Privacy-preserving dual splitting distributed optimization with Application to load flattening in California

Francois Belletti[a], Caroline Le Floch[b], Scott Moura[b], Alexandre M. Bayen[a,b]

***Abstract*— This article presents a dual splitting technique for a class of strongly convex optimization problems whose constraints are block-wise independent. The average-based input in the objective is the only binding element. A dual splitting strategy enables the design of distributed and privacy preserving algorithms. Theoretical convergence bounds and numerical experiments show this method successfully applies to the problem of charging electric devices so as to even out the daily energy demand in California. The solution we provide is a privacy enforced algorithm readily implementable in a network of smart electric vehicle chargers. It can reach any arbitrary precision for the common optimization goal while relying on randomly perturbed information at the agent level. We show that, provided the community is large enough, an averaging effect enables the group to learn its global optimum faster than individual information is leaked. A limited number of messages are sent out in the distributed implementation which prevents adversary statisticians from having low theoretical Mean Square Errors for their estimates.**

I. Introduction

The approach developed in this article is tailored to strongly convex block constrained optimization in which the objective is the sum of two terms. The first one is the compound of an averaging function and a strongly convex cost. The second one is a fully decoupled regularization factor. Although constraints are block-wise independent, the averaging term is binding which leads to the use of a dual splitting method in order to solve the problem on a swarm of distributed computing devices.

Many techniques have been developed to parallelize convex optimization in the case of separable objectives and binding constraints. The ADMM [1], [2] and graph projection [3] algorithms are two instances of the Douglas-Rachford algorithm [4] which belongs to the larger family of proximal point methods [5]. Other approaches have been developed for splitting similar problems such as Springarn's partial inverse algorithm [6], distributed subgradient descents [7] and the proximal center technique [8]. Closer to the setting under consideration here, the primal-dual approaches developed in [9] and [10] respectively deal with block constrained problems where the objective is known partially to the computation nodes and with optimization over a common constraint set that consists of the intersection of local constraint sets.

[a]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, United States, francois.belletti@berkeley.edu

[b]Department of Civil and Environmental Engineering, University of California, Berkeley, United States

However efficient and provably fast, these techniques do not leverage the particular structure of the problem under study here. Therefore, we designed a novel dual splitting strategy tailored to the average-based input in the objective. Gradient methods are used to find the optimum of the dual problem. We consider agents should not give out their own information to the rest of the community in order to preserve their privacy. Convergence rates are calculated for stochastic algorithms that enforce privacy by signal obfuscation. In particular, the proofs below extend previous work on privacy-aware optimization [11], [12], [13]. In this competition to derive estimates of the optimal programs the agents will undertake, we prove that introducing random perturbations in the dual gradient method is much more detrimental to an eavesdropping statistician adversary than it is for the community.

The algorithms designed here straightforwardly apply to evening out electrical power imbalances thanks to a smart grid [14]. In contrast to [15] where no convergence rates were given for the proximal method that had been developed, we show that the precision of our privacy aware procedure increases. Preserving secrecy is indeed crucial in the context of smart metering as pointed out in [16].

The general class of problems under consideration will be presented in Section II. Section III will derive a dual splitting reformulation paving the way to a holistic dual gradient method and an incremental counterpart. The number of iterations preserving secrecy in these algorithms is derived from the attacker model in Section IV. Finally, Section V presents numerical experiments with actual data for re-balancing power consumption in California thanks to electric vehicles.

II. General problem formulation

We consider a group of N agents in a collaborative setting. Each agent, indexed by n, decides on d actions characterized by a vector $u_n \in \mathbb{R}^d = (u_{k,n})_{k\in\{1\dots d\}}$. For each agent, these actions belong to a convex closed set $\mathcal{C}_n \subset \mathbb{R}^d$ corresponding to strictly feasible constraints. Let $u \in \mathbb{R}^{N\times d}$ be the concatenated vector of all individual actions $(u_n)_{n\in\{1\dots N\}}$,and $(o_k)_{k\in\{1\dots d\}}$ the common effort the agents attempt to replicate. The aim for the community is to optimize the collaborative objective below.

Main problem:

$$\min_u \left[\sum_{k=1}^{d} \ell_k\left(o_k - \frac{1}{N}\sum_{n=1}^{N} u_{k,n}\right) + \frac{1}{N}\sum_{n=1}^{N} \mathrm{r}_n\left(u_n\right)\right] \quad (1)$$
$$\text{st } \forall n \in \{1\dots N\},\ u_n \in \mathcal{C}_n$$

Costs are represented by the family of strongly convex functions $(\ell_k(\cdot))_{k\in\{1\dots d\}}$ from $\mathbb{R}$ onto $\mathbb{R}$. All the elements of the family of regularizing functions $(\mathrm{r}(\cdot)_n)_{n\in\{1\dots N\}}$ from $\mathbb{R}^d$ onto $\mathbb{R}$ are strongly convex.

