \leq \beta I_d$. Also,

$$\begin{aligned}|\nabla((1-\lambda)h_1(x) + \lambda h_2(x))| &\leq (1-\lambda)|\nabla h_1(x)| + \lambda |\nabla h_2(x)| \\ &\leq (1-\lambda)\overline{u} + \lambda \overline{u} \leq \overline{u},\end{aligned}$$

for all $x \in D$. To establish closedness, let

$$\mathcal{S}_1 = \{h \in \mathcal{S}_0 \mid \alpha I_d \leq \nabla^2 h(x) \leq \beta I_d,\ \forall x \in D^o\},$$
$$\mathcal{S}_2 = \{h \in \mathcal{S}_0 \mid |\nabla h(x)| \leq \overline{u},\ \forall x \in D\}.$$

Since $\mathcal{S} = \mathcal{S}_1 \cap \mathcal{S}_2$, it is enough to show that $\mathcal{S}_1$ and $\mathcal{S}_2$ are both closed subsets of $\mathcal{S}_0$.

To show that $\mathcal{S}_1$ is closed, let $\{h_k\}_{k=1}^\infty$ be a sequence of functions in $\mathcal{S}_1$ such that $h_k \xrightarrow{\|\cdot\|_2} h \in \mathcal{S}_0$. We show that $h \in \mathcal{S}_1$. Since $h_k - \frac{\alpha}{2}|x|^2 \xrightarrow{\|\cdot\|_2} h - \frac{\alpha}{2}|x|^2$ and L_2 convergence implies pointwise convergence of a subsequence almost everywhere, there exists $\{h_{k_\ell}\}_{\ell=1}^\infty$ and $Y \subset D$ such that $m(D \setminus Y) = 0$ and $h_{k_\ell}(x) - \frac{\alpha}{2}|x|^2 \to h(x) - \frac{\alpha}{2}|x|^2$ for all $x \in Y$. It is straightforward to verify that Y is dense in D and therefore $Y \cap D^o$ is dense in D^o. Then, by [28, Theorem 10.8], $h - \frac{\alpha}{2}|x|^2$ is convex on D^o, so $\alpha I_2 \leq \nabla^2 h(x)$ for all $x \in D^o$. Similarly, one can show that $\nabla^2 h(x) \leq \beta I_d$ for all $x \in D^o$. Therefore, $h \in \mathcal{S}_1$.

The proof of closedness of $\mathcal{S}_2$ is more involved. By contradiction, assume that $\{h_k\}_{k=1}^\infty$ is a sequence of functions in $\mathcal{S}_2$ such that $h_k \xrightarrow{\|\cdot\|_2} h \in \mathcal{S}_0$ but $h \notin \mathcal{S}_2$. Therefore, there exist $x_0 \in D^o$ such that $|\nabla h(x_0)| > \overline{u}$ and, by continuity of ∇h, $\delta_0 > 0$ and $\upsilon_0 > 0$ such that

$$|\nabla h(x)| \geq \overline{u} + \upsilon_0, \quad \forall x \in B(x_0, \delta_0) \subseteq D.$$

Let $u_0 = \frac{\nabla h(x_0)}{|\nabla h(x_0)|}$. By continuity of ∇h, for all $\upsilon_1 > 0$ there exists $\delta_1 \in (0, \delta_0]$ such that

$$\nabla h(x) \cdot u_0 \geq (1 - \upsilon_1)|\nabla h(x)|, \quad \forall x \in B(x_0, \delta_1).$$

As mentioned above, L_2 convergence implies pointwise convergence of a subsequence $\{h_{k_\ell}\}_{\ell=1}^\infty$ almost everywhere. In turn, this subsequence converges to h almost uniformly, i.e., for all $\upsilon_2 > 0$ and all $\upsilon_3 > 0$, there exist $E \subset D$ and $L \in \mathbb{N}$ such that $m(E) < \upsilon_2$ and

$$|h_{k_\ell}(x) - h(x)| < \upsilon_3, \quad \forall x \in D \setminus E \text{ and } \ell \geq L. \tag{16}$$

For ease of notation, let $\delta_2 = \delta_1/2$. Using the fundamental theorem of line integrals [29], for all $x \in B(x_0, \delta_2) \setminus E$,

$$\begin{aligned} h(x + \delta_2 u_0) - h(x) &= \int_x^{x+\delta_2 u_0} \nabla h \cdot dr \\ &= \int_x^{x+\delta_2 u_0} \nabla h \cdot u_0 |dr| \geq \int_x^{x+\delta_2 u_0} (1 - \upsilon_1)|\nabla h||dr| \\ &\geq (1 - \upsilon_1)(\overline{u} + \upsilon_0)\delta_2. \end{aligned} \tag{17}$$

