Distributed Convex Stochastic Optimization under Few Constraints in Large Networks

Romain Couillet[1], Pascal Bianchi[2], and Jérémie Jakubowicz[2]
[1] *Centrale-Supélec-EDF Chair on System Sciences and the Energy Challenge, Gif sur Yvette, France.*
[2] *Institut Télécom - Télécom ParisTech - CNRS/LTCI, France.*

***Abstract*—This article introduces a distributed convex optimization algorithm in a constrained multi-agent system composed by a large number of nodes. We focus on the case where each agent seeks to optimize its own local parameter under few coupling equality and inequality constraints. The objective function is of the power flow type and can be decoupled as a sum of elementary functions, each of which assumed (imperfectly) known by only one node. Under these assumptions, a cost-efficient decentralized iterative solution based on Lagrangian duality is derived, which is provably converging. This new approach alleviates several limitations of algorithms proposed in the stochastic optimization literature. Applications are proposed to decentralized power flow optimization in smart grids.**

I. Introduction

In many fields of network engineering, networks tend to grow extremely large and complex. As a telling example, we will focus in this article on smart grid networks, our interest being on power flow optimization. Such networks integrate millions of energy generators and consumers (wind turbines, electrical vehicles etc.) connected to a common electricity grid. To optimize operability and ensure stability, it is often assumed that a central entity controls the system as a whole, which supposes that this entity is knowledgeable about all system parameters and is capable of solving large dimensional problems, sometimes in real-time. This clearly comes along with several issues: computational burden, heavy information feedback from the network, as well as security issues.

These problems led research in the direction of decentralized algorithms, initiated by Tsitsiklis [2] on parallel computing, to distribute the computational effort into multiple communicating processors. In the framework of multi-agent systems, a central issue is to determine distributed algorithm allowing to search for minimizers of a global objective function, equal to the sum of local utility functions of the agents. The network communication structure is usually based on incremental [9] or gossip-based [8] data exchanges. Notice however that in the case where the network has to cope with possible link failures, more robust approaches should be used [11], [5]. In the incremental approach, a message travels across the network, updating at each time instant the status of a *single* receiver. In this paper, we focus on gossip-based approach such as [8]: agents iteratively update their own local estimates of the minimizer, and simultaneously merge their estimates in order to eventually reach a consensus on the value of the minimizer.

This work has been partially funded by ANR program C-FLAM.

Consider a network composed of N agents, and assume that each agent $i = 1, \dots, N$ seeks to optimize a local scalar value x_i. We consider the following optimization problem:

$$\min_{(x_1,\dots,x_N)\in\mathcal{D}} \sum_{i=1}^{N} f_i(x_i)\,, \tag{1}$$

where f_i is a real convex function and $\mathcal{D}$ is a convex domain of $\mathbb{R}^N$ determined by a set of equality and inequality constraints. The set $\mathcal{D}$ is assumed perfectly known by all agents. Each agent $i = 1, \dots, N$ ignores the utility functions f_j of other agents $j \neq i$. In addition, agent i observes its utility function f_i only up to a possible random perturbation[1]: we thus investigate stochastic approximation methods.

The above problem could be naively solved using a traditional "primal" gossip-based optimization approach: This method would require that each agent has its own estimate of a global minimizer $(x_1^*, \dots, x_N^*)$ in $\mathbb{R}^N$, and that connected agents exchange their estimates in order to achieve a consensus (see for instance Section V in [3] for an illustration). However, in terms of both communication burden and convergence speed, it is often unreasonable to assume that any agent $i = 1, \dots, N$ maintains an estimate of the parameters x_j of *all* other agents $j \neq i$, especially if N is large. In order to circumvent this problem, we consider a system in which only few (equality and inequality) constraints relate the decision variables. In the power flow application, equality constraints represent demand-response energy satisfaction while inequality constraints are used to model some critical power line outage constraints. Using convex duality, we solve the joint optimization problem from its dual, the data exchanged now being Lagrangian multipliers. This follows closely the ideas from [6]. However, in [6], the proposed data exchange method is not gossip-based, the considered gradient updating method is deterministic, and only equality constraints are considered. The tools required for these various generalizations are based on the recent results from [3] on constrained decentralized optimization methods.