Such a problem, with strongly convex objectives and strictly feasible convex constraints can be solved in a generic but non-scalable manner by standard optimization solvers. The present article develops specific algorithms which efficiently leverage the structure of problem (1).

We also consider that the constraints $u_n \in \mathcal{C}_n$ of each agent cannot be broadcast as they are privacy sensitive. This does not have implications on the formulation (1) of the problem but will be essential in the algorithms to solve it. Let u^* be the solution of the minimization problem (1). The optimal program $u_n^* \in \mathbb{R}^d$ of each agent also contains sensitive information and has to be difficult for an adversary to estimate (the meaning of this statement will be explicitly defined later in the article).

III. Distributed optimization

This section introduces additional variables before using a dual reformulation that manages to split the problem with respect to the number of agents.

A. Dual splitting for independent constraints

As N is the largest scale factor in the problem, it is the most suitable axis for parallelization. We leverage independence between blocks of constraints to render a low-memory and privacy preserving algorithm.

1) Introducing additional variables: For each $k \in \{1\dots d\}$, consider $z_k \in \mathbb{R}^d$. We form the Lagrangian with d real dual variables $(\lambda_k)_{t\in\{1\dots d\}}$ corresponding to constraints $\forall k \in \{1\dots d\}$, $z_k = o_k - \frac{1}{N}\sum_{n=1}^N u_{k,n}$. Slater's conditions hold with the assumptions above (see [17] for details). Minimization and maximization can therefore be swapped in the Lagrangian. This proves that problem (1) is equivalent to

$$\begin{aligned}\max_{\lambda}\min_{u,z}\Bigg[&\sum_{k=1}^d \ell_k(z_k)+\lambda_k z_k\\ &+\sum_{k=1}^d \lambda_k\left(-o_k+\frac{1}{N}\sum_{n=1}^N u_{k,n}\right)\\ &+\frac{1}{N}\sum_{n=1}^N \mathrm{r}_n(u_n)\Bigg]\\ &\text{st } \lambda\in\mathbb{R}^d,\ z\in\mathbb{R}^d,\ \forall n\in\{1\dots N\},\ u_n\in\mathcal{C}_n\end{aligned}$$

2) Block constraints and distribution of min operators: The key step now is that we can distribute the operator min with respect to each z_k and each u_n. Indeed, z and u are decoupled and the constraints $u_n \in \mathcal{C}_n$ are independent by assumption.

Furthermore, considering the Fenchel-Legendre transform $\ell_k^*(\lambda_k) = \sup_{z_k\in\mathbb{R}} l_k(z_k) + \lambda_k z_k$ of $\ell_k(\cdot)$, one has $\min_{z\in\mathbb{R}^d}\sum_{k=1}^d \ell_k(z_k) + \lambda_k z_k = -\sum_{k=1}^d \ell^*(-\lambda_k)$. Also, denoting $\Pi_{n=1}^N \mathcal{C}^n$ the Cartesian product of the constraint sets, $\min_{u\in\Pi_{n=1}^N\mathcal{C}^n}\frac{1}{N}\sum_{n=1}^N \lambda^T u_n + \mathrm{r}(u_n) = \frac{1}{N}\sum_{n=1}^N \min_{u_n\in\mathcal{C}^n}\lambda^T u_n + \mathrm{r}(u_n)$. This proves that problem (1) is equivalent to

$$\begin{aligned}\max_{\lambda}\Bigg[&-\sum_{k=1}^d \ell_k^*(-\lambda_k)-\lambda^T o\\ &+\frac{1}{N}\sum_{n=1}^N \min_{u_n\in\mathcal{C}_n}\lambda^T u_n+\mathrm{r}_n(u_n)\Bigg]\end{aligned}\tag{2}$$

3) Extended value regularization functions: For each $n \in \{1\dots N\}$, let $\overline{r_n}(\cdot)$ be the extended value function that equals $r_n(u_n)$ whenever $u_n \in \mathcal{C}_n$ and $+\infty$ otherwise. With the assumptions above, $r_n(\cdot)$ is proper, closed and lower semi continuous. It is not differentiable in general but is strongly convex by assumption. Let σ_n be its strong convexity constant. Generic convex analysis (see [18] for details) allows us to show that the Fenchel-Legendre transform of $\overline{r_n(\cdot)}$, denoted $\overline{r_n(\cdot)}^*$, is differentiable and has a Lipschitz gradient with constant $\frac{1}{\sigma_n}$. It is also trivially convex. For any $n \in \{1\dots N\}$, the strong convexity assumption on r_n also guarantees uniqueness of $u_n^*(\lambda) = \mathrm{argmin}_{u_n\in\mathcal{C}_n}(\lambda^*)^T u_n + \mathrm{r}_n(u_n)$ where λ^* is the unique solution of problem (2) (see [17] for details).

4) Formulating an optimal price: The problem is now equivalent to the unconstrained minimization below.

