23rd International Symposium on Mathematical Theory of Networks and Systems
Hong Kong University of Science and Technology, Hong Kong, July 16-20, 2018

Stochastic Mirror Descent Algorithm Over Distributed Time-varying Network

Yinghui Wang, Hongbing Zhou and Yiguang Hong

***Abstract*— In this paper, we propose a stochastic mirror descent algorithm for solving distributed convex optimization problem over a distributed time-varying network. We adopt Bregman divergence rather than Euclidean distance as the augmented distance measuring function to solve the distributed onvex optimization problem. With a fixed step-size, we analyze the convergence rate of our algorithm, which is also the best known convergence rate for distributed first-order algorithms.**

***Index Terms*— Distributed convex optimization, mirror descent, time-varying network, convergence rate**

I. Introduction

Currently, many optimization problems in data science (for instance, machine learning, data mining and statistics) include an increasing scale of data, which can hardly be handled through a single processor or a center. Still, the conventional parallel optimization in a network also requires a central unit to collect all the network data and then assign jobs to the full network. This urges the development of distributed designs and algorithms for multi-agent network, based on local computation and neighborhood communication in many areas, including sensor networks [1], [2], machine learning [3], [4] and power systems [5], [6].

Over the past years, there has been a considerable work on distributed convex optimization, including distributed projected subgradient method [1], [7], [8], distributed primal-dual subgradient method [9]–[11], and distributed ADMM method [12], [13]. These algorithms only need first-order subgradient information of the objective function and the Euclidean projection to the local constrained set, which can be applied to large-scale optimization problems. However, the aforementioned algorithms are based on Euclidean projection, provided that the local projections can be easily computed.

In these Euclidean projection cases, the local constraints sets can only be described by simple sets, such as hyperplanes, balls, bounded constraints, etc. Mirror descent algorithms based on Bregman divergence [14], [15] were developed to solve optimization problems with complex constraints sets. For example, [16], [17] studied the distributed convex optimization algorithms that use the mirror descent technologies. However, these algorithms are established with diminishing step-sizes. Additionally, [16] provided only asymptotic convergence results for distributed convex optimization, while [17] investigated the convergence rate of the proposed algorithm for distributed strongly convex cases.

The objective of this paper is to collectively solve the following distributed general (non-differentiable) convex problem:

$$\begin{aligned} \min \;\; F(x) &= \sum_{i=1}^{m} f_i(x) \\ s.\,t.\; x \in X &= \cap_{i=1}^{m} X_i \end{aligned} \tag{1}$$

where f_i is the non-differentiable local objective function of agent i and $X_i \in \mathcal{R}^n$ is the local closed convex constraint set known only by agent i. Here, the intersection X of X_is is non-empty, and agents can communicate over a given time-varying network.

With this background, we develop a distributed mirror descent algorithm based on Bregman divergence and moreover, analyze the convergence rate of distributed mirror descent algorithm for distributed optimization problem (1). The contributions of this paper are summarized as follows:

- We propose a distributed stochastic mirror descent algorithm in this paper, which is intended to minimize the distance between primal and dual spaces of problem (1) in an infinite dimensional setting using non-Euclidean distance. Our algorithm is a generalization and simplification to traditional Euclidean based distributed algorithms in the cases that the local constraint set are complex, like the unit simplex as an example or the cases that the distance is defined in the non-Euclidean sense, like the Kullback-Leibler (KL) divergence widely used in the measurement of the distance between two probability distributions.
- Convergence rate is given to distributed stochastic mirror descent algorithm intended to solve a distributed convex optimization problem with local constraint sets. We show that, for a total number of T iterations, the algorithm achieves an $O(\frac{1}{T})$ convergence rate with an error bound for fixed step-size, noting that [16] only provided convergence results for the same problem and [17] provided an $O(\frac{1}{T})$ convergence rate for solving distributed strongly convex optimization with global constraint sets known to all the agents.
- Our proposed algorithm recovers the best known convergence rate for algorithms with fixed step-size ([18], [13]) with an error bound in solving problem (1) over

Y. Wang, H. Zhou, and Y. Hong are with School of Mathematical Sciences, University of Chinese Academy of Sciences, and Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. e-mails: wangyinghuisdu@163.com; e-mail: zhouhongbing13@126.com; e-mail: yghong@iss.ac.cn.