Similarly, for all $x \in B(x_0, \delta_2) \setminus E$ and all $\ell \in \mathbb{N}$,

$$\begin{aligned} h_{k_\ell}(x + \delta_2 u_0) - h_{k_\ell}(x) &= \int_x^{x+\delta_2 u_0} \nabla h_{k_\ell} \cdot dr \\ &\leq \int_x^{x+\delta_2 u_0} |\nabla h_{k_\ell}||dr| \leq \overline{u}\delta_2. \end{aligned} \tag{18}$$

Putting (17), (18), and (16) together and choosing $\upsilon_3 = \upsilon_1 \delta_2 \overline{u}$, we have for all $x \in B(x_0, \delta_2) \setminus E$ and all $\ell \geq L$,

$$\begin{aligned} h(x + \delta_2 u_0) - h_{k_\ell}(x + \delta_2 u_0) & \\ \geq h(x) - h_{k_\ell}(x) + \delta_2(1 - \upsilon_1)(\overline{u} + \upsilon_0) - \delta_2 \overline{u} & \\ \geq \delta_2(1 - \upsilon_1)(\overline{u} + \upsilon_0) - \delta_2(1 + \upsilon_1)\overline{u} \triangleq \upsilon_4. & \end{aligned} \tag{19}$$

The quantity υ_4 can be made strictly positive choosing $\upsilon_1 = \frac{\upsilon_0}{4\overline{u} + 3\upsilon_0} > 0$. Let $E^+ = E + \delta_2 u_0$ and $x_1 = x_0 + \delta_2 u_0$. Then, (19) can be rewritten as

$$h(x) - h_{k_\ell}(x) \geq \upsilon_4, \quad \forall x \in \mathcal{N}_{\delta_2}(x_1) \setminus E^+ \text{ and } \ell \geq L,$$

which, by choosing $\upsilon_2 = \frac{1}{2} m(B(x_1, \delta_2))$, implies

$$\begin{aligned} &\int_{\mathcal{N}_{\delta_2}(x_1) \setminus E^+} |h(x) - h_{k_\ell}(x)|^2 dx \geq \upsilon_4^2 \cdot m(B(x_1, \delta_2) \setminus E^+) \\ &\Rightarrow \|h - h_{k_\ell}\| \geq \upsilon_4 \sqrt{m(B(x_1, \delta_2))/2} > 0. \end{aligned}$$

This contradicts $h_{k_\ell} \xrightarrow{\|\cdot\|_2} h$, so $\mathcal{S}_2$ must be closed. ■

Given the result in Proposition VI.2, the best approximation in $\mathcal{S}$ of a function $h \in \mathcal{S}_0$ is its unique projection onto $\mathcal{S}$, i.e.,

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

By definition, $\tilde{h}$ has bounded gradient and Hessian.

B. Algorithm Design and Analysis

We address here the challenge (iii) laid out in Section IV-B and put together the discussion above to propose a class of differentially private, distributed optimization algorithms that solve Problem 1. Unlike the message-perturbing algorithms where agents use the original objective functions in the computations and rely on perturbing the inter-agent messages, here we propose that agents locally perturb their objective functions and use them in their computations, without adding any additional noise to the inter-agent messages. Therefore, we require each agent $i \in \{1, \ldots, n\}$ to first compute

$$\hat{f}_i = \mathcal{M}(f_i, \boldsymbol{\eta}_i) = f_i + \Phi(\boldsymbol{\eta}_i), \tag{20a}$$

where $\boldsymbol{\eta}_i$ is a sequence of Laplace noise generated by i according to (10) with the choice (14), then select $\hat{f}_i^s \in \mathcal{S}_0$ such that

$$\|\hat{f}_i - \hat{f}_i^s\| < \varepsilon_i, \tag{20b}$$

and finally compute

$$\tilde{f}_i = \text{proj}_{\mathcal{S}}(\hat{f}_i^s). \tag{20c}$$

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

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

denote, respectively, the output of the distributed algorithm and the optimizer for the original optimization problem (with objective functions $\{f_i\}_{i=1}^n$). The following result establishes the connection between the algorithm's accuracy and the design parameters.

