Adaptive Stepsize Rule for Consensus Optimization by Supervisory Control Architecture

Yoshifumi Kusunoki†, Naoki Hayashi†, Takeshi Hatanaka† and Keiji Tatsumi†

We propose an adaptive stepsize rule for multi-agent based consensus optimization, to overcome drawbacks of the conventional diminishing stepsize rules. The proposed stepsize rule is based on an agreement degree of agent-wise gradients. It can archive both of fast approach and convergence at the early and last stages of iteration, respectively. Since the stepsize of each iteration is computed using a part of the global information of agents, a supervisory control architecture is required. We prove that the sequence generated by the consensus optimization algorithm using the proposed supervisory stepsize rule converges to the optimum solution. Moreover, we experimentally show that the proposed rule has better performance than the diminishing and constant stepsize rules.

1. Introduction

Technologies of distributed control and optimization are required in many real world issues, for example reduction of communication and/or computational load in networked systems, enhancement of robustness in harsh environment, handling adversaries in systems and so on. For that purpose, many researchers have developed theories based on multi-agent systems [1–8], and demonstrated their usefulness in real world applications [9,10]. In particular, we focus on consensus based optimization algorithms [10–15], which search an agreement of a multi-agent system in which each agent has a different objective function. For important classes of optimization, e.g. minimization of a sum of convex functions, algorithms of the consensus optimization successfully reach an optimal solution [11]. There are several applications of the consensus optimization: for example, estimation of a physical model in a sensor network [5,6], adaptation of mobile networks [16], object detection using multiple cameras [10], matrix factorization for big and distributed data [17] and so on.

In iterative optimization algorithms, adjustment of stepsizes is important for the reason that it determines the speed of convergence. In the framework of consensus optimization, furthermore, we should resolve two conflicting aims using a single stepsize parameter: one is fast approach of agent states to the optimal point at the early stages of iteration, and the other is to make them convergent (agreement) at the last stages. Diminishing stepsize rules can address this conflict, and they are favored because of their theoretical convergent properties. However, they cannot perform fast convergence in practice, and their parameters sensitively affect the convergent speed. Moreover, diminishing stepsize rules schedule stepsizes in advance of running an algorithm. As a result, the algorithm is vulnerable to disturbance.

To overcome these drawbacks, we propose a stepsize rule which is adaptive to the evolution of agent states. In this paper, we particularly focus on minimization of a sum of differentiable convex objective functions, in which each function is associated with an agent. In our proposed rule, the stepsize of each of iteration is defined by an agreement degree of the gradients of agents’ objective functions. This means that we need a supervisor or mechanism which calculates and distributes the stepsize. It may destroy a part of distributed architecture of the consensus optimization algorithm. However, this supervisory architecture is inevitable in order to achieve both of fast and accurate convergence. Additionally, in the literature of cooperative distributed optimization, popular algorithms, such as the dual decomposition [6] and ADMM [18], are equipped with information aggregation mechanisms.

We remark that our approach only controls stepsizes and does not change the distributed nature of the consensus optimization algorithm. Moreover, the stepsize of each of iteration is identical for all agents, so they are transmitted from the supervisor to the agents in a broadcast way, which requires a less communication cost comparing distinct controls of agents.
This broadcast control architecture is similar to that proposed in [4].

As a theoretical contribution, we show a sufficient condition for a scale factor of our stepsize rule in which the algorithm generates a convergent sequence to the global optimal solution. Additionally, through numerical experiments for minimization of a sum of convex quadratic functions, we verify that our stepsize rule shows better convergence performance than diminishing and constant stepsize rules.

This paper is organized as follows. In Section 2, we describe problem settings which are supposed throughout this paper, and the consensus optimization algorithm studied in this paper. Moreover, we explain the supervisory stepsize rule. In Section 3, we discuss descent properties of Perron-matrix update that is an agreement algorithm used in the consensus optimization algorithm. That discussion is important to simplify an analysis of the next section. In Section 4, we show a convergence analysis of the consensus optimization algorithm with our stepsize rule. In Section 5, through numerical experiments, we show advantages of our stepsize rule. Finally, in Section 6, we provide concluding remarks.

2. Consensus Optimization and Supervisory Stepsize Rule

2.1 Optimization Problem and Consensus Algorithm

We consider an unconstrained optimization problem whose objective function is an additive convex function.

\[
\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x),
\]

where \( f_1, \ldots, f_n : \mathbb{R}^m \to \mathbb{R} \) are differentiable convex functions. We assume the mean \( f \) is strongly convex, i.e., for some \( \sigma > 0 \) it holds that for every \( x, y \in \mathbb{R}^m \),

\[
(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \sigma \|x - y\|^2.
\]

Moreover, the gradient of \( f \), denoted by \( \nabla f \), is Lipschitz continuous, i.e., for some \( A \geq \sigma \) it holds that for every \( x, y \in \mathbb{R}^m \),

\[
\|\nabla f(x) - \nabla f(y)\| \leq A \|x - y\|.
\]

Furthermore, we assume each of \( f_i \) is also Lipschitz continuous with a constant \( A_i \). Let \( A_{\text{max}} = \max\{A_1, \ldots, A_n\} \). Because of the Lipschitz condition, \( f_1, \ldots, f_n \) and \( f \) are continuously differentiable. Let \( x^* \) be the unique optimal solution of \( f \), and \( p^* \) be its optimal value.

We solve the optimization problem (1) using a multi-agent system. First we separate objective functions \( f_1, \ldots, f_n \) by introducing variables \( x_1, \ldots, x_n \).

\[
\min_{x_1, \ldots, x_n \in \mathbb{R}^n} \frac{1}{n} \sum_{i=1}^{n} f_i(x_i)
\]

subject to \( x_1 = \cdots = x_n \).

The indices \( 1, \ldots, n \) are regarded as agents, and each \( x_i \) is a state of the \( i \)-th agent. Let \( N = \{1, \ldots, n\} \). We suppose that each agent \( i \) has exclusive access to the \( i \)-th objective function \( f_i \). Additionally, each agent can communicate with a certain subset of agents. The communication structure is modeled by a (undirected) graph \( G = (N, E) \), where \( E \subseteq \{(i,j) \mid i,j \in N, i \neq j\} \) be a set of edges. We assume \( G \) is connected in the sense of graph theory.