Dual split reformulation:

$$\min_{\lambda\in\mathbb{R}^d} f(\lambda) = \min_{\lambda\in\mathbb{R}^d}\left[\sum_{k=1}^d \ell_k^*(-\lambda_k)+\lambda^T o+\frac{1}{N}\sum_{n=1}^N \overline{\mathrm{r}_n}^*(-\lambda)\right]\tag{3}$$

For each $k \in \{1\dots d\}$ we denote L_k the Lipschitz constant of the gradient of $\ell_k(\cdot)$ and m_k the strong concavity constant of the function. As in [18], $\ell_k^*(\cdot)$ has a Lipschitz gradient with constant $\frac{1}{m_k}$ and is strongly convex with constant $\frac{1}{L_k}$. Therefore, f is strongly convex with constant $m = \sum_{k=1}^d \frac{1}{L_k}$ and has a Lipschitz continuous gradient with constant $L = \sum_{k=1}^d \frac{1}{m_k} + \frac{1}{N}\sum_{n=1}^N \frac{1}{\sigma_n}$. The strong convexity property shows in particular that there is a unique price vector λ^* that synthesizes the information contained in the common objective and the constraints. Indeed, if agent n is given λ^* it is sufficient for it to individually solve $\min_{u_n\in\mathcal{C}_n}(\lambda^*)^T u_n + \mathrm{r}_n(u_n)$ in order to retrieve the optimal action u_n^* that contributes to the overall objective best. In particular, this shows privacy sensitive constraints do not have to be shared with other participants in the system.

5) Holistic deterministic gradient descent: Gradient descent and momentum methods are both straightforward ways to minimize f in practice. We have $\nabla f(\lambda) = -\sum_{k=1}^d \nabla\ell_k^*(-\lambda_k) + o - \frac{1}{N}\sum_{n=1}^N \nabla\overline{\mathrm{r}_n}^*(-\lambda)$.

Usual theorems for differentiating maxima of functions (see [19] for details) give, $\forall n \in \{1\dots N\}$, $\nabla\overline{\mathrm{r}_n}^*(\lambda) =$

$u_n^*(-\lambda)$, therefore

$$\nabla f(\lambda) = -\sum_{k=1}^{d} \nabla \ell_k^*(-\lambda_k) + o - \frac{1}{N}\sum_{n=1}^{N} u_n^*(\lambda) \quad (4)$$

.

From [18], we know $O\left(\log\left(\frac{L}{m}\frac{1}{\varepsilon}\right)\right)$ iterations are sufficient for the distributed gradient descent algorithm below (Algorithm 1) to achieve an ε precision in the value of the function we are trying to minimize.

Data: Constraints $(\mathcal{C}_n)_{n\in\{1\dots N\}}$, target $o \in \mathbb{R}^d$
Result: Optimal dual price ($\operatorname{argmin}_{\lambda\in\mathbb{R}^d} f(\lambda)$)
decide on initial value $\lambda^{(0)} \in \mathbb{R}^d$.
for $i \leftarrow 2$ ***to*** *maximum number of steps* **do**
  broadcast $\lambda^{(i)}$
  compute optimal response $u_n^*\left(\lambda^{(i)}\right)$
  broadcast $u_n^*\left(\lambda^{(i)}\right)$
  gather and compute concatenated $u^*\left(\lambda^{(i)}\right)$
  compute $\lambda^{(i+1)} = \lambda^{(i)} - s^{(i)}\nabla f\left(\lambda^{(i)}\right)$
end

Algorithm 1: Holistic distributed gradient descent

It is worth mentioning here that the computation of the gradient is based on independent calculations by the agents. Indeed, calculating $\nabla\overline{\mathrm{r}_n}^*(-\lambda) = u_n^*(\lambda)$ is the only point where agents' constraints are to be taken into account and they are completely decoupled here. In particular, if each agent computes this step locally, it does not have to give any information about its individual constraints to others. This means that the reformulation above yields intrinsically privacy preserving gradient descents.

6) Incremental stochastic gradient method: A standard way of dealing with the optimization of the sum of differentiable functions is the incremental stochastic gradient method. Each holistic computation of the gradient is replaced by a small stochastic correction that on average descends towards the optimum. One can write $f(\lambda) = \frac{1}{N}\sum_1^N f_n(\lambda)$ with

$$f_n(\lambda) = \sum_{k=1}^{d} \ell_k^*(-\lambda_k) + \lambda^T o + \overline{\mathrm{r}_n}^*(-\lambda) \quad (5)$$

For any $n \in \{1\dots N\}$, let L_n^* be the Lipschitz constant of the gradient of f_n, $L_n^* \leq \sum_{k=1}^d \frac{1}{m_k} + \frac{1}{\sigma_n}$, let $\lambda^{n,*}$ be the minimum of f_n over $\mathbb{R}^d$.