Gossip-based schemes use matrices $\mathbf{W}_n = ([\mathbf{W}_n]_{ij}) \in \mathbb{C}^{N\times N}$ to weight networks exchanges, with $[\mathbf{W}_n]_{ij}$ measuring the importance given by node i to the data received from node j at time n. In [8], matrices $\mathbf{W}_n$ are assumed doubly-stochastic. However, double stochasticity is often difficult to ensure in practice, because it implicitely requires additional

[1] In the smart grid context, the production cost at a renewable source may be imperfectly known due to the real-time evolving environment.

communication overhead in the network. For instance, it prevents from using some natural message passing schemes such as the broadcast method, which consists for a given node to asynchronously spread its data towards its neighbors. Instead, [7], [3] weaken the double-stochasticity assumption and analyze the broadcast method.

The remainder of this article is organized as follows: in Section II, we introduce the system model under study; in Section III, we introduce the decentralized optimization method; in Section IV, we provide application and simulation results; finally, in Section V, we conclude this article.

Notations: Uppercase boldface characters denote matrices. Lowercase boldface characters denote (column) vectors, with $\mathbf{1}$ the vector filled with ones. The notation $(\cdot)^\mathsf{T}$ stands for vector transpose. We note $[x]^+ = \max(x, 0)$. The notation $\Pi_G[\mathbf{x}]$ for $\mathbf{x} \in \mathbb{C}^N$ and $G \subset \mathbb{C}^N$ is the projection of $\mathbf{x}$ onto the convex G for the classical Euclidean norm $\|\mathbf{x}\|$. We denote $\otimes$ the Kronecker matrix product. The symbol ∇ denotes the gradient operator. The symbol $\mathrm{E}[\cdot]$ denotes expectation. The function $\delta(A)$ is the Kronecker delta function.

II. General convex optimization model

We start by considering the following optimal DC power flow example, motivated by alternative decentralized power flow approaches such as [1], [4]. The overall system is composed of N energy sources connected by a given network. Source k produces a quantity $0 \leq x_k \leq x_{k,\max}$ of energy at a cost per unit $f_k(x_k)$ for a given cost function f_k defined on $[0, x_{k,\max}]$ and assumed strictly convex, and $x_{k,\max}$ source k-production capacity. The total energy demand is denoted D and must satisfy $\sum_{k=1}^N x_k = D$. We denote $\mathbf{x} = (x_1, \ldots, x_N)^\mathsf{T} \in \mathbb{R}_+^N$ the vector of all powers produced and $\mathbf{x}_{\max} = (x_{1,\max}, \ldots, x_{N,\max})^\mathsf{T} \in \mathbb{R}_+^N$ the plant capacities. Furthermore, we assume a scenario where local power line constraints are inactive, i.e. the maximum power flowing in a given line does not reach its allowable maximum, but we must ensure that some inter-regional critical energy transfers are achieved. This is a usual assumption which already allows for a reduction of complexity of centralized power flow calculus [10]. The number K of power line constraints is therefore assumed small compared to N. These constraints are expressed as a (point-wise) vector inequality $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ for a given matrix $\mathbf{A} \in \mathbb{C}^{K \times N}$ and a vector $\mathbf{b} \in \mathbb{C}^K$. This problem enters the general framework of the following optimization problem, which we will refer to as the *primal problem*:

$$\min_{\substack{\mathbf{x} \in I \\ g_j(\mathbf{x}) = 0,\ j \in \{1,\ldots,J\} \\ h_k(\mathbf{x}) \leq 0,\ k \in \{1,\ldots,K\}}} \sum_{i=1}^N f_i(x_i) \tag{2}$$

for which we take the following assumptions:

