

RESEARCH

Open Access



Distributed constrained optimization via continuous-time mirror design

Rui Sheng¹ and Wei Ni^{1*} 

*Correspondence:
niwei@ncu.edu.cn

¹School of Science, Nanchang University, Nanchang, China

Abstract

Recently, distributed convex optimization using a multiagent system has received much attention by many researchers. This problem is frequently approached by combing the consensus algorithms in the multiagent literature and the gradient algorithms in the convex optimization literature. Compared with unconstrained distributed optimization, the constrained case is more challenging, and it is usually tackled by the projected gradient method. However, the projected gradient algorithm involves projection nonlinearity and thus is hard to analyze. To avoid gradient projection, in this paper, we present a novel distributed convex optimization algorithm in continuous time by using mirror design. The resulting optimization dynamics is smooth without using gradient projection and is designed in a primal-dual framework, where the primal and dual dynamics are respectively aided by the mirror descent and the mirror ascent. As for the merit of mirror design in our paper, it avoids gradient projection in the optimization dynamics design and removes the difficulty of analyzing projection nonlinearity. Furthermore, the mirror base primal-dual optimization dynamics facilitates more convenience construction of Lyapunov functions in the stability analysis.

Keywords: Distributed convex optimization; Mirror descent; Multiagent; Constrained optimization

1 Introduction

Optimization is an important field in mathematics, and many engineering applications can be converted into optimization problems [1–10]. Recent years have witnessed an increasing attention on distributed convex optimization using multiagent systems [11, 12], which is motivated by the emergence of large-scale networks such as internet networks, wireless sensor networks, and mobile ad hoc networks. Distributed convex optimization refers to minimizing the aggregate sum of N convex cost functions by designing N dynamics, where each dynamics is distributed on one node and has only access to the information of one cost function and the state from its neighboring dynamics. The objective is that all the states of the N dynamics consensually and asymptotically converge to the minimizer of the total objective function. The optimization is solved in a distributed way since each local optimization dynamics, termed as a node on the network, uses information from its neighbors. The distributed optimization problem has been investigated from different perspectives; refer [11, 13–15] and references therein for details.

For the distributed optimization problem, there were many useful algorithms reported in the literature, such as distributed primal-dual gradient algorithm [16], nonsmooth-analysis-based algorithm [17], and approximate-projection-based algorithm [18]. Of particular interest among them is the distributed gradient projection method [11, 19–22], which requires computing the projection of the gradient. To overcome this difficulty, we propose a novel distributed convex optimization algorithm utilizing the mirror ascent/descent method. The original mirror ascent/descent method was proposed by Nemirovski and Yudin [23] and later evolved into a series of papers [24, 25]. However, most of these algorithms were discrete-time algorithms [20, 26–28], with relatively few putting attention on the case of continuous time [29]. As for distributed mirror descent algorithm, the continuous-time case, compared with discrete-time case, is more attractive since it can facilitate the use of the elegant Lyapunov argument [30] to aid the convergence analysis and allow the tool of differential geometry to be used in optimizing constrained problem [31].

Along the line of utilizing continuous-time mirror descent algorithms for distributed convex optimization problem, the work [32] was one of the earliest contributions; however, no proof of convergence was given there. Later, the authors in [33] presented a proof by using tools from nonsmooth analysis and set-valued dynamical systems, but with a limitation of only considering the simple case of unconstrained optimization, leaving the challenging case of constrained optimization untouched. The first contribution of this paper is to tackling the hard problem of constrained optimization under the framework of continuous-time mirror descent.

Generalization of the above works on continuous-time distributed mirror descent algorithms from the unconstrained case to the constrained one is not a simple task. By reviewing some commonly used constrained optimization algorithms and by pointing out their inherent disadvantages, in this paper, we apply the continuous-time mirror descent to design a novel optimization algorithm, which overcomes these disadvantages. It is well known that for distributed optimization problem, it is not a simple task when constraints in the optimization problem are taken into account. To handle the constraints, a variety of approaches, including the logarithmic barrier method [32], Lagrangian multiplier method [15], projected consensus algorithm [11], penalty-based method [34], and global linearization approach [9], can be resorted to. Among them, the primal-dual method is the most popular. It transforms the constrained optimization problem into an equivalent unconstrained one by designing the primal and dual dynamics. We note that the dual dynamics designed in this way is nonsmooth due to the projection operator involved to keep the evolution of the Lagrangian multipliers within the nonnegative orphan, and therefore it is difficult to analyze. As another contribution of this paper, we pursue a novel line of designing the dual dynamics via the mirror descent method. The merit of our design is avoiding gradient projection, and furthermore the resulting optimization dynamics is smooth.

Aside from modifying the nonsmooth dual dynamics (i.e., the dynamics for the Lagrangian multipliers) in the existing literature into a smooth version by borrowing the idea of mirror descent algorithm, we also redesign the primal dynamics in the existing literature by using the mirror ascend. For our smooth dual dynamics, it is designed in such a way that if the initial value of the dual variable is positive, then the value of the multipliers stays positive all the time with the evolution of the dual dynamics. Such a design has the benefit that the positive system theory can be utilized for convergence analysis.

Also, we redesign the primal dynamics in the existing literature in a mirror descent way so that the theory of Bregman divergence and Fenchel coupling can be utilized in the stability analysis. The stability of our resulting primal-dual optimization dynamics is analyzed by constructing a Lyapunov function, which is exactly the Fenchel coupling of the Bregman function; for details, refer the explicit form of our Lyapunov function in the proof of Theorem 3. The construction of a new Lyapunov function and the corresponding stability analysis constitute the last contribution of this paper.

In conclusion, the main novelties of this paper are as follows. Firstly, we obtain a continuous-time and distributed version of mirror descent algorithm, which complements the existing distributed optimization discrete-time algorithms. Secondly, the results in this paper consider the more challenging case of constrained distributed mirror descent algorithms, extending the existing results, which only deal with the unconstrained case. The third superiority of our method in comparison with existing results lies in the fact that it avoids using gradient projection in the optimization algorithm design. Therefore, it removes the difficulty of analyzing the resulting nonsmooth optimization dynamics and makes the simulation easier. Fourthly, the frequently used primal-dual algorithm for the optimization problem in the existing literature is modified in our paper via the mirror descent method, giving rise to new primal and dual dynamics. The modified primal dynamics facilitates the use of Bregman divergence and Fenchel coupling in the stability analysis, and the redesigned dual dynamics for the evolution of positive Lagrange multiplier does not include projection and therefore reduces the complexity of convergence analysis. Also, the construction of a Lyapunov function and the corresponding stability analysis are novel.

