Distributed convex optimization via proportional-integral-differential algorithm

Wei Zhu and Haibao Tian

Abstract
This paper studies the distributed convex optimization problem, where the global utility function is the sum of local cost functions associated to the individual agents. Only using the local information, a novel continuous-time distributed algorithm based on proportional-integral-differential (PID) control strategy is proposed. Under the assumption that the global utility function is strictly convex and local utility functions have locally Lipschitz gradients, the exponential convergence of the proposed algorithm is established with undirected and connected graph among these agents. Finally, numerical simulations are presented to illustrate the effectiveness of theoretical results.

Keywords
Distributed, convex optimization, proportional-integral-differential, exponential convergence

Introduction
Recent years have witnessed an increasing interest in distributed optimization and its widely applications in various fields, including energy internet, intelligent manufacturing and machine learning. The aim of distributed optimization problem is to minimize global objective function with a sum of individual objective function, which computes and exchanges information among neighboring agents. The research on distributed optimization can be traced back to the pioneering work and then many valuable algorithms for this topic have been proposed, which can mainly be classified as discrete-time algorithm and continuous-time algorithm.

The discrete-time algorithm has attracted much attention since the distributed gradient descent (DGD) algorithm was proposed by Nedić and Ozdaglar. But, due to the diminishing step-size, the convergence rate of DGD algorithm was slow. Then, Shi et al. proposed an exact first-order algorithm (EXTRA) with fixed step-size, which converges to the optimal solution precisely. In order to eliminate the requirement of the double random weight matrix, the push-sum algorithm under the directed graph was proposed by Nedić and Olshevsky. And then, the combination of push-sum protocol and distributed subgradient algorithm were proposed for unconstrained distributed optimization problem in the directed time-varying topology. By introducing projection operator, various distributed projection algorithms for constrained distributed optimization problem were proposed in Lou et al., Lin et al., Liu et al., Wang et al., and Liang et al. To deal with equality and inequality constraints, Chang et al. proposed the consensus-based distributed primal-dual perturbation (PDP) algorithm, which was applied to smooth inequality constraints and non-smooth constraints. For the case of non-smooth optimization problems, a smooth double proximal primal-dual algorithm was present in Wei et al. Moreover, without convexity assumptions on the objectives, a chebyshev proxy and consensus-based algorithm (CPCA) was given in He et al., which had low computational costs.

As a counterpart, much attention has also been paid to the continuous-time algorithm. Several distributed optimization algorithms with the single integrator dynamics were proposed in Lin et al., which considered state-dependent gradient gains and nonuniform gradient gains. Considering second-order multi-agent...
networks, a double-integrator distributed optimization algorithm under the undirected graph was presented in Zhang and Hong.\textsuperscript{23} Based on the Karush-Kuhn-Tucker condition and the saddle point property, a new distributed algorithm for constrained optimization was developed in Yi et al.\textsuperscript{24} With impulsive communication framework, an average quasi-consensus algorithm for distributed constrained optimization was designed in He et al.\textsuperscript{25} In order to reduce communication costs and energy consumption, the event-triggered control idea was introduced to design the distributed optimization algorithm. In Chen and Ren,\textsuperscript{26} the event-triggered zero-gradient-sum algorithm was established. In Li et al.,\textsuperscript{27} distributed optimization algorithm with event-triggered communication via input feedforward passivity was presented, and the event-triggered distributed fixed-time optimization algorithm over a directed network was developed in Yu et al.\textsuperscript{28}

Due to the progressive control techniques, various of algorithms based on proportional-integral (PI) control strategy for distributed optimization have attracted an increasing attention recently. Wang and Elia\textsuperscript{29} proposed a distributed optimization algorithm based on PI control strategy, where each agent used an auxiliary state to correct the error caused by different local gradient. Based on PI control strategy, a continuous-time coordination algorithm with discrete-time communication was investigated in Kia et al.,\textsuperscript{30} which can reduce communication costs and energy consumption. Combining projection output feedback with PI protocol, the distributed convex optimization subject to general constraints was studied in Yang et al.\textsuperscript{31} It is well known that proportional-integral-derivative (PID) control strategy is a classic control approach, which has simple structures, clear physical meaning of parameters, and strong robustness. Thus, it has been applied in various fields, such as, active disturbance rejection control (ADRC),\textsuperscript{32} formation control,\textsuperscript{33} and so on. But, to the best knowledge of ours, the PID control strategy for distributed optimization problem has not been discussed.