Theorem VI.3. *(Accuracy of the proposed class of distributed, differentially private coordination algorithms): Consider a group of n agents which perturb their local objective functions according to* (20) *with Laplace noise* (10) *of variance* (14)*, where $q_i > 1$, $\gamma_i > 0$, and $p_i \in \left(\frac{1}{2}, q_i - \frac{1}{2}\right)$ for all $i \in \{1, \ldots, n\}$. Let the agents participate in any distributed coordination algorithm that asymptotically converges to the optimizer $\tilde{x}^*$ of the perturbed aggregate objective function. Then, ϵ_i-differential privacy of each agent i's original objective function is preserved with $\epsilon_i = \sqrt{\zeta(2(q_i - p_i))}/\gamma_i$ and*

$$\mathbb{E}\,|\tilde{x}^* - x^*| \leq \sum_{i=1}^n \kappa_n\left(\gamma_i \sqrt{\zeta(2p_i)}\right) + \kappa_n(\varepsilon_i),$$

where the function $\kappa_n \equiv \kappa_{n\alpha, n\beta} \in \mathcal{K}_\infty$ is defined in Proposition A.2.

Proof: Since the distributed algorithm is a post-processing step on the perturbed functions, privacy preservation of the objective functions follows from Theorem V.2 and Proposition IV.3. For convenience, let

$$\Delta = \mathbb{E}\,|\tilde{x}^* - x^*| = \mathbb{E}\Big|\operatorname*{argmin}_{x\in X}\sum_{i=1}^{n}\tilde{f}_i - \operatorname*{argmin}_{x\in X}\sum_{i=1}^{n} f_i\Big|.$$

Since $\mu_{n\alpha,n\beta}$ is convex and belongs to class $\mathcal{K}_\infty$ (so is monotonically increasing), κ_n is concave and belongs to class $\mathcal{K}_\infty$ and so is subadditive. Therefore, using Proposition A.2,

$$\begin{aligned}\Delta &\le \mathbb{E}\Big[\kappa_n\Big(\Big\|\sum_{i=1}^{n}\tilde{f}_i - \sum_{i=1}^{n} f_i\Big\|\Big)\Big]\\ &\le \mathbb{E}\Big[\kappa_n\Big(\sum_{i=1}^{n}\|\tilde{f}_i - f_i\|\Big)\Big] \le \sum_{i=1}^{n}\mathbb{E}\big[\kappa_n\big(\|\tilde{f}_i - f_i\|\big)\big].\end{aligned}$$

Then, by the non-expansiveness of projection, we have

$$\begin{aligned}\Delta &\le \sum_{i=1}^{n}\mathbb{E}\big[\kappa_n\big(\|\hat{f}_i^s - f_i\|\big)\big]\\ &\le \sum_{i=1}^{n}\mathbb{E}\big[\kappa_n\big(\|\hat{f}_i^s - \hat{f}_i\|\big) + \kappa_n\big(\|\hat{f}_i - f_i\|\big)\big]\\ &\le \sum_{i=1}^{n}\big(\kappa_n(\varepsilon_i) + \mathbb{E}\big[\kappa_n\big(\|\boldsymbol{\eta}_i\|\big)\big]\big). \qquad (21)\end{aligned}$$

By invoking Jensen's inequality twice, for all $i \in \{1,\dots,n\}$,

$$\begin{aligned}&\mathbb{E}\big[\kappa_n(\|\boldsymbol{\eta}_i\|)\big] \le \kappa_n\big(\mathbb{E}\big[\|\boldsymbol{\eta}_i\|\big]\big) = \kappa_n\big(\mathbb{E}\big[\sqrt{\|\boldsymbol{\eta}_i\|^2}\big]\big) \qquad (22)\\ &\le \kappa_n\Big(\sqrt{\mathbb{E}\big[\|\boldsymbol{\eta}_i\|^2\big]}\Big) = \kappa_n\Big(\sqrt{\textstyle\sum_{k=1}^{\infty} b_{i,k}^2}\Big) = \kappa_n\Big(\gamma_i\sqrt{\zeta(2p_i)}\Big).\end{aligned}$$

The result follows from (21) and (22). ■

The following result describes the trade-off between accuracy and privacy. The proof follows by direct substitution.

Corollary VI.4. *(Privacy-accuracy trade-off): Under the hypotheses of Theorem VI.3, if $p_i = \frac{q_i}{2}$ in* (14) *for all i, then*

$$\mathbb{E}\,|\tilde{x}^* - x^*| \le \sum_{i=1}^{n}\kappa_n\Big(\frac{\zeta(q_i)}{\epsilon_i}\Big) + \kappa_n(\varepsilon_i). \qquad (23)$$

In Corollary VI.4, q_i and ϵ_i are chosen independently, which in turn determines the value of γ_i according to (15). Also, it is clear from (23) that in order for the accuracy of the coordination algorithm not to be affected by the smoothening step, each agent $i \in \{1,\dots,n\}$ has to take the value of ε_i sufficiently small so that it is negligible relative to $\zeta(2p_i)/\epsilon_i$. In particular, this procedure can be executed for any arbitrarily large value of ϵ_i, so that in the case of no privacy requirements at all, perfect accuracy is recovered, as specified in Problem 1.