The rest of this paper is organized as follows. The problem is formulated in Sect. 3. In Sect. 4, we introduce the distributed mirror descent algorithm and use it in the primal and dual dynamics design and in the corresponding convergence analysis. The simulation results supporting our theoretical results are presented in Sect. 5. In Sect. 6, we summarize the paper.

2 Preliminaries

A. Notation. By \mathbb{R} and \mathbb{R}_+ we denote the sets of real and nonnegative real numbers, respectively; \mathbb{R}^n and \mathbb{R}_+^n are the sets of n -dimensional vectors and n -dimensional vectors with nonnegative components, respectively. The norm of a vector $x \in \mathbb{R}^n$ is denoted by $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$. By $x < 0$ ($x \leq 0$) for a vector x we mean that each entry of x is less than (less than or equal to) zero. For two vectors $x, y \in \mathbb{R}^n$, their inner product is defined as $\langle x, y \rangle = x^T y$. We denote $\mathbf{1}_n = (1, 1, \dots, 1)^T \in \mathbb{R}^n$ and $\mathbf{0}_n = (0, 0, \dots, 0)^T \in \mathbb{R}^n$. For a set of vectors $x_1, \dots, x_N \in \mathbb{R}^n$, we denote $\text{col}\{x_1, \dots, x_N\} = (x_1^T, \dots, x_N^T)^T$. For a number $a \in \mathbb{R}$, the projection $P_+[a]$ is defined to be zero if $a < 0$ and a if $a \geq 0$. For a vector x , its projection $P_+[x]$ is defined componentwise. The n -dimensional identity matrix is denoted by I_n . For arbitrary matrices A and B , $A \otimes B$ denotes the Kronecker product of matrices. The eigenvalues of a matrix $A \in \mathbb{R}^{n \times n}$ are denoted by $\lambda_i(A)$ for $i \in 1, \dots, n$. For an integer $k \geq 1$, we denote by C^k the set of k times continuously differentiable functions. For $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $\nabla f(x) = (\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n})^T$ is its gradient, and for $L: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, we denote $\nabla_\lambda L(x, \lambda) = (\frac{\partial L(x, \lambda)}{\partial \lambda_1}, \frac{\partial L(x, \lambda)}{\partial \lambda_2}, \dots, \frac{\partial L(x, \lambda)}{\partial \lambda_n})^T$. We say that $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function if for any $x, y \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$, it satisfies $f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y)$. The convexity

of f implies $\nabla f(x)(y-x) \leq f(y) - f(x)$. Furthermore, when $x \neq y$, the strict convexity of f is equivalent to $\nabla f(x)(y-x) < f(y) - f(x)$ or to $\nabla^2 f(x) > 0$.

B. Graph theory. Consider a graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \dots, N\}$ is the set of nodes representing N agents, and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges of the graph. An edge of G is denoted by (i, j) , which means that agents i and j can exchange information between them. A graph is undirected if the edges (i, j) and (j, i) in \mathcal{V} are considered to be the same; otherwise, the graph is directed. The set of neighbors of node i is denoted by $\mathcal{N}_i = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}, j \neq i\}$. The adjacency matrix $A = [a_{ij}]$ of a graph G on vertex $\{1, \dots, N\}$ is the $N \times N$ matrix with off-diagonal elements defined by specifying $a_{ij} = 1$ if (i, j) is an edge of G and $a_{ij} = 0$ otherwise and with diagonal elements defined as $a_{ii} = -\sum_{j \in \mathcal{N}_i} a_{ij}$. The Laplacian matrix $\mathcal{L} \in \mathbb{R}^{N \times N}$ of a graph $G = (\mathcal{V}, \mathcal{E})$ is defined as follows: if $i = j$, then $\mathcal{L}_{ij} = \sum_{j \in \mathcal{N}_i} a_{ij}$, and if $i \neq j$, then $\mathcal{L}_{ij} = -a_{ij}$. For any undirected graph, its Laplacian is symmetric positive semidefinite and satisfies $\mathcal{L} \cdot \mathbf{1}_N = 0 \cdot \mathbf{1}_N$. We say that the graph is strongly connected if there is a path between any pair of vertices. Furthermore, the graph G is connected if and only if its Laplacian matrix has a simple zero eigenvalue.

3 Problem formulation

Consider a network described by a graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \dots, N\}$ represents the set of N nodes, and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ denotes the set of edges of the graph. For each node $i \in \mathcal{V}$, there are a convex cost function $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, a set of inequality constraints $g_{ij} \leq 0, j = 1, \dots, r_i$, and a set of equality constraints $h_{ij} = 0, j = 1, \dots, s_i$, where r_i, s_i are positive integers, and $g_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}, j = 1, \dots, r_i$, and $h_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}, j = 1, \dots, s_i$, are all convex functions. If there are no constraints for agent i , we set $g_{ij}(x) \equiv 0, j = 1, \dots, s_i$, and $h_{ij} \equiv 0, j = 1, \dots, r_i$. The global network cost function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined as $f(x) = \sum_{i=1}^N f_i(x)$. In this paper, we consider the following optimization problem:

$$\begin{cases} \text{minimize} & f(x) = \sum_{i=1}^N f_i(x), \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \dots, N, \\ & h_i(x) = 0, \quad i = 1, \dots, N, \end{cases} \tag{3.1}$$

where $g_i = (g_{i1}, \dots, g_{ir_i})^T$ and $h_i = (h_{i1}, \dots, h_{is_i})^T$. Obviously, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}^r$ and $h_i : \mathbb{R}^n \rightarrow \mathbb{R}^s$, where $r = r_1 + \dots + r_N$ and $s = s_1 + \dots + s_N$. The optimization problem is to find $x^* \in \mathbb{R}^n$ such that the objective function $f(x)$ is minimized and the constraints $g_i(x^*) \leq 0$ and $h_i(x^*) = 0$ are satisfied. Such x^* is called the optimal solution, and the corresponding value $f^* = f(x^*)$ is called the optimal value.

This problem is usually solved by introducing a Lagrangian function and its saddle point, which are defined as follows.