[0] The work is supported by NSFC under grants 61733018, 61573344, 61333001.

a time-varying network. We also provided the convergence results of our algorithm to a fixed bound, where the result in [18] becomes a special case where the measuring function is the Euclidean divergence without constraint sets.

The rest of the paper is as follows. In Section II, we establish mathematical preliminaries and propose assumptions of this paper. Next, a Distributed Stochastic Mirror Descent Algorithm is proposed to solve the distributed convex problem (1) and we give the basis convergence rate analysis is given for the proposed algorithm.in Section III. Section IV concludes the paper.

II. Preliminaries and Assumptions

A. Graph Theory

Consider a directed graph $\mathcal{G}(k) = (\mathcal{M}, \mathcal{E}(k), W(k))$, where $M = \{1, 2, \dots, m\}$ represents the agent set, $\mathcal{E}(k)$ stands for the set of activated communication links between the agents at time k, and $W(k) = [w_{ij}^k]_{ij}$ represents the communication pattern at time k. We make the following assumption on $\mathcal{G}(k) = (\mathcal{M}, \mathcal{E}(k), W(k))$ [18]:

Assumption 2.1: The graph $\mathcal{G}(k)$ satisfies:

(a) There exists a constant c with $0 < c < 1$ such that, $\forall k \geqslant 0$ and $\forall i, j$, $w_{ii}^k \geqslant c$; $w_{ij}^k \geqslant c$ if $(j, i) \in \mathcal{E}(k)$.
(b) $W(k)$ is doubly stochastic, i. e. $\sum_{i=1}^m w_{ij}^k = 1$ and $\sum_{j=1}^m w_{ij}^k = 1$.
(c) There is an integer $\kappa \geqslant 1$ such that $\forall k \geqslant 0$ and $\forall (j, i) \in \mathcal{M} \times \mathcal{M}$,

$$(j, i) \in \mathcal{E}(k) \cup \mathcal{E}(k+1) \cup \cdots \cup \mathcal{E}(k+\kappa-1).$$

B. Convex Theory

We introduce the definition and properties of subgradient and strongly convex function according to [19].

Define $\triangledown f(x) \in \mathcal{R}^n$ as the subgradient of a (non-differentiable) convex function $f(x) : \mathcal{R}^n \to \mathcal{R}$ at a given vector $x :\in dom(f) \in \mathcal{R}^n$. Then $\triangledown f(x_2) \in \mathcal{R}^n$ is the subgradient of the function $f(x)$ at x_2 when for any $x_1, x_2 \in dom(f)$ the following inequality holds:

$$f(x_1) - f(x_2) - \langle \triangledown f(x_2), x_1 - x_2 \rangle \geqslant 0. \qquad (2)$$

A differentiable function $f : \mathcal{R}^n \to \mathcal{R}$ is called σ-strongly convex on $\mathcal{R}^n$ if, for any $x_1, x_2 \in \mathcal{R}^n$, the subgradient $\triangledown f(\cdot)$ of $f(\cdot)$ satisfies

$$f(x_2) \geqslant f(x_1) + \langle \triangledown f(x_1), x_2 - x_1 \rangle + \frac{1}{2}\sigma ||x_2 - x_1||_2^2, \qquad (3)$$

where μ is a constant.

Next, we give the definition of the Bregman divergence, which is used in mirror descent designs.

Definition 2.1: (Bregman divergence [14]) Given a strongly convex and differentiable funcion $\phi : \mathcal{R}^n \to \mathcal{R}$. We define the Bregman divergence $B(\cdot, \cdot)$ introduced by function ϕ as:

$$B(x, z) = \phi(x) - \phi(z) - \langle \triangledown \phi(z), x - z \rangle \geqslant 0.$$

We make the following assumption on Bregman divergence:

Assumption 2.2: (a) ϕ is σ_ϕ-strongly convex with respect to Euclidean norm, where for any two point $x_1, x_2 \in \mathcal{R}^n$:

$$\phi(x_2) \geqslant \phi(x_1) + \langle \triangledown \phi(x_1), x_2 - x_1 \rangle + \frac{1}{2}\sigma_\phi ||x_2 - x_1||_2^2,$$

or equally,

$$\langle \triangledown \phi(x_1), x_2 - x_1 \rangle \geqslant \sigma_\phi ||x_2 - x_1||_2^2.$$

(b) The Bregman divergence $B(x_1, x_2)$ is convex in x_1 for fixed x_2.