A1 1) $I = I_1 \times \cdots \times I_N \subset \mathbb{R}^N$ is a compact rectangle.
2) $(f_i)_i$ are strictly convex and continuously differentiable.
3) For any $\mathbf{x}$, $g_j(\mathbf{x}) = \sum_{i=1}^N g_{j,i}(x_i)$ for *affine* functions $g_{j,i}$
4) For any $\mathbf{x}$, $h_k(\mathbf{x}) = \sum_{i=1}^N h_{k,i}(x_i)$ for some convex continuously differentiable functions $h_{k,i}$.
5) The following Weak Slater condition is satisfied:

$$\exists \mathbf{x}_0 \in I, \quad \begin{array}{l} \forall j \in \{1,\ldots,J\}, g_j(\mathbf{x}_0) < 0, \\ \forall k \in \{1,\ldots,K\}, h_k(\mathbf{x}_0) = 0\ . \end{array}$$

We first proceed to a Lagrangian relaxation of (2). Denote $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)^\mathsf{T} \in \mathbb{R}^J$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_K)^\mathsf{T} \in \mathbb{R}_+^K$ and

$$\mathcal{L}_i(x_i; \boldsymbol{\lambda}, \boldsymbol{\nu}) \triangleq f_i(x_i) + \sum_{j=1}^J \lambda_j g_{j,i}(x_i) + \sum_{k=1}^K \nu_k h_{k,i}(x_i).$$

The Lagrangian of (2) writes $\mathcal{L}(\mathbf{x}; \boldsymbol{\lambda}, \boldsymbol{\nu}) \triangleq \sum_i \mathcal{L}_i(x_i; \boldsymbol{\lambda}, \boldsymbol{\nu})$. Also define

$$F_i(\boldsymbol{\lambda}, \boldsymbol{\nu}) \triangleq \inf_{x_i \in I_i} \mathcal{L}_i(x_i; \boldsymbol{\lambda}, \boldsymbol{\nu}),$$

$x_i^\star(\boldsymbol{\lambda}, \boldsymbol{\nu})$ the argument of the minimum, and $F = \sum_i F_i$. From standard convex analysis, Assumption **A1** implies *strong duality*. Hence, using separability, denoting by $\mathbf{x}^\star$ the unique solution of (2), and $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = \arg\max_{\boldsymbol{\lambda} \in \mathbb{R}^J, \boldsymbol{\nu} \in \mathbb{R}_+^K} F(\boldsymbol{\lambda}, \boldsymbol{\nu})$, one has:

$$\sum_{i=1}^N f_i(x_i^\star) = \sum_{i=1}^N F_i(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star).$$

Example 1: (*Convex optimization under quadratic costs and affine constraints*). Consider the scenario where $f_i(x) = f_{a_i} x^2 + f_{b_i} x$, $g_{j,i}(x) = g_{a_{j,i}} x + g_{b_{j,i}}$, and $h_{k,i}(x) = h_{a_{k,i}} x + h_{b_{k,i}}$; for $f_{a_i} > 0$, and $I_i = [0, x_{\max}]$ for some large $x_{\max}$. Then, after simple algebraic manipulations, we obtain

$$x^{i,\star}(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \left[-\frac{f_{b_i} + \sum_{j=1}^J \lambda_j g_{a_{j,i}} + \sum_{k=1}^K \nu_k h_{a_{k,i}}}{2 f_{a_i}} \right]^+$$

if the right-hand side is smaller than $x_{\max}$ and $x^{i,\star}(\boldsymbol{\lambda}, \boldsymbol{\nu}) = x_{\max}$ otherwise. We also have

$$F_i(\boldsymbol{\lambda}, \boldsymbol{\nu}) = -f_{a_i} x^{i,\star}(\boldsymbol{\lambda}, \boldsymbol{\nu})^2 + \sum_{j=1}^J \lambda_j g_{b_{j,i}} + \sum_{k=1}^K \nu_k h_{b_{k,i}}.$$

The objective of this article is to propose a decentralized programming method to solve the dual convex optimization

$$\max_{\substack{\boldsymbol{\lambda} \in \mathbb{R}^J \\ \boldsymbol{\nu} \in \mathbb{R}_+^K}} \sum_{i=1}^N F_i(\boldsymbol{\lambda}, \boldsymbol{\nu}) \tag{3}$$

based on recent techniques from stochastic optimization and gossiping methods. This is the target of the subsequent section.