Definition 1 (Lagrangian function [35]) The Lagrangian function associated with problem (3.1) is defined as a mapping $L : \mathbb{R}^n \times \mathbb{R}_+^r \times \mathbb{R}^s \rightarrow \mathbb{R}$ specified by

$$L(x, \lambda, \nu) = \sum_{i=1}^N f_i(x) + \sum_{i=1}^N \sum_{j=1}^{r_i} \lambda_{ij} g_{ij}(x) + \sum_{i=1}^N \sum_{j=1}^{s_i} \nu_{ij} h_{ij}(x), \tag{3.2}$$

where $\lambda = \text{col}\{\lambda_1, \dots, \lambda_N\}$ with $\lambda_i = \text{col}\{\lambda_{i1}, \dots, \lambda_{ir_i}\} \in \mathbb{R}_+^{r_i}$, and $\nu = \text{col}\{\nu_1, \dots, \nu_N\}$ with $\nu_i = \text{col}\{\nu_{i1}, \dots, \nu_{is_i}\} \in \mathbb{R}^{s_i}$. Obviously, $\lambda \in \mathbb{R}_+^r$ and $\nu \in \mathbb{R}^s$.

Definition 2 (Saddle point [36]) A couple $(x^*, (\lambda^*, v^*)) \in \mathbb{R}^n \times (\mathbb{R}_+^r \times \mathbb{R}^s)$ is a saddle point of the Lagrangian function L if it satisfies

$$L(x^*, (\lambda, v)) \leq L(x^*, (\lambda^*, v^*)) \leq L(x, (\lambda^*, v^*)).$$

To the Lagrangian function $L(x, \lambda, v)$, there corresponds the Lagrange dual function $\Omega : \mathbb{R}_+^r \times \mathbb{R}^s \rightarrow \mathbb{R}$ defined as

$$\Omega(\lambda, v) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, v).$$

Obviously, $\Omega(\lambda, v) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, v) \leq L(x^*, \lambda, v) \leq f(x^*) = f^*$. So the dual function provides a lower bound for the optimal value. We hope that the best lower bound

$$\rho^* = \sup_{\lambda \succeq 0, v \in \mathbb{R}^s} \Omega(\lambda, v).$$

The couple of values (λ^*, v^*) satisfying $\Omega(\lambda^*, v^*) = \rho^*$ is called the dual optimal solution, whereas x^* achieving $f(x^*) = f^*$ is called the primal optimal solution. The case $f^* = \rho^*$ can be guaranteed by imposing, for example, the Slater condition as follows.

Definition 3 (Slater’s constraint qualification certificate [36]) The Slater constraint qualification certificate is satisfied by (3.1) if there exists $x \in \mathbb{R}^n$ such that $g_i(x) < 0$ and $h_i(x) = 0$ for $i = 1, \dots, N$.

It can be shown that the saddle point $(x^*, (\lambda^*, v^*))$ of the Lagrangian $L(x, \lambda, v)$ associated with problem (3.1) provides an optimal solution x^* to the optimization problem (3.1), but conversely, the primal optimal solution x^* together with the dual optimal solution (λ^*, v^*) does not provide a saddle point $(x^*, (\lambda^*, v^*))$ for $L(x, \lambda, v)$. To achieve this, the following theorem, which can be found in [36, 37], is useful.

Theorem 1 *If the pair $(x^*, (\lambda^*, v^*))$ is a saddle point for $L(x, \lambda, v)$, then x^* is an optimal solution to problem (3.1). Conversely, if x^* is an optimal solution to problem (3.1), then there exists a couple of points $(\lambda^*, v^*) \in \mathbb{R}_+^r \times \mathbb{R}^s$ such that $(x^*, (\lambda^*, v^*))$ is a saddle point for $L(x, \lambda, v)$.*

Remark 1 According to this theorem, finding an optimal solution x^* of problem (3.1) transforms to seeking a saddle point $(x^*, (\lambda^*, v^*))$ of the Lagrangian $L(x, \lambda, v)$ in (3.2). The latter amounts to minimizing the Lagrangian with respect to x and maximizing the Lagrangian with respect to (λ, v) .

Another useful concept characterizing the optimal solution to problem (3.1) is the KKT conditions stated in the following theorem [36].

Theorem 2 *Suppose that the Slater conditions in Definition 3 are satisfied and $f_i, g_i, h_i, i = 1, \dots, N$, are convex. Then x^* is a solution of (3.1) if and only if there exists $(\lambda^*, v^*) \in$*

$\mathbb{R}_+^r \times \mathbb{R}^s$ such that the following conditions (called the KKT conditions) hold:

$$\text{KKT: } \begin{cases} g_{ij}(x^*) \leq 0, & j = 1, \dots, r_i, \\ h_{ij}(x^*) = 0, & j = 1, \dots, s_i, \\ \lambda_{ij}^* \geq 0, & j = 1, \dots, r_i, \\ \lambda_{ij}^* g_{ij}(x^*) = 0, & j = 1, \dots, r_i, \\ \sum_{i=1}^N \nabla f_i(x^*) + \sum_{i=1}^N \sum_{j=1}^{r_i} \lambda_{ij}^* \nabla g_{ij}(x^*) + \sum_{i=1}^N \sum_{j=1}^{s_i} v_{ij}^* \nabla h_{ij}(x^*) = 0. \end{cases} \tag{3.3}$$

The point (x^*, λ^*, v^*) obtained in Theorem 2 is called the KKT point. Motivated by this theorem, we will tackle the constrained optimization problem (3.1) by adopting the dynamical system approach. We formulate the problem in detail as follows.

Problem formulation: In what follows, we will design in a distributed way a continuous-time dynamics for (x, λ, v) such that:

- the dynamics is smooth by avoiding projection;
- the equilibrium of this dynamics is exactly the saddle point or the KKT point;
- the equilibrium of this dynamics is asymptotically stable;
- the λ -subdynamics remains nonnegative all the time for a nonnegative initial condition.

Remark 2 In the literature, the gradient method is used to design the primal dynamics $\dot{x} = -\nabla_x L(x, \lambda, v)$, and the projected gradient method is applied to design the dual dynamics $\dot{\lambda} = P_+[\nabla_\lambda L(x, \lambda, v)]$, $\dot{v} = \nabla_v L(x, \lambda, v)$. The λ -dynamics is obviously nonsmooth. To overcome the difficulty of the nonsmoothness in the dual dynamics, this paper proposes a mirror descent method, rather than the projected gradient method, to design a smooth λ -dynamics. This paper also extends the gradient x -dynamics to a mirror descent setup and the distributed framework.