Moreover, the following assumption holds for every f_i:

Assumption 2.3: (a) $f_i(\cdot)$ is a convex function for any $i = 1, 2, \dots, m$.

(b) For any $x \in dom(f_i)$, the stochastic subgradient d_i of f_i satisfies:

$$||d_i(x)||_2 \leqslant G, \quad \forall i.$$

All the assumptions are widely used in the literature.

III. Distributed Algorithm and Main Results

The Distributed Stochastic Mirror Descent Algorithm is given in Algorithm 1. Then we analyze the disagreement between agents in the network and also the convergence rate of Algorithm 1.

Algorithm 1 Distributed Stochastic Mirror Descent Algorithm

Input: Total numbers of iteration T, fixed step-size η_i for each agent i
Initialize: $x_i^1 \in X_i$ for all $i = 1, 2, \dots m$.

for $k = 1, 1, \dots T$ **do**
Get a random subgradient d_i^k such that $\mathbb{E}[d_i^k] = \triangledown f_i(x_i^k)$ where $\triangledown f_i(x_i^k)$ is the subgradient of f_i at x_i^k.
Mirror Descent Step:

$$v_i^k = \arg\min_{v_i \in X_i} \left\{ \langle v_i, d_i^k \rangle + \frac{1}{\eta_i} B(v_i, x_i^k) \right\}. \qquad (4)$$

Weighted Step:

$$x_i^{k+1} = \sum_{i=1}^N w_{ij}^k v_j^k. \qquad (5)$$

end for

Remark 3.1: The selection of d_i^k:

(a)

$$d_i(k) = \triangledown f_i(x_i^k) + \epsilon_i^k,$$

where ϵ_i^k is chosen to be martingale-difference sequences or white Gaussian noise with zero-means.

(b) Each agent has access to a stream of data subsets: $x_{ij}, j = 1, 2 \dots p$, and $\triangledown f_i(x_i^k) = \frac{1}{p}\sum_{j=1}^p \triangledown f_i(x_{ij}^k)$. In this case, we can randomly choose $q \in \{1, 2 \dots, p\}$, such that $d_i(k) = \triangledown f_i(x_{iq}^k)$.

Define $\Phi(k,t) = W(k)W(k-1)\cdots W(t)$ as the transition matrices of $W(\cdot)$ for all $k \geqslant t \geqslant 0$.

Lemma 3.1: With Assumption 2.1, for any i, j and all $k \geqslant t \geqslant 0$, we have $|[\Phi(k,t)]_{ij} - \frac{1}{m}| \leqslant \alpha\beta^{k-t+1}$, where $\alpha = (1-\frac{c}{4m^2})^{-2}$ and $\beta = (1-\frac{c}{4m^2})^{\frac{1}{\kappa}}$.

Define $s_i^k = \arg\min_{x_i \in X_i} \left\{ \langle x_i, d_i^k \rangle + \frac{1}{\eta_i} B(x_i, v_i^k) \right\} - x_i^k$ as the error between v_i^k and x_i^k. The following lemma gives an expected error bound of s_i^k.

Lemma 3.2: Let Assumptions 2.1 2.2, and 2.3 hold. For any i, $j \in \mathcal{M}$, we have

$$\mathbb{E}||s_i^k||_2 \leqslant \frac{\eta_i G}{\sigma_\phi}.$$

Proof: According to [Theorem 3.34, [20]], for any $x_i \in X_i$, we have

$$\langle x_i - v_i^k, \eta_i d_i^k + \nabla\phi(v_i^k) - \nabla\phi(x_i^k)\rangle \geqslant 0. \tag{6}$$

According to the doubly stochasticity of the weight matrix $W(k)$ in Assumption 2.1 and the convexity of X_i, we get $x_i^k \in X_i$. Thus,

$$\langle v_i^k - x_i^k, \eta_i d_i^k + \nabla\phi(v_i^k) - \nabla\phi(x_i^k)\rangle \leqslant 0, \tag{7}$$