Each agent \( i \) is controlled to minimize its own objective function and to make a consensus with the other agents. Let \( \{x_1, \ldots, x_n\} \) be a set of current states. We update each state \( x_i \) to \( x_i' \) by the following rule:

\[
\dot{x}_i = x_i - r \sum_{(j,i) \in E} (x_i - x_j),
\]

\[
x_i' = \dot{x}_i - s \nabla f_i(\dot{x}_i),
\]

where \( r \) and \( s \) are nonnegative values. In the first equation, each agent \( i \) updates its state to agree with the states of the other agents in its neighborhood. The second equation is regarded as the update rule of the gradient descent applied to the \( i \)-th objective function.

We express the above update using the graph Laplacian. Let \( L \in \mathbb{R}^{n \times n} \) be the graph Laplacian of \( G \), i.e., for each \( i,j \in N \) we define

\[
L_{ij} = \begin{cases} 
-1 & \{i,j\} \in E, \\
|\{e \in E \mid e \ni i\}| & i = j, \\
0 & \text{otherwise.}
\end{cases}
\]

Given a nonnegative parameter \( r \), we consider a Perron matrix \( P \in \mathbb{R}^{n \times n} \), defined by

\[
P = I - rL.
\]

Then we have

\[
\dot{x}_i = \sum_{j=1}^{n} P_{ij} x_j.
\]

The above discussion is summarized in Algorithm 1. In this algorithm, we have a possibility to change the stepsize \( s \) for each of iteration. On the other hand, \( r \) should be constant.

2.2 Supervisory Stepsize Rule

In conventional consensus optimization, diminishing stepsize rules are often used to define \( s_k \). We can archive convergence of the consensus optimization algorithm by using those simple rules, however they have some drawbacks. One is that they cannot perform fast convergence in practice, especially the
Algorithm 1 Consensus Optimization Algorithm

1: An initial set of states $x_1^n,\ldots,x_n^n \in \mathbb{R}^m$ is provided.
2: Let $k \leftarrow 0$.
3: while a stopping criterion is not fulfilled do
4:     Calculate $x_i^{k+1}$ for $i = 1,\ldots,n$ as follows:
5:         $\hat{x}_i^k \leftarrow \sum_{j=1}^n P_{ij}x_j^k$,
6:         $x_i^{k+1} \leftarrow \hat{x}_i^k - s_k \nabla f_i(x_i^k)$.
7:     $k \leftarrow k + 1$.
8: end while

convergence is slow when their parameter tuning is missed. The other is that the rules are not robust to disturbance. That is, the stepsize sequence is scheduled in advance of running of the algorithm, and it continues to decrease even if the multi-agent system is disturbed.

To overcome the drawbacks, in this paper, we propose the following adaptive stepsize rule. Given $\alpha > 0$, $s_k$ of the $k$-th iteration is defined by

$$s_k = \alpha \beta_k,$$

where

$$\beta_k = \frac{\|\frac{1}{n} \sum_{i=1}^n \nabla f_i(\hat{x}_i^k)\|}{\sqrt{\frac{1}{n} \sum_{i=1}^n \|\nabla f_i(\hat{x}_i^k)\|^2}},$$

and $\beta_k = 0$ when the denominator is 0. We intuitively explain adaptiveness of the stepsize rule. When many of agents are far from the region including the minimums of the objective function $f$ as well as the agents’ objective functions $f_i$, their gradients have agreement. Hence, in that case, $\beta_k$ is close to 1 and $s_k$ is a long stepsize. On the other hand, When many of agents are near to the minimum of $f$, their gradients have disagreement. Hence in that case, the numerator of $\beta_k$ is close to 0 and it makes $s_k$ short. Then, the sequence becomes convergent.

A physical architecture of our proposed stepsize rule is explained as follows. We assume that a multi-agent system involves a supervisor. Fig. 1 shows such a configuration, where a graph $G = (N,E)$ is defined by $N = \{1,2,3,4,5\}$ and $E = \{\{1,2\},\{1,3\},\{2,3\},\{2,4\},\{3,4\},\{4,5\}\}$, and node 0 is the supervisor. At each $k$ of iteration, the supervisor gathers all gradients of agents and calculate $s_k$, then broadcasts it to the agents. Here, we remark that the stepsize is identical for all agents. This broadcast architecture is similar to control framework proposed by Azuma et al. [4].

In order to deal with two conflict situations: long stepsizes at the early stages of iteration and short ones for convergence, we need to observe the evolution in iteration and control stepsizes adequately. Therefore, the supervisor is regarded as a necessary factor of an adaptive stepsize rule for the consensus optimization algorithm. Our approach improves the performance of the algorithm at the expense of a part of its distributed architecture.

We remark that, using specific communication protocols, the proposed adaptive stepsizes can be computed and shared among agents without a supervisor at the expense of delay in updating the stepsizes. For example, Hirayama et al. [19] proposed a protocol which computes adaptive stepsizes for a Lagrangian relaxation approach in distributed architecture.

3. Descent Property of the Perron-matrix Update

This section is a preparation of the convergence analysis of the proposed rule that is shown in Section 4. Here, we focus on the agreement part of the algorithm, i.e., the Perron-matrix update, ignoring the minimization of the objective function. We introduce the following assumption, which simplifies the convergence analysis and clarifies effect of parameters of the algorithm. Let $\hat{x}^k = \frac{1}{n} \sum_{i=1}^n x_i^k$.

Assumption 1 There exists a constant $\rho \in (0, 1]$ such that

$$\frac{1}{n} \sum_{i=1}^n \|x_i^k - \hat{x}^k\| \leq (1 - \rho) \frac{1}{n} \sum_{i=1}^n \|x_i^k - \hat{x}^k\|,$$

for all $k = 1,2,\ldots$.

In other words, we assume that the sum for the distance between each state $x_i$ and their average $\hat{x}$ is reduced with a ratio $\rho$ by the Perron-matrix update. Estimation of the constant $\rho$ is important to evaluate a convergence rate of the algorithm. Unfortunately, this assumption is not fulfilled for any set of states. That is, there is a case in which the sum of distances does not strictly decrease. Consider the following situation.

[Example 1] Consider the graph involving 5 agents that is defined in Fig. 1, and 1-dimensional states $(x_1,x_2,x_3,x_4,x_5) = (0,1,1,1,2)$. We have $\bar{x} = (0 + 1 + 1 + 1 + 2)/5 = 1$ and $\bar{x} = Px = (2r,1-r,1-r,1+r,2-r)$. Consider the sum of distances when
0 < r < 1/2. We have \( \sum_{i=1}^{n} |x_i - \bar{x}| = 1 + 0 + 0 + 0 + 1 = 2 \) and \( \sum_{i=1}^{n} |\bar{x}_i - \bar{x}| = (1 - 2r) + (1 - 1 + r) + (1 + r - 1) + (2 - r - 1) = 2 \). Hence, \( \sum_{i=1}^{n} |\bar{x}_i - \bar{x}| < \sum_{i=1}^{n} |x_i - \bar{x}| \) fails even if the set of states is not the minimum of \( \sum_{i=1}^{n} |x_i - \bar{x}| \) and r is small.

