# Approximate Projections for Decentralized Optimization with SDP Constraints

Soomin Lee* and Michael M. Zavlanos†

*Abstract*— **We consider distributed convex optimization problems that involve a separable objective function and nontrivial convex local constraints, such as Linear Matrix Inequalities (LMIs). We propose a decentralized, computationally inexpensive algorithm to solve such problems over time-varying directed networks of agents, that is based on the concept of approximate projections. Our algorithm is one of the consensus based methods in that, at every iteration, every agent performs a consensus update of its decision variables followed by an optimization step of its local objective function and local constraint. Unlike other methods, the last step of our method is not a projection to the feasible set, but instead a subgradient step in the direction that minimizes the local constraint violation. We show that the algorithm converges almost surely, i.e., every agent agrees on the same optimal solution, when the objective functions and constraint functions are nondifferentiable and their subgradients are bounded.**

## I. Introduction

Decentralized optimization has been extensively studied in recent years due to a variety of applications in machine learning, signal processing, and control for robotic networks, sensor networks, power networks, and wireless communication networks [1]–[4]. A number of problems arising in these areas can be cast as distributed convex optimization problems over multi-agent networks, where individual agents cooperatively try to minimize a common cost function over a common constraint set in the absence of full knowledge about the global problem structure. The main feature of carrying these optimizations over networks is that the agents can only communicate with their neighboring agents. This communication structure can be cast as a graph, often directed and/or time-varying.

The literature on distributed optimization methods is vast and involves first-order methods in the primal domain, the dual domain, augmented Lagrangian methods, or Newton methods, to name a few. In this paper, we discuss methods that are more closely related to the method under consideration. Among those, one of the most well-studied techniques are the so called *consensus-based* optimization algorithms [5]–[13], where the goal is to repeatedly average the estimates of all agents in a decentralized fashion in order to obtain a network-wide consensus. Between the averaging steps, each agent usually performs a single local optimization step and the goal is to use the local information to cooperatively steer the consensus point toward the optimal set of the global problem.

Decentralized algorithms that fall in this class of methods can be distinguished based on which averaging scheme or optimization method is used, and in which space (primal or dual) the iterates are maintained. All these algorithms often require expensive optimization steps or projections on a complicated constraint set at every iteration. Such intensive computations, however, shorten the lifespan of mobile network systems, such as wireless sensor networks or robotic networks.

In this work, we describe and prove convergence of a new decentralized algorithm with low computational complexity. Our work in this paper is an extension of the author's previous work [14]. Specifically, we use the same local information exchange model and gradient descent algorithm as in [14], but a different projection method motivated by the work in [15]. In contrast to [14], our contribution can be summarized as follows: (1) Instead of using a projection, we approximate the projection by measuring the constraint violation and taking a subgradient step minimizing this violation; (2) We show convergence under milder assumptions. Specifically, we remove the smoothness assumption in the objective functions.

Considering that projections have a closed form solution only in a few special cases of constraints, our new algorithm is more general and can be applied to a wider class of problems including Semidefinite Programming (SDP), where the constraints are represented by Linear Matrix Inequalities (LMIs). It is well known that even finding a feasible point that satisfies a handful of LMIs is a difficult problem on its own. Also, the computational overhead of this algorithm in each iteration is extremely cheap as the additional subgradient step requires only $O(n)$ calculations ($n$ is the number of decision variables). Therefore, our algorithm can also handle SDPs in a decentralized fashion with an extremely cheap per-iteration computational overhead.

The work in the paper is also related to the centralized random projection algorithms for convex constrained optimization [16] and convex feasibility problems [17]. Other related work is [18]–[20], where optimization problems with uncertain constraints have been considered by finding probabilistic feasible solutions through random sampling of constraints.

**Notation:** All vectors are viewed as column vectors. We write $x^T$ to denote the transpose of a vector $x$. The scalar product of two vectors $x$ and $y$ is $\langle x, y\rangle$. For vectors

This work is supported by ONR under grant #N000141410479.

*Soomin Lee is with the Dept. of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332. This work was conducted when she was at Duke University. email: soomin.lee@isye.gatech.edu.

†Michael M. Zavlanos are with the Dept. of Mechanical Engineering and Materials Science, Duke University, Durham, NC, 27708, USA, e-mail: michael.zavlanos@duke.edu.

associated with agent $i$ at time $k$, we use subscripts $i,k$ such as, for example, $p_{i,k}$, $x_{i,k}$, etc. Unless otherwise stated, $\|\cdot\|$ represents the standard Euclidean norm. For a matrix $A \in \mathbb{R}^{n\times n}$, we use $[A]_{ij}$ to denote the entry of the $i$-th row and $j$-th column and $\|A\|_F$ to denote the Frobenius norm $\|A\|_F = \left(\sum_{i,j=1}^n ([A]_{ij})^2\right)^{1/2}$. We use $\mathsf{Tr}A$ to denote the trace of $A$, i.e., $\mathsf{Tr}A = \sum_{i=1}^n [A]_{ii}$. We denote $\mathbb{S}_m$ and $\mathbb{S}_m^+$ the space of $m\times m$ real symmetric and real symmetric positive semidefinite matrices, respectively. The matrix inequality $A \preceq 0$ means $-A$ is positive semidefinite. We use $\mathbf{1}$ to denote a vector with all entries equal to 1. The identity matrix is denoted by $I$. We use $\mathsf{Pr}\{Z\}$ and $\mathsf{E}[Z]$ to denote the probability and the expectation of a random variable $Z$. We write $\operatorname{dist}(x,\mathcal{X})$ for the distance of a vector $x$ from a closed convex set $\mathcal{X}$, i.e., $\operatorname{dist}(x,\mathcal{X}) = \min_{v\in\mathcal{X}} \|v - x\|$. We use $\Pi_{\mathcal{X}}[x]$ for the projection of a vector $x$ on the set $\mathcal{X}$, i.e., $\Pi_{\mathcal{X}}[x] = \arg\min_{v\in\mathcal{X}} \|v - x\|^2$. We often abbreviate *almost surely* and *independent identically distributed* as *a.s.* and *i.i.d.*, respectively.