III. Decentralized stochastic optimization

From the decoupled formulation of the dual problem (3) and the fact that $J + K \ll N$, it is reasonable to consider the following decentralized iterative optimization method. We suppose that each agent $i \in \{1, \ldots, N\}$ in the system only knows the function F_i, possibly up to some noise. We now assume that at time n, agent i possesses an estimate $(\boldsymbol{\lambda}^{i,n}, \boldsymbol{\nu}^{i,n})$ of $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$, from which an estimate $x^{i,\star}(\boldsymbol{\lambda}^{i,n}, \boldsymbol{\nu}^{i,n})$ of $x_i^\star$ can be evaluated. First, from its (possibly noisy) knowledge of F_i, agent i may improve the received $(\boldsymbol{\lambda}^{i,n}, \boldsymbol{\nu}^{i,n})$ into $(\tilde{\boldsymbol{\lambda}}^{i,n+1}, \tilde{\boldsymbol{\nu}}^{i,n+1})$, using a gradient ascent. In a second step, agent i may then receive updated information from M neighboring agents $j_1, \ldots, j_M$, so to further update his evaluation as a weighted average $(\boldsymbol{\lambda}^{i,n+1}, \boldsymbol{\nu}^{i,n+1})$ of $(\tilde{\boldsymbol{\lambda}}^{i,n+1}, \tilde{\boldsymbol{\nu}}^{i,n+1})$ and $(\tilde{\boldsymbol{\lambda}}^{j_1,n+1}, \tilde{\boldsymbol{\nu}}^{j_1,n+1}), \ldots, (\tilde{\boldsymbol{\lambda}}^{j_M,n+1}, \tilde{\boldsymbol{\nu}}^{j_M,n+1})$. If performed adequately, such algorithms can be shown to converge to $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ as $n \to \infty$ for each i.

The idea behind this article is to use the recent results of [3] to provide a practical algorithm for decentralized resolution of (2) under few constraints on the information exchanges between neighboring agents, and under the realistic assumption that agent i may not know perfectly its own cost function $f_i(x)$. In the power flow context, this statement translates the fact that production costs of renewable energy systems are environment-dependent and may be difficult to appreciate at any time instant. Define the following compact set of $\mathbb{R}^{J+K}$:

$$G = \{(\boldsymbol{\lambda}, \boldsymbol{\nu}) : \forall (k,j),\ |\lambda_j| \leq \lambda_{\max},\ , 0 \leq \nu_k \leq \nu_{\max}\}$$

where $(\lambda_{\max}, \nu_{\max})$ are arbitrarily large constants. For each i, consider the series $\boldsymbol{\lambda}^{i,0}, \boldsymbol{\lambda}^{i,1}, \ldots \in \mathbb{R}^J$, $\boldsymbol{\nu}^{i,0}, \boldsymbol{\nu}^{i,1}, \ldots \in \mathbb{R}^K_+$, and denote $\boldsymbol{\theta}^{i,n} = (\boldsymbol{\lambda}^{i,n\mathsf{T}}, \boldsymbol{\nu}^{i,n\mathsf{T}})^\mathsf{T}$. We assume $\boldsymbol{\lambda}^{i,0}, \boldsymbol{\nu}^{i,0}$ given and, for $n \geq 1$, define the series:

$$\tilde{\boldsymbol{\theta}}^{i,n+1} = \Pi_G \left[\boldsymbol{\theta}^{i,n} + \gamma_{n+1} \mathbf{y}_{i,n+1} \right] \tag{4}$$

$$\boldsymbol{\theta}^{i,n+1} = \sum_{j=1}^{N} [\mathbf{W}_{n+1}]_{ij} \tilde{\boldsymbol{\theta}}^{j,n+1} \tag{5}$$