4 Distributed mirror descent algorithm for constrained optimization

According to Theorems 1 and 2, the constrained optimization problem (3.1) can be transformed to solving the unconstrained optimization problem of minimizing the Lagrangian $L(x, \lambda, v)$ with respect to x and maximizing $L(x, \lambda, v)$ with respect to (λ, v) . Let us review the traditional way to tackle these minimization and maximization problems:

- Minimization of $L(x, \lambda, v)$ with respect to x can be realized by designing a dynamics following the gradient descent as $\dot{x} = -\nabla_x L(x, \lambda, v)$, extended in a distributed way by including a consensus term to the following form:

$$\dot{x}_i = -\nabla f_i(x_i) - \sum_j^{r_i} \lambda_{ij} \nabla g_{ij}(x_i) - \sum_j^{s_i} v_{ij} \nabla h_{ij}(x_i) + \sum_{j \in \mathcal{N}_i} (x_j - x_i). \tag{4.1}$$

- Likewise, maximization of $L(x, \lambda, v)$ with respect to (λ, v) can be achieved by resorting the projected gradient ascent method $\dot{\lambda} = P_+[\nabla_\lambda L(x, \lambda, v)]$, $\dot{v} = \nabla_v L(x, \lambda, v)$, or more specifically,

$$\dot{\lambda}_{ij} = P_+[g_{ij}(x_i)], \quad j = 1, \dots, r_i, \tag{4.2}$$

$$\dot{v}_{ij} = h_{ij}(x_i), \quad j = 1, \dots, s_i, \tag{4.3}$$

where the projection P_+ is used to keep positive $\lambda_{ij}(t)$ all the time.

In this section, we will borrow the mirror method to redesign the projected λ -dynamics in (4.2) such that the redesigned λ -dynamics is smooth and positive invariant with respect to $\Lambda = \{\lambda \mid \lambda_{ij} \geq 0, j = 1, \dots, r_i, i = 1, \dots, N\}$. More specifically, Sect. 4.1 is devoted to the general theory of continuous-time mirror descent, which is used to redesign the dual λ -dynamics in Sect. 4.2 and x -dynamics in Sect. 4.3.

4.1 General theory on mirror descent

The mirror descent algorithm is devoted to the constrained minimization problem

$$\min_{x \in \mathcal{X}} F(x), \tag{4.4}$$

where \mathcal{X} is a convex compact set in \mathbb{R}^n . To explain this algorithm, the following definitions are reviewed for later use.

Definition 4 (Distance-generating function [29]) A function $\phi : \mathcal{X} \rightarrow \mathbb{R}$ is called a distant-generating function modulus $\alpha > 0$ with respect to $\|\cdot\|$ if ϕ is convex and continuous on \mathcal{X} , the set $\mathcal{X}^\circ = \{x \in \mathcal{X} \mid \nabla\phi(x) \neq \emptyset\}$ is convex (note that \mathcal{X}° contains the relative interior of \mathcal{X}) and ϕ restricted to \mathcal{X}° is continuously differentiable and strongly convex with parameter α with respect to $\|\cdot\|$ in the sense that $(y - x)^T(\nabla\phi(y) - \nabla\phi(x)) \geq \alpha\|y - x\|^2$ for all $x, y \in \mathcal{X}^\circ$.

Definition 5 (Bregman divergence [38]) A function $B_\phi : \mathcal{X}^\circ \times \mathcal{X} \rightarrow \mathbb{R}_+$ defined by

$$B_\phi(x, y) = \phi(y) - \phi(x) - \nabla\phi(x)^T(y - x)$$

is called the Bregman divergence (or prox-function) associated with ϕ .

In what follows, we use the Bregman divergence associated with $\phi(x) = \sum_{i=1}^N x_i \log(x_i)$; in this case, we easily calculate

$$B_\phi(x, x') = \sum_{i=1}^N x_i \log \frac{x_i}{x'_i} + \sum_{i=1}^N (x'_i - x_i).$$

The definition of conjugate and its properties is very important in the analysis of mirror image descent.

Remark 3 Let $\mathcal{Z} = \{z \in \mathbb{R}^n \mid z = \nabla\phi(x), x \in \mathcal{X}\}$. We can define the so-called Fenchel coupling $F(x^*, z) = \phi(x^*) + \phi^*(z) - \langle z, x^* \rangle$ for $x^* \in \mathcal{X}$ and $z \in \mathcal{Z}$, which is nonnegative and strictly convex in both arguments.

Definition 6 (Legendre–Fenchel conjugate [39]) For a distance-generating function $\phi : \mathcal{X} \rightarrow \mathbb{R}$, its Legendre–Fenchel conjugate convex function ϕ^* is defined as $\phi^*(\omega) = \sup_{x \in \mathcal{X}} \{(x, \omega) - \phi(x)\}$, which can be shown to be strictly convex and twice differentiable.

In our notation, the general continuous-time mirror ascent algorithm for the constrained optimization problem $\max_{x \in \mathcal{X}} F(x)$ takes the form

$$\dot{z} = \nabla F(x), \tag{4.5}$$

Let x^* be the optimal solution of the optimization problem (3.1), and let λ_{ij}^* and v_{ij}^* be defined as in Theorem 2. Define $X = (x_1, \dots, x_N)^T$ and $X^* = (x^*, \dots, x^*)$. Then by Theorem 2 we can see that (X^*, λ^*, v^*) is the equilibrium of the dynamical system (4.8a)–(4.8d). Therefore, if we can prove the asymptotic stability of the equilibrium (X^*, λ^*, v^*) of the dynamical system (4.8a)–(4.8d), then $X(t) \xrightarrow{t \rightarrow \infty} X^*$, which implies that the states of all agents can estimate the optimal solution x^* consensually and asymptotically. The following theorem presents a convergence analysis for system (4.8a)–(4.8d).