Remark VI.5. *(Accuracy bound for sufficiently large domains):* One can obtain a less conservative bound than (23) on the accuracy of the proposed class of algorithms if the minimizers of all the agents' objective functions are sufficiently far from the boundary of X. This can be made precise via Corollary A.3. If the aggregate objective function satisfies (28) and the amount of noise is also sufficiently small so that the minimizer of the sum of the perturbed objective functions satisfies this condition, then invoking Corollary A.3, we have

$$\begin{aligned}\mathbb{E}\,|\tilde{x}^* - x^*| &\le \frac{L}{n^2}\sum_{i=1}^{n}\Big(\gamma_i^{\frac{2}{d+4}}\zeta(2p_i)^{\frac{1}{d+4}} + \varepsilon_i^{\frac{2}{d+4}}\Big)\\ &= \frac{L}{n^2}\sum_{i=1}^{n}\Big[\Big(\frac{\zeta(q_i)}{\epsilon_i}\Big)^{\frac{2}{d+4}} + \varepsilon_i^{\frac{2}{d+4}}\Big],\end{aligned}$$

where the equality holds under the assumption that $p_i = \frac{q_i}{2}$ in (14) for all $i \in \{1,\dots,n\}$. □

VII. Simulations

In this section, we report simulation results of our algorithm for Example III.2 with $D = X = [-5,5]^2$, $n = 10$, $N_d = 100$, and $\lambda = 0.01$. The orthonormal basis of $L_2(D)$ is constructed from the Gram-Schmidt orthogonalization of the Taylor functions and the series is truncated to the second, sixth, and fourteenth orders, resulting in 15, 28, and 120-dimensional coefficient spaces, respectively. This truncation also acts as the smoothening step described in Section VI-A1, where higher truncation orders result in smaller ε. We evaluate the projection operator in (20c) by numerically solving the convex optimization problem $\min_{\tilde{f}_i\in\mathcal{S}}\|\tilde{f}_i - \hat{f}_i^s\|$, where $\hat{f}_i^s$ is the result of the truncation. The parameters of $\mathcal{S}$ are given by $\alpha = N_d\lambda$, $\beta = N_d\lambda + N_d r_D\sqrt{2} + e^{2r_D}$, and $\overline{u} = \sqrt{2}N_d(\lambda r_D + e^{2r_D})$ where $r_D = 5$. Rather than implementing any specific distributed coordination algorithm, we use an iterative interior-point algorithm on $\tilde{f}$ and f to find the perturbed $\tilde{x}^*$ and original x^* optimizers, respectively (these points correspond to the asymptotic behavior of any provably correct distributed optimization algorithm with the perturbed and original functions, respectively).

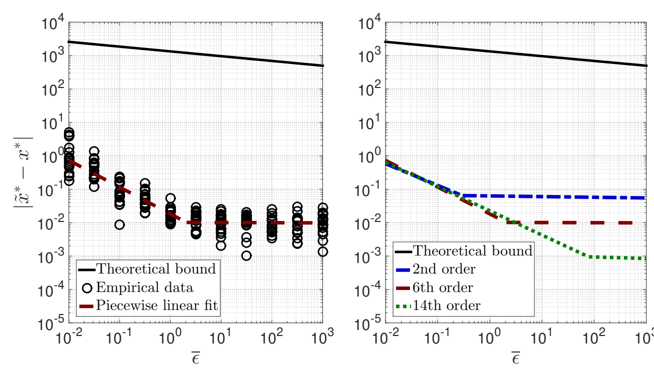


Fig. 2. Privacy-accuracy trade-off curve of the proposed class of distributed, differentially private algorithms in Section VI-B for Example III.2 (with the same data as Figure 1) and different truncation orders. Left: empirical data and its best piecewise linear fit for 6th-order truncation of the function expansions, together with the theoretical upper bound of Corollary VI.4. Right: piecewise linear fit of empirical data for 2nd, 6th, and 14th order truncations as well as the theoretical upper bound. Accuracy improves with the truncation order.