The full gradient update step is replaced by a partial increment that only requires the computation of

$$\nabla^{u_{n_0}} f(\lambda) = -\sum_{k=1}^{d} \nabla \ell_k^*(-\lambda_k) + o - \nabla\overline{\mathrm{r}_{n_0}}^*(-\lambda) \quad (6)$$

where n_0 is chosen uniformly at random in $\{1\dots N\}$. Let us recall that $\nabla\overline{\mathrm{r}_{n_0}}^*(-\lambda) = u_{n_0}^*(-\lambda)$ therefore the constraints of u_{n_0} do not need to be broadcast.

One can show, as in [20], that $\mathbb{E}\left[\|\nabla^{u_{n_0}} f(\lambda)\|_2^2\right] \leq A_{\text{inc}}^2\|\lambda - \lambda^*\|_2^2 + B_{\text{inc}}^2$ where $A_{\text{inc}}^2 = \frac{2}{N}\sum_{n=1}^N {L_n^*}^2$ and $B_{\text{inc}}^2 = \max\left(\frac{2}{N}\sum_{n=1}^N {L_n^*}^2\|\lambda^{n,*} - \lambda^*\|_2^2, \frac{A_{\text{inc}}^2 D_0}{2}\right)$ with $D_0 = \mathbb{E}\left[\|\lambda^{(0)} - \lambda^*\|_2^2\right]$

Data: Constraints $(\mathcal{C}_n)_{n\in\{1\dots N\}}$, target $o \in \mathbb{R}^d$
Result: Optimal dual price ($\operatorname{argmin}_{\lambda\in\mathbb{R}^d} f(\lambda)$)
decide on initial value $\lambda^{(0)} \in \mathbb{R}^d$.
for $i \leftarrow 2$ ***to*** *maximum number of steps* **do**
  select n_0 at random in $\{1\dots N\}$
  send $\lambda^{(i)}$ to node n_0
  compute optimal response $u_{n_0}^*\left(\lambda^{(i)}\right)$
  compute $\lambda^{(i+1)} = \lambda^{(i)} - s^{(i)}\nabla^{u_{n_0}} f\left(\lambda^{(i)}\right)$
end

Algorithm 2: Incremental stochastic gradient method

From [20], we deduce the lemma below.

Lemma 1: If one uses decreasing step size $s^{(i)} = 2\left(m\left(2\frac{A_{\text{inc}}^2}{m^2} + i\right)\right)^{-1}$ after i iterations of the incremental stochastic gradient method

$$\mathbb{E}\left[f(\lambda^{(i)}) - f(\lambda^*)\right] \leq \left(\frac{\sum_{n=1}^N L_n^*}{\sum_{n=1}^N {L_n^*}^2}\right)\left(2 + \frac{i}{T_{\text{inc}}}\right)^{-1} \quad (7)$$

where $T_{\text{inc}} = \frac{2}{N}\frac{\sum_{n=1}^N {L_n^*}^2}{m^2}$.

Here the convergence rate of the algorithm presented below, Algorithm 2, is much slower than before. However, the computation burden for each step is considerably lower as compared to the holistic case as only one optimal program needs to be computed and broadcast at each iteration.

IV. Privacy preservation

In the procedures above, the aim of enforcing the confidentiality of individual's constraints has been successfully achieved. This section will focus on obfuscating information that would help infer those indirectly.

A. Privacy model

Algorithm 1 is interpreted as a survey in which, at each iteration i, all agents are queried for their optimal action $u_n^*\left(\lambda^{(i)}\right)$. Agents do not have to send out their personal sets of constraints, $\mathcal{C}^n$, for the dual optimum λ^* to be estimated. However, they send out vectors $u_n^*(\lambda)$ which correspond to the optimal series of actions to undertake with respect to a given signal vector λ. This information is considered privacy sensitive as it can help infer the agents' constraints. In the context of smart metering for instance, $u_n^*(\lambda)$ will correspond to the power consumption profile of a given household. Burglars trying to infer when the house is unoccupied will most likely want to identify low consumption periods of the day. Therefore, one considers a framework close to that of [21], in which participants in a survey are reluctant to give out personal data. It is possible here to leverage the averaging behavior of the dual gradient in order to compute the common optimum of the whole community without jeopardizing individual's privacy.

1) Adding noise to broadcast: Adding noise to data has successfully enabled differential privacy in databases in [22] and in filtering [23] for example. Inspired by this work, we design an algorithm where, instead of sending $u_n^{*(i)} = u_n^*\left(\lambda^{(i)}\right)$, agent n broadcasts $\widetilde{u_n^*}^{(i)} = u_n^{*(i)} + \nu_n^{(i)}$ in which the d-dimensional white noise sequences $(\nu_n^{(i)})_{i\in\mathbb{N}}$ are all mutually independent and have variance η^2 for each of their d components. This framework where only blurry observations of the gradient are available has also been studied in [11]. The approach presented here diverges in that it intrinsically leverages the effect of having many distributed processors taking part in the computation of the gradient. In particular a high value of N is core to obtaining good precision and at the same time privacy enforcement.