Theorem 3 *For the constrained optimization problem (3.1), let the Slater’s constraint qualification certificate in Definition 3 be satisfied. Suppose there are N agents whose dynamics are given by (4.8a)–(4.8d) and connected by a fixed network. Then for any initial condition with $\lambda_{ij}(0) \geq 0$, we have $\lambda_{ij}(t) \geq 0$ and $\lim_{t \rightarrow \infty} \|x_i(t) - x^*\| = 0$.*

Proof We use the Lyapunov method to prove the stability. To this end, construct a Lyapunov candidate $V(X, \lambda, v) = V_1 + V_2 + V_3 + V_4$ as

$$\begin{aligned} V_1 &= \sum_{i=1}^N [\phi(x^*) + \phi^*(z_i) - \langle z_i, x^* \rangle], \quad z_i = \nabla \phi(x_i), \\ V_2 &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^{r_i} \alpha (\lambda_{ij} - \lambda_{ij}^*)^2, \\ V_3 &= \sum_{i=1}^N \sum_{j=1}^{r_i} \beta \left[(\lambda_{ij} - \lambda_{ij}^*) - \sum_{i,j \in \cup} \lambda_{ij}^* (\ln \lambda_{ij} - \ln \lambda_{ij}^*) \right], \\ V_4 &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^{s_i} (v_{ij} - v_{ij}^*)^2, \end{aligned}$$

where $\cup = \{i, j \mid \lambda_{ij}^* \neq 0\}$. According to the third equation in (4.8a)–(4.8d), we see that $\lambda_{ij}(t) \geq 0$ if $\lambda_{ij}(0) > 0$; this equation also implies $\dot{\lambda}_{ij} \leq 0$, which means that $\lambda_{ij}(t)$ is a decreasing function of t , and consequently $\lambda_{ij}(t) \geq \lambda_{ij}^* > 0$ for $i, j \in \cup$. Thus, V_3 is well defined. Further, a simple calculation shows that $V_3 \geq 0$. (In fact, V_3 is the Bregman divergence induced by $\phi(x) = \sum_i x_i \ln x_i$ with respect to norm $\|\cdot\|_1$.) Therefore, $V \geq 0$ and $V(X, \lambda, v) = 0$ if and only if $(X, \lambda, v) = (X^*, \lambda^*, v^*)$.

We now calculate the time derivative of the Lyapunov function V along the trajectories of system (4.8a)–(4.8d). For V_1 , a straightforward calculation shows that

$$\begin{aligned} \dot{V}_1 &= \sum_{i=1}^N [\nabla \phi^*(z_i) \dot{z}_i - \langle \dot{z}_i, x^* \rangle] \\ &= \sum_{i=1}^N [\langle \dot{z}_i, \nabla \phi^*(z_i) \rangle - \langle \dot{z}_i, x^* \rangle] \\ &= \sum_{i=1}^N [\langle \dot{z}_i, x_i - x^* \rangle] \\ &= \sum_{i=1}^N \left[\left\langle x_i - x^*, -\nabla f_i(x_i) + \sum_{j \in \mathcal{N}_i} (x_j - x_i) - \sum_{j=1}^{r_i} \lambda_{ij} \nabla g_{ij}(x_i) - \sum_{j=1}^{s_i} v_{ij} \nabla h_{ij}(x_i) \right\rangle \right]. \end{aligned}$$

Since $(y - x)^T \nabla f(x) \leq f(y) - f(x)$ and $\mathcal{L}X^* = 0$, we get

$$\begin{aligned} \dot{V}_1 &= \sum_{i=1}^N (x_i - x^*)^T (-\nabla f_i(x_i)) + \sum_{i=1}^N \left\langle x_i - x^*, \sum_{j \in \mathcal{N}_i} (x_j - x_i) \right\rangle \\ &\quad - \sum_{i=1}^N \sum_{j=1}^{r_i} \lambda_{ij} (x_i - x^*)^T \nabla g_{ij}(x_i) - \sum_{i=1}^N \sum_{j=1}^{s_i} v_{ij} (x_i - x^*)^T \nabla h_{ij}(x_i) \\ &\leq -(X - X^*)^T \mathcal{L}(X - X^*) + \sum_{i=1}^N f_i(x^*) - \sum_{i=1}^N f_i(x_i) \\ &\quad + \sum_{i=1}^N \sum_{j=1}^{r_i} \lambda_{ij} g_{ij}(x^*) - \sum_{i=1}^N \sum_{j=1}^{r_i} \lambda_{ij} g_{ij}(x_i) + \sum_{i=1}^N \sum_{j=1}^{s_i} v_{ij} h_{ij}(x^*) - \sum_{i=1}^N \sum_{j=1}^{s_i} v_{ij} h_{ij}(x_i). \end{aligned}$$

Furthermore,

$$\begin{aligned} \dot{V}_2 + \dot{V}_3 &= \sum_{i=1}^N \sum_{j=1}^{r_i} \frac{\alpha \lambda_{ij} (\lambda_{ij} - \lambda_{ij}^*)}{\beta + \alpha \lambda_{ij}} g_{ij}(x_i) + \sum_{i=1}^N \sum_{j=1}^{r_i} \frac{\beta \lambda_{ij}}{\beta + \alpha \lambda_{ij}} g_{ij}(x_i) - \sum_{i,j \in \cup} \frac{\beta \lambda_{ij}^*}{\beta + \alpha \lambda_{ij}} g_{ij}(x_i) \\ &= \sum_{i=1}^N \sum_{j=1}^{r_i} (\lambda_{ij} - \lambda_{ij}^*) g_{ij}(x_i), \\ \dot{V}_4 &= \sum_{i=1}^N \sum_{j=1}^{s_i} (v_{ij} - v_{ij}^*) h_{ij}(x_i). \end{aligned}$$

Combining these calculations together yields

$$\begin{aligned} \dot{V} &\leq \sum_{i=1}^N f_i(x^*) + \sum_{i=1}^N \sum_{j=1}^{r_i} \lambda_{ij} g_{ij}(x^*) + \sum_{i=1}^N \sum_{j=1}^{s_i} v_{ij} h_{ij}(x^*) \\ &\quad - \sum_{i=1}^N f_i(x_i) - \sum_{i=1}^N \sum_{j=1}^{r_i} \lambda_{ij}^* g_{ij}(x_i) - \sum_{i=1}^N \sum_{j=1}^{s_i} v_{ij}^* h_{ij}(x_i) \\ &\quad - (X - X^*)^T \mathcal{L}(X - X^*). \end{aligned}$$