The privacy levels are chosen the same for all agents, i.e., $\epsilon = (\overline{\epsilon}, \overline{\epsilon}, \cdots, \overline{\epsilon})$, and $\overline{\epsilon}$ is swept logarithmically over $[10^{-2}, 10^3]$. For each $i \in \{1,\dots,n\}$, we set $q_i = 2p_i = 1.1$ and $\gamma_i = \sqrt{\zeta(2(q_i - p_i))}/\overline{\epsilon}$. For each value of $\overline{\epsilon}$ and truncation order, the simulations are repeated 20 times to capture the stochasticity of the solutions. Figure 2 illustrates the error $|\tilde{x}^* - x^*|$ as a function of $\overline{\epsilon}$ for different truncation orders, together with the best linear fit of $\log|\tilde{x}^* - x^*|$ against $\log\bar{\epsilon}$, and the upper bound obtained in Corollary VI.4. The conservative nature of this upper bound can be explained by noting the approximations leading to the computation of κ_n in Proposition A.2, suggesting that there is room for refining this bound. Figure 2 shows that accuracy keeps improving as the privacy requirement is relaxed until the ε-term (resulting from the smoothening/truncation) dominates the error. This "saturation" value can be decreased by increasing the truncation order (which comes at the expense of more computational complexity), in contrast with the behavior of message-perturbing algorithms, cf. Figure 1. It is important to mention that the respective error values for a fixed ϵ cannot be compared between Figures 1 and 2 because, in [17], ϵ is defined as the total exponent in (5), i.e., $\epsilon_{i_0}\|f_{i_0} - f'_{i_0}\|_{\mathcal{V}}$. However, it can be seen that the accuracy in Figure 1 is almost indifferent to the value of ϵ and is in the same order as $r_D = 5$. This is explained by the impossibility result of Proposition IV.2: since the noise-free algorithm of [17] is not asymptotically stable, depending on the specific application, its accuracy may not be desirable regardless of the value of ϵ. In contrast, the accuracy in Figure 2 keeps improving as ϵ is increased (with an appropriate choice of truncation order).

VIII. Conclusions and Future Work

We have studied the distributed optimization of the sum of local strongly convex objective functions by a group of agents subject to the requirement of differential privacy. We have established the incompatibility between differential privacy and the asymptotic stability of the underlying (noise-free) dynamics for coordination strategies based on the perturbation of the messages among agents. This has motivated us to develop a new framework for differentially private release of functions from the L_2 space that can be applied to arbitrary separable Hilbert spaces. We have also carefully described how perturbed functions with the desired regularity requirements can be provided to any distributed coordination algorithm while preserving privacy. Finally, we upper bounded the accuracy of the resulting class of distributed, differentially private coordination algorithms. Future work will analyze how to relax the smoothness and convexity assumptions on the local objective functions and the compactness hypothesis on their domains, characterize the *optimal* privacy-accuracy trade-off curve of distributed coordination algorithms based on objective perturbation, characterize the expected suboptimality gap of distributed coordination algorithms for a given desired level of privacy, provide procedures for choosing the truncation order of the functional expansion based on the objective function properties (which must itself be differentially private), compare the numerical efficiency of different orthonormal bases for the L_2 space, and further understand the appropriate scale of the privacy parameter for specific application domains.

Acknowledgments

The authors would like to thank the anonymous reviewers for helpful comments and suggestions that helped improve the presentation. This research was partially supported by NSF Award CNS-1329619 and Award FA9550-15-1-0108.

References

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

[2] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[3] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM Journal on Control and Optimization*, vol. 20, no. 3, pp. 1157–1170, 2009.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] D. Kifer, A. Smith, and A. Thakurta, "Private convex empirical risk minimization and high-dimensional regression," in *25th Annual Conference on Learning Theory*, vol. 23, 2012, pp. 25.1–25.40.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] S. Han, U. Topcu, and G. J. Pappas, "Differentially private distributed constrained optimization," *IEEE Transactions on Automatic Control*, 2016, to appear.

[19] 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.

[20] 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.

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

[22] C. Cai and A. R. Teel, "Results on input-to-state stability for hybrid systems," in *44th IEEE Conference on Decision and Control and European Control Conference*. Seville, Spain: IEEE, Dec. 2005, pp. 5403–5408.

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

[24] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, Oct 2013, pp. 429–438.

[25] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, 3rd ed. Cambridge University Press, 1999.
[26] J. L. Ny and G. J. Pappas, "Differentially private filtering," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 341–354, 2014.
[27] R. Durrett, *Probability: Theory and Examples*, 4th ed., ser. Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010.
[28] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
[29] R. E. Williamson and H. F. Trotter, *Multivariable Mathematics*, 4th ed. Pearson Education, Inc., June 2003.

APPENDIX A
$\mathcal{K}$-LIPSCHITZ PROPERTY OF THE argmin MAP

Here, we establish the Lipschitzness of the argmin map under suitable assumptions. This is a strong result of independent interest given that argmin is not even continuous for arbitrary C^2 functions. Our accuracy analysis for the proposed class of distributed, differentially private algorithms in Section VI-B relies on this result. We begin with a lemma stating a geometric property of balls contained in convex, compact domains.