Based on the above discussions, a new continuous-time algorithm with PID control strategy for distributed optimization under an undirected communication graph is studied in this paper. The main contributions of this paper are as follows. (I) The PID algorithm is firstly presented for distributed optimization, where each agent uses only local information, that is, the proposed algorithm is fully distributed. (II) Using the matrix transform and the inequality technique, the exponential convergence and the estimation for convergence rate of the proposed algorithm are established under the assumption that the global utility functions are strictly convex and local utility functions have locally Lipschitz gradients. (III) The constrictions on the initial value is removed, which has less conservation than that in Zhang and Hong\textsuperscript{23} and Kia et al.\textsuperscript{30}

The remainder of this paper is organized as follows. Section 2 is devoted to mathematical preliminaries on distributed optimization problem. The continuous-time distributed PID algorithm and its convergence are presented in Section 3. Numerical examples are given to illustrate the performance of the proposed algorithms in Section 4. Finally, conclusions are drawn in Section 5.

**Notations:** Let $R^n$ be the n-dimensional real vector space. $||.||$ denotes the Euclidean norm or induced matrix norm for a vector or a matrix. $\otimes$ is the matrix Kronecker product. $I_n$ is the identity matrix in $R^{n \times n}$. $0_N$ and $1_N$ represent the column vectors with $N$ zeros and ones, respectively. $\det(\cdot)$ is the determinant of a matrix.

**Preliminaries**

**Graph theory**

A weighted undirected graph is a triplet $G = (\mathcal{V}, \mathcal{E}, A)$, where $\mathcal{V} = \{v_1, \ldots, v_N\}$ is a vertex set, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is an undirected edge set, and $A = \{a_{ij}\}_{N \times N}$ is an adjacency matrix with nonnegative elements $a_{ij}$, where $a_{ij} > 0$ if and only if $(v_i, v_j) \in \mathcal{E}$, otherwise $a_{ij} = 0$. We assume that there is no self-loops, that is, $a_{ii} = 0$ for all $i \in \mathcal{V}$. The degree of node $i$ is defined by $d_i = \sum_{j=1}^{N} a_{ij}$. The Laplacian of graph $G$ is defined as $L_N = D_{N} - A$, where $D_{N} = \text{diag}(d_1, \ldots, d_N)$.

An undirected edge $e_{ij}$ in graph $G$ is denoted by the unordered pair of nodes $(v_i, v_j)$, which means that nodes $v_i$ and $v_j$ can exchange information with each other and $A$ is a symmetric matrix. A path between nodes $v_i$ and $v_j$ in graph $G$ is a sequence of edges $(v_i, v_{j1}), (v_{j1}, v_{j2}), \ldots, (v_{j_{k-1}}, v_{j_k})$ and $v_{j_k} \in \mathcal{V}$. An undirected graph $G$ is connected if there exists a path between any pair of distinct nodes $v_i$ and $v_j$ $(i, j = 1, 2, \ldots, N)$, $L_N$ has a simple zero eigenvalue and all the other eigenvalues are positive if and only if the graph $G$ is connected, and the eigenvalues of $L_N$ can be ordered as $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_N$.