According to definition (1), we now define another Lagrangian $\bar{L} : \mathbb{R}^{nN} \times \mathbb{R}_+^r \times \mathbb{R}^s \rightarrow \mathbb{R}$ as follows:

$$\bar{L}(X, \lambda, v) = \sum_{i=1}^N f_i(x_i) + \sum_{i=1}^N \sum_{j=1}^{r_i} \lambda_{ij} g_{ij}(x_i) + \sum_{i=1}^N \sum_{j=1}^{s_i} v_{ij} h_{ij}(x_i). \tag{4.9}$$

Then we get

$$\begin{aligned} \dot{V} &\leq \bar{L}(X^*, \lambda, v) - \bar{L}(X, \lambda^*, v^*) - (X - X^*)^T \mathcal{L}(X - X^*) \\ &\leq 0. \end{aligned}$$

We finally prove the asymptotic stability of the equilibrium. To this end, letting $\dot{V} = 0$ yields

$$\bar{L}(X^*, \lambda, \nu) - \bar{L}(X, \lambda^*, \nu^*) = 0, \tag{4.10}$$

$$(X - X^*)^T \mathcal{L}(X - X^*) = 0. \tag{4.11}$$

Since the graph is connected, it then follows from (4.11) that $X = X^*$. Also noting that $(X^*, (\lambda^*, \nu^*))$ is a saddle point of the Lagrangian \bar{L} , we have $\bar{L}(X^*, \lambda, \nu) = \bar{L}(X, \lambda^*, \nu^*) = \bar{L}(X^*, \lambda^*, \nu^*)$. This yields $\sum_{i=1}^N \sum_{j=1}^{r_i} (\lambda_{ij} - \lambda_{ij}^*) g_{ij}(x^*) = 0$. Since $g_{ij}(x^*) \leq 0$ and $\lambda_{ij} \geq \lambda_{ij}^*$, it follows that $(\lambda_{ij} - \lambda_{ij}^*) g_{ij}(x^*) = 0$. The KKT condition $\lambda_{ij}^* g_{ij}(x^*) = 0$ implies that, for all $\lambda_{ij} \in \{\lambda_{ij} \mid g_{ij}(x^*) < 0\}$, $\lambda_{ij} = 0$. Then from equation (4.8a) we have $\nabla f_i(x^*) + \sum_j^{r_i} \lambda_{ij} \nabla g_{ij}(x^*) + \sum_j^{s_i} \nu_{ij} \nabla h_{ij}(x^*) = 0$. Therefore the KKT condition is satisfied, and by uniqueness $\lambda_{ij} = \lambda_{ij}^*$, $\nu_{ij} = \nu_{ij}^*$. Therefore $(X, \lambda, \nu) = (X^*, \lambda^*, \nu^*)$. The application of the Lasalle invariance principle in [40] yields that (X^*, λ^*, ν^*) of system (4.8a)–(4.8d) is asymptotically stable almost surely. \square

5 Simulation

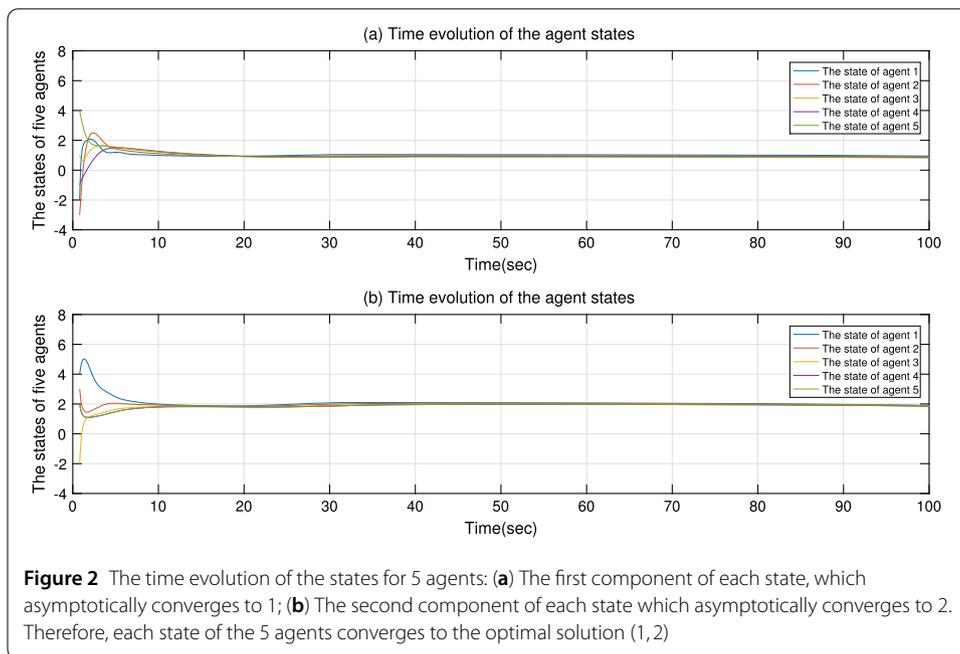
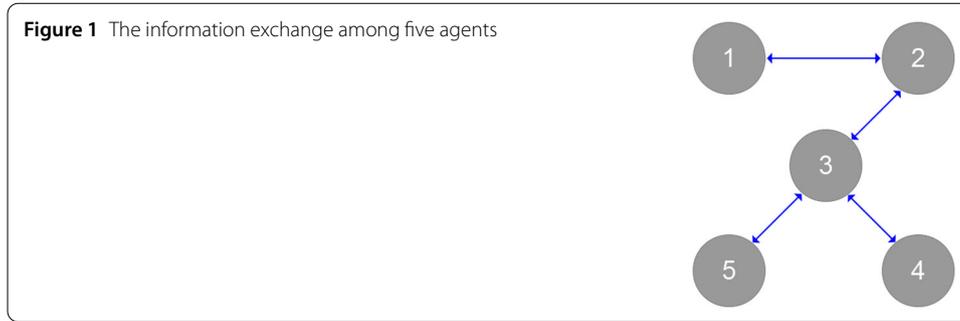
Consider the optimization problem (3.1) on a network with five agents. The five local cost functions for five agents are as follows: $f_1(x_1, x_2) = 4x_1^2 + 2x_2$, $f_2(x_1, x_2) = 2x^2$, $f_3(x_1, x_2) = 4x_1$, $f_4(x_1, x_2) = 2x_2$, $f_5(x_1, x_2) = 3x_1 + x_2$. Assume that agent 1 has both inequality and equality constraints with constraint functions $g_1(x_1, x_2) = (x_1 - 2)^2 - x_2 + 1$ and $h_1(x_1, x_2) = 2x_1 - x_2$. Agent 2 has the inequality constraint $g_2(x_1, x_2) = -x_1 + x_2 - 2$, whereas there are no constraints for agents 3, 4, 5. We can check that all functions mentioned are convex and the constrained set is nonempty. The true optimal solution and optimal value of this problem are $(x_1^*, x_2^*) = (1, 2)$ and $\tilde{f}(x_1^*, x_2^*) = 23$.