With Assumption 2.2, we get

$$\begin{aligned}\langle \eta_i d_i^k, x_i^k - v_i^k\rangle &\geqslant \langle \nabla\phi(x_i^k) - \nabla\phi(v_i^k), x_i^k - v_i^k\rangle \\ &\geqslant \sigma_\phi ||x_i^k - v_i^k||_2^2,\end{aligned} \tag{8}$$

This leads to

$$\begin{aligned}\mathbb{E}||s_i^k||_2 &= \mathbb{E}||x_i^k - v_i^k||_2 \\ &\leqslant \frac{\eta_i}{\sigma_\phi}\mathbb{E}||d_i^k||_2 \\ &\leqslant \frac{\eta_i G}{\sigma_\phi},\end{aligned} \tag{9}$$

which completes the proof. ■

Based on Lemmas 3.1 and 3.2, we establish the disagreement analysis of the Distributed Stochastic Mirror Descent Algorithm (Algorithm 1) in expectation.

Theorem 3.1: (Disagreement) Let Assumption 2.1, 2.2 and 2.3 hold. Then for any i, $j \in \mathcal{M}$, we have

$$\frac{1}{T}\sum_{k=1}^{T}\sum_{i=1}^{m}\mathbb{E}||x_i^k - x_j^k||_2 \leqslant \frac{B_1}{T} + B_2,$$

where $B_1 = (\dfrac{2\alpha\beta}{1-\beta}+1)\sum_{i=1}^m \mathbb{E}||x_i^1||_2$, $B_2 = \frac{2\alpha\beta m^2\eta G}{\sigma_\phi(1-\beta)}$, and $\eta = \max\{\eta_1, \ldots, \eta_m\}$.

Proof: According to the definition of s_i^k, we get

$$v_i^k = x_i^k + s_i^k. \tag{10}$$

Define $\bar{x}^k = \frac{1}{m}\sum_{i=1}^m x_i^k$. According to the doubly stochasticity of $W(k)$ in Assumption 2.1, the following equation holds:

$$\begin{aligned}\bar{x}^k &= \bar{x}^{k-1} + \frac{1}{m}\sum_{i=1}^m s_i^k \\ &= \bar{x}^1 + \sum_{i=2}^k \frac{1}{m}\sum_{i=1}^m s_i^t.\end{aligned} \tag{11}$$

With a similar analysis,

$$x_i^k = \sum_{j=1}^m [\Phi(k-1,1)]_{ij} x_j^1 + \sum_{t=2}^k\sum_{j=1}^m [\Phi(k-1,t)]_{ij} s_j^t. \tag{12}$$

According to Lemmas 3.1 and 3.2,

$$\begin{aligned}\mathbb{E}||x_i^k - \bar{x}^k||_2 \leqslant{}& \sum_{j=1}^m |[\Phi(k-1,1)]_{ij} - \frac{1}{m}|\mathbb{E}||x_j^1||_2 \\ &+ \sum_{t=2}^k\sum_{j=1}^m |[\Phi(k-1,t)]_{ij} - \frac{1}{m}|\mathbb{E}||s_j^t||_2 \\ \leqslant{}& \alpha\beta^{k-1}c^1(x) + \sum_{t=2}^k \alpha\beta^{k-t}\sum_{j=1}^m \mathbb{E}||s_j^t||_2 \\ \leqslant{}& \alpha\beta^{k-1}c^1(x) + \frac{\alpha m\eta G}{\sigma_\phi}\sum_{t=2}^k \beta^{k-t},\end{aligned} \tag{13}$$

where $c^1(x) = \sum_{i=1}^m \mathbb{E}||x_i^1||_2$ and $\eta = \max\{\eta_1, \ldots, \eta_m\}$. Summing the preceding inequalities over $k = 1, 2 \ldots T$,

$$\begin{aligned}\sum_{k=1}^T \mathbb{E}||x_i^k - x_j^k||_2 ={}& \mathbb{E}||x_i^1 - x_j^1||_2 + \sum_{k=2}^T \mathbb{E}||x_i^k - x_j^k||_2 \\ \leqslant{}& c^1(x) + 2\alpha\sum_{k=2}^T \beta^{k-1}c^1(x) \\ &+ \frac{2\alpha m\eta G}{\sigma_\phi}\sum_{k=2}^T\sum_{t=2}^k \beta^{k-t} \\ \leqslant{}& (\frac{2\alpha\beta}{1-\beta}+1)c^1(x) + \frac{2\alpha\beta mT\eta G}{\sigma_\phi(1-\beta)},\end{aligned} \tag{14}$$

where the last inequality is based on the following argument:

$$\sum_{k=2}^T\sum_{t=2}^k \beta^{k-t} \leqslant \sum_{k=2}^T\sum_{t=2}^T \beta^t \leqslant \frac{\beta T}{1-\beta}. \tag{15}$$

