LINEARLY CONVERGENT DECENTRALIZED CONSENSUS OPTIMIZATION WITH THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS

Wei Shi[*] *Qing Ling*[*] *Kun Yuan*[*] *Gang Wu*[*] *Wotao Yin*[†]

[*] Department of Automation, University of Science and Technology of China, Hefei, Anhui, China
[†] Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA

ABSTRACT

In the decentralized consensus optimization problem, a network of agents minimizes the summation of their local objective functions on a common set of variables, allowing only information exchange among neighbors. The alternating direction method of multipliers (ADMM) has been shown to be a powerful tool for solving the problem with empirically fast convergence. This paper establishes the linear convergence rate of the ADMM in decentralized consensus optimization. The theoretical convergence rate is a function of the network topology, properties of the local objective functions, and the algorithm parameter. This result not only gives a performance guarantee for the ADMM but also provides a guideline to accelerate its convergence rate for the decentralized consensus optimization problems.

Index Terms— Network consensus optimization, alternating direction method of multipliers, linear convergence

1. INTRODUCTION

Recent advances in signal processing and decision-making of large-scale networked multi-agent systems have motivated much research interest in decentralized optimization [1, 2]. Considering scalability and robustness issues, centralized optimization is no longer a proper choice and we prefer letting the agents accomplish their task in a decentralized manner. In decentralized optimization, no fusion center collects data and executes computation; contrarily, each agent holds its own data, computes by itself, and exchange information with its neighbors to collaboratively minimize an overall objective function. Typical applications include networked control systems [2], wireless sensor networks [3], and smart grids [4].

In this paper, we focus on *decentralized consensus optimization*, an important class of decentralized optimization. This problem considers a network of L agents which cooperatively minimize a separable objective function

$$\min \quad \sum_{i=1}^{L} f_i(\tilde{x}), \qquad (1)$$

where $f_i(\tilde{x}) : \mathcal{R}^N \to \mathcal{R}$ is the local objective function known by agent i only, and all the agents have a common optimization variable $\tilde{x}$. This formulation arises in averaging [5], estimation [3, 4, 6, 7], and machine learning [8]. The form of $f_i(\tilde{x})$ can be least squares [5, 6], regularized least squares [3, 7], or more general ones [8]. Among various decentralized approaches for solving (1), the alternating direction method of multipliers (ADMM) [9] demonstrates fast convergence in many applications, e.g., [3, 6, 7]. However, how fast it converges and what factors affect the speed are both unknown. This paper addresses these issues theoretically.

1.1. Our Contributions

The contributions of this paper are twofold. First, we establish the linear convergence rate of the ADMM in decentralized consensus optimization when each local objective function is strongly convex. This result gives a performance guarantee for the ADMM and validates the observation from prior literature. Second, the theoretical convergence rate is a function of the network topology, properties of the local objective functions, and the algorithm parameter. Thus, it provides a guideline of adjusting the network topology, constructing the optimization problem, and tuning the algorithm parameter to accelerate its convergence.

1.2. Related Work

Besides the ADMM, existing methods for (1) include belief propagation [8], incremental optimization [10], and distributed subgradient descent [11]. Belief propagation and incremental optimization require to predefine a tree or loop structure in the network, while the advantage of the ADMM and distributed subgradient descent is that they do not rely on any predefined structures. Distributed subgradient descent fits for asynchronous networks, but suffers from slow convergence. The descent rate of objective value is typically $O(\log(k)/k)$ where k is the number of iterations [12]. The ADMM generally needs synchronous steps taken by all the agents, but has much faster empirical convergence.

Existing convergence rate analysis of the ADMM is restricted to the classic, centralized computation. Generally, when the objective function is strongly convex, the constraints

are linear and non-redundant, the ADMM guarantees linear convergence. However, that is not the case for decentralized consensus optimization, as we will see in Section 2. In [13] an ADMM with restricted stepsizes is proposed and proved to be linearly convergent for certain types of non-strongly convex objective functions. A recent paper [14] shows linear convergence with a strong convexity assumption, and our paper extends the analysis tools therein to the decentralized regime.

A notable work of convergence rate analysis is [15] which proves linear convergence of the ADMM when (1) is an average consensus problem. That is, each local objective function is a simple least squares (for agent i, $f_i(\tilde{x}) = ||\tilde{x} - y_i||_2^2$ where y_i is the measurement of agent i). Since [15] treats the ADMM from the state transition equation perspective, when $f(\tilde{x})$ has a more general form, the matrix spectral analysis on the state transition equation is no longer viable.