where $\gamma_1, \gamma_2, \ldots$ are positive, $\mathbf{y}_{i,1}, \mathbf{y}_{i,2}, \ldots \in \mathbb{C}^{J+K}$ are random vectors and $\mathbf{W}_1, \mathbf{W}_2, \ldots \in \mathbb{R}^{N \times N}$ form a series of random matrices. Random variable $\mathbf{y}_{i,n+1}$ should be interpreted as a noisy version of $\nabla F_i(\boldsymbol{\theta}^{i,n})$ as made clear by Assumption **A3** below. This means that (4) can be viewed as a local stochastic gradient step, locally performed by each agent. On the opposite, (5) represents the gossip step where connected nodes merge their temporary estimates. In the sequel, we note $\boldsymbol{\theta}^n = ([\boldsymbol{\theta}^{1,n}]^\mathsf{T}, \ldots, [\boldsymbol{\theta}^{N,n}]^\mathsf{T})^\mathsf{T}$ and $\mathbf{y}_n = (\mathbf{y}_{1,n}^\mathsf{T}, \ldots, \mathbf{y}_{N,n}^\mathsf{T})^\mathsf{T}$. Denote by $\mathbf{1}$ the $N \times 1$ vector whose elements are all equal to one. We take the following assumptions on matrices $(\mathbf{W}_n)_n$:

A2 1) $(\mathbf{W}_n)_n$ is a sequence of iid matrices.
2) $\mathbf{W}_n \mathbf{1} = \mathbf{1}$ and $\mathbf{1}^\mathsf{T} \mathrm{E}[\mathbf{W}_n] = \mathbf{1}^\mathsf{T}$.
3) the spectral radius of $\mathrm{E}[\mathbf{W}_n^\mathsf{T}(\mathbf{I}_N - \frac{1}{N}\mathbf{1}\mathbf{1}^\mathsf{T})\mathbf{W}_n]$ is < 1.

Assumption **A2**-2) is extremely useful and one of the main applicative interests of this article, as it loosens the classical assumption $\mathbf{1}^\mathsf{T}\mathbf{W}_n = \mathbf{1}^\mathsf{T}$ found e.g. in [8]. With double-stochasticity only in expectation, it is possible to consider broadcast schemes, where nodes send data to arbitrarily many neighbors, e.g. $[\mathbf{W}_n]_{i,j} = 1/2$, $[\mathbf{W}_n]_{j,j} = 1/2$ for all neighbors j of a randomly selected node i and $[\mathbf{W}_n]_{kl} = 0$ otherwise, which is an asymmetrical updating method corresponding to node i sending its data to all its neighbors node, but not being fed back by updated data from its neighbors. We now make some assumption on the probabilistic model.

A3 Denote $\mathcal{F}_n$ the σ-field generated by $(\theta_0, \mathbf{y}_1, \ldots, \mathbf{y}_n, \mathbf{W}_1, \ldots, \mathbf{W}_n)$ and assume that there exists a measure $\mu_{\boldsymbol{\theta}_n}$ indexed by $\boldsymbol{\theta}_n$ s.t. $\mathbb{P}(\mathbf{y}_{n+1} \in A | \mathcal{F}_n) = \mu_{\boldsymbol{\theta}_n}(A)$ for any Borel set A. Then:

1) $\mathrm{E}_{\boldsymbol{\theta}_n}[\mathbf{y}] = (\nabla F_1(\boldsymbol{\theta}^{1,n})^\mathsf{T}, \ldots, \nabla F_N(\boldsymbol{\theta}^{N,n})^\mathsf{T})^\mathsf{T}$
2) $\sup_{\boldsymbol{\theta} \in G^N} \mathrm{E}_{\boldsymbol{\theta}}[\mathbf{y}^\mathsf{T}\mathbf{y}] < \infty$
3) $(\mu_{\boldsymbol{\theta}})_{\boldsymbol{\theta} \in G^N}$ is tight.

Of importance are **A3**-1) and **A3**-2) which state that the updating scheme follows a gradient ascent up to a zero mean random noise with bounded variations. The noise term can be understood as an estimation error of the gradient. In practice, this suggests that agent i only knows approximately ∇F_i. Finally, we need some classical assumptions regarding the sequence $\gamma_1, \gamma_2, \ldots$:

A4 The following holds: $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$.