Lemma A.1. *(Minimum portion of balls contained in convex compact domains): Assume $D \subset \mathbb{R}^d$ is convex, compact, and has nonempty interior and let $r_D > 0$ denote its inradius. Then, there exists $\lambda_D \in (0,1)$ such that,*

$$m(B(x,r) \cap D) \geq \lambda_D m(B(x,r)),$$

for any $x \in D$ and $r \leq r_D$.

Proof: Let $B(c_D, r_D)$ be the inball of D, i.e., the largest ball contained in D. If this ball is not unique, we pick one arbitrarily. Since $D^o \neq \varnothing$, $r_D > 0$. Let R_D be the radius of the largest ball centered at c_D that contains D. Since D is compact, $R_D < \infty$. For any $x \in D$ that is on or outside of $B(c_D, r_D)$, let Σ be the intersection of $B(c_D, r_D)$ and the hyperplane passing through c_D and perpendicular to $c_D - x$. Consider the cone $C = \operatorname{conv}(\Sigma \cup \{x\})$ where conv denotes convex hull. Since D is convex, $C \subseteq D$. Note that C has half angle $\theta_x = \tan^{-1} \frac{r_D}{|x - c_D|}$ so the solid angle at its apex is

$$\Omega_{\theta_x} = \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_0^{\theta_x} \sin^{d-2}(\phi) d\phi. \tag{24}$$

Therefore, for any $r \leq r_D$, the proportion $\frac{\Omega_{\theta_x}}{\Omega_d}$ of $B(x,r)$ is contained in D where $\Omega_d = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}$ is the total d-dimensional solid angle. For any x inside $B(c_D, r_D)$, the same argument holds with

$$\theta_x = \max_{|x - c_D| \geq r_D} \tan^{-1} \frac{r_D}{|x - c_D|} = \frac{\pi}{4}.$$

Therefore, for arbitrary $x \in D$, the statement holds with

$$\lambda_D = \min_{x \in D} \frac{\Omega_{\theta_x}}{\Omega_d} = \frac{1}{\Omega_d} \Omega_{\tan^{-1}(r_D/R_D)}. \qquad \square$$

■

We are now ready to establish the $\mathcal{K}$-Lipschitzness of the argmin map.

Proposition A.2. *($\mathcal{K}$-Lipschitzness of* argmin*): For any two functions $f, g \in \mathcal{S}$,*

$$\Big| \operatorname*{argmin}_{x \in X} f - \operatorname*{argmin}_{x \in X} g \Big| \leq \kappa_{\alpha,\beta}(\|f - g\|), \tag{25}$$

where $\kappa_{\alpha,\beta} \in \mathcal{K}_\infty$ is given by

$$\kappa_{\alpha,\beta}^{-1}(r) = \frac{\alpha^2 \pi^{\frac{d}{2}}}{d 2^{d+3} \Gamma(\frac{d}{2})} \lambda_D \left(\frac{r_D}{d_D}\right)^d r^4 \mu_{\alpha,\beta}^d(r), \quad \forall r \in [0, \infty),$$

r_D and λ_D are as in Lemma A.1, d_D is the diameter of D, and $\mu_{\alpha,\beta} \in \mathcal{K}_\infty$ is defined for all $r \in [0,\infty)$ by

$$\mu_{\alpha,\beta}(r) = \frac{\alpha r^2}{2\sqrt{\alpha\beta r^2 + 2(\beta + \alpha)\overline{u} r + 4\overline{u}^2}}.$$

Proof: We consider the case where $a = \operatorname{argmin}_{x \in X} f(x) \neq \operatorname{argmin}_{x \in X} g(x) = b$ since the statement is trivial otherwise. Let $m_a = f(a)$, $m_b = g(b)$, $m = m_a - m_b$, $u_a = \nabla f(a)$, and $u_b = \nabla g(b)$. Without loss of generality, assume $m \geq 0$. Define,

$$f_l(x) = \frac{\alpha}{2}|x - a|^2 + u_a^T(x - a) + m_a,$$
$$g_u(x) = \frac{\beta}{2}|x - b|^2 + u_b^T(x - b) + m_b,$$

for all $x \in D$. Since $f, g \in \mathcal{S}$, we can integrate $\nabla^2 f \geq \alpha I_d$ and $\nabla^2 g \leq \beta I_d$ twice to get,

$$\forall x \in D \quad f_l(x) \leq f(x) \text{ and } g(x) \leq g_u(x). \tag{26}$$