2. DECENTRALIZED CONSENSUS OPTIMIZATION

Throughout the paper, we make the following assumptions on the network topology and the local objective functions:

Assumption 1. The network of L agents is bidirectionally connected with E edges (and hence $2E$ directional arcs). We describe it as a symmetric directed graph $\mathcal{G}_d = \{\mathcal{V}, \mathcal{A}\}$ or an undirected graph $\mathcal{G}_u = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is the set of vertexes with $|\mathcal{V}| = L$, $\mathcal{A}$ is the set of arcs with $|\mathcal{A}| = 2E$, and $\mathcal{E}$ is the set of edges with $|\mathcal{E}| = E$.

Assumption 2. Local objective functions are differentiable and strongly convex. For agent i, $\langle \nabla f_i(\tilde{x}_a) - \nabla f_i(\tilde{x}_b), \tilde{x}_a - \tilde{x}_b \rangle \geq m_{f_i} ||\tilde{x}_a - \tilde{x}_b||_2^2$ for any $\tilde{x}_a, \tilde{x}_b \in \mathcal{R}^N$ with $m_{f_i} > 0$.

Assumption 3. Gradients of local objective functions are Lipschitz continuous. For agent i, $||\nabla f_i(\tilde{x}_a) - \nabla f_i(\tilde{x}_b)||_2 \leq M_{f_i} ||\tilde{x}_a - \tilde{x}_b||_2$ for any $\tilde{x}_a, \tilde{x}_b \in \mathcal{R}^N$ with $M_{f_i} > 0$.

To apply the ADMM in solving (1), we reformulate it as

$$\min \quad \sum_{i=1}^{L} f_i(x_i), \qquad (2)$$
$$s.t. \quad x_i = z_{ij}, x_j = z_{ij}, \forall(i,j) \in \mathcal{A}.$$

Here x_i is the local copy of $\tilde{x}$ at agent i and z_{ij} is an auxiliary variable imposing the consensus constraint on two neighboring agents i and j. The optimization variables in (2) are x_i's and z_{ij}'s. Defining $x \in \mathcal{R}^{LN}$ as a vector concatenating all x_i's, $z \in \mathcal{R}^{2EN}$ as a vector concatenating all z_{ij}'s, and $f(x) = \sum_{i=1}^{L} f_i(x_i)$, (2) can be written as a matrix form

$$\min \quad f(x), \qquad (3)$$
$$s.t. \quad Ax + Bz = 0.$$

Here $A = [A_1^T, A_2^T]^T$; $A_1, A_2 \in \mathcal{R}^{2EN \times LN}$ are both composed of $2E \times L$ blocks of $N \times N$ matrices. If $(i,j) \in \mathcal{A}$ and z_{ij} is the pth block of z, then the (p, i)th block of A_1 and the (p, j)th block of A_2 are $N \times N$ identity matrix I_N; otherwise the corresponding blocks are $N \times N$ zero matrix 0_N. And $B = [-I_{2EN}, -I_{2EN}]^T$ with I_{2EN} being a $2EN \times 2EN$ identity matrix.

Under Assumptions 1-3, we have the following lemma.

Lemma 1. Under Assumption 1, (3) is equivalent to (1); under Assumption 2, $f(x)$ is strongly convex, $\langle \nabla f(x_a) - \nabla f(x_b), x_a - x_b \rangle \geq m_f ||x_a - x_b||_2^2$ for any $x_a, x_b \in \mathcal{R}^{LN}$ with $m_f = \min_i m_{f_i}$; under Assumption 3, gradient of $f(x)$ is Lipschitz continuous, $||\nabla f(x_a) - \nabla f(x_b)||_2 \leq M_f ||x_a - x_b||_2$ for any $x_a, x_b \in \mathcal{R}^{LN}$ with $M_f = \max_i M_{f_i}$.

The augmented Lagrangian function of (3) is

$$L_c(x, z, \lambda) = f(x) + \langle \lambda, Ax + Bz \rangle + \frac{c}{2}||Ax + Bz||_2^2,$$

where $\lambda \in \mathcal{R}^{4EN}$ is the Lagrange multiplier and c is a positive penalty factor. At iteration $k + 1$, the ADMM firstly minimizes $L_c(x, z^k, \lambda^k)$ to get x^{k+1}, secondly minimizes $L_c(x^{k+1}, z, \lambda^k)$ to get z^{k+1}, and finally updates λ^{k+1} from x^{k+1} and z^{k+1}. The algorithm is outlined as