One remedy for violation of the assumption is an inner loop in which the Perron-matrix update is repeated until the sum of distances sufficiently decreases. Since the sum of “squared” distances is linearly converges to 0, the repetition is terminated in finite time.

Regardless of the above discussion, in this study, we analyze the consensus optimization algorithm under Assumption 1. In the rest of this section, we investigate on descent properties of the Perron-matrix update, in order to confirm that the assumption is not unrealistic. First, we show that the sum of distances does not increase if r is sufficiently small as shown in Example 1.

**Proposition 1**
Let \( 0 < r \leq \frac{1}{\max_i \{ \lambda_i \}} \). Then, we have
\[
\frac{1}{n} \sum_{i=1}^{n} ||x_i - \bar{x}|| \leq \frac{1}{n} \sum_{i=1}^{n} ||x_i - \bar{x}||,
\]
for any \( x_1, \ldots, x_n \in \mathbb{R}^n \), \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \).

(Proof) Note that \( P_{ij} \geq 0 \) for all \( i, j \in N \). We have
\[
\sum_{i=1}^{n} ||x_i - \bar{x}|| = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} P_{ij} ||x_j - \bar{x}|| \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij} ||x_j - \bar{x}|| \leq \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij} ||x_j - \bar{x}|| = \sum_{j=1}^{n} ||x_j - \bar{x}||. \]

Next, we consider the asymptotic convergence factor that was studied in [20]. In that paper, the case of \( m = 1 \) was considered. Let \( x \in \mathbb{R}^n \) and \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \). The factor is defined by
\[
\phi = \sup_{x \neq \bar{x}} \lim_{k \to \infty} \left( \frac{||P^k x - \bar{x}||}{||x - \bar{x}||} \right)^{1/k},
\]
where \( \bar{x} = [1 \cdots 1]^T \in \mathbb{R}^n \).

We discuss the asymptotic convergence factor for the sum of distances in the case of arbitrary finite \( m \). First of all, we introduce some properties of eigenvalues and eigenvectors of the graph Laplacian \( L \) and the Perron matrix \( P \). Without loss of generality, let \( \lambda_1 \geq \cdots \geq \lambda_{n-1} > \lambda_n = 0 \) be the eigenvalues of \( L \). The nullspace of \( L \) is \( N = \{ [a \cdots a]^T \in \mathbb{R}^n | a \in \mathbb{R} \} \). We easily see that the eigenvalues of \( P \) are
\[
1 - r \lambda_1 \leq \cdots \leq 1 - r \lambda_{n-1} < 1 - r \lambda_n = 1,
\]
and the eigenspace of 1 is \( N \). It can be proved that \( \phi \) is identical to the spectral radius of \( P - I_n \) if
\[
r < \frac{1}{\sqrt{n}}.
\]
Moreover, \( \phi = \max \{|1 - r \lambda_1|,|1 - r \lambda_{n-1}|\} < 1 \).

Then, we obtain the following proposition.

**Proposition 2** Let \( r < \frac{1}{\sqrt{n}} \). For \( x_1, \ldots, x_n \in \mathbb{R}^m \), let \( x = [x_1^T \cdots x_n^T]^T \in \mathbb{R}^{mn} \). We define
\[
\psi = \sup_{x \neq \bar{x}} \lim_{k \to \infty} \left( \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |[P^k]_{ij} x_j - \bar{x}|}{\sum_{i=1}^{n} ||x_i - \bar{x}||} \right)^{1/k}
\]
(15)
where \( \otimes \) is the Kronecker product, and \( [P^k]_{ij} \) is the \( ij \)-th element of matrix \( P^k \). Then, we have \( \psi = \max \{|1 - r \lambda_1|,|1 - r \lambda_{n-1}|\} \).

(Proof) Consider the following asymptotic convergence factor:
\[
\phi' = \sup_{x \neq \bar{x}} \lim_{k \to \infty} \left( \frac{||P^k x - \bar{x}||}{||x - \bar{x}||} \right)^{1/k},
\]
where \( I \in \mathbb{R}^{n \times m} \) is the identity matrix. We have the following equality:
\[
(P^k \phi')^{1/k} = \left( \frac{P - I_n}{n} \right)^k (x - \bar{x}).
\]

Here, we remark that \( (A \otimes I)^k = A^k \otimes I \) for any matrix \( A \). \( P^{11^n} / n = I^{11^n} / n \) and \( (11^n / n)^2 = 11^{2n} / n \). Let \( Q = (P - I^{11^n} / n) \otimes I \) and \( E = \{ x - \bar{x} | x \in \mathbb{R}^{mn} \} \). Note that \( x - \bar{x} \in E \). Furthermore, let \( \phi(Q) \) be the spectral radius of \( Q \). By the Gelfand formula [21], we have \( \phi(Q) = \lim_{k \to \infty} \sup_{0 \neq e \in E} ||Q^k e||/||e||^{1/k} \). Additionally, we can easily say \( \sup_{0 \neq e \in E} \lim_{k \to \infty} \left( ||Q^k e||/||e|| \right)^{1/k} \leq \lim_{k \to \infty} \sup_{0 \neq e \in E} \left( ||Q^k e||/||e|| \right)^{1/k} \). Let \( e' \) be one of the eigenvectors corresponding to the spectral radius of \( Q \). We have \( \phi(Q) = \lim_{k \to \infty} \left( ||Q^k e'||/||e'|| \right)^{1/k} \). Moreover, since every row vector of \( 11^n / n \otimes I \) is an eigenvector of \( Q \) corresponding to eigenvalue 0, we have \( (11^n / n \otimes I)^k e' = 0 \), and consequently \( e' = (I - 11^n / n \otimes I) e' \in E \). In consequence, we obtain \( \phi(Q) \leq \phi' = \sup_{0 \neq e \in E} \lim_{k \to \infty} \left( ||Q^k e||/||e|| \right)^{1/k} \). By inequalities between \( L_1 \) and \( L_2 \) norms, we have for \( x \neq \bar{x} \),
\[
\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |[P^k]_{ij} x_j - \bar{x}|}{\sum_{i=1}^{n} ||x_i - \bar{x}||} \leq \left[ \frac{1}{\sqrt{n}} \right] \times \frac{||P^k x - \bar{x}||}{||x - \bar{x}||}.
\]
It implies \( \psi = \phi' = \phi(Q) \). Finally, since the spectral radii of \( Q = (P - I^{11^n} / n) \otimes I \) and \( P - I^{11^n} / n \) coincide, we obtain \( \psi = \phi(Q) = \max \{|1 - r \lambda_1|,|1 - r \lambda_{n-1}|\} \).