In [3], it is proved that, under the above conditions, the iterations (4)-(5) converge to a point in the set of KKT points of the function F. Since F is strictly concave, $\boldsymbol{\theta}^{i,n}$ converges to the global maximizer of $\boldsymbol{\theta}^\star$ of F. Define $\boldsymbol{\theta}^\star = ((\boldsymbol{\lambda}^\star)^\mathsf{T}, (\boldsymbol{\nu}^\star)^\mathsf{T})^\mathsf{T}$.

Theorem 1 (Theorem 1 of [3]): Assume **A2**-**A4** and that $\boldsymbol{\theta}^\star \in G$. For every $i \in \{1, \ldots, N\}$, the following holds w.p.1.:

$$\|\boldsymbol{\theta}^{i,n} - \boldsymbol{\theta}^\star\| \to 0\ .$$

Theorem 1 states precisely that, not only will the agents reach a consensus on their values of $\boldsymbol{\theta}^{i,n}$ as n grows large, but also the achieved value is optimal for the dual problem (3). A necessary condition is of course that constants $\lambda_{\max}$ and $\nu_{\max}$ are chosen large enough s.t. $\boldsymbol{\theta}^\star \in G$. As a consequence, from the fact that primal and dual problems have zero duality gap, we obtain the following corollary. Denote $x^{i,n} = x^{i,\star}(\boldsymbol{\lambda}^{i,n}, \boldsymbol{\nu}^{i,n})$.

Corollary 1: Under the stated assumptions, $x^{i,n} - x_i^\star \to 0$ for each $i \in \{1, \ldots, N\}$, almost surely, as $n \to \infty$.

In the following section, we provide simulation examples in a particular applicative case, which is made simple for the sake of understanding and simple reproducibility.

IV. Application

We consider a network of $N = 20$ power production units, with total production constrained to equal D energy units. We take $D = 1$. The network is divided in two areas of $N/2$ units each, with area 1 being constrained not to produce more than $1/4$ energy unit in total, and therefore with area 2 constrained to produce at least $3/4$ energy units in total. This situation may arise when the total energy demand in area 1 equals $1/4 - \varepsilon$, for some $0 < \varepsilon < 1/4$, and that the capacity for carrying power from area 1 to area 2 equals ε. On the opposite, area 2 can transport as much power as required into area 1. Denoting x_i the production of unit i, we consider the optimization problem of Example 1 where $g_{a_{j,i}} = 1$, $g_{b_{j,i}} = -D/N$, for all (i,j), and $h_{a_{k,i}} = \delta(k \leq N/2)$, $h_{b_{k,i}} = -D/(4N)$ for all (i,k). The functions f_i are determined by choosing uniform random realizations for $f_{a_i} \in [0,1]$ and $f_{b_i} \in [-1/2, 1/2]$. As such, there is one equality and one inequality constraints. Only two sets of Lagrangian multipliers $\lambda^{i,n} \in \mathbb{R}$ and $\nu^{i,n} > 0$, $i \in \{1, \ldots, N\}$ and $n \in \mathbb{N}$ need then be exchanged. We will denote $\boldsymbol{\theta}^{i,n} = (\lambda^{i,n}, \nu^{i,n})^\mathsf{T}$.

We first simulate the iterative sequence $\boldsymbol{\theta}^n$ for $50\,000$ iterations with $\boldsymbol{\theta}^{i,0} = \mathbf{0}$ and $\gamma_n = 0.5n^{-0.8}$ for the updating phase. For the gossiping phase, we consider two strategies. The first strategy, referred to as the *broadcast* scheme, consists at time n in selecting uniformly a random unit $i \in \{1, \ldots, N\}$ and in communicating the value $\tilde{\boldsymbol{\theta}}^{i,n+1}$ to units $i-1$ and $i+1$. Upon reception, unit $j \in \{i-1, i+1\}$ evaluates $\boldsymbol{\theta}^{j,n+1} = \frac{1}{2}(\tilde{\boldsymbol{\theta}}^{i,n+1} + \tilde{\boldsymbol{\theta}}^{j,n+1})$, while other units $k \notin \{i-1, i+1\}$ take $\boldsymbol{\theta}^{k,n+1} = \tilde{\boldsymbol{\theta}}^{k,n+1}$. The second strategy, the *joint averaging* strategy consists for the three units $i-1, \ldots, i+1$ in a