**Problem formulation**

Consider a network of $N$ agents and each agent has a local private differentiable function $f_i(x): R^n \rightarrow R$. The communication topology among agents is described by an undirected graph $G$. The following convex optimization problem will be studied.

$$\min_{x \in R^n} f(x) = \sum_{i=1}^{N} f_i(x).$$ (1)

The global utility function $f(x) = \sum_{i=1}^{N} f_i(x)$ is assumed to be strictly convex, that is,
\[(x - y)^T(\nabla f(x) - \nabla f(y)) > 0, \forall x, y \in R^n \text{ and } x \neq y. \]  

Moreover, a function \( f(x): R^n \rightarrow R \) is Lipschitz with constant \( m > 0 \) if
\[
\|f(x) - f(y)\| \leq m\|x - y\|, \forall x, y \in R^n, \tag{3}
\]
and for each \( i \in V \), the gradient of \( f^i(x) \) is \( l^i \)-Lipschitz if
\[
\|\nabla f^i(x) - \nabla f^i(y)\| \leq l_i \|x - y\|, \forall x, y \in R^n. \tag{4}
\]

In this paper, we focus on designing a distributed proportional-integral-differential (PID) algorithm such that each agent obtains the global minimizer \( x^* \in (-\infty, \infty) \) of the feasible optimization problem
\[
x^* = \arg \min_{x \in R^n} f(x) \tag{5}
\]
only using its own and its neighbors’ information.

**Distributed PID algorithm**

In this section, we propose a distributed PID algorithm and obtain the convergence result.

For each agent \( i \in \{1, \ldots, N\} \), the distributed PID algorithm for the problem (1) is described as follows.

\[
\dot{x}_i = v_i,
\]
\[
\dot{v}_i = -a_i \sum_{j=1}^{N} a_{ij}(x_i - x_j) - b_i \sum_{j=1}^{N} a_{ij}(v_i - v_j) - c_i \sum_{j=1}^{N} a_{ij}(y_i - y_j) - k_i \nabla f^i(x_i),
\]
\[
\dot{y}_i = a^2_i k_i \sum_{j=1}^{N} a_{ij}(x_i - x_j),
\]
where \( x_i, v_i, y_i \in R^n \) represent position, velocity, and the auxiliary state of the \( i \)-th agent, respectively. \( x_i \) is a solution to problem (1) for agent \( i \). Positive constants \( a_i, b_i, c_i > 0 \) are the proportional parameter, differential parameter, integral parameter, respectively. \( k_i > 0 \) is the constant stepsize.

Denote \( f(x) = \sum_{i=1}^{N} f^i(x_i), \quad x = ((x_1)^T, \ldots, (x_N)^T)^T \in R^{Nn}, \quad v = ((v_1)^T, \ldots, (v_N)^T)^T \in R^{Nn}, \quad y = ((y_1)^T, \ldots, (y_N)^T)^T \in R^{Nn} \). Then, the algorithm (6) can be rewritten as the following compact form
\[
\begin{align*}
\dot{x} &= v \\
\dot{v} &= -aLx - bLv - cLy - k\nabla f(x), \\
\dot{y} &= a^2kLx,
\end{align*}
\]
where \( L = L_N \otimes I_n \).

**Lemma 1.** Suppose the communication topology \( G \) is undirected and connected, the equilibrium point of system (7) is the optimal solution of problem (1).

**Proof.** Since the communication graph is undirected and connected, we have \( L_N^T L_N = 0 \).

Assume that \( (x^*, v^*, y^*) \) is the equilibrium point of system (7), we have
\[
\begin{align*}
v^* &= 0_N, \\
-ax^* - bLv^* - cy^* - k\nabla f(x^*) &= 0_N, \\
a^2kLx^* &= 0_N.
\end{align*}
\]

Since \( G \) is connected, there is a \( x^* \in R^n \) such that
\[
x^* = 1_N \otimes x^*.
\]

Next, it will be shown that \( x^* \) is the optimal solution of problem (1). Multiply \((1_N \otimes I_n)^T\) on the both sides of the second equation in (8), one can derive that
\[
(1_N \otimes I_N)^T \nabla f(x^*) = 0,
\]
and then
\[
\sum_{i=1}^{N} \nabla f^i(x^*) = 0,
\]
which implies \( x^* \) is the optimal solution of problem (1). The proof is completed.

**Theorem 1.** Assume that the communication topology \( G \) is connected and the global utility function \( f(x) \) is strictly convex. The local utility function is continuously differentiable and its gradient is \( l^i \)-Lipschitz on \( R^n \). Then, with the PID algorithm (6), where \( a_i, c_i > 0, b_i > a_i \), and \( k_i > 0 \) is bounded, \( x^*(t) \) exponentially converges to the global minimizer \( x^* \) for each \( i \in V \) with any initial values.