## II. Problem Definition

Consider a multiagent network system whose communication at time $k$ is governed by a digraph $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$, where $\mathcal{V} = \{1,\ldots,N\}$ and $\mathcal{E}_k \subseteq \mathcal{V}\times\mathcal{V}$. If there exists a directed link from agent $j$ to $i$ which we denote by $(j,i)$, agent $j$ may send its information to agent $i$. Thus, each agent $i \in \mathcal{V}$ can directly receive information only from the agents in its neighborhood

$$\mathcal{N}_{i,k} = \{j \in \mathcal{V} \mid (j,i) \in \mathcal{E}_k\} \cup \{i\}, \tag{1}$$

where we assume that there always exists a self-loop $(i,i)$ for all agents $i \in \mathcal{V}$.

### A. Problem Statement

Our goal is to let the network of agents cooperatively solve the following minimization problem:

$$\begin{aligned} \min_x \;\; & f(x) \triangleq \sum_{i\in\mathcal{V}} f_i(x) \\ \text{s.t. } & x \in \mathcal{X}, \quad \mathcal{X} \triangleq \mathcal{X}_0 \bigcap \left(\textstyle\bigcap_{i\in\mathcal{V}} \mathcal{X}_i\right), \end{aligned} \tag{2}$$

where only agent $i$ knows the function $f_i : \mathbb{R}^n \to \mathbb{R}$ and the constraint set $\mathcal{X}_i \subseteq \mathbb{R}^n$. The set $\mathcal{X}_0 \subseteq \mathbb{R}^n$ is common to all agents and assumed to have some simple structure in the sense that the projection onto $\mathcal{X}_0$ can be made easily (e.g., a box, ball, probability simplex, or even $\mathbb{R}^n$). Note that the local constraints may overlap, i.e., $\mathcal{X}_i \cap \mathcal{X}_j$ for $i \neq j$ may be nonempty. Also, note that the common constraint set, i.e., $\mathcal{X}_0 = \mathcal{X}_i$ for all $i \in \mathcal{V}$, is a special case of this problem definition. We assume that the set of optimal solutions $\mathcal{X}^* = \arg\min_{x\in\mathcal{X}} f(x)$ is nonempty.

### B. Problems of Interest

Each local constraint set $\mathcal{X}_i$ consists of one or more inequalities, which we denote by

$$\mathcal{X}_i = \{x \in \mathbb{R}^n \mid g(x,\omega) \le 0, \text{ for } \omega \in \Omega_i\},$$

where $\Omega_i$ is a finite collection of indices. From this definition, the global constraint set $\mathcal{X}$ can be represented as $\mathcal{X} = \{x \in \mathbb{R}^n \mid x \in \mathcal{X}_0 \text{ and } g(x,\omega) \le 0, \text{ for } \omega \in \Omega_i, \; i \in \mathcal{V}\}$.

Problems of particular interest are those involving nontrivial constraints on which projections are computationally expensive, such as Linear Matrix Inequalities (LMIs):

$$\mathcal{X}_i = \left\{x \in \mathbb{R}^n \mid A_0(\omega) + \sum_{j=1}^n x_j A_j(\omega) \preceq 0, \text{ for } \omega \in \Omega_i\right\}, \tag{3}$$

where $A_j(\omega) \in \mathbb{S}_m$ for $j = 0,1,\ldots,n$ and $\omega \in \Omega_i$ are given matrices. The inequalities in (3) are referred to as *linear matrix inequalities* (LMIs). A semidefinite programming (SDP) problem has one or more LMI constraints. Finding a feasible point of the set (3) is often a difficult problem on its own.

Note that the inequalities in (3) can represent a wide variety of convex constraints (see [21] for more details). For example, quadratic inequalities, inequalities involving matrix norms, and various inequality constraints arising in control such as Lyapunov and quadratic matrix inequalities can be all cast as LMIs in (3). When all matrices $A_j(\omega)$ in (3) are diagonal, the equalities reduce to linear inequalities.

## III. Algorithm, Assumptions, and Main Results

Our goal is to design a decentralized protocol by which all agents $i \in \mathcal{V}$ maintain a sequence of local copies $\{x_{i,k}\}_{k\ge 0}$ converging to the same point in $\mathcal{X}^*$ as $k$ goes to infinity.

Since we assume that the local constraint sets $\mathcal{X}_i$'s are nontrivial, we do not find an exact projection onto $\mathcal{X}_i$ at each step of the algorithm. Instead, at iteration $k$, each agent $i$ randomly generates $\omega_{i,k} \in \Omega_i$ and makes an *approximate projection* on the selected constraint function $g(\cdot,\omega_{i,k}) \le 0$.