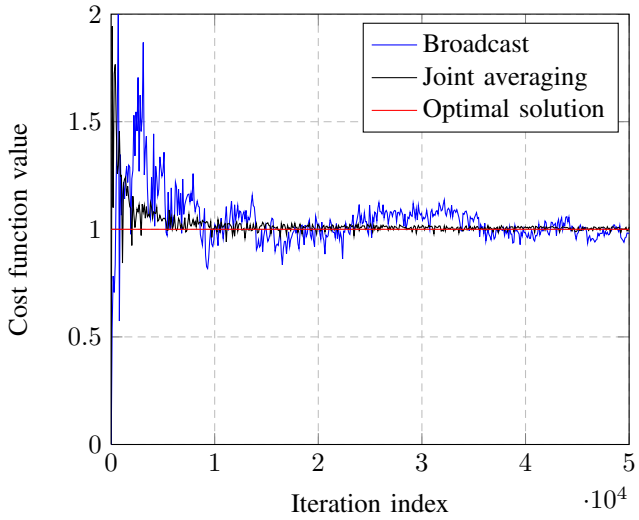

Fig. 1. Normalize cost function against optimal, for the broadcast and broadcast-feedback schemes, $N = 20$, gossip with 4 closest neighbors.

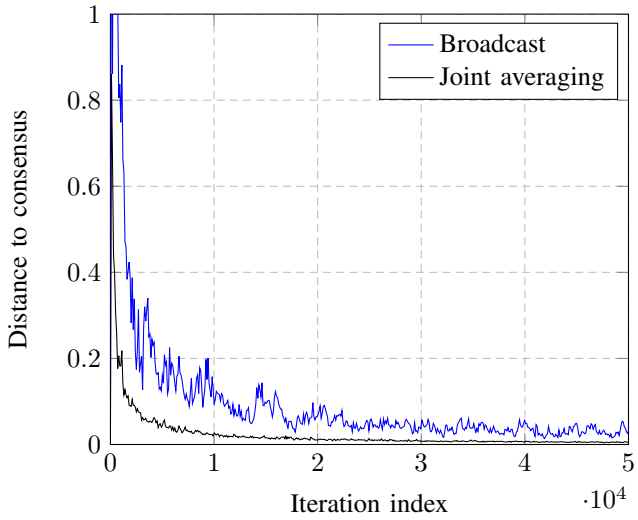

Fig. 2. Distance to consensus, for the broadcast and broadcast-feedback schemes, $N = 20$, gossip with 4 closest neighbors.

simple averaging of their transmitted values. That is, for each $j \in \{i-1, i, i+1\}$, $\boldsymbol{\theta}^{j,n+1} = \frac{1}{3}\sum_{k=i-1}^{i+1} \tilde{\boldsymbol{\theta}}^{k,n+1}$. It is clear that **A2** is met under these conditions: for the joint averaging scheme, the associated transfer matrix $\mathbf{W}_n$ is even doubly stochastic, i.e. $\mathbf{1}^\mathsf{T}\mathbf{W}_n = \mathbf{1}^\mathsf{T}$.

In Figure 1, we depict the evolution of the cost $\sum_{i=1}^N f_i(x^{i,n})$ normalized by the optimal cost $\sum_{i=1}^N f_i(x_i^\star)$. We observe that, although both schemes converge to the optimal cost, the joint averaging scheme converges much faster. In Figure 2, we depict the distance to consensus in the variables $\boldsymbol{\theta}^{i,n}$, which corresponds to the evolution of $\sum_i \|\boldsymbol{\theta}^{i,n} - \frac{1}{N}\mathbf{1}^\mathsf{T}\boldsymbol{\theta}^{i,n}\mathbf{1}\|$. Here also, the distance to consensus decreases faster to zero for the joint averaging scheme. We then consider in Figure 3 a scenario where a randomly selected value (f_{a_i}, f_{b_i}) changes every 12 500 iterations. Upon cost changes, we run a new instance of the decentralized algorithm. We reinitialize $\gamma_n = 0.5n^{-0.8}$ to 1 at each change. It is observed that the iterative algorithm follows accurately the changes in the optimal $\mathbf{x}^\star$ solution. This way, we simulate adequately a tracking version of the iterative algorithm.