**Proof.** Without loss of generality, for simplicity, let \( n = 1 \). The case for \( n > 1 \) can also be proved by the property of Kronecker product only with complex computation.

Let \( \tilde{x} = x - x^*, \tilde{v} = v - v^*, \tilde{y} = y - y^* \). Then, (7) can be rewritten in the following matrix form
\[
\begin{align*}
\dot{X}(t) &= AX(t) - G_1(t),
\end{align*}
\]
where
\[
X = \begin{pmatrix}
x \\ v \\ y
\end{pmatrix}, \quad A = \begin{bmatrix}
0_N & I_N & 0_N \\
-aLN & -bLN & -CLN \\
a^2kLN & 0_N & 0_N
\end{bmatrix},
\]
\[
G_1(t) = \begin{bmatrix}
0_N \\
-k \cdot g(t) \\
0_N
\end{bmatrix}, \quad g(t) = \nabla f(x + x^*) - \nabla f(x^*).
\]

Denote \( E = [-1_{N-1} I_{N-1}] \) and \( M_1 = \text{diag}\{EEE\} \). Let \( Y(t) = M_1 X(t) \), in term of (9), we have...
\[ \hat{Y}(t) = BY(t) - G_2(t), \quad (10) \]

where

\[
B = \begin{bmatrix}
0_{N-1} & I_{N-1} & 0_{N-1} \\
-aL_{N-1} & -bL_{N-1} & -cL_{N-1} \\
-a^2kL_{N-1} & 0_{N-1} & 0_{N-1}
\end{bmatrix},
\]

\[
L_{22} = \begin{bmatrix}
d_2 & -a_{23} & \cdots & -a_{2N} \\
-a_{32} & d_3 & \cdots & -a_{3N} \\
\vdots & \vdots & \ddots & \vdots \\
-a_{N2} & -a_{N3} & \cdots & d_N
\end{bmatrix},
\]

\[
G_2(t) = \begin{bmatrix}
0_{N-1} \\
k \cdot E \cdot g(t) \\
0_{N-1}
\end{bmatrix}.
\]

If \( \theta = 0 \), that is, 0 is one of the eigenvalues of \( M \). According to the Laplace theorem for a partitioned matrix, by (12), we have \( \text{det}(J) = 0 \), which contradicts with that all the eigenvalues of \( J \) are positive. Thus, recall (12), we have

\[
\left| \theta^2 I_{N-1} + b\theta J + aJ + \frac{a^2ck}{\theta} J^2 \right| \neq 0.
\]

It follows that

\[
\left| \theta^2 I_{N-1} + b\theta J + aJ + \frac{a^2ck}{\theta} J^2 \right| = 0,
\]

which is a contradiction. Thus, there is at least a \( i \in \{2, \cdots, N\} \) such that

\[
\theta^3 + b\theta^2\lambda_i + a\theta\lambda_i + a^2ck\lambda_i^2 = 0. \quad (13)
\]

Since the communication topology \( \mathcal{G} \) is connected, it is clear that all the eigenvalues of \( L_{N-1} \), that is, \( \lambda_2, \lambda_3, \ldots, \lambda_N \) are positive. Moreover, there is an invertible matrix \( S \) such that \( S^{-1}L_{N-1}S = J \). Thus, all the eigenvalues of \( J \) are positive.

Let \( M_2 = \text{diag}\{S^{-1}S^{-1}\} \) and \( Z(t) = M_2 Y(t) \). Then, it follows from (10) that

\[ \dot{Z}(t) = MZ(t) - G(t), \quad (11) \]

where

\[
M = \begin{bmatrix}
0_{N-1} & I_{N-1} & 0_{N-1} \\
-aJ & -bJ & -cJ \\
-a^2kJ & 0_{N-1} & 0_{N-1}
\end{bmatrix},
\]

\[
G(t) = \begin{bmatrix}
0_{N-1} \\
k \cdot S^{-1} \cdot E \cdot g(t) \\
0_{N-1}
\end{bmatrix}.
\]