By Proposition 2, we expect that \( \rho \) in Assumption 1 is approximated by \( 1 - \max \{|1 - r \lambda_1|,|1 - r \lambda_{n-1}|\} \) for large \( k \), if stepsizes \( s_k \) are sufficiently small and the effect of \( s_k \nabla f_i(x_i^k) \cdots s_k \nabla f_n(x_n^k) \) vanishes in the update of \( x^{k+1} \).
4. Convergence Analysis for the Supervisory Stepsize Rule

In this section, we analyze the convergence of the consensus optimization algorithm with the stepsize rule of (10), under Assumption 1 and an additional assumption. That is, under those assumptions, we prove that sequences \( \{x_k^1\}, \ldots, \{x_k^n\} \) given by stepsize \( \{s_k = \alpha \beta_k\} \) converge to the minimum vector \( x^* \) if \( \alpha \) is sufficiently small.

For every \( k = 1, 2, \ldots \), we define

\[
\zeta_k = \|x^k - x^*\|, \quad \delta_k = \frac{1}{n} \sum_{i=1}^{n} \|x^k_i - x^k\|. \tag{16}
\]

Moreover, we define

\[
\tilde{\gamma}_k = \frac{1}{n} \sum_{i=1}^{n} \left\| \nabla f_i(x^k) \right\|, \quad \bar{\gamma}_k = \|\nabla f(x^k)\|. \tag{17}
\]

Here, we introduce a technical assumption in which a convergence speed of \( \{\tilde{\gamma}_k\} \) is at most as much as that of \( \{\gamma_k\} \).

**Assumption 2** There exists a constant \( \eta \geq 0 \) such that

\[
\tilde{\gamma}_k \leq \eta \gamma_k, \tag{19}
\]

for all \( k = 1, 2, \ldots \).

First, we provide the well-known contraction property of the gradient descent.

**Lemma 1** [22] Let \( f: \mathbb{R}^m \rightarrow \mathbb{R} \) be a differentiable function. Additionally, there exist constants \( 0 < \sigma \leq \Lambda \) such that for any \( x, y \in \mathbb{R}^m \) the following inequalities hold:

\[
\left( \nabla f(x) - \nabla f(y) \right)^\top (x - y) \geq \sigma \|x - y\|^2, \tag{20}
\]

\[
\|\nabla f(x) - \nabla f(y)\| \leq \Lambda \|x - y\|. \tag{21}
\]

Then, for any \( s \geq 0 \), we have

\[
\|s \nabla f(x) - (y - s \nabla f(y))\| \\
\leq \max\{|1 - s\Lambda|, |1 - s\sigma|\} \|x - y\|. \tag{22}
\]

The next lemma plays the central role of our convergence analysis. In this lemma, we propose that decrement of \( \max\{\zeta_k, \theta \delta_k\} \) should be considered to show the convergence of \( \{x_1^k\}, \ldots, \{x_n^k\} \) to \( x^* \), where the constant \( \theta \) is regarded as an expression of the balance of the speeds of the sequences \( \{\zeta_k\} \) and \( \{\delta_k\} \).

**Lemma 2** Suppose \( \alpha \in \left( 0, \frac{1}{2\tau^2} \right) \). Let \( \theta > 0 \). We have the following inequality:

\[
\max\{\zeta_{k+1}, \theta \delta_{k+1}\} \\
\leq (1 - \min\{\mu_k(\theta), \nu_k(\theta)\}) \max\{\zeta_k, \theta \delta_k\}, \tag{23}
\]

where \( \mu_k \) and \( \nu_k \) are defined by

\[
\mu_k(\theta) = s_k \sigma \left( 1 - \frac{A_{\max}}{\sigma} \frac{1 - \rho}{\theta} \right), \tag{24}
\]

\[
\nu_k(\theta) = \rho \left( 1 - \alpha \Lambda \eta \frac{\theta}{\rho} \sqrt{1 - \beta_k^2} \right). \tag{25}
\]

(Proof) First, we show \( \zeta_{k+1} \leq (1 - \mu_k(\theta)) \max\{\zeta_k, \theta \delta_k\} \). By the triangle inequality,

\[
\zeta_{k+1} = \left\| x^{k+1} - x^* \right\| \\
\leq \left\| x^k - s_k \nabla f(x^k) - x^* \right\| \\
+ s_k \left( \frac{1}{n} \sum_{i=1}^{n} \left\| \nabla f_i(x^k) - \nabla f(x^k) \right\| \right).
\]

Consider the first term in the right hand side. Using Lemma 1, we have

\[
\|x^k - s_k \nabla f(x^k) - x^*\| \\
= \left\| x^k - s_k \nabla f(x^k) - (x^* - s\nabla f(x^*))\right\| \\
\leq \max\{|1 - s_k\Lambda|, |1 - s_k\sigma|\} \|x^k - x^*\| \\
\leq (1 - s_k\sigma) \|x^k - x^*\| = (1 - s_k\sigma) \zeta_k.
\]

Here, to show the third inequality, we use \( |1 - s_k\sigma| \geq |1 - s_k\Lambda| \) and \( 1 - s_k\sigma \geq 0 \) derived from \( s_k = \alpha \beta_k \leq \frac{1}{\sqrt{\tau^2}} \).