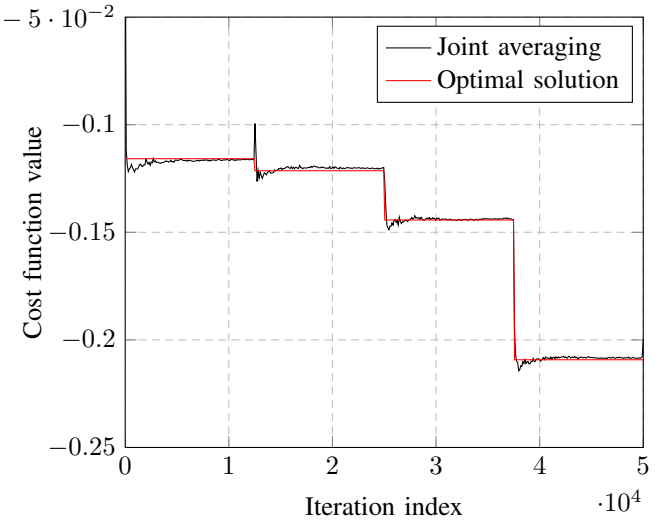

Fig. 3. Cost function against optimal, with four state changes, for the joint averaging scheme, $N = 20$, gossip with 4 closest neighbors.

V. Conclusion

In this article, we derived an iterative decentralized scheme to solve a certain type of convex optimization problems in large networks under few coupling equality and inequality constraints. This method, which solves the dual problem of a convex optimization, relies on stochastic optimization and gossiping, and is proved to converge asymptotically to the desired solution. This scheme is in particular applicable to power flow optimization for smart grid networks.

References

[1] A. G. Bakirtzis and P. N. Biskas, "A decentralized solution to the dc-opf of interconnected power systems," *Power Systems, IEEE Transactions on*, vol. 18, no. 3, pp. 1007–1013, 2003.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation*. Old Tappan, NJ (USA): Prentice Hall Inc., 1989.

[3] P. Bianchi and J. Jakubowicz, "On the convergence of a multi-agent projected stochastic gradient algorithm," *IEEE Trans. on Autom. Control*, 2011, submitted, preprint arXiv:1107.2526.

[4] G. Hug-Glanzmann and G. Andersson, "Decentralized optimal power flow control for overlapping areas in power systems," *Power Systems, IEEE Transactions on*, vol. 24, no. 1, pp. 327–336, 2009.

[5] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1383–1400, 2010.

[6] D. Mosk-Aoyama, T. Roughgarden, and D. Shah, "Fully distributed algorithms for convex optimization problems," *Lecture Notes in Computer Science*, 2007.

[7] A. Nedic, "Asynchronous Broadcast-Based Convex Optimization Over a Network," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1337–1351, june 2011.

[8] S. S. Ram, A. Nedic, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of optimization theory and applications*, vol. 147, no. 3, pp. 516–545, 2010.

[9] S. Ram, A. Nedic, and V. Veeravalli, "Incremental Stochastic Subgradient Algorithms for Convex Optimization," *SIAM Journal on Optimization*, vol. 20, no. 2, pp. 691–717, 2009.

[10] A. B. Rodrigues and M. G. D. Silva, "Probabilistic assessment of available transfer capability based on monte carlo method with sequential simulation," *Power Systems, IEEE Transactions on*, vol. 22, no. 1, pp. 484–492, 2007.

[11] I. Schizas, A. Ribeiro, and G. Giannakis, "Consensus in ad hoc WSNs with noisy links-Part I: Distributed estimation of deterministic signals," *IEEE Trans. on Signal Processing*, vol. 56, no. 1, pp. 350–364, 2008.