Assume \( \theta \) is an eigenvalue of \( M \). Then

\[
\text{det}(\theta I_{N-1} - M) = \begin{vmatrix}
\theta I_{N-1} - aJ & \theta I_{N-1} + bJ & cJ \\
-a^2kJ & 0_{N-1} & \theta I_{N-1} \\
\end{vmatrix} = 0,
\]

\[
\left| \theta^2 I_{N-1} + b\theta J + aJ + \frac{a^2ck}{\theta} J^2 \right| \neq 0.
\]

Denote \( \theta = \theta_1 + \sqrt{-1}\theta_2 \), where \( \theta_1, \theta_2 \in R \). Then

\[
\begin{cases}
\theta_1^3 + 3\theta_1\theta_2^2 + b\lambda_i(\theta_1^2 - \theta_2^2) + a\lambda_i\theta_1 + a^2ck\lambda_i^2 = 0 \\
-\theta_2^3 + (3\theta_1^2 + 2b\lambda_i\theta_1 + a\lambda_i)\theta_2 = 0
\end{cases} \quad (14)
\]

Case 1: \( \theta_2 = 0 \), which implies that

\[ \theta_1^3 + b\lambda_i\theta_1^2 + a\lambda_i\theta_1 + a^2ck\lambda_i^2 = 0. \quad (15) \]

Thus, by the Routh-Hurwitz theorem, \( \theta_1 < 0 \).

Case 2: \( \theta_2 \neq 0 \). It follows from (14) that

\[ 8\theta_1^3 + 8b\lambda_i\theta_1^2 + (2a\lambda_i + 2b^2\lambda_i^2)\theta_1 + ab\lambda_i^2 - a^2ck\lambda_i^2 = 0. \quad (16) \]
Since \( b > ack \), by the Routh-Hurwitz theorem, \( \theta_1 < 0 \).

Thus, all the eigenvalues of \( M \) are located at the open left half plane. Then, it is clear that there exist constants \( \mu > 0 \) and \( \alpha > 0 \) such that
\[
\| e^{M(t-t_0)} \| \leq \mu e^{-\alpha(t-t_0)}, \quad t \geq t_0. \tag{17}
\]

Noticing (11), by the variation of parameter formula, we have
\[
Z(t) = e^{M(t-t_0)}Z(t_0) + \int_{t_0}^{t} e^{M(t-s)}G(s)ds,
\]
and then
\[
\| Z(t) \| \leq \| e^{M(t-t_0)} \| \cdot \| Z(t_0) \| + \int_{t_0}^{t} \| e^{M(t-s)} \| \| G(s) \| ds. \tag{18}
\]

According to (4), it can be obtained that
\[
\| G(s) \| = \| -k \cdot S^{-1} \cdot E \cdot (\nabla \tilde{\mathbf{x}} + \mathbf{x}^*) - \nabla (\mathbf{x}^*) \| \\
\leq kl \| S^{-1} \cdot E \| \cdot \| \tilde{\mathbf{x}}(s) \|,
\]
where \( l \) is the maximum value of \( l(i = 1, \ldots, N) \). By (17–19), it follows that
\[
\| Z(t) \| \leq \mu e^{-\alpha(t-t_0)} \| Z(t_0) \| + \mu kl \| S^{-1} \cdot E \| \\
\cdot \int_{t_0}^{t} e^{-\alpha(t-s)} \| \tilde{\mathbf{x}}(s) \| ds \\
\leq \mu e^{-\alpha(t-t_0)} \| Z(t_0) \| + \mu kl \| S^{-1} \cdot E \| \\
\cdot \| E^{-1} \cdot S \| \int_{t_0}^{t} e^{-\alpha(t-s)} \| Z(s) \| ds \\
= \mu e^{-\alpha(t-t_0)} \| Z(t_0) \| + \mu klq \int_{t_0}^{t} e^{-\alpha(t-s)} \| Z(s) \| ds,
\]
where \( q = \| S^{-1} \cdot E \| \cdot \| E^{-1} \cdot S \|. \)