Consider the second term. Using the Lipschitz continuity of \( \nabla f_1, \ldots, \nabla f_n \) and Assumption 1, we have

\[
s_k \left( \frac{1}{n} \sum_{j=1}^{n} \left\| \nabla f_j(x^k) - \nabla f(x^k) \right\| \right) \\
= s_k \left( \frac{1}{n} \sum_{j=1}^{n} \left\| \nabla f_j(x^k) - \nabla f_j(x^k) \right\| \right) \\
\leq s_k \left( \frac{1}{n} \sum_{j=1}^{n} \|x^k_j - x^k\| \right) \\
\leq s_k A_{\max} \left( \frac{1}{n} \sum_{j=1}^{n} \|x^k_j - x^k\| \right) \leq s_k A_{\max} \|x^k - x^k\| \\
\leq s_k A_{\max} \|x^k - x^k\| \leq s_k A_{\max} (1 - \rho) \delta_k.
\]

Therefore, we obtain

\[
\zeta_{k+1} \leq (1 - s_k\sigma) \zeta_k + s_k A_{\max} (1 - \rho) \delta_k \\
\leq \left( 1 - s_k\sigma \left( 1 - \frac{A_{\max}}{\sigma} \frac{1 - \rho}{\theta} \right) \right) \max\{\zeta_k, \theta \delta_k\}.
\]

Next, we show \( \theta \delta_{k+1} \leq (1 - \nu_k(\theta)) \max\{\zeta_k, \theta \delta_k\} \). By the triangle inequality and Assumption 1, we have

\[
\mu_k(\theta) = s_k \sigma \left( 1 - \frac{A_{\max}}{\sigma} \frac{1 - \rho}{\theta} \right), \tag{24}
\]

\[
\nu_k(\theta) = \rho \left( 1 - \alpha \Lambda \eta \frac{\theta}{\rho} \sqrt{1 - \beta_k^2} \right). \tag{25}
\]
\( \delta_{k+1} = \frac{1}{n} \sum_{i=1}^{n} \|x_{k+1}^{i} - x^{k+1}\| \leq \frac{1}{n} \sum_{i=1}^{n} \|x_{k}^{i} - x_{k}\| + \frac{1}{n} \sum_{i=1}^{n} \left| \nabla f_{i}(x_{k}^{i}) - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j}(x_{k}^{j}) \right| \) 

\( = (1-\rho)\delta_{k} \) 

\( + s_{k} \frac{1}{n} \sum_{i=1}^{n} \left| \nabla f_{i}(x_{k}^{i}) - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j}(x_{k}^{j}) \right| \)

Considering the second term of the right hand side, we have 

\( s_{k} \frac{1}{n} \sum_{i=1}^{n} \left| \nabla f_{i}(x_{k}^{i}) - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j}(x_{k}^{j}) \right| \leq s_{k} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left| \nabla f_{i}(x_{k}^{i}) - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j}(x_{k}^{j}) \right|^{2}} \)

\( = s_{k} \frac{1}{n} \sum_{i=1}^{n} \left| \nabla f_{i}(x_{k}^{i}) \right|^{2} - \left( \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j}(x_{k}^{j}) \right)^{2} \)

\( = \alpha_{k} \beta_{k}^{2} \frac{1}{n} \sum_{i=1}^{n} \left| \nabla f_{i}(x_{k}^{i}) \right|^{2} - \beta_{k}^{2} \gamma_{k} \)

\( = \alpha_{k} \beta_{k}^{2} \frac{1}{n} \sum_{i=1}^{n} \left| \nabla f_{i}(x_{k}^{i}) \right|^{2} - \beta_{k}^{2} \gamma_{k} \)

The next lemma shows conditions in which \( \{x_{k}^{1}\}, \ldots, \{x_{k}^{n}\} \) converge to \( x^{*} \).

[Lemma 3] If \( \lim_{k \to \infty} s_{k} = 0 \) and \( \{x_{k}^{1}\}, \ldots, \{x_{k}^{n}\} \) are bounded, then all sequences converge to \( x^{*} \).

Proof. First, since \( \{x_{k}^{1}\}, \ldots, \{x_{k}^{n}\} \) as well as \( \{x_{k}^{1}\}, \ldots, \{x_{k}^{n}\} \) are bounded, we see that \( \{\nabla f_{i}(x_{k}^{1})\}, \ldots, \{\nabla f_{i}(x_{k}^{n})\} \) are also bounded.

Since \( \beta_{k} = \alpha_{k} \beta_{k} \), \( \lim_{k \to \infty} s_{k} = 0 \) implies \( \lim_{k \to \infty} \beta_{k} = 0 \). It further implies two cases that \( \sum_{i=1}^{n} \| \nabla f_{i}(x_{k}^{i}) \|^{2} \) diverges or \( \lim_{k \to \infty} \sum_{i=1}^{n} \| \nabla f_{i}(x_{k}^{i}) \| = 0 \). However, the first case is impossible, since \( \{\nabla f_{i}(x_{k}^{1})\}, \ldots, \{\nabla f_{i}(x_{k}^{n})\} \) are bounded. Hence, we have the second case.

We consider convergence of \( \delta_{k} \). We show that for any \( \varepsilon > 0 \), there exists \( k' \) such that \( \delta_{k} < 2\varepsilon \) for any \( k > k' \). By the same discussion of the proof of Lemma 2, we have 

\( \delta_{k+1} \leq (1-\rho)\delta_{k} + s_{k} \frac{1}{n} \sum_{i=1}^{n} \left| \nabla f_{i}(x_{k}^{i}) - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j}(x_{k}^{j}) \right| \)

Since \( \{x_{k}^{1}\}, \ldots, \{x_{k}^{n}\} \) are bounded, without loss of generality, there exists \( c_{1} \) and \( c_{2} \) such that 

\( \frac{1}{n} \sum_{i=1}^{n} \left| \nabla f_{i}(x_{k}^{i}) - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j}(x_{k}^{j}) \right| \leq c_{1} \),

and \( \delta_{k} \leq c_{2} \) for all \( k \). Since \( \rho \in (0,1) \), there exists \( l \) such that \( (1-\rho)^{l}c_{2} < \varepsilon \). Moreover, since \( s_{k} \) converges to 0, for sufficiently large \( l' \), we have \( s_{k'} < \frac{\varepsilon}{2\varepsilon} \) for all \( k \geq l' \). Consider \( \delta_{l'+1} \).

\( \delta_{l'+1} \leq (1-\rho)^{l}c_{2} + c_{1} \frac{l'+l-1}{k=l'} (1-\rho)^{l'+l-1-k} s_{k} \)

\( < \varepsilon + \varepsilon \rho \sum_{k=l'+1}^{l'+l-1} (1-\rho)^{l'+l-1-k} < \varepsilon + \varepsilon = 2\varepsilon \).

Finally, consider an upper bound of 

\( \frac{1}{n} \sum_{i=1}^{n} \| \nabla f(x_{k}^{i}) \| \).