It follows that, for all $x \in D$,

$$|f(x) - g(x)| \geq [f_l(x) - g_u(x)]^+ \geq [f_l(x) - g_u(x) - m]^+,$$

where $[z]^+ = \max\{z, 0\}$ for any $z \in \mathbb{R}$. After some computations, we get

$$f_l(x) - g_u(x) - m = -\frac{\beta - \alpha}{2}\Big(|x - c|^2 - r^2 \Big),$$

where

$$c = \frac{\beta b - \alpha a + u_a - u_b}{\beta - \alpha},$$
$$r^2 = \frac{\alpha\beta|a - b|^2}{(\beta - \alpha)^2} + \frac{|u_a - u_b|^2}{(\beta - \alpha)^2} + \frac{2(\beta u_a - \alpha u_b)^T(b - a)}{(\beta - \alpha)^2}.$$

Therefore, the region where $f_l - g_u - m \geq 0$ is $B(c, r)$. Next, we seek to identify a subset inside this ball where we can determine a strictly positive lower bound of $f_l - g_u$ that depends on the difference $|a - b|$. To this effect, note that $b \in B(c, r)$, since

$$r^2 - |c - b|^2 = \frac{\alpha}{\beta - \alpha}|a - b|^2 + \frac{2}{\beta - \alpha} u_a^T(b - a),$$

and, by the convexity of the problem, $u_a^T(b - a) \geq 0$. Let $\underline{r} = r - |c - b| > 0$ be the radius of the largest ball centered at b and contained in $B(c, r)$. We have,

$$r^2 - |c - b|^2 = (r - |c - b|)(r + |c - b|) \geq \frac{\alpha}{\beta - \alpha}|a - b|^2$$
$$\Rightarrow \underline{r} \geq \frac{\frac{\alpha}{\beta - \alpha}|a - b|^2}{r + |c - b|} \geq \frac{\frac{\alpha}{\beta - \alpha}|a - b|^2}{2r} \geq \mu_{\alpha,\beta}(|a - b|),$$

where in the last inequality, we have used $|u_a|, |u_b| \leq \overline{u}$. Next, note that for all $x \in B(c, \frac{r + |c - b|}{2})$,

$$f_l(x) - g_u(x) - m \geq - \frac{\beta - \alpha}{2}\Big(\frac{r^2 + |c - b|^2 + 2r|c - b|}{4} - r^2\Big). \tag{27}$$

Using the bound $2r|c-b| \le r^2 + |c-b|^2$, we get after some simplifications,

$$(f_l - g_u)(x) - m \ge \frac{\alpha}{4}|a-b|^2 + \frac{1}{2}u_a^T(b-a) \ge \frac{\alpha}{4}|a-b|^2,$$

for all $x \in B(b, \frac{r}{2}) \subset B(c, \frac{r+|c-b|}{2})$. Therefore,

$$\begin{aligned}
\|f-g\|^2 &= \int_D |f(x)-g(x)|^2 dx \\
&\ge \int_D ([f_l(x) - g_u(x) - m]^+)^2 dx \\
&\ge \int_{B(b,\frac{r}{2})\cap D} (f_l(x) - g_u(x) - m)^2 dx \\
&\ge \frac{\alpha^2}{16}|a-b|^4 m\left(B(b, \tfrac{r}{2}) \cap D\right) \\
&\ge \frac{\alpha^2}{16}|a-b|^4 m\left(B(b, \tfrac{\mu_{\alpha,\beta}(|a-b|)}{2}) \cap D\right).
\end{aligned}$$

Now, we invoke Lemma A.1 to lower bound the last term. Note that $\mu_{\alpha,\beta}(|a-b|) \le |a-b| \le d_D$ for all $a, b \in D$. Therefore, $\frac{r_D}{d_D}\frac{\mu_{\alpha,\beta}(|a-b|)}{2} \le \min\{r_D, \mu_{\alpha,\beta}(|a-b|)/2\}$, so by Lemma A.1,

$$\begin{aligned}
\|f-g\|^2 &\ge \frac{\alpha^2}{16}|a-b|^4 m\left(B\left(b, \frac{r_D \mu_{\alpha,\beta}(|a-b|}{2d_D}\right) \cap D\right) \\
&\ge \frac{\alpha^2}{16}|a-b|^4 \lambda_D \frac{2\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})}\frac{r_D^d}{2^d d_D^d}(\mu_{\alpha,\beta}(|a-b|))^d,
\end{aligned}$$

which yields (25). ■

The next result shows that if the minimizers of f and g are sufficiently far from the boundary of D, then their gradients need not be uniformly bounded and yet one can obtain a less conservative characterization of the $\mathcal{K}$-Lipschitz property of the argmin map.