2) Learning rate on personal information: In this setting, the system itself cannot be trusted and there is competition between the speed at which the community discovers λ^* and the rate at which a spying statistician can learn individual information. This attacker is trying to estimate $\hat{u^*}_n^{(i)}$ based on a series of i observations $\left(\hat{u^*}_n^{(j)}\right)_{j\in\{1\dots i\}}$ for a given agent n that is targeted as an individual. Classically, when trying to estimate a vector from a series of linearly perturbed measurements, empirical mean estimators or Kalman filters yield a Mean Squared Error (MSE) that will scale proportionally to the variance $d\eta^2$. Therefore we assume the attacker's estimator for $u_n^{*(i)}$ is unbiased and has variance $\mathbb{E}\left[\left\|\widehat{u_n^*}^{(i)} - u_n^{*(i)}\right\|_2^2\right] = \frac{d\eta^2\kappa}{i^\gamma}$ where κ is a constant that depends on the estimation technique adopted by the adversary.

The privacy enforcement criterion here is that the MSE of the estimator of the attacker remains above a certain lower bound $\kappa_{\min}$. This implies the optimization program has an iteration budget $i^{\max} = \left(\frac{d\kappa\eta^2}{\kappa_{\min}}\right)^{\frac{1}{\gamma}}$.

The most favorable case for the attacker occurs when the sequence $\left(u_n^{*(j)}\right)_{j=1\dots i}$ remains constant. The law of large numbers guarantees a convergence rate $\gamma = 1$ for the empirical mean estimator. Thus, from hereon, we will assume $\gamma \leq 1$.

B. Noisy descents

The privacy enforcing strategies below aim at converging towards an optimal scheduling price λ^* faster than the attacker increases its precision in the estimation of u_n^*.

1) Noisy holistic descent strategy: A first strategy to preserve privacy in the distributed gradient computation is to run the deterministic holistic descent above with noisy broadcasts from the agents. The update of λ in the descent becomes $\lambda^{(i+1)} = \lambda^{(i)} - s^{(i)}\widetilde{\nabla f}\left(\lambda^{(i)}\right)$ where

$$\widetilde{\nabla f}\left(\lambda^{(i)}\right) = -\sum_{k=1}^{d}\nabla\ell_k^*\left(-\lambda_k^{(i)}\right) + o - \frac{1}{N}\sum_{n=1}^{N}\widetilde{u^*}_n^{(i)}\left(-\lambda^{(i)}\right).$$

Recalling that $\widetilde{u_n^*}^{(i)} = u_n^{*(i)} + \nu_n^{(i)}$, as $\nu_n^{(i)}$ is a white noise whose variance trace is $d\eta^2$, $\mathbb{E}\left[\widetilde{\nabla f}\left(\lambda^{(i)}\right)\right] = \nabla f\left(\lambda^{(i)}\right)$ and

$$\mathbb{E}\left[\left\|\widetilde{\nabla f}\left(\lambda^{(i)}\right)\right\|_2^2\right] \leq A_{\text{hol}}^2\left\|\lambda^{(i)} - \lambda^*\right\|_2^2 + B_{\text{hol}}^2. \quad (8)$$

where $A_{\text{hol}} = L = \sum_{k=1}^{d}\frac{1}{m_k} + \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\sigma_n}$ and $B_{\text{hol}}^2 = \frac{d}{N}\eta^2$.

Lemma 2: If one uses step size $s^{(i)} = \left(m\left(2\frac{A_{\text{hol}}^2}{m^2} + i\right)\right)^{-1}$, after i iterations of the noisy holistic descent,

$$\mathbb{E}\left[f(\lambda^{(i)}) - f\left(\lambda^*\right)\right] \leq \frac{2B_{\text{hol}}^2}{L\left(2 + \frac{im^2}{A_{\text{hol}}^2}\right)} \leq 2\frac{\frac{d}{N}\eta^2 L}{im^2}. \quad (9)$$

2) Noisy incremental stochastic descent strategy: In such a case we get a similar bound for the expected value error with constants $A_{\text{inc noisy}} = A_{\text{inc}}$ and $B_{\text{inc noisy}}^2 = B_{\text{inc}}^2 + \eta^2$.