$$
\begin{aligned}
x: \quad & \nabla f(x^{k+1}) + A^T\lambda^k + cA^T(Ax^{k+1} + Bz^k) = 0, \\
z: \quad & B^T\lambda^k + cB^T(Ax^{k+1} + Bz^{k+1}) = 0, \qquad (4) \\
\lambda: \quad & \lambda^{k+1} - \lambda^k - c(Ax^{k+1} + Bz^{k+1}) = 0,
\end{aligned}
$$

which can be significantly simplified (to (7) below). First, (4) can be alternatively expressed as

$$
\begin{aligned}
& \nabla f(x^{k+1}) + A^T\lambda^{k+1} + cA^TB(z^k - z^{k+1}) = 0, \\
& B^T\lambda^{k+1} = 0, \qquad (5) \\
& \lambda^{k+1} - \lambda^k - c(Ax^{k+1} + Bz^{k+1}) = 0.
\end{aligned}
$$

Letting $\lambda^T = [\beta^T, \gamma^T]^T$ where $\beta, \gamma \in \mathcal{R}^{2EN}$, we know $\beta^{k+1} = -\gamma^{k+1}$ from the second equation of (5). If choosing $\beta^0 = -\gamma^0$ such that $B^T\lambda^0 = 0$ and choosing $z^0 = \frac{c}{2}M_+^Tx^0$, the updates in (5) is equivalent to

$$
\begin{aligned}
& \nabla f(x^{k+1}) + M_-\beta^{k+1} - cM_+(z^k - z^{k+1}) = 0, \\
& \beta^{k+1} - \beta^k - \frac{c}{2}M_-^Tx^{k+1} = 0, \qquad (6) \\
& \frac{1}{2}M_+^Tx^k - z^k = 0.
\end{aligned}
$$

Here $M_+ = A_1^T + A_2^T$ and $M_- = A_1^T - A_2^T$. We often choose β^0 in the column space of M_-^T (e.g., $\beta^0 = 0$), such that β^{k+1} also lies in the column space of M_-^T; the reason will be given in Section 3.

If further letting $W \in \mathcal{R}^{LN \times LN}$ with its (i,i)th block being the degree of agent i multiplying I_N and other blocks being 0_N, $\alpha = M_-\beta$, $L_+ = \frac{1}{2}M_+M_+^T$, and $L_- = \frac{1}{2}M_-M_-^T$ [7], we can get a simple decentralized algorithm

$$
\begin{aligned}
x: \quad & \nabla f(x^{k+1}) + \alpha^k + 2cWx^{k+1} - cL_+x^k = 0, \\
\alpha: \quad & \alpha^{k+1} - \alpha^k - cL_-x^{k+1} = 0. \qquad (7)
\end{aligned}
$$

In practice, we use the simpler updates (7); for the analysis purpose, this paper considers the equivalent form (6).

Remark 1. The newly introduced matrices M_+, M_-, L_+, and L_- are determined by network topology. Note that $L_+ = \frac{1}{2}M_+M_+^T$ is an extended signless Laplacian matrix of $\mathcal{G}_u$ and $L_- = \frac{1}{2}M_-M_-^T$ is an extended Laplacian matrix of $\mathcal{G}_u$; by extended we mean replacing 1's by I_N's, -1's by

$-I_N$'s, and 0's by 0_N's [16, 17]. Denote $\sigma_{\max}(M_+)$ and $\tilde{\sigma}_{\min}(M_-)$ as the largest singular value of M_+ and the smallest nonzero singular value of M_-, respectively; $\sigma_{\max}(M_+) = \sqrt{\sigma_{\max}(2L_+)}$ and $\tilde{\sigma}_{\min}(M_-) = \sqrt{\tilde{\sigma}_{\min}(2L_-)}$ are both measures of connectedness. Larger $\sigma_{\max}(M_+)$ and $\tilde{\sigma}_{\min}(M_-)$ mean better network connectivity.

Remark 2. Convergence of (4) to the global optimal solution of (3) can be established from the convergence property of the ADMM [9]. However, deriving its rate of convergence in our setting is nontrivial. In (3), the objective function $f(x)$ is strongly convex with respect to x, but not strongly convex with respect to x and z. Further, the linear constraints are always redundant when the network is connected.

3. CONVERGENCE RATE ANALYSIS

Taking $k \to +\infty$ in (6) yields the KKT condition of (3)

$$
\begin{aligned}
\nabla f(x^*) + M_- \beta^* &= 0, \\
M_-^T x^* &= 0, \\
\tfrac{1}{2} M_+^T x^* - z^* &= 0,
\end{aligned}
\tag{8}
$$

Where (x^*, z^*) is the unique primal optimal solution (uniqueness comes from the strong convexity of $f(x)$) and $\lambda^* = [\beta^{*T}, -\beta^{*T}]^T$ is one of the dual optimal solutions of (3).