In the literature the method of projected gradient descent is widely used to solve this problem where the projected gradient is hard to compute. For example, if the gradient projection algorithm in [37] is used here, then we should solve the following iterative problem: $x_{k+1} = \Pi_{\mathcal{X}}(x_k - r_k f'(x_k))$, where $\Pi_{\mathcal{X}}(x)$ is the projection of a vector $x \in \mathbb{R}^n$ onto the convex compact set $\mathcal{X} \subset \mathbb{R}^n$ in the sense of $\Pi_{\mathcal{X}}(x) = \inf_{y \in \mathcal{X}} \|y - x\|$. In our case, $\mathcal{X} = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, h_i(x) = 0, i = 1, \dots, N\}$. Obviously, computation of $\Pi_{\mathcal{X}}(x_k - r_k f'(x_k))$ is a heavy job, and additional algorithms such as [11, 41] should be adopted. By contrast, our algorithm (4.8a)–(4.8d) avoids using projection and thus is easier to carry out.

Now we apply our distributed convex optimization algorithm (4.8a)–(4.8d) by using five agents that are connected as in Fig. 1 to find the optimal solution. The Laplacian matrix of this undirect graph is

$$\mathcal{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix}.$$

Set the initial values of the five agents states as $x_1(0) = (-2, 4)^T$, $x_2(0) = (-3, 3)^T$, $x_3(0) = (1, -2)^T$, $x_4(0) = (-1, 2)^T$, $x_5(0) = (4, 2)^T$, $\lambda_1(0) = 3$, $\lambda_2(0) = 3$, $\nu(0) = 3$, respectively. The time evolution of the states for five agents is illustrated in Fig. 2, where subfigure (a) shows



the first component of each state, which asymptotically converges to 1, and subfigure (b) shows the second component of each state, which asymptotically converges to 2. Therefore, each state of the five agents asymptotically converges to the optimal solution (1, 2).

Some comparisons of our method with existing algorithm are included here. For example, in the spirit of the primal-dual algorithm in [30], the dynamics for the Lagrange multiplier should be designed as $\dot{\lambda} = \Gamma P_+[-\frac{\partial L}{\partial \lambda}] = \Gamma P_+[g(x)]$ with any positive definite matrix Γ . To make the value of multiplier corresponding to the inequality constraint stay positive all the time, the projection $P_+[g(x)]$ is also used, and thus the corresponding dynamics is nonsmooth. This obviously causes difficulties in simulation and convergence analysis. In contrast, our algorithm uses equations (4.8c)–(4.8d) for the evolution of the Lagrange multipliers, which are smooth dynamics and easy to simulate.

6 Conclusion

When considering distributed convex optimization using multiagent systems, a consensus-based distributed method is usually adopted. It is well known that, for a distributed optimization problem, it is a challenging problem if the constraint is taken into account. One popular method to deal with constraints is based on the projection, which causes the optimization dynamics to be nonsmooth. To overcome this difficulty, we propose a novel

distributed convex optimization algorithm utilizing the mirror descent method. Although there are reported results in the field of the mirror descent for the distributed optimization problem, most of them are presented in the discrete-time form. For the results on continuous-time mirror descent, they work only for unconstrained optimization. Our algorithm is valid for the more challenging case of constrained optimization. The superiority of our method in comparison with existing results lies in the fact that our method avoids using gradient projection in the optimization algorithm design. Therefore, it removes the difficulty of analyzing the resulting nonsmooth optimization dynamics and makes the simulation easier. By the aid of mirror descent, in our paper, we modify the frequently used primal-dual algorithm for the optimization problem, giving rise to new primal and dual dynamics. The modified primal dynamics facilitates more convenient use of Bregman divergence and Fenchel coupling in the stability analysis, and the redesigned dual dynamics for the evolution of positive Lagrange multiplier does not include projection and therefore reduces the complexity of convergence analysis.

Note that avoiding projection (and therefore avoiding using nonsmooth optimization dynamics) is a key in our paper. This forms a sharp contrast to the nonsmooth optimization problem as in [9]. A direct connection and adoption of our method to the nonsmooth case as in [9] however needs further investigation.

Acknowledgements

The authors would like to thank the referees for their careful reviews.

Funding

This work is supported by the NNSF of China under the grants 61663026, 61473098, 61563033, 11361043, 61603175.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Both authors contributed equally to the writing of this paper. Both authors of the manuscript have read and agreed to its content and are accountable for all aspects of the accuracy and integrity of the manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 26 January 2018 Accepted: 24 July 2018 Published online: 18 October 2018