### A. Decentralized Algorithm with Approximate Projections

Each agent $i$ maintains a sequence $\{x_{i,k}\}_{k\ge 0}$. The element $x_{i,k}$ of the sequence is the agent $i$'s estimate of the decision variable $x$ at time $k$. Let $g^+(x,\omega)$ denote the function that measures the violation of the constraint $g(\cdot,\omega)$ at $x$, i.e., $g^+(x,\omega) \triangleq \max\{g(x,\omega), 0\}$.

At $k = 0$, the estimates $x_{i,0}$ are locally initialized such that $x_{i,0} \in \mathcal{X}_0$. At time step $k$, all agents $j$ broadcast their previous estimates $x_{j,k-1}$ to all of the nodes in their neighborhood, i.e., to all agents $i$ such that $(j,i) \in \mathcal{E}_k$. Then, each agent $i$ updates $x_{i,k}$ as follows:

$$p_{i,k} = \sum_{j\in\mathcal{V}} [W_k]_{ij} x_{j,k-1} \tag{4a}$$

$$v_{i,k} = \Pi_{\mathcal{X}_0}\left[p_{i,k} - \alpha_k s_{i,k}\right] \tag{4b}$$

$$x_{i,k} = \Pi_{\mathcal{X}_0}\left[v_{i,k} - \lambda_k d_{i,k}\right], \tag{4c}$$

where $W_k$ is a nonnegative $N\times N$ weight matrix, $\{\alpha_k\}$ is a positive sequence of nonincreasing stepsizes; $s_{i,k}$ is a subgradient of the function $f_i$ at $p_{i,k}$; $\omega_{i,k}$ is a random variable taking values in the index set $\Omega_i$; $d_{i,k}$ is chosen such that $d_{i,k} \in \partial g^+(v_{i,k},\omega_{i,k})$ if $g^+(v_{i,k},\omega_{i,k}) > 0$ and $d_{i,k} = d$ for some $d \neq 0$, if $g^+(v_{i,k},\omega_{i,k}) = 0$. We let

$\lambda_k \triangleq g^+(v_{i,k}, \omega_{i,k})/\|d_{i,k}\|^2$, which is a variant of the Polyak stepsize [22]. Note that $\lambda_k$ is always well-defined as the vector $d_{i,k}$ is nonzero if $g^+(v_{i,k}, \omega_{i,k}) > 0$.

In (4a), each agent $i$ calculates a weighted average of the received messages (including its own message $x_{i,k-1}$) to obtain $p_{i,k}$. Specifically, $[W_k]_{ij} \geq 0$ is the weight that agent $i$ allocates to the message $x_{j,k-1}$. This communication step is decentralized since the weight matrix $W_k$ respects the topology of the graph $\mathcal{G}_k$, i.e., $[W_k]_{ij} > 0$ only if $(j,i) \in \mathcal{E}_k$ and $[W_k]_{ij} = 0$, otherwise.

In (4b), each agent $i$ adjusts the average $p_{i,k}$ in the direction of the negative subgradient of its local objective $f_i$ to obtain $v_{i,k}$. The adjusted average is projected to the simple set $\mathcal{X}_0$.

In (4c), agent $i$ observes a random realization of $\omega_{i,k} \in \Omega_i$ and measures the feasibility violation of the selected component constraint $g(\cdot, \omega_{i,k})$ at $v_{i,k}$. If $g^+(v_{i,k}, \omega_{i,k}) > 0$, it calculates a subgradient $d_{i,k}$ and takes an additional subgradient step with the stepsize $\lambda_k$ to minimize this violation. The adjusted vector is then projected back to the set $\mathcal{X}_0$. This guarantees $x_{i,k} \in \mathcal{X}_0$ for all $k \geq 0$ and $i \in \mathcal{V}$, but does not necessarily guarantee $x_{i,k} \in \mathcal{X}$. In Section IV, we show that $x_{i,k}$ for all $i \in \mathcal{V}$ asymptotically achieve feasibility, i.e., $\lim_{k\to\infty} \mathrm{dist}(x_{i,k}, \mathcal{X}) = 0$ for all $i \in \mathcal{V}$.

To further explain the step (4c), let us consider the example mentioned in Section II-B. Let us define the projection $A^+$ of a real symmetric matrix $A$ onto the cone of positive semidefinite matrices. For any $A \in \mathbb{S}^m$, we can find an eigenvalue decomposition $A = B\Lambda B^T$, where $B$ is an orthogonal matrix and $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_m)$. Then, its projection is given by $A^+ = B\Lambda^+ B^T$, where $\Lambda^+ = \mathrm{diag}(\lambda_1^+, \dots, \lambda_m^+)$ with $\lambda_i^+ = \max\{0, \lambda_i\}$ [23]. Let us define

$$A(x,\omega) \triangleq A_0(\omega) + \sum_{j=1}^{n} x_j A_j(\omega).$$

Then, the amount of violation of the corresponding LMI constraint $A(x,\omega) \preceq 0$ can be measured by the following scalar function:

$$g^+(x,\omega) = \|A^+(x,\omega)\|_F.$$

By direct calculations, it is not difficult to see that its subgradient is given by

$$\partial g^+(x,\omega) = \frac{1}{g^+(x,\omega)} \begin{pmatrix} \mathsf{Tr} A_1 A^+(x,\omega) \\ \vdots \\ \mathsf{Tr} A_n A^+(x,\omega) \end{pmatrix}, \qquad (5)$$

if $g^+(x,\omega) > 0$. Otherwise, $\partial g^+(x,\omega) = d$.