Corollary A.3. *($\mathcal{K}$-Lipschitzness of* argmin *for sufficiently large domains): If f and g belong to $\mathcal{S}_1 = \{h \in \mathcal{S}_0 \mid \alpha I_d \le \nabla^2 h(x) \le \beta I_d, \ \forall x \in D^o\}$ and*

$$\operatorname*{argmin}_{x\in X} f(x), \ \operatorname*{argmin}_{x\in X} g(x) \in B(c_D, \underline{r}_D) \cap X^o, \tag{28}$$

where $\underline{r}_D = \frac{\beta-\alpha}{\alpha+\beta+2\sqrt{\alpha\beta}} r_D$ and $B(c_D, r_D) \subset D$, then

$$\Big|\operatorname*{argmin}_{x\in X} f - \operatorname*{argmin}_{x\in X} g\Big| \le L\|f-g\|^{\frac{2}{d+4}},$$

where

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

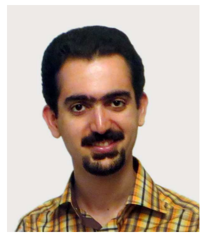
Proof: The proof follows in the same lines as the proof of Proposition A.2 (and we use here the same notation). Since the minimizers of f and g lie in the interior of X, $u_a = u_b = 0$. The main difference here is that due to (28), we have for all $x \in B(c, r)$ that

$$\begin{aligned}
|x - c_D| &\le |x-c| + |c-b| + |b-c_D| \\
&\le \frac{\alpha+\sqrt{\alpha\beta}}{\beta-\alpha}2\underline{r}_D + \underline{r}_D = r_D,
\end{aligned}$$

so $B(c,r) \subset D$. Therefore, one can integrate $(f_l - g_u - m)^2$ on the whole $B(c,r)$ instead of its lower bound (27) on the smaller ball $B(c, \frac{r+|c-b|}{2})$. To explicitly calculate the value of the resulting integral, one can use the change of variables $x_i = c_i + r y_i^{1/2}, i \in \{1, \dots, d\}$ and then use the formula

$$\forall a_i > -1, \ \int_{S_d} y_1^{a_1}\cdots y_d^{a_d} dy = \frac{\Gamma(a_1+1)\cdots\Gamma(a_d+1)}{\Gamma(a_1+\cdots+a_d+d+1)},$$

where $S_d = \{y \in \mathbb{R}^d \mid \sum_{i=1}^d y_i = 1 \text{ and } y_i \ge 0, \ i \in \{1, \dots, d\}\}$. The result then follows from straightforward simplifications of the integral. ■



Erfan Nozari received his B.Sc. degree in Electrical Engineering-Controls from Isfahan University of Technology, Isfahan, Iran in 2013 and M.Sc. in Mechanical Engineering from University of California, San Diego, CA, USA in 2015. He is currently pursuing his Ph.D. degree in Mechanical Engineering from University of California, San Diego, CA, USA. He has been the recipient of the Campus-wide Best Undergraduate Student Award in 2013 from Isfahan University of Technology and the Mechanical and Aerospace Engineering Recruitment Fellowship in 2014 from the University of California, San Diego. His research interests include complex networks, brain networks, networked and distributed control systems, and differential privacy.

Pavankumar Tallapragada received the B.E. degree in Instrumentation Engineering from SGGS Institute of Engineering & Technology, Nanded, India in 2005, M.Sc. (Engg.) degree in Instrumentation from the Indian Institute of Science, Bangalore, India in 2007 and the Ph.D. degree in Mechanical Engineering from the University of Maryland, College Park in 2013. He is currently a Postdoctoral Scholar in the Department of Mechanical and Aerospace Engineering at the University of California, San Diego. His research interests include event-triggered control, networked control systems, distributed control and networked transportation systems.

Jorge Cortés received the Licenciatura degree in mathematics from Universidad de Zaragoza, Zaragoza, Spain, in 1997, and the Ph.D. degree in engineering mathematics from Universidad Carlos III de Madrid, Madrid, Spain, in 2001. He held postdoctoral positions with the University of Twente, Twente, The Netherlands, and the University of Illinois at Urbana-Champaign, Urbana, IL, USA. He was an Assistant Professor with the Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA, USA, from 2004 to 2007. He is currently a Professor in the Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA, USA. He is the author of Geometric, Control and Numerical Aspects of Nonholonomic Systems (Springer-Verlag, 2002) and co-author (together with F. Bullo and S. Martínez) of Distributed Control of Robotic Networks (Princeton University Press, 2009). He is an IEEE Fellow and an IEEE Control Systems Society Distinguished Lecturer. His current research interests include distributed control, networked games, opportunistic state-triggered control and coordination, power networks, distributed optimization, spatial estimation, and geometric mechanics.