References

1. Baleanu, D., Inc, M., Yusuf, A., Aliyu, A.I.: Time fractional third-order evolution equation: symmetry analysis, explicit solutions, and conservation laws. *J. Comput. Nonlinear Dyn.* **13**, 021011 (2017)
2. Baleanu, D., Jajarmi, A., Hajipour, M.: A new formulation of the fractional optimal control problems involving Mittag-Leffler nonsingular kernel. *J. Optim. Theory Appl.* **175**, 718–737 (2017)
3. Baleanu, D., Inc, M., Yusuf, A., Aliyu, A.: Lie symmetry analysis, exact solutions and conservation laws for the time fractional Caudrey–Dodd–Gibbon–Sawada–Kotera equation. *Commun. Nonlinear Sci. Numer. Simul.* **59**, 222–234 (2017)
4. Bai, Y., Baleanu, D., Wu, G.C.: Existence and discrete approximation for optimization problems governed by fractional differential equations. *Commun. Nonlinear Sci. Numer. Simul.* **59**, 338–348 (2018)
5. Farnad, B., Jafarian, A., Baleanu, D.: A new hybrid algorithm for continuous optimization problem. *Appl. Math. Model.* **55**, 652–673 (2018)
6. Hajipour, A., Malek, A.: High accurate modified Weno method for the solution of Black–Scholes equation. *Comput. Appl. Math.* **34**, 125–140 (2015)
7. Jajarmi, A., Hajipour, M.: An efficient finite difference method for the time-delay optimal control problems with time-varying delay. *Asian J. Control* **19**, 554–563 (2017)
8. Razminia, A., Baleanu, D., Majd, V.: Conditional optimization problems: fractional order case. *J. Optim. Theory Appl.* **156**, 45–55 (2013)
9. Vaziri, A., Kamyad, A., Jajarmi, A., Effati, S.: A global linearization approach to solve nonlinear nonsmooth constrained programming problems. *Comput. Appl. Math.* **30**, 427–443 (2011)
10. Inc, M., Yusuf, A., Aliyu, A.I., Baleanu, D.: Lie symmetry analysis, explicit solutions and conservation laws for the space–time fractional nonlinear evolution equations. *Phys. A, Stat. Mech. Appl.* **496**, 371–383 (2018)

11. Nedic, A., Ozdaglar, A.: Distributed subgradient method for multi-agent optimization. *IEEE Trans. Autom. Control* **54**, 48–61 (2009)
12. Yi, P., Hong, Y.: Quantized subgradient algorithm and data-rate analysis for distributed optimization. *IEEE Trans. Control Netw. Syst.* **1**, 380–392 (2014)
13. Duchi, J.C., Agarwal, A., Wainwright, M.J.: Dual averaging for distributed optimization: convergence analysis and network scaling. *IEEE Trans. Autom. Control* **57**, 592–606 (2012)
14. Deng, Z., Hong, Y., Wang, X.: Distributed optimisation design with triggers for disturbed continuous-time multi-agent systems. *IET Control Theory Appl.* **11**, 282–290 (2017)
15. Zhu, M., Martinez, S.: On distributed convex optimization under inequality and equality constraints. *IEEE Trans. Autom. Control* **57**, 151–164 (2012)
16. Yi, P., Hong, Y., Liu, F.: Distributed gradient algorithm for constrained optimization with application to load sharing in power systems. *Syst. Control Lett.* **83**, 45–52 (2015)
17. Zeng, X., Yi, P., Hong, Y.: Distributed continuous time algorithm for constrained convex optimizations via nonsmooth analysis approach. *IEEE Trans. Autom. Control* **62**, 5227–5233 (2016)
18. Lou, Y., Hong, Y., Wang, S.: Distributed continuous-time approximate projection protocols for shortest distance optimization problems. *Automatica* **69**, 289–297 (2016)
19. Calafiore, G., Carlone, L., Wei, M.: A distributed gradient method for localization of formations using relative range measurements. In: *IEEE Int. Symp. on Computer-Aided Control System Design*, pp. 1146–1151 (2010)
20. Chen, J.S., Sayed, A.H.: Diffusion adaptation strategies for distributed optimization and learning over network. *IEEE Trans. Signal Process.* **60**, 4289–4305 (2012)
21. Neglia, G., Reina, G., Alouf, S.: Distributed gradient optimization for epidemic routing: a preliminary evaluation. In: *IFIP Conference on Wireless Days*, pp. 1–6 (2009)
22. Ram, S., Nedic, A.: Distributed subgradient method for multi-agent optimization. *J. Optim. Theory Appl.* **147**, 516–545 (2010)
23. Nemirovski, A.S., Yudin, D.B.: *Problem Complexity and Method Efficiency in Optimization*. Wiley, New York (1983)
24. Beck, A., Teboulle, M.: Mirror descent and nonlinear projected subgradient methods for convex optimization. *Oper. Res. Lett.* **13**, 167–175 (2003)
25. Ben-Tal, A., Margalit, T., Nemirovski, A.: The ordered subsets mirror descent optimization method with applications to tomography. *SIAM J. Optim.* **12**(1), 79–108 (2006)
26. Li, J.Y., Chen, G.: Distributed mirror descent method for multi-agent optimization with delay. *Neurocomputing* **177**, 643–650 (2016)
27. Wibisono, A., Wilson, A.C.: On accelerated methods in optimization. <http://arXiv.org/abs/1509.03616> (2015)
28. Xi, C., Wu, Q., Khan, U.A.: Distributed mirror descent over directed graphs. <http://arXiv.org/abs/1412.5526> (2014)
29. Raginsky, M., Bouvrie, J.: Continuous-time stochastic mirror descent on a network: variance reduction, consensus, convergence. In: *Decision and Control*, pp. 6793–6800 (2012)
30. Feijer, P.F.D.: Stability of primal-dual gradient dynamics and applications to network optimization. *Automatica* **46**, 1974–1981 (2010)
31. Tanabe, K.: A geometric method in nonlinear programming. *J. Optim. Theory Appl.* **30**, 181–210 (1980)
32. Wang, J., Elia, N.: A control perspective for centralized and distributed convex optimization. In: *Decision and Control and European Control Conference*, pp. 3800–3805 (2011)
33. Ghahesifard, B., Cortes, J.: Distributed continuous time convex optimization on weighted balanced digraphs. *IEEE Trans. Autom. Control* **59**, 781–786 (2014)
34. Towfic, Z.J., Sayed, A.: Adaptive penalty-based distributed stochastic convex optimization. *IEEE Trans. Signal Process.* **62**, 3924–3938 (2014)
35. Arrow, K.J., Hurwicz, L., Uzawa, H.: *Studies in Linear and Nonlinear Programming*. Mathematical Studies in the Social Sciences (1958)
36. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
37. Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific (1999)
38. Brègma, L.M.: Relaxation method for finding a common point of convex sets and its application to optimization problems. *Comput. Math. Math. Phys.* **48**(2), 1019–1022 (1966)
39. Hiriart-Urruty, J.B., Lemaréchal, C.: *Fundamentals of Convex Analysis*. Springer, Berlin (2001)
40. Mao, X.: Stochastic versions of the LaSalle theorem. *J. Differ. Equ.* **153**, 175–195 (1999)
41. Nedic, A., Ozdaglar, A., Parrilo, P.: Constrained consensus and optimization in multi-agent network. *IEEE Trans. Autom. Control* **55**, 922–938 (2010)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