It is worth mentioning that the algorithm (4a)-(4c) includes the method that has been proposed in [14] as a special case. In order to see this, let $\mathcal{X}_0 = \mathbb{R}^n$ and $g(x, \omega_{i,k}) = \mathrm{dist}(x, \mathcal{X}_i^{\omega_{i,k}})$, where $\mathcal{X}_i^{\omega_{i,k}} = \{x \mid g(x, \omega_{i,k}) \leq 0\}$. Then, it is not difficult to see that

$$d_{i,k} = \frac{v_{i,k} - \Pi_{\mathcal{X}_i^{\omega_{i,k}}}[v_{i,k}]}{\|v_{i,k} - \Pi_{\mathcal{X}_i^{\omega_{i,k}}}[v_{i,k}]\|}$$

and since $\mathrm{dist}(v_{i,k}, \Pi_{\mathcal{X}_i^{\omega_{i,k}}}) = \|v_{i,k} - \Pi_{\mathcal{X}_i^{\omega_{i,k}}}[v_{i,k}]\|$, the steps (4b)-(4c) reduce to

$$x_{i,k} = \Pi_{\mathcal{X}_i^{\omega_{i,k}}}[p_{i,k} - \alpha_k s_{i,k}],$$

which is exactly the algorithm in [14].

### B. Assumptions

For the optimization problem (2), we make the following assumptions on the set $\mathcal{X}_0$, the objective functions $f_i(x)$ for $i \in \mathcal{V}$, and the constraint functions $g(x,\omega)$ for $\omega \in \Omega_i$ and $i \in \mathcal{V}$.

*Assumption 1:* (a) The set $\mathcal{X}_0$ is nonempty, closed and convex.

(b) The function $f_i(x)$, for each $i \in \mathcal{V}$, is defined and convex (not necessarily differentiable) over some open set that contains $\mathcal{X}_0$.

(c) The subgradients $s \in \partial f_i(x)$ are uniformly bounded over the set $\mathcal{X}_0$. That is, for all $i \in \mathcal{V}$, there is a scalar $C_{f_i}$ such that for all $s \in \partial f_i(x)$ and $x \in \mathcal{X}_0$,

$$\|s\| \leq C_{f_i}.$$

(d) The function $g(x,\omega)$, for each $\omega \in \Omega_i$ and $i \in \mathcal{V}$, is defined and convex in $x$ (not necessarily differentiable) over some open set that contains $\mathcal{X}_0$.

(e) The subgradients $d \in \partial g^+(x,\omega)$ are uniformly bounded over the set $\mathcal{X}_0$. That is, there is a scalar $C_g$ such that for all $d \in \partial g^+(x,\omega)$, $x \in \mathcal{X}_0$, $\omega \in \Omega_i$, and $i \in \mathcal{V}$,

$$\|d\| \leq C_g.$$

By Assumption 1, the subdifferentials $\partial f_i(x)$ and $\partial g^+(x,\omega)$ are nonempty over $\mathcal{X}_0$. It also implies that for any $i \in \mathcal{V}$ and $x, y \in \mathcal{X}_0$,

$$|f_i(x) - f_i(y)| \leq C_{f_i}\|x - y\|, \qquad (6)$$

and for any $\omega \in \Omega_i$, $i \in \mathcal{V}$, and $x, y \in \mathcal{X}_0$,

$$|g^+(x,\omega) - g^+(y,\omega)| \leq C_g\|x - y\|. \qquad (7)$$

One sufficient condition for Assumption 1(c) and 1(e) is that the set $\mathcal{X}_0$ is compact.

At each iteration $k$, each agent $i$ randomly generates $\omega_{i,k} \in \Omega_i$. We assume that they are *i.i.d.* samples from some probability distribution on $\Omega_i$. The next two assumptions are for the random variables $\omega \in \Omega_i$.

*Assumption 2:* Each $\Omega_i$ is a finite set and each element of $\Omega_i$ is generated with nonzero probability, i.e., for any $\omega \in \Omega_i$ and $i \in \mathcal{V}$

$$\pi_i(\omega) \triangleq \mathsf{Pr}\{\omega \mid \omega \in \Omega_i\} > 0$$

*Assumption 3:* For all $\omega \in \Omega_i$ and $i \in \mathcal{V}$, there exists a constant $c > 0$ such that for all $x \in \mathcal{X}_0$

$$\mathrm{dist}^2(x, \mathcal{X}) \leq c\mathsf{E}\left[(g^+(x,\omega))^2\right].$$

The upper bound in Assumption 3 is known as *global error bound* and is crucial for the convergence analysis of our method (4a)-(4c). Sufficient conditions for this bound have been shown in [24] and [25], which require the existence of a Slater point, i.e., there exists a point $\tilde{x}$ such that $\tilde{x} \in \mathcal{X}_0$ and $g(\tilde{x}, \omega) < 0$ for all $\omega$. In the case when each function $g(\cdot, \omega)$

is either a linear equality or inequality, Assumption 3 is called *linear regularity* and can be shown to hold by using the results in [26] and [27]. Also, if the random sequences $\{\omega_{i,k}\}$ are stationary Markov processes (which are independent of the distribution of the initial points $x_{i,0}$), then Assumption 3 holds when the feasible set $\mathcal{X}$ has a nonempty interior [28].

The inter-agent communication relies on the time-varying graph sequence $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$, for $k \geq 0$. Two key assumptions on these communication graphs are the following:

*Assumption 4:* For all $k \geq 0$,

(a) $[W_k]_{ij} \geq 0$ for all $i, j \in \mathcal{V}$ and there exists a scalar $\gamma \in (0, 1)$ such that $[W_k]_{ij} \geq \gamma$ only if $j \in \mathcal{N}_{i,k}$

(b) $\sum_{j\in\mathcal{V}}[W_k]_{ij} = 1$ for all $i \in \mathcal{V}$, and $\sum_{i\in\mathcal{V}}[W_k]_{ij} = 1$ for all $j \in \mathcal{V}$.