Now, for any \( m > 1 \) and \( \gamma \in (0, \alpha) \), it will be proved that
\[
\| Z(t) \| < m\mu \| Z(t_0) \| e^{-\gamma(t-t_0)} = h(t), \quad t \geq t_0. \tag{21}
\]
If (21) does not hold, by continuity, there must exist a \( t^* > t_0 \) such that \( \| Z(t^*) \| = h(t^*) \) and \( \| Z(t) \| < h(t) \) for \( t \in [t_0, t^*) \). Then, by (20), we can obtain

\[
h(t^*) = \| Z(t^*) \| \leq m\mu \| Z(t_0) \| e^{-\gamma(t-t_0)} < h(t),
\]
which implies that \( Z(t) \) can exponentially converge to 0, that is, \( \mathbf{x}'(t) \to \mathbf{x}^* \) as \( t \to \infty \), for all \( i \in \{1, \ldots, N\} \), and the convergence rate is estimated by \( \gamma \).

The proof is completed.

**Numerical simulations**

In the following, the proposed PID distributed algorithm is illustrated by some numerical simulations. We assume that there are six agents in the undirected and connected communication graph, which is described as Figure 1.

Case 1: All local cost functions are convex.
The trajectory of velocity in Case 1 with algorithm (6).

Figure 3.

The trajectory of state in Case 1 with algorithm (6).

Figure 2.

The trajectory of cost function in Case 1 with algorithm (6).

Figure 4.

The trajectory of state in Case 1 with PI algorithm.

Figure 5.

Assume that \( f(x) = x^T P_k x + b_k x + c_k, i = 1, 2, \ldots, 6, \) \( x \in \mathbb{R}^2, \) where

\[
P_1 = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 20 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 4 & 0.2 \\ 0.2 & 4 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 0.3 & 1 \\ 1 & 2 \end{bmatrix},
\]

\[
P_4 = \begin{bmatrix} 5 & 0.1 \\ 0.1 & 20 \end{bmatrix}, \quad P_5 = \begin{bmatrix} 0.7 & 1 \\ 1 & 0.2 \end{bmatrix}, \quad P_6 = \begin{bmatrix} 9 & 0.1 \\ 0.1 & 0.1 \end{bmatrix},
\]

\[b_1 = [1, 8], b_2 = [10, 1], b_3 = [3, 1], b_4 = [50, 12],
\]

\[b_5 = [4, 2], \quad b_6 = [70, 13], \quad c_1 = 1, c_2 = 0.2, \]

\[c_3 = 3, c_4 = 0.4, c_5 = 5, c_6 = 6.\]

The optimal solution of \( x^* = \arg \min_{x \in \mathbb{R}^2} f(x), \) where \( f(x) = \sum_{i=1}^{6} f_i(x) \) can be given explicitly by

\[x^* = -\frac{1}{b_k} \left( \sum_{i=1}^{6} P_i \right)^{-1} \left( \sum_{i=1}^{6} b_i \right)^T = [-3.5645 - 0.2246]^T.\]

Let \( a = 1, b = 15, c = 1, k = 1. \) Based on Theorem 1, the state \( x^i \) of each agent \( i (i = 1, \ldots, 6) \) will converge to the global optimal solution of the objective function

\[f(x) = \sum_{i=1}^{6} f_i(x) \]

for any initial values. The numerical simulation results using distributed PID algorithm (6) are depicted in Figures 2 to 4, which show that \( x^i \rightarrow [-3.5645 - 0.2246]^T, \) \( x^i \rightarrow 0, \) and the global cost function converges to the optimal value \(-234.5056.\)

In order to show the advantage of PID algorithm, the numerical simulation based on PI algorithm, that is, let \( b = 0 \) in (6), is shown in Figure 5, which shows that the state \( x^i \) of each agent \( i (i = 1, \ldots, 6) \) can not converges to the global optimal solution.

Case 2: Some of the local cost functions are non-convex.