\( \frac{1}{n} \sum_{i=1}^{n} \| \nabla f(x_{k}^{i}) \| \leq \frac{1}{n} \sum_{i=1}^{n} \| \nabla f(x_{k}^{i}) - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j}(x_{k}^{j}) \| + \frac{1}{n} \sum_{i=1}^{n} \| \nabla f_{i}(x_{k}^{i}) \| 

\)

\( \leq \frac{1}{n} \sum_{i=1}^{n} \| \nabla f_{i}(x_{k}^{i}) \| + \frac{1}{n} \sum_{i=1}^{n} \| \nabla f_{i}(x_{k}^{i}) \| + \gamma_{k} \)

\( < A_{k} d_{k} + A_{\max} (1-\rho) \delta_{k} + \gamma_{k} \).

Hence, by \( \lim_{k \to \infty} d_{k} = 0 \) and \( \lim_{k \to \infty} \delta_{k} = 0 \), we obtain \( \lim_{k \to \infty} \frac{1}{n} \sum_{i=1}^{n} \| \nabla f(x_{k}^{i}) \| = 0 \). It means \( \lim_{k \to \infty} x_{k}^{i} = x^{*} \) for all \( i = 1,\ldots,n \).

Using Lemmas 2 and 3, we prove that all of \( \{x_{k}^{1}\}, \ldots, \{x_{k}^{n}\} \) converge to \( x^{*} \) under a certain condition for \( \alpha \) of (10). The condition ensures that there exists \( \theta \) such that both of \( \mu_{k}(\theta) \) and \( \nu_{k}(\theta) \) are nonnegative for all \( k \).

[Theorem 1] Suppose that Assumptions 1 and 2 hold. Let \( \{x_{k}^{1}\}, \ldots, \{x_{k}^{n}\} \) be the sequences that are generated by Algorithm 1 with the stepsize rule \( s_{k} = \alpha \beta_{k} \). If \( \alpha \in \left( 0, \frac{2}{\tau_{\max}} \right) \) is in the following interval:

\( 0 < \alpha < \frac{\rho}{1-\rho \Lambda \max \eta} \)

Then, all of \( \{x_{k}^{1}\}, \ldots, \{x_{k}^{n}\} \) converge to the minimum \( x^{*} \) of \( f \).

(Proof) Let \( \theta = \frac{\rho}{\max \eta} \). We show that \( \max \{\xi_{k}, \delta_{k}\} \) in Lemma 2 is a non-increasing sequence. In other words, we show \( \mu_{k}(\theta) = 0 \) and \( \nu_{k}(\theta) = 0 \). Consider \( \mu_{k}(\theta) \). Remark the condition of \( \alpha \), we obtain...
\[
\mu_k(\theta) = s_k \sigma \left( 1 - \frac{A_{\max}}{\sigma} \frac{1 - \rho}{\theta} \right),
\]
\[
= s_k \alpha \sigma \left( \frac{1}{\alpha} - \frac{A_{\max}}{\sigma} \frac{1 - \rho}{\theta} \right) \geq 0.
\]

Consider \( \nu_k(\theta) \). We obtain
\[
\nu_k(\theta) = \rho \left( 1 - \frac{\alpha A_{\max} \theta}{\rho} \sqrt{1 - \beta_k^2} \right)
\geq \rho \left( 1 - \frac{\theta \alpha A_{\max}}{\rho} \right) = 0.
\]

Hence, \( \{\max\{\xi_k, \theta \delta_k\}\} \) is non-increasing.

Next, we show \( \lim_{k \to \infty} x_k^i = x^* \), \( i = 1, \ldots, n \). Since \( \{\max\{\xi_k, \theta \delta_k\}\} \) is non-increasing and its level sets are bounded, we see that \( \{x_1^i, \ldots, x_n^i\} \) are bounded. Moreover, since the sequence is non-increasing and bounded below, it should hold that
\[
\lim_{k \to \infty} \max \{\xi_k, \theta \delta_k\} = 0,
\]
or
\[
\lim_{k \to \infty} \min \{\xi_k, \theta \delta_k\} = 0.
\]

In the first case, we immediately obtain \( \lim_{k \to \infty} x_k^i = x^* \). We consider the latter case. Suppose \( \lim_{k \to \infty} \xi_k(\theta) = 0 \). Since \( \alpha < \frac{\rho}{\sqrt{1 - \beta_k^2}} \), we have \( 1 - \frac{A_{\max}}{\sigma} \frac{1 - \rho}{\theta} > 0 \). Hence, \( \lim_{k \to \infty} s_k = 0 \) should hold. By Lemma 3, we obtain \( \lim_{k \to \infty} x_k^i = x^* \). On the other hand, suppose \( \lim_{k \to \infty} \nu_k(\theta) = 0 \). Remark that \( \rho > 0 \), we see that it should hold that \( \lim_{k \to \infty} \beta_k = 0 \). It implies \( \lim_{k \to \infty} s_k = 0 \). Therefore, by Lemma 3, we obtain \( \lim_{k \to \infty} x_k^i = x^* \).

In the proof of this theorem, we show that \( \{\max\{\xi_k, \theta \delta_k\}\} \) is a decreasing sequence. The rate of its convergence depends on \( 1 - \min \{\xi_k(\theta), \nu_k(\theta)\} \). When the states \( x_1^i, \ldots, x_n^i \) are far from the minimum \( x^* \), \( \beta_k \) is close to 1 and by taking a large \( \theta \), we have \( 1 - \min \{\xi_k(\theta), \nu_k(\theta)\} \approx 1 - \min \{\alpha \sigma, \rho\} \). Hence, the sequences \( \{x_1^i\}, \ldots, \{x_n^i\} \) linearly converge to \( x^* \) in rate of \( 1 - \min \{\alpha \sigma, \rho\} \). On the other hand, when \( x_1^i, \ldots, x_n^i \) approach to \( x^* \), \( \beta_k \) approaches to 0. As a consequence, \( 1 - \min \{\xi_k(\theta), \nu_k(\theta)\} \) approaches to 1. It shows a sublinear of the sequences \( \{x_1^i\}, \ldots, \{x_n^i\} \). Additionally, \( \theta \) in the proof is a finite nonzero value. It means that \( \xi_k \) and \( \delta_k \) are expected to converge in the same order.