Assumption 4(a) ensures that the weight matrices $W_k$ respects the underlying topology $\mathcal{G}_k$ for every $k$ so that the communication is indeed decentralized. The lower boundedness of the weights is required to show consensus among all agents (see [29] for more details) but the agents need not know the $\gamma$ value in running the algorithm. Assumption 4(d) (doubly stochasticity of the matrices $W_k$) is required for balanced communication. Here assuming doubly stochasticity of $W_k$ for all $k \geq 1$ might be too strong as finding such $W_k$ usually requires a global view (unless the underlying graph is regular or fully connected) and not all directed graphs admit a doubly stochastic matrix [30]. However, we will stick to this assumption because removing doubly stochasticity can introduce bias in the optimization [31]. Such a drawback can be easily overcome by adopting some other communication protocols like asynchronous gossip [32] (we leave this as our future work).

*Assumption 5:* There exists a scalar $Q$ such that the graphs $\left(\mathcal{V}, \bigcup_{\ell=0,\ldots,Q-1} \mathcal{E}_{k+\ell}\right)$ are strongly connected for all $k \geq 0$.

Assumption 5 ensures that there exists a path from one agent to every other agent within any bounded interval of length $Q$. We say that such a sequence of graphs is $Q$-strongly connected.

### C. Main Results

We demonstrate the correctness of the algorithm (4a)-(4c) by proving its convergence. In what follows, the end result is stated, which holds under the assumptions we have laid out above.

*Proposition 1:* Let Assumptions 1 - 5 hold. Let the stepsizes be such that $\sum_{k=1}^{\infty} \alpha_k = \infty$ and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$. Assume that problem (2) has a nonempty optimal set $\mathcal{X}^*$. Then, with the choice of $\lambda_k = g^+(v_{i,k}, \omega_{i,k})/\|d_{i,k}\|^2$ for all $k \geq 1$ and $i \in \mathcal{V}$, the iterates $\{x_{i,k}\}$ generated by each agent $i \in \mathcal{V}$ via method (4a)-(4c) converge almost surely to the same point in the optimal set $\mathcal{X}^*$, i.e., for a random point $x^\star \in \mathcal{X}^*$

$$\lim_{k\to\infty} x_{i,k} = x^\star \quad \text{for all } i \in \mathcal{V} \quad a.s.$$

## IV. Convergence Analysis

In this section, we are concerned with demonstrating the convergence results stated in Proposition 1. We first review some lemmas from existing literature that are necessary in our analysis.

### A. Preliminary Results

The following lemma states a non-expansiveness property of the projection operator (see [33] for its proof).

*Lemma 1:* Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a nonempty closed convex set. The function $\Pi_{\mathcal{X}} : \mathbb{R}^d \to \mathcal{X}$ is nonexpansive, i.e.,

$$\|\Pi_{\mathcal{X}}[x] - \Pi_{\mathcal{X}}[y]\| \leq \|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n.$$

In our analysis of the algorithm, we also make use of the following convergence result due to Robbins and Siegmund (see [34, Lemma 10-11, p. 49-50]).

*Theorem 1:* Let $\{v_k\}$, $\{u_k\}$, $\{a_k\}$ and $\{b_k\}$ be sequences of non-negative random variables such that

$$\mathsf{E}[v_{k+1}|\mathcal{F}_k] \leq (1 + a_k)v_k - u_k + b_k \quad \text{for all } k \geq 0 \quad a.s.$$

where $\mathcal{F}_k$ denotes the collection $v_0, \ldots, v_k$, $u_0, \ldots, u_k$, $a_0, \ldots, a_k$ and $b_0, \ldots, b_k$. Also, let $\sum_{k=0}^{\infty} a_k < \infty$ and $\sum_{k=0}^{\infty} b_k < \infty$ *a.s.* Then, we have $\lim_{k\to\infty} v_k = v$ for a random variable $v \geq 0$ *a.s.*, and $\sum_{k=0}^{\infty} u_k < \infty$ *a.s.*

In the following lemma, we show a relation of $p_{i,k}$ and $x_{i,k-1}$ associated with any convex function $h$ which will be often used in the analysis. For example, $h(x) = \|x - a\|^2$ for some $a \in \mathbb{R}^n$ or $h(x) = \text{dist}^2(x, \mathcal{X})$.

*Lemma 2:* Let Assumption 4 hold. Then, for any convex function $h : \mathbb{R}^n \to \mathbb{R}$, we have

$$\sum_{i\in\mathcal{V}} h(p_{i,k}) \leq \sum_{i\in\mathcal{V}} h(x_{i,k-1}).$$

Lastly, for the convergence proof of our algorithm, we use a result from [35] which shows the averaged iterates can still arrive at consensus if the errors behave nicely.

*Lemma 3:* Let Assumptions 4 and 5 hold. Consider the iterates generated by

$$\theta_{i,k} = \sum_{j\in\mathcal{V}} [W_k]_{ij}\theta_{j,k-1} + e_{i,k} \text{ for all } i \in \mathcal{V}. \tag{8}$$

Suppose there exists a nonnegative nonincreasing scalar sequence $\{\alpha_k\}$ such that

$$\sum_{k=1}^{\infty} \alpha_k \|e_{i,k}\| < \infty \text{ for all } i \in \mathcal{V}.$$

Then, for all $i, j \in \mathcal{V}$,

$$\sum_{k=1}^{\infty} \alpha_k \|\theta_{i,k} - \theta_{j,k}\| < \infty.$$

### B. Lemmas

We need a series of lemmas for proving Proposition 1. The proofs are omitted here due to the space limit; interested readers may refer to the supplementary material [36].