The convergence rate is unsurprisingly slower. However participating agents send out sensitive data less often as only one individual selected uniformly at random broadcasts its response vector $u_n^{*(i)} = u_n^*\left(\lambda^{(i)}\right)$ at each iteration. Let $\xi^{(i)}$ be the number of times the agent selected at the first iteration has sent out its optimal response (including that of the first step). The random variable $\xi^{(i)} - 1$ is the sum of $i-1$ Bernoulli trials with parameter $\frac{1}{N}$. Therefore $\xi^{(i)} - 1$ has a binomial distribution with parameters $\frac{1}{N}$ and $i-1$. Jensen inequality yields

$$\mathbb{E}\left[\left\|\widehat{u^*}_n^{(i)} - u_n^{*(i)}\right\|_2^2\right] = E\left[\frac{d\kappa\eta^2}{\left(\xi^{(i)}\right)^\gamma}\right] \geq \frac{d\kappa\eta^2}{\left(1 + \frac{i-1}{N}\right)^\gamma}.$$

Proposition 1: After $N\left(\left(\frac{d\kappa\eta^2}{\kappa_{\min}}\right)^{\frac{1}{\gamma}} - 1\right)$ iterations of the incremental noisy stochastic descent method the MSE of the adversary, $E\left[\left\|\widehat{u^*}_n^{(i)} - u_n^{*(i)}\right\|_2^2\right]$, is bounded by below by $\kappa_{\min}$ and the privacy preservation constraint is respected.

V. Application to load flattening in California

A. Problem formulation

Let $(o_k)_{k=1\dots T}$ be the normalized hourly excess of electrical production in California ($T = 24$). The objective here is to use the electric consumption of N electrical vehicles so as to alleviate the "valley" effect [24]. For a given device n, the daily utilization vector, $u_n = (u_{k,n})_{k=1\dots T}$ is constrained and belongs to a convex closed set $\mathcal{C}_n \subset \mathbb{R}^T$ representing the periods of the day during which the owner is driving the car. The convex program that should be solved collaboratively by the swarm of smart chargers for electric vehicles is

$$\min_c \sum_{k=1}^{T}\frac{1}{2}\left(o_k - \frac{1}{N}\sum_{n=1}^{N}u_{k,n}\right)^2 + \frac{1}{N}\frac{1}{2}\sigma\sum_{n=1}^{N}\|u_n\|_2^2$$
$$\text{st } \forall n \in \{1\dots N\},\ u_n \in \mathcal{C}_n$$

This formulation takes into account both evening out the electrical excess in its primary objective and regularizing the devices' utilization vectors so as to preserve their life span [25]. It is similar to (1) with $\ell_k(x) = \frac{1}{2}x^2$ for any $k \in \{1 \dots T\}$, $\mathrm{r}_n(u_n) = \|u_n\|_2^2$. Therefore, $m = 1$ is a strong convexity constant of the dual split objective. Similarly an admissible Lipschitz continuity constant for its gradient is $L = 1 + \frac{1}{\sigma}$.

1) Iteration budget for the noisy holistic descent: From the derivations above, with the assumptions formulated on the attacker's learning rate, the iteration budget is $i^{\max} = \left(\frac{T\kappa\eta^2}{\kappa_{\min}}\right)^{\frac{1}{\gamma}}$. In this setting, one has $A_{\text{hol}} = L = 1 + \frac{1}{\sigma}$ and $B_{\text{hol}}^2 = T\eta^2$. From (9), one deduces the following theorem.

Theorem 1: The privacy safe precision of the noisy holistic descent is bounded by

$$\mathbb{E}\left[f\left(\lambda^{(i^{\max})}\right) - f\left(\lambda^*\right)\right] \leq 2\left(\frac{\kappa_{\min}}{T\kappa}\right)^{\frac{1}{\gamma}} \frac{T}{N}\left(1 + \frac{1}{\sigma}\right)\eta^{2\left(1-\frac{1}{\gamma}\right)} \tag{10}$$

If $\gamma < 1$, increasing the magnitude of the noise, η, increases the precision that can be attained with a privacy safe iteration budget. A particular case occurs when the attacker has an optimal learning rate $\gamma = 1$. In such a setting we leverage the fact that the reachable precision is proportional to $\frac{1}{N}$. An attacker targets individuals whereas the gradient descent computes the average optimal response at each measurements. Therefore, as more computation nodes are involved in the scheme, noise is more problematic to the attacker than it is to the collaborating swarm.

2) Numerical experiments with noisy holistic method: In the following, actual data of utilization pattern of electric vehicles in California gives a set of $3T$ affine convex constraints for each of the N agents. The prescribed step size is $s^{(i)} = \frac{2}{2\left(1+\frac{1}{\sigma}\right)^2 + i}$. The noise magnitude η is multiplied by the standard deviation of the primal optimum u^* in order to be of the same scale as the signal it perturbs. On the figures below, the value of η before scaling is denominated "normalized η". We run 100 instances of the gradient descent for different values of N, with different magnitudes of η for $\sigma = 1$. Noise has a Laplace distribution here as in [22]. Only problems with relatively small crowds of agents are considered as the primal solution is also computed to provide a comparison baseline. Satisfying regularization is a second order problem here. While respecting the constraints, the effort vector $\left(\frac{1}{N}\sum_{n=1}^{N} u_{k,n}\right)_{k\in\{1\dots T\}}$ needs to replicate $(o_k)_{k\in\{1\dots T\}}$ which is displayed in Figure 1. The effort vector resulting from the dual distributed algorithm is compared to that of the primal solution (calculated by CVX). Filled areas represent the bands between percentiles of empirical distributions. Figure 2 highlights that, with the same iteration budget (20), having more agents involved in the optimization yields a better replication precision. We show that the effect of averaging has an impact on the gradient based search as well. Figure 3 indicates that the empirical MSE is inversely