5. Numerical Experiments

5.1 Experimental Settings

Through numerical experiments, we empirically show that the proposed stepsize rule of (10) has better performance than conventional diminishing and constant stepsize rules. We compared the following 4 stepsize rules:
\[
s_k^1 = \alpha \frac{1}{k}, \quad s_k^2 = \alpha \frac{1}{\sqrt{k}}, \quad s_k^3 = \alpha \beta_k, \quad s_k^4 = \alpha, \quad (27)
\]
where
\[
\alpha = \alpha' \frac{2}{A_{\max}}, \quad (28)
\]
and \( \alpha' \in (0, 1] \). Note that the reason why we used the constant \( \frac{2}{A_{\max}} \) is that the sequence \( \{z_k^i\} \) generated by \( z_{k+1}^i = z_k^i - s \nabla f_i(z_k^i) \) does not diverge for all \( i = 1, \ldots, n \) if \( s \leq \frac{2}{A_{\max}} \). As assuming in Theorem 1, \( \alpha \leq \frac{2}{A_{\max}} \) holds in the experiments, however it does not necessarily hold in general.\(^1\) The first and second ones are diminishing stepsize rules.\(^2\) The third one is the proposed stepsize rule using \( \beta_k \) of (11). The last one is the constant stepsize rule. The parameter \( \alpha' \) was selected from 1, 0.1, 0.01, 0.001 in the experiments. From Theorem 1, if \( \alpha' \) is sufficiently small then it is ensured that a sequence with stepsize \( \{s_k^i\} \) converges.

We considered a regression problem. The \( i \)-th objective function was given by
\[
f_i(x) = \frac{1}{2} \|A_i x - b_i\|^2, \quad (29)
\]
where \( A_i \in \mathbb{R}^{l_i \times m} \) and \( b_i \in \mathbb{R}^{l_i} \). We describe how we generated \( A_i \) and \( b_i \) for \( i = 1, \ldots, n \). \( A_i \) and \( b_i \) were given by \( a_i A'_i (\text{diag}(w_i)) \) and \( c_i b'_i \), respectively. \( A'_i \) was an \( l_i \times m \) matrix whose elements were independently obtained from the normal standard distribution. \( a_i \) and \( c_i \) were values obtained from the uniform distribution in \([0, 1] \). \( w_i \) was an \( m \)-dimensional vector whose elements were independently obtained from the uniform distribution in \([0, 1] \), and \( \text{diag}(w_i) \) means the diagonal matrix whose elements are \( w_i \). \( b'_i \) was an \( l_i \)-vector which was uniformly obtained from the unit ball. In this paper, we used one instance of this optimization problem, where \( n = 50, m = 50 \) and \( l_i = 50 \) for \( i = 1, \ldots, n \).

The graph \( G \) was also randomly generated. Each edge \( \{i, j\} \), where \( i, j \in N \), was independently connected with probability \( p = 0.1 \). We check whether the obtained graph was connected. The parameter \( r \) of (8) was provided as follows:
\[
r = r' \frac{2}{\lambda_1 + \lambda_{n-1}}, \quad (30)
\]
Remark that the constant \( r = \frac{2}{\lambda_1 + \lambda_{n-1}} \) minimizes the asymptotic convergence factor \( \psi = \max \{|1 - r \lambda_1|, |1 - r \lambda_{n-1}| \} \) of Proposition 2. The parameter \( r' \) was selected from 1, 0.1, 0.01, 0.001 in the experiments.

For each \( i = 1, \ldots, n \), the \( i \)-th initial states \( x_k^i \) were given by a random point in a ball whose center and radius are the minimum \( x^* \) and 1000, respectively.

5.2 Comparison of Stepsize Rules

To show that the proposed stepsize rule has better performance than conventional ones, we compared the stepsize rules of (27) in 3 situations: \( (\alpha', r') = \)
\(^1\)The condition \( \alpha \leq \frac{2}{A_{\max}} \) can be weakened as \( \alpha < \frac{2}{A_{\max}} \) by a minor revision of Lemma 2.
\(^2\)However, \( \sum_{k=1}^{\infty} (s_k^4) < \infty \) is not fulfilled.
Fig. 2 Comparison of stepsize rules

(a) Relative error \((\alpha' = 1, r' = 1)\)

(b) Relative error \((\alpha' = 1, r' = 0.1)\)

(c) Relative error \((\alpha' = 0.1, r' = 1)\)

(d) Trajectory of \(\{ (\delta_k, \zeta_k) \} \) \((\alpha' = 1, r' = 1)\)

(e) Trajectory of \(\{ (\delta_k, \zeta_k) \} \) \((\alpha' = 1, r' = 0.1)\)

(f) Trajectory of \(\{ (\delta_k, \zeta_k) \} \) \((\alpha' = 0.1, r' = 1)\)

5.3 Analysis for the Supervisory Step-size Rule

Using Fig. 3, we investigate the effects of the parameters \(\alpha'\) and \(r'\) for the proposed stepsize rule. The first two figures: Figs. 3(a) and 3(b) show changes of relative errors with different values of \(\alpha'\) and \(r'\), respectively. Second, Figs. 3(c) and 3(d) show trajectories of \(\{ (\delta_k, \zeta_k) \} \). Third, Figs. 3(e) and 3(f) show \(1 - \hat{\delta}_k^\delta\), which evaluates \(\rho\). The last two figures: Figs. 3(g) and 3(h) show \(\gamma_k \hat{\gamma}_k\), which evaluates \(\eta\).

As expected from the convergence rate \(1 - \min\{\mu_k(\theta), \nu_k(\theta)\}\) in Lemma 2, \(\alpha'\) and \(r'\) (it controls \(\rho\)) affect the speed of error reduction. The larger \(\alpha'\) and \(r'\) have better convergence performance. However, too large values result in divergence of sequences.

Moreover, \(\alpha'\) changes the balance between the speed of \(\{ \delta_k \} \) and \(\{ \zeta_k \} \), where the balance appears as positions of the 45° line in figures. This phenomenon may be explained using \(\theta = \frac{\alpha'}{\alpha} \) in the proof of Theorem 1. That is, a value of \(\theta\) represents the balance between \(\{ \delta_k \} \) and \(\{ \zeta_k \} \), and it depends on \(\alpha\). This explanation can be applied for the reason why the balance was barely changed for different values of \(r'\). It is because \(\rho\) and \(\eta\) have a correlation, as shown in (1.1), (1.0.1), (0.1.1). Results are shown in Fig. 2.
Figs 3(f) and 3(h), and θ was barely changed too.
Finally, from Figs. 3(g)–3(h), we can see Assumptions 1 and 2 are fulfilled in the experiments.

6. Concluding Remarks
In this paper, we have proposed a supervisory stepsize rule for consensus optimization. It resolves the conflict between the fast approach and the convergence of the sequence of solutions, using the agreement degree of the agents’ gradients.
In Section 4, we have shown a sufficient condition of the stepsize factor $\alpha$ in which the sequence generated by the proposed stepsize rule converges. Moreover, the convergence analysis clarified the linear convergence of the sequence at the early stages of iteration, and its sublinear property near the minimum. The rate of convergence depends on both of the parameters $\alpha$ and $r$. Furthermore, the convergence analysis expects that $\delta_k$ and $\zeta_k$ converge in the same order.