We first state an auxiliary lemma that will be later used to relate two consecutive iterates $x_{i,k}$ and $x_{i,k-1}$. This lemma can be shown by combining two existing results in [22] and [16], but we include it here for completeness.

*Lemma 4:* Let Assumption 1 hold. Let the sequences $p_{i,k}$, $v_{i,k}$ and $x_{i,k}$ be generated by the algorithm (4a)-(4c). Then, we have almost surely for any $\check{x}, z \in \mathcal{X}_0$, $\omega_{i,k} \in \Omega_i$, $i \in \mathcal{V}$ and $k \geq 1$,

$$\|x_{i,k} - \check{x}\|^2 \leq \|p_{i,k} - \check{x}\|^2 - 2\alpha_k(f_i(z) - f_i(\check{x})) - \frac{\tau - 1}{\tau C_g^2}\left(g^+(p_{i,k}, \omega_{i,k})\right)^2 + \frac{1}{4\eta}\|p_{i,k} - z\|^2 + D_{\tau,\eta}\alpha_k^2,$$

where $D_{\tau,\eta} = (\tau + 4\eta + 1)C_{f_i}^2$ and $\eta, \tau > 0$ are arbitrary.

Since we use an approximate projection, we cannot guarantee the feasibility of the sequences $x_{i,k}$ or $p_{i,k}$. In the next lemma, we prove that $p_{i,k}$ for all $i \in \mathcal{V}$ asymptotically achieve feasibility.

*Lemma 5:* Let Assumptions 1-4 hold. Let the sequence $\{\alpha_k\}$ be such that $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$. Then, the iterates $\{p_{i,k}\}$ generated by each agent $i \in \mathcal{V}$ via method (4a)-(4c) almost surely satisfy:

$$\sum_{k=1}^{\infty} \mathrm{dist}^2(p_{i,k}, \mathcal{X}) < \infty, \quad \text{for all } i \in \mathcal{V}.$$

To complete the proof, we show in the first part of the next lemma that the error $e_{i,k} = x_{i,k} - p_{i,k}$ (perturbations made after the consensus step (4a)) eventually converges to zero for all $i \in \mathcal{V}$. This result combined with Lemma 5 implies that the sequence $x_{i,k}$ for all $i \in \mathcal{V}$ also achieve asymptotic feasibility. Furthermore, since the error $e_{i,k}$ for all $i \in \mathcal{V}$ converge to zero, we can invoke Lemma 3 to show the iterate consensus. Let us define $z_{i,k} \triangleq \Pi_{\mathcal{X}}[p_{i,k}]$ and $\bar{z}_k \triangleq \frac{1}{N}\sum_{i\in\mathcal{V}} z_{i,k}$. In the second part of the next lemma, we bound the network error term $\|z_{i,k} - \bar{z}_k\|$ for any $i \in \mathcal{V}$.

*Lemma 6:* Let Assumptions 1-5 hold. Let the sequence $\{\alpha_k\}$ is such that $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$. Define $e_{i,k} = x_{i,k} - p_{i,k}$ for all $i \in \mathcal{V}$ and $k \geq 1$. Then, we have almost surely

(a) $\sum_{k=1}^{\infty} \|e_{i,k}\|^2 < \infty$ for all $i \in \mathcal{V}$,
(b) $\sum_{k=1}^{\infty} \alpha_k \|z_{i,k} - \bar{z}_k\| < \infty$ for all $i \in \mathcal{V}$.

In the next lemma, we use standard convexity analysis to lower-bound the term $\sum_{i\in\mathcal{V}}(f_i(z_{i,k}) - f_i(\check{x}))$ with a network error term and a global term.

*Lemma 7:* Let Assumption 1 hold. Then, for all $\check{x} \in \mathcal{X}$, we have

$$\sum_{i\in\mathcal{V}}(f_i(z_{i,k}) - f_i(\check{x})) \geq -C_f \sum_{i\in\mathcal{V}}\|z_{i,k} - \bar{z}_k\| + f(\bar{z}_k) - f(\check{x}),$$

where $C_f = \max_{i\in\mathcal{V}} C_{f_i}$.

### C. Proof of Proposition 1

We invoke Lemma 4 with $z = z_{i,k} = \Pi_{\mathcal{X}}[p_{i,k}]$, $\tau = 4$ and $\eta = cC_g^2$. We also let $\check{x} = x^*$ for an arbitrary $x^* \in \mathcal{X}^*$. Therefore, for any $x^* \in \mathcal{X}^*$, $\omega_{i,k} \in \Omega_i$, $i \in \mathcal{V}$ and $k \geq 1$, we almost surely have

$$\|x_{i,k} - x^*\|^2 \leq \|p_{i,k} - x^*\|^2 - 2\alpha_k(f_i(z_{i,k}) - f_i(x^*)) - \frac{3}{4C_g^2}\left(g^+(p_{i,k}, \omega_{i,k})\right)^2 + \frac{1}{4cC_g^2}\mathrm{dist}^2(p_{i,k}, \mathcal{X}) + (5 + 4cC_g^2)C_{f_i}^2\alpha_k^2.$$