Assume that \( f_1(x) = \sin(x - \frac{\pi}{3}), f_2(x) = -\cos(x - 2) - 0.2, f_3(x) = (x + 0.3)^2 + e^{0.3x}, f_4(x) = 0.2e^{-0.3x} + 0.4e^{0.4x}, f_5(x) = x^2(x^2 + 0.5)^2, f_6(x) = \frac{x^2}{x^2 + 6}, x \in \mathbb{R}, \) and \( f(x) = \sum_{i=1}^{6} f_i(x), \) which was discussed in Wu et al. Note that the local functions of agents 1 and 2 are non-convex. It can be verified the strict convexity of the global cost function, which ensures the uniqueness of the optimal solution. Assume that the initial values are randomly selected. Then, the simulation results are presented in Figures 6 to 8 with \( a = 1, b = 15, c = 1, k = 1, \) which show that...
\( x^i \rightarrow -0.14272, \ x^i \rightarrow 0, \) and the global cost function converges to the optimal value 1.7159.

**Conclusions**

In this paper, a continuous-time PID algorithm is proposed for solving the distributed optimization problem, where the global utility function is strictly convex and local utility functions have locally Lipschitz gradients. Using matrix transform and inequality technique, the exponential convergence of the proposed algorithm is established with undirected and connected graph. The estimation of the convergence rate is also given. It should be pointed out that this paper only consider the undirected communication graph and ideal communication. The directed communication graph and event-triggered PID control for distributed convex optimization will be discussed in future study.

**Acknowledgements**

The authors would like to thank the Associate Editor and reviewers for their insightful comments and suggestions based on which the presentation of this paper has been greatly improved.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported partly by National Natural Science Foundation of China under Grant No. 61673080, 61876200 partly by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJZD-K20200601, partly by Natural Science Foundation of Chongqing under Grant cstc2019jcyj-msxmX0102, partly by Venture & Innovation Support Program for Chongqing Overseas Returnees under Grant cx2017099.

**ORCID iD**

Wei Zhu https://orcid.org/0000-0003-1120-9750

**References**

1. Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found Trends Mach Learn* 2010; 3(1): 1–122.

2. Huang B, Wang Y, Yang C, et al. A neurodynamic-based distributed energy management approach for integrated local energy systems. *Int J Electr Power Energy Syst* 2021; 128: 106737.

3. Li Y, Gao DW, Gao W, et al. Double-mode energy management for multi-energy system via distributed dynamic event-triggered Newton-Raphson algorithm. *IEEE Trans Smart Grid* 2020; 11(6): 5339–5356.

4. Huang B, Liu L, Zhang H, et al. Distributed optimal economic dispatch for microgrids considering communication delays. *IEEE Trans Syst Man Cybern Syst* 2019; 49(8): 1634–1642.

5. Tsitsiklis JN. Problems in decentralized decision making and computation. *Technical report, Massachusetts Inst of Tech Cambridge Lab for Information and Decision Systems*, 1984.
6. Nagumey A. Book review: parallel and distributed computation: numerical methods. *Int J Supercomput Appl* 1989; 3(4): 73–74.

7. Nedić A and Ozdaglar A. Distributed subgradient methods for multi-agent optimization. *IEEE Trans Autom Control* 2009; 54(1): 48–61.

8. Shui W, Ling Q, Wu G, et al. Extra: an exact first-order algorithm for decentralized consensus optimization. *SIAM J Optim* 2015; 25(2): 944–966.

9. Nedić A and Olshevsky A. Distributed optimization over time-varying directed graphs. *IEEE Trans Autom Control* 2015; 60(3): 601–615.

10. Nedić A and Olshevsky A. Stochastic gradient-push for strongly convex functions on time-varying directed graphs. *IEEE Trans Autom Control* 2016; 61(12): 3936–3947.

11. Nedić A, Olshevsky A and Shi W. Achieving geometric convergence for distributed optimization over time-varying graphs. *SIAM J Optim* 2017; 27(4): 2597–2633.

12. Nedić A, Olshevsky A, Shi W, et al. Uribe. Geometrically convergent distributed optimization with uncoordinated step-sizes. In: *Proceedings of the American control conference*, St. Louis, MO, USA, pp.3950–3955, 2017.