These theoretical results were observed in numerical experiments where minimization of the sum of quadratic functions was considered. Assumptions 1 and 2 were also verified in the experiments. Moreover, the experiments showed that the proposed rule had better performance than the diminishing and constant stepsize rules. We only compared their convergence speeds with respect to numbers of iteration, because the time elapsed for each iteration depends on a hardware configuration. However, we believe that the stepsize selection based on the proposed rule is useful for fast convergence of the algorithm in real-world experiments. Some of the readers consider that applications of the proposed method are limited, because of the assumptions of this paper. However, since we only proposed a stepsize rule and the other parts of the consensus optimization algorithm are untouched, the applicability of the algorithm remains largely intact. The assumptions are required to explain convergence situations of the algorithm theoretically. Furthermore, since the proposed stepsizes (10) do not exceed the given parameter $\alpha$, conventional results regarding stability and optimality for the constant stepsize rule may be applied to the proposed rule.

One of the future directions of our study is to extend this analysis to constrained optimization.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP17K14701 and JST-Mirai Program Grant Number JPMJMI17B4, Japan.

References

[1] R. Olfati-Saber, J. A. Fax and R. M. Murray: Consensus and cooperation in networked multi-agent systems; Proceedings of the IEEE, pp. 215–233 (2007)
[2] F. Bullo, J. Cortès and S. Martínez: Distributed Control of Robotic Networks: A Mathematical Approach to Motion Coordination Algorithms, Princeton University Press (2009)
[3] M. Mesbahi and M. Egerstedt: Graph Theoretic Methods in Multiagent Networks, Princeton University Press (2010)
[4] S. Azuma, R. Yoshimura and T. Sugie: Broadcast control of multi-agent systems; Automatica, Vol. 49, No. 8, pp. 2307–2316 (2013)
[5] A. H. Sayed, S. Tu, J. Chen, X. Zhao and Z. J. Towfic: Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior; IEEE Signal Processing Magazine, Vol. 30, No. 3, pp. 155–171 (2013)
[6] S. Azuma, N. Nagahara, H. Ishii, N. Hayashi, K. Sakurama and T. Hatanaka: Control of Multi-agent Systems, in Japanese, Corona Publishing (2015)
[7] T. Hatanaka, N. Chopra, M. Fujita and M. W. Spong: Passivity-Based Control and Estimation in Networked Robotics, Springer International Publishing (2015)
[8] K. Sakurama, S. Azuma and T. Sugie: Distributed controllers for multi-agent coordination via gradient-flow approach; IEEE Transactions on Automatic Control, Vol. 60, No. 6, pp. 1471–1485 (2015)
[9] W. Ren and R. W. Beard: Distributed Consensus in Multi-vehicle Cooperative Control, Springer (2008)
[10] T. Hatanaka, R. Funada, G. Gezer and M. Fujita: Distributed visual 3-D localization of a human using pedestrian detection algorithm: A passivity-based approach; IFAC-PapersOnLine, Vol. 49, No. 22, pp. 210–215 (2016)
[11] A. Nedić and A. Ozdaglar: Distributed subgradient methods for multi-agent optimization; IEEE Transactions on Automatic Control, Vol. 54, No. 1, pp. 48–61 (2009)
[12] A. Nedić, 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)
[13] M. Zhu and S. Martinez: On distributed convex optimization under inequality and equality constraints; IEEE Transactions on Automatic Control, Vol. 57, No. 1, pp. 151–164 (2012)
[14] T. Chang, A. Nedić and A. Scaglione: Distributed constrained optimization by consensus-based primal-dual perturbation method; IEEE Transactions on Automatic Control, Vol. 59, No. 6, pp. 1524–1538 (2014)
[15] I. Masubuchi, T. Wada, T. Assai, L. T. H. Nguyen, Y. Ohta and Y. Fujisaki: Distributed multi-agent optimization based on an exact penalty method with equality and inequality constraints; SICE Journal of Control, Measurement, and System Integration, Vol. 9, No. 4, pp. 179–186 (2016)
[16] S. Tu and A. H. Sayed: Mobile adaptive networks; IEEE Journal of Selected Topics in Signal Processing, Vol. 5, No. 4, pp. 649-664 (2011)
[17] H. Raja and W. U. Bajwa: Cloud K-SVD: A collaborative dictionary learning algorithm for big, distributed data; IEEE Transactions on Signal Processing, Vol. 64, No. 1, pp. 173–188 (2016)
[18] I. D. Schizas, G. Mateos and G. B. Giannakis: Distributed LMS for consensus-based in-network adaptive processing; IEEE Transactions on Signal Processing, Vol. 57, No. 6, pp. 2365–2382 (2009)
[19] K. Hirayama, T. Matsui and M. Yokoo: Adaptive price update in distributed lagrangian relaxation protocol; Transactions of the Japanese Society for Artificial Intelligence, Vol. 26, No. 1, pp. 59–67, in Japanese (2011)
[20] L. Xiao and S. Boyd: Fast linear iterations for
Authors

Yoshifumi Kusunoki (Member)
He received the Ph.D. degree from Osaka University in 2010. He is currently an Assistant Professor at Osaka University. His research research includes machine learning and cooperative game theory, connecting with optimization. He is a member of the ORSJ, SICE and SOFT.

Naoki Hayashi (Member)
He received the Ph.D. degree from Osaka University in 2011. He was a Research Assistant at Kyoto University in 2011. He is currently an Assistant Professor at Osaka University. His research interests include cooperative control and distributed optimization. He is a member of SICE, IEICE, and IEEE.

Takeshi Hatanaka (Member)
He received the Ph.D. degree in applied mathematics and physics from Kyoto University in 2007. He then joined Tokyo Institute of Technology in 2007, where he held positions as an assistant and associate professor. Since April 2018, he is an associate professor at Osaka University. His research interests include networked robotics and energy management systems. He is an AE for IEEE TSCT and SICE JCMSI, and a member of the Conference Editorial Board of IEEE CSS.

Keiji Tatsumi (Member)
He received the Ph.D. degree in informatics from Kyoto University in 2006. Since 1998, he has been working at Osaka University, where he is an Associate Professor. His current research includes metaheuristics for global optimization and machine learning. He is a member of the ORSJ, SICE and IEEE.