Therefore,

$$\begin{aligned}&\sum_{k=1}^T\sum_{i=1}^m \mathbb{E}||x_i^k - x_j^k||_2 \\ &\leqslant m(\frac{2\alpha\beta}{1-\beta}+1)\sum_{i=1}^m \mathbb{E}||x_j^1||_2 + \frac{2\alpha\beta m^2 T\eta G}{\sigma_\phi(1-\beta)},\end{aligned} \tag{16}$$

which completes the proof. ■

Based on Lemmas 3.2, we then analyse the expected convergence rate of the Distributed Stochastic Mirror Descent Method Algorithm (Algorithm 1), by showing the expected convergence of the average point $\frac{1}{T}\sum_{k=1}^T f_i(x_i^k)$ to the optimal point $f_i(x^*)$, where $x^* = \arg\min_{x\in X}\sum_{i=1}^m f_i(x)$.

Theorem 3.2: With Assumptions 2.1, 2.2 and 2.3(a) for all i, $j \in \mathcal{M}$, we have

$$\frac{1}{T}\sum_{k=1}^T\sum_{i=1}^m \mathbb{E}\big[f_i(x_i^k) - f_i(x^*)\big] \leqslant \frac{B_3}{T} + B_4,$$

where $B_3 = \frac{1}{\eta_i}\big[B(x^*, x_i^0) - B(x^*, x_i^T)\big]$ and $B_4 = \frac{\eta_i m G^2}{\sigma_\phi}$.

Proof: Through simple proof, we have

$$\langle d_i^k, v_i^k - x^* \rangle \leqslant \frac{1}{\eta_i}\big[B(x^*, x_i^k) - B(v_i^k, x_i^k) - B(x^*, v_i^k)\big]. \tag{17}$$

Since $f_i(\cdot)$ is convex, we obtain

$$\begin{aligned}\mathbb{E}\big[f_i(x_i^k) - f_i(x^*)\big] &\leqslant \mathbb{E}\langle d_i^k, v_i^k - x^* \rangle + \mathbb{E}\langle d_i^k, x_i^k - v_i^k \rangle \\ &\leqslant \frac{1}{\eta_i}\big[B(x^*, x_i^k) - B(v_i^k, x_i^k) - B(x^*, v_i^k)\big] \\ &\quad + \frac{\eta_i G^2}{\sigma_\phi}.\end{aligned} \tag{18}$$

As a result,

$$\begin{aligned}&\sum_{i=1}^m \mathbb{E}\big[f_i(x_i^k) - f_i(x^*)\big] \\ \leqslant &\sum_{i=1}^m \frac{1}{\eta_i}\big[B(x^*, x_i^k) - B(v_i^k, x_i^k) - B(x^*, v_i^k)\big] \\ &+ \frac{\eta_i m G^2}{\sigma_\phi}.\end{aligned} \tag{19}$$

By Assumptions 2.1 and 2.2, we conclude that

$$\begin{aligned}\sum_{i=1}^m B(x^*, x_i^k) &= \sum_{i=1}^m B(x^*, \sum_{j=1}^m w_{ij}^{k-1} v_j^{k-1}) \\ &\leqslant \sum_{i=1}^m \sum_{j=1}^m w_{ij}^{k-1} B(x^*, v_j^{k-1}) \\ &= \sum_{i=1}^m B(x^*, v_i^{k-1}).\end{aligned} \tag{20}$$