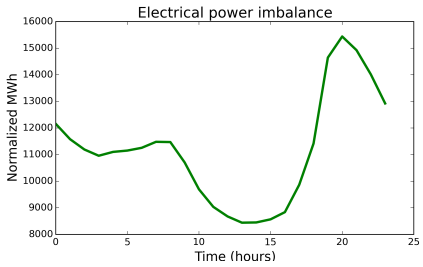

Fig. 1. Objective to collectively replicate.

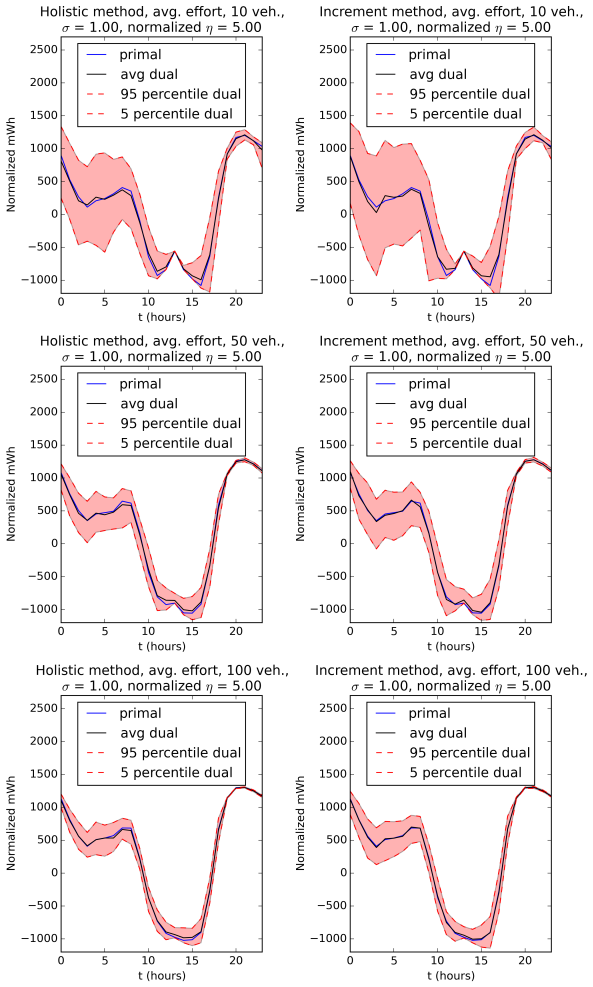

Fig. 2. Errors in aggregated effort.

proportional to N for both the objective and the normalized distance to the optimum $\mathbb{E}\left[\frac{1}{N}\|u - u^*\|_2^2\right]$ (normalizing by N takes into account that u has $N \times T$ elements). Lower value for the latter shows that the averaging gradient helps reach a better precision in both the objective value and the action vector.

3) Iteration budget for the noisy incremental descent: The iteration budget here is $i^{\max} = N\left(\left(\frac{T\kappa\eta^2}{\kappa_{\min}}\right)^{\frac{1}{\gamma}} - 1\right)$. With the derivations above, $A_{\text{inc noisy}}^2 = 2\left(1 + \frac{1}{\sigma}\right)^2$ and $B_{\text{inc noisy}}^2 = 2\beta^2\left(1 + \frac{1}{\sigma}\right)^2 + T\eta^2$ where $\beta^2 = \max\left(\frac{1}{N}\sum_{n=1}^{N}\|\lambda^{n,*} - \lambda^*\|_2^2, \frac{D_0}{2}\right)$.

Theorem 2: The privacy safe expected precision of the

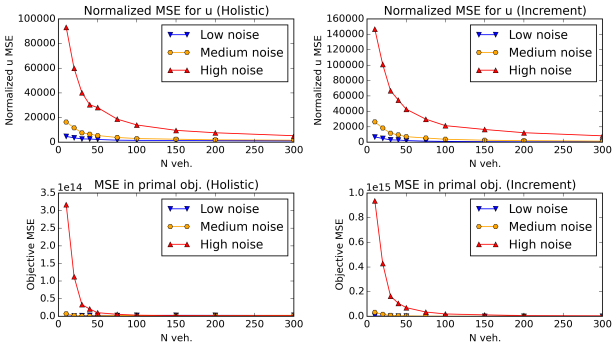

Fig. 3. Errors in response and objective.

noisy incremental descent is bounded by

$$\mathbb{E}\left[f\left(\lambda^{(i)}\right)-f\left(\lambda^{*}\right)\right] \leq 2\frac{\left(1+\frac{1}{\sigma}\right)\left(T\eta^{2}+2\beta^{2}\left(1+\frac{1}{\sigma}\right)^{2}\right)}{N\left(\left(\frac{T\kappa\eta^{2}}{\kappa_{\min}}\right)^{\frac{1}{\gamma}}-1\right)} \tag{11}$$

If $\gamma < 1$, any precision can be reached while guaranteeing secrecy enforcement. If $\gamma = 1$ the average privacy secure precision increases linearly with N. This corresponds to the intuition that, in the incremental descent case, it becomes more unlikely for the attacker to intercept repeated occurrences of the optimal program of a given agent if N is large.