Let $\mathcal{F}_k$ denote the algorithm's history up to time $k$. i.e.,

$$\mathcal{F}_k = \{x_{i,0},\ (\omega_{i,t},\ 1 \leq t \leq k),\ i \in \mathcal{V}\},$$

and $\mathcal{F}_= \{x_{i,0},\ i \in \mathcal{V}\}$. By taking the expectation conditioned on $\mathcal{F}_{k-1}$ in the above relation and summing this over $i \in \mathcal{V}$, we obtain

$$\sum_{i\in\mathcal{V}} \mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}\right] \leq \sum_{i\in\mathcal{V}} \|p_{i,k} - x^*\|^2 - 2\alpha_k \sum_{i\in\mathcal{V}}(f_i(z_{i,k}) - f_i(x^*)) - \frac{3}{4C_g^2}\sum_{i\in\mathcal{V}} \mathsf{E}\left[\left(g_i^+(p_{i,k}, \omega_{i,k})\right)^2 \mid \mathcal{F}_{k-1}\right] + \frac{1}{4cC_g^2}\sum_{i\in\mathcal{V}} \mathrm{dist}^2(p_{i,k}, \mathcal{X}) + DN\alpha_k^2,$$

where $D = (5 + 4cC_g^2)C_f^2$ with $C_f = \max_{i\in\mathcal{V}} C_{f_i}$. Now we use Lemma 2 with $h(x) = \|x - x^*\|^2$, Assumption 3 and Lemma 7 with $\check{x} = x^*$ to further estimate the terms on the right-hand side. From these, obtain almost surely for any $k \geq 1$ and $x^* \in \mathcal{X}^*$,

$$\sum_{i\in\mathcal{V}} \mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathcal{F}_{k-1}\right] \leq \sum_{i\in\mathcal{V}} \|x_{i,k-1} - x^*\|^2 - 2\alpha_k \sum_{i\in\mathcal{V}}(f(\bar{z}_k) - f(x^*)) - \frac{1}{2cC_g^2}\sum_{i\in\mathcal{V}} \mathrm{dist}^2(p_{i,k}, \mathcal{X}) + 2\alpha_k C_f \sum_{i\in\mathcal{V}} \|z_{i,k} - \bar{z}_k\| + DN\alpha_k^2.$$

Since $\bar{z}_k \in \mathcal{X}$, we have $f(\bar{z}_k) - f(x^*) \geq 0$. Thus, under the assumption $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$ and Lemma 6(b), the above relation satisfies all the conditions of the convergence Theorem 1. Using this theorem, we have the following results.

*Result 1:* The sequence $\{\|x_{i,k} - x^*\|\}$ is convergent *a.s.* for all $i \in \mathcal{V}$ and every $x^* \in \mathcal{X}^*$.

*Result 2:* For every $x^* \in \mathcal{X}^*$,

$$\sum_{k=1}^{\infty} \alpha_k(f(\bar{z}_k) - f(x^*)) < \infty \quad a.s.$$

As a direct consequence of *Result 1*, we know that the sequences $\{\|p_{i,k} - x^*\|\}$ and $\{\|z_{i,k} - x^*\|\}$ are also convergent *a.s.* (This is straightforward in view of relation (4a) and Lemma 5 with the knowledge that $\mathrm{dist}^2(p_{i,k}, \mathcal{X}) = \|p_{i,k} - z_{i,k}\|^2$.) Since $\|\bar{z}_k - x^*\| \leq \frac{1}{N}\sum_{i\in\mathcal{V}} \|z_{i,k} - x^*\|$, we further know that the sequence $\{\|\bar{z}_k - x^*\|\}$ is also convergent *a.s.*

As a direct consequence of *Result 2*, since $\sum_{k=1}^{\infty} \alpha_k = \infty$, we know that

$$\liminf_{k\to\infty}(f(\bar{z}_k) - f(x^*)) = 0 \quad a.s.$$

From this relation and the continuity of $f$, it follows that the sequence $\{\bar{z}_k\}$ must have one accumulation point in the set $\mathcal{X}^*$ *a.s.* This and the fact that $\{\|\bar{z}_k - x^*\|\}$ is convergent *a.s.* for every $x^* \in \mathcal{X}^*$ imply that for a random point $x^\star \in \mathcal{X}^*$,

$$\lim_{k\to\infty} \bar{z}_k = x^\star \quad a.s. \tag{9}$$

Also, by Lemma 6(b), we have

$$\liminf_{k\to\infty} \|z_{i,k} - \bar{z}_k\| = 0 \quad \text{for all } i \in \mathcal{V} \quad a.s. \tag{10}$$

The fact that $\{\|z_{i,k} - x^*\|\}$ is convergent *a.s.* for all $i \in \mathcal{V}$, together with (9) and (10) implies that

$$\lim_{k\to\infty} z_{i,k} = x^\star \quad \text{for all } i \in \mathcal{V} \quad a.s. \tag{11}$$

Since $\|x_{i,k} - z_{i,k}\| \le \|x_{i,k} - p_{i,k}\| + \|p_{i,k} - z_{i,k}\|$, by invoking Lemma 5 and 6(a), we have

$$\lim_{k\to\infty} \|x_{i,k} - z_{i,k}\| = 0 \quad \text{for all} i \in \mathcal{V} \quad a.s.$$

This and relation (11) give us the desired result, which is

$$\lim_{k\to\infty} x_{i,k} = x^\star \quad \text{for all } i \in \mathcal{V} \quad a.s.$$

## V. Conclusion

We have studied a distributed optimization problem defined on a multi-agent network which involves nontrivial constraints like LMIs. We have proposed a decentralized algorithm based on random feasibility updates, where we approximate the projection with an additional subgradient step. The proposed algorithm is efficiently applicable for solving any distributed optimization problems which involve lots of computationally prohibitive constraints, for example, decentralized SDPs. We have established the almost sure convergence of our method.