Therefore,

$$\begin{aligned}&\sum_{k=1}^T \sum_{i=1}^m \frac{1}{\eta_i}\big[B(x^*, v_i^{k-1}) - B(v_i^k, x_i^k) - B(x^*, v_i^k)\big] \\ \leqslant &\sum_{k=1}^T \sum_{i=1}^m \frac{1}{\eta_i}\big[B(x^*, x_i^{k-1}) - B(x^*, x_i^k)\big] \\ = &\sum_{i=1}^m \frac{1}{\eta_i}\big[B(x^*, x_i^0) - B(x^*, x_i^T)\big]\end{aligned} \tag{21}$$

Thus,

$$\begin{aligned}&\frac{1}{T}\sum_{k=1}^T \sum_{i=1}^m \big[f_i(x_i^k) - f_i(x^*)\big] \\ \leqslant &\frac{1}{T}\sum_{i=1}^m \frac{1}{\eta_i}\big[B(x^*, x_i^0) - B(x^*, x_i^T)\big] + \frac{\eta_i m G^2}{\sigma_\phi} \\ = &O(\frac{1}{T}) + \frac{\eta_i m G^2}{\sigma_\phi}.\end{aligned} \tag{22}$$

■

Remark 3.2: According to Theorem 3.2, as $T \to \infty$, the algorithm converges to a set with a fixed bound, which is consistent with the results in [18] for fixed step-size and without constraint sets.

IV. Conclusions

In this paper, we studied a distributed general convex problem over a multi-agent network. We have proposed a Distributed Stochastic Mirror Descent Algorithm with fixed step-sizes when the objective functions are general (non-differentiable) convex and the communication topology is assumed to be time-varying. The algorithm recovers the best convergence rate with an error bound. Moreover, we gave some simulation to demonstrate the effectiveness of the proposed algorithm.

References

[1] A. Nedic and A. Olshevsky, "Stochastic gradient-push for strongly convex functions on time-varying directed graphs," *arXiv preprint arXiv:1406.2075*, 2014.

[2] W. Shi, Q. Ling, G. Wu, and W. Yin, "Extra: An exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.

[3] Y. Wang, P. Lin, and Y. Hong, "Distributed regression estimation with incomplete data in multi-agent networks," *Science China Information Sciences*, vol. 61, no. 9, p. 092202, 2018.

[4] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributed Computing*, vol. 67, no. 1, pp. 33–46, 2007.

[5] S. Kar and G. Hug, "Distributed robust economic dispatch in power systems: A consensus+ innovations approach," in *Power and Energy Society General Meeting, 2012 IEEE*. IEEE, 2012, pp. 1–8.

[6] P. Yi, Y. Hong, and F. Liu, "Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems," *Automatica*, vol. 74, pp. 259–269, 2016.

[7] Y. Lou, Y. Hong, and S. Wang, "Distributed continuous-time approximate projection protocols for shortest distance optimization problems," *Automatica*, vol. 69, pp. 289–297, 2016.

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

[9] D. Yuan, S. Xu, and H. Zhao, "Distributed primal–dual subgradient method for multiagent optimization via consensus algorithms," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 6, pp. 1715–1724, 2011.

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

[11] J. Lei, H.-F. Chen, and H.-T. Fang, "Primal–dual algorithm for distributed constrained optimization," *Systems & Control Letters*, vol. 96, pp. 110–117, 2016.

[12] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the linear convergence of the admm in decentralized consensus optimization." *IEEE Trans. Signal Processing*, vol. 62, no. 7, pp. 1750–1761, 2014.

[13] A. Makhdoumi and A. Ozdaglar, "Convergence rate of distributed admm over networks," *IEEE Transactions on Automatic Control*, 2017.

[14] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR computational mathematics and mathematical physics*, vol. 7, no. 3, pp. 200–217, 1967.

[15] A. Beck and M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," *Operations Research Letters*, vol. 31, no. 3, pp. 167–175, 2003.

[16] C. Xi, Q. Wu, and U. A. Khan, "Distributed mirror descent over directed graphs," *arXiv preprint arXiv:1412.5526*, 2014.

[17] D. Yuan, Y. Hong, D. W. Ho, and G. Jiang, "Optimal distributed stochastic mirror descent for strongly convex optimization," *arXiv preprint arXiv:1610.04702*, 2016.

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

[19] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.

[20] A. P. Ruszczyński, *Nonlinear optimization*. Princeton university press, 2006, vol. 13.