4) Numerical experiments with noisy incremental method: The same numerical experiments are conducted as for the holistic method. However, the new increased iteration budget is taken into account. Figure 2 shows how the incremental is slightly more sensitive to noise than its holistic counterpart. Figure 3 shows that, as in the holistic case, the MSE for the objective value and the agents' actions is inversely proportional to N. The higher number of steps in the descent helps reach a value of u closer to the global optimum u^*.

VI. Conclusion

For the class of problems considered above, splitting the binding objective thanks to a dual reformulation enabled the full separation of independent block constraints. This result has strong implications in terms of enforcing privacy. The individual constraints of the agents remain local and noise can be added to communications that hinders information leakage. Convergence rates indicate that this approach becomes more efficient as more agents are involved. This has been confirmed by numerical experiments conducted with actual data on electricity production and consumption in California. This new scheme is readily applicable to smart electric vehicle chargers whose information are highly privacy sensitive. Further work should focus on asynchronous implementations.

Acknowledgments

The authors wish to thank PhD student Walid Krichene, Professor Laurent El Ghaoui and Professor Ben Recht for their support.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[2] D. Goldfarb and S. Ma, "Fast multiple-splitting algorithms for convex optimization," *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 533–556, 2012.

[3] N. Parikh and S. Boyd, "Block splitting for distributed optimization," *Mathematical Programming Computation*, vol. 6, no. 1, pp. 77–102, 2014.

[4] J. Eckstein and D. P. Bertsekas, "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, 1992.

[5] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM journal on control and optimization*, vol. 14, no. 5, pp. 877–898, 1976.

[6] J. E. Spingarn, "Applications of the method of partial inverses to convex programming: Decomposition," *Mathematical Programming*, vol. 32, pp. 199–223, 1985.

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

[8] I. Necoara and J. A. Suykens, "Application of a smoothing technique to decomposition in convex optimization," *Automatic Control, IEEE Transactions on*, vol. 53, no. 11, pp. 2674–2679, 2008.

[9] T.-H. Chang, A. Nedic, and A. Scaglione, "Distributed constrained optimization by consensus-based primal-dual perturbation method," 2014.

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

[11] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy, data processing inequalities, and minimax rates," *arXiv preprint arXiv:1302.3203*, 2013.

[12] P. Chathuranga Weeraddana, G. Athanasiou, M. Jakobsson, C. Fischione, and J. S. Baras, "On the Privacy of Optimization Approaches," *ArXiv e-prints*.

[13] M. J. Wainwright, M. I. Jordan, and J. C. Duchi, "Privacy aware learning," in *Advances in Neural Information Processing Systems*, 2012, pp. 1430–1438.

[14] Z. Ma, D. Callaway, and I. Hiskens, "Decentralized charging control for large populations of plug-in electric vehicles: Application of the nash certainty equivalence principle," in *Control Applications (CCA), 2010 IEEE International Conference on*, Sept 2010, pp. 191–195.

[15] L. Gan, U. Topcu, and S. Low, "Optimal decentralized protocol for electric vehicle charging," *Power Systems, IEEE Transactions on*, vol. 28, no. 2, pp. 940–951, May 2013.

[16] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on cyber security for smart grid communications," *Communications Surveys Tutorials, IEEE*, vol. 14, no. 4, pp. 998–1010, Fourth 2012.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*, ser. Berichte über verteilte messysteme. Cambridge University Press, 2004.

[18] R. Rockafellar, *Convex Analysis*, ser. Convex Analysis. Princeton University Press, 1997. [Online]. Available: http://books.google.com/books?id=1TiOka9bx3sC

[19] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, Sept. 1999. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/1886529000

[20] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1574–1609, 2009.

[21] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[22] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proceedings of the Third Conference on Theory of Cryptography*, ser. TCC'06, 2006, pp. 265–284.

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

[24] "Plug-in hybrid electric vehicle charge pattern optimization for energy cost and battery longevity," *Journal of Power Sources*, vol. 196, no. 1, pp. 541 – 549, 2011.

[25] "On the aggregate grid load imposed by battery health-conscious charging of plug-in hybrid electric vehicles," *Journal of Power Sources*, vol. 196, no. 20, pp. 8747 – 8754, 2011.