13. Lou Y, Shi G, Johansson KH, et al. Approximate projected consensus for convex intersection computation: convergence analysis and critical error angle. *IEEE Trans Autom Control* 2014; 59(7): 1722–1736.

14. Lin P, Ren W and Song Y. Distributed multi-agent optimization subject to nonidentical constraints and communication delays. *Automatica* 2016; 65: 120–131.

15. Liu S, Qiu Z and Xie L. Convergence rate analysis of distributed optimization with projected subgradient algorithm. *Automatica* 2017; 83: 162–169.

16. Liu Q, Yang S and Hong Y. Constrained consensus algorithms with fixed step size for distributed convex optimization over multi-agent networks. *IEEE Trans Autom Control* 2017; 62(8): 4259–4265.

17. Wang P, Lin P, Ren W, et al. Distributed subgradient-based multiagent optimization with more general step sizes. *IEEE Trans Autom Control* 2018; 63(7): 2295–2302.

18. Liang S, Wang LY and Yin G. Distributed smooth convex optimization with coupled constraints. *IEEE Trans Autom Control* 2020; 65(1): 347–353.

19. Chang TH, Nedić A and Scaglione A. Distributed constrained optimization by consensus-based primal-dual perturbation method. *IEEE Trans Autom Control* 2014; 59(6): 1524–1538.

20. Wei Y, Fang H, Zeng X, et al. A smooth double proximal primal-dual algorithm for a class of distributed non-smooth optimization problem. *IEEE Trans Autom Control* 2018; 65(4): 1800–1806.

21. He Z, He J, Chen C, et al. CPCA: a Chebyshev proxy and consensus based algorithm for general distributed optimization. In: *2020 American control conference (ACC)*, Denver, CO, USA, pp.94–99, 2020.

22. Lin P, Ren W and Farrell JA. Distributed continuous-time optimization: nonuniform gradient gains, finite-time convergence, and convex constraint set. *IEEE Trans Autom Control* 2017; 62(5): 2239–2253.

23. Zhang Y and Hong Y. Distributed optimization design for high-order multi-agent systems. In: *Chinese control conference, CCC*, Hangzhou, China, pp.7251–7256, 2015.

24. Yi P, Hong Y and Liu F. Distributed gradient algorithm for constrained optimization with application to load sharing in power systems. *Syst Control Lett* 2015; 83: 45–52.

25. He X, Yu J, Huang T, et al. Average quasi-consensus algorithm for distributed constrained optimization: impulsive communication framework. *IEEE Trans Control Cybern* 2020; 50(1): 351–360.

26. Chen W and Ren W. Event-triggered zero-gradient-sum distributed consensus optimization over directed networks. *Automatica* 2016; 65: 90–97.

27. Li M, Su L and Liu T. Distributed optimization with event-triggered communication via input feedforward passivity. *IEEE Control Syst Lett* 2021; 5(1): 283–288.

28. Yu Z, Yu S, Jiang H, et al. Distributed fixed-time optimization for multi-agent systems over a directed network. *Nonlinear Dyn* 2021; 103(1): 775–789.

29. Wang J and Elia N. Control approach to distributed optimization. In: *2010 48th annual allerton conference on communication, control, and computing, allerton 2010*, Monticello, Va, USA, pp.557–561, 2010.

30. Kia SS, Cortés J and Martínez S. Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. *Automatica* 2015; 55: 254–264.

31. Yang S, Liu Q and Wang J. A multi-agent system with a proportional-integral protocol for distributed constrained optimization. *IEEE Trans Autom Control* 2017; 62(7): 3461–3467.

32. Han J. From PID to active disturbance rejection control. *IEEE Trans Ind Electron* 2009; 56(3): 900–906.

33. Qin L, He X and Zhou D. Distributed proportion-derivative formation control for second-order multi-agent systems with communication time delays. *Neurocomputing* 2017; 267: 271–282.

34. Wu Z, Li Z, Ding Z, et al. Distributed continuous-time optimization with scalable adaptive event-based mechanisms. *IEEE Trans Syst Man Cybern Syst* 2020; 50(9): 3252–3257.