## References


[1] S. S. Ram, V. V. Veeravalli, and A. Nedić, "Distributed non-autonomous power control through distributed convex optimization," in *IEEE INFOCOM*, 2009, pp. 3001–3005.

[2] M. Rabbat and R. D. Nowak, "Distributed optimization in sensor networks," in *IPSN*, 2004, pp. 20–27.

[3] 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, March 2010.

[4] A. Jadbabaie, J. Lin, and A. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988 – 1001, June 2003.

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

[6] A. Nedić, "Asynchronous broadcast-based convex optimization over a network," *IEEE Trans. Automat. Contr.*, vol. 56, no. 6, pp. 1337–1351, 2011.

[7] K. Srivastava and A. Nedić, "Distributed asynchronous constrained stochastic optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 772–790, 2011.

[8] I. Lobel, A. Ozdaglar, and D. Feijer, "Distributed multi-agent optimization with state-dependent communication," *Mathematical Programming*, vol. 129, no. 2, pp. 255–284, 2011.

[9] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, March 2015.

[10] K. Tsianos, S. Lawlor, and M. Rabbat, "Push-sum distributed dual-averaging for convex optimization," in *Proceedings of the 51st IEEE Conference on Decision and Control*, Maui, Hawaii, December 2012, pp. 5453–5458.

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

[12] P. Bianchi, W. Hachem, and F. Iutzeler, "A stochastic coordinate descent primal-dual algorithm and applications to large-scale composite optimization," 2014, http://arxiv.org/pdf/1407.0898.

[13] W.Shi, Q. Ling, G. Wu, and W. Yin, "Extra: An exact first-order algorithm for decentralized consensus optimization," 2014, http://arxiv.org/pdf/1404.6264.

[14] S. Lee and A. Nedić, "Distributed random projection algorithm for convex optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, pp. 221–229, April 2013.

[15] B. Polyak, "Random algorithms for solving convex inequalities," in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, ser. Studies in Computational Mathematics, Y. C. Dan Butnariu and S. Reich, Eds. Elsevier, 2001, vol. 8, pp. 409 – 422.

[16] A. Nedić, "Random algorithms for convex minimization problems," *Mathematical Programming - B*, vol. 129, pp. 225–253, 2011.

[17] A. Nedić, "Random projection algorithms for convex set intersection problems," in *Proc. of the 49th IEEE Conference on Decision and Control*, 2010, pp. 7655–7660.

[18] T. Alamo, R. Tempo, and E. Camacho, "Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2545 –2559, Nov 2009.

[19] G. C. Calafiore, "Random convex programs," *SIAM J. Optimiz.*, vol. 20, no. 6, pp. 3427–3464, Dec. 2010.

[20] G. Calafiore and M. Campi, "Uncertain convex programs: Randomized solutions and confidence levels," *Mathematical Programming*, vol. 102, pp. 25–46, 2005.

[21] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, ser. Studies in Applied Mathematics. Philadelphia, PA: SIAM, Jun. 1994, vol. 15.

[22] B. Polyak, "Minimization of unsmooth functionals," *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 3, pp. 14 – 29, 1969.

[23] ——, "Gradient methods for solving equations and inequalities," *USSR Comput. Math. and Math. Phys.*, vol. 4, no. 6, pp. 17–32, 1964.

[24] F. Facchinei and J.-S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer, New York, 2003.

[25] A. Lewis and J.-S. Pang, "Error bounds for convex inequality systems," in *Generalized Convexity*. Kluwer Academic Publishers, 1996, pp. 75–110.

[26] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.*, vol. 38, no. 3, pp. 367–426, Sep. 1996.

[27] J. V. Burke and M. C. Ferris, "Weak sharp minima in mathematical programming," *SIAM Journal on Control and Optimization*, vol. 31, pp. 1340–1359, 1993.

[28] L. Gubin, B. Polyak, and E. Raik, "The method of projections for finding the common point of convex sets," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 6, pp. 1 – 24, 1967.

[29] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis, "Distributed subgradient methods and quantization effects," in *Proceedings of 47th IEEE Conference on Decision and Control*, December 2008, pp. 4177 –4184.

[30] B. Gharesifard and J. Cortes, "When does a digraph admit a doubly stochastic adjacency matrix?" in *Proceedings of the American Control Conference*, Baltimore, MD, 2010, pp. 2440–2445.

[31] K. I. Tsianos and M. G. Rabbat, "Distributed dual averaging for convex optimization under communication delays," in *Proceedings of the American Control Conference*, Montreal, QC, 2012, pp. 1067–1072.

[32] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[33] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex analysis and optimization*. Athena Scientific, 2003.

[34] B. Polyak, *Intro. to optimization*. Optimization software, Inc., Publications division, New York, 1987.

[35] S. S. Ram, A. Nedić, and V. V. Veeravalli, "A new class of distributed optimization algorithms: application to regression of distributed data," *Optimization Methods and Software*, vol. 27, no. 1, pp. 71–88, 2012.

[36] S. Lee and M. Zavlanos, "Approximate projections for decentralized optimization with SDP constraints," 2015, http://arxiv.org/abs/1509.08007.