Let us introduce

$$
u = \begin{pmatrix} z \\ \beta \end{pmatrix}, G = \begin{pmatrix} cI_{2EN} & 0_{2EN} \\ 0_{2EN} & \frac{1}{c}I_{2EN} \end{pmatrix}.
$$

The following lemma shows Q-linear convergence of the sequence $\{\|u^k - u^*\|_G = \sqrt{(u^k - u^*)^T G(u^k - u^*)}\}$ to 0.

Theorem 1. For any $\mu > 1$, there exists a positive number

$$
\delta = \min\left\{\frac{(\mu-1)\tilde{\sigma}_{\min}^2(M_-)}{\mu\sigma_{\max}^2(M_+)}, \frac{m_f}{\frac{c}{4}\sigma_{\max}^2(M_+) + \frac{\mu}{c}M_f^2\tilde{\sigma}_{\min}^{-2}(M_-)}\right\}
\tag{9}
$$

such that

$$
(1+\delta)\|u^{k+1} - u^*\|_G^2 \leq \|u^k - u^*\|_G^2.
\tag{10}
$$

The primal sequence $\{x^k\}$ is R-linear convergent to x^* with

$$
m_f\|x^{k+1} - x^*\|_2^2 \leq \|u^k - u^*\|_G^2.
\tag{11}
$$

Proof. Subtracting the three equations in (8) from the corresponding equations in (6) yields

$$
\nabla f(x^{k+1}) - \nabla f(x^*) = cM_+(z^k - z^{k+1}) - M_-(\beta^{k+1} - \beta^*),
\tag{12}
$$

$$
\tfrac{c}{2}M_-^T(x^{k+1} - x^*) = \beta^{k+1} - \beta^k,
\tag{13}
$$

$$
\tfrac{1}{2}M_+^T(x^{k+1} - x^*) = z^{k+1} - z^*,
\tag{14}
$$

respectively. Therefore, we can bound $\|x^{k+1} - x^*\|_2^2$ with

$$
\begin{aligned}
& m_f\|x^{k+1} - x^*\|_2^2 \\
\leq\ & \langle x^{k+1} - x^*, \nabla f(x^{k+1}) - \nabla f(x^*) \rangle \\
=\ & \langle x^{k+1} - x^*, cM_+(z^k - z^{k+1}) - M_-(\beta^{k+1} - \beta^*) \rangle \\
=\ & \langle x^{k+1} - x^*, cM_+(z^k - z^{k+1}) \rangle \\
& + \langle x^{k+1} - x^*, -M_-(\beta^{k+1} - \beta^*) \rangle \\
=\ & 2c\langle z^k - z^{k+1}, z^{k+1} - z^* \rangle + \tfrac{2}{c}\langle \beta^k - \beta^{k+1}, \beta^{k+1} - \beta^* \rangle
\end{aligned}
$$

$$
\begin{aligned}
=\ & 2(u^k - u^{k+1})^T G(u^{k+1} - u^*) \\
=\ & \|u^k - u^*\|_G^2 - \|u^{k+1} - u^*\|_G^2 - \|u^k - u^{k+1}\|_G^2,
\end{aligned}
\tag{15}
$$

where the inequality follows from the strong convexity of $f(x)$; the first equality follows from (12); the third equality comes from (13) and (14).

Next we show

$$
\|u^k - u^{k+1}\|_G^2 + m_f\|x^{k+1} - x^*\|_2^2 \geq \delta\|u^{k+1} - u^*\|_G^2
\tag{16}
$$

or equivalently

$$
\begin{aligned}
& c\|z^k - z^{k+1}\|_2^2 + \tfrac{1}{c}\|\beta^k - \beta^{k+1}\|_2^2 + m_f\|x^{k+1} - x^*\|_2^2 \\
& \geq \delta(c\|z^{k+1} - z^*\|_2^2 + \tfrac{1}{c}\|\beta^{k+1} - \beta^*\|_2^2).
\end{aligned}
\tag{17}
$$

From the third equation of (14) we have

$$
\begin{aligned}
& \|z^{k+1} - z^*\|_2^2 \\
=\ & \tfrac{1}{4}\|M_+^T(x^{k+1} - x^*)\|_2^2 \\
\leq\ & \tfrac{1}{4}\sigma_{\max}^2(M_+)\|(x^{k+1} - x^*)\|_2^2.
\end{aligned}
\tag{18}
$$

On the other hand,

$$
\begin{aligned}
& c^2\sigma_{\max}^2(M_+)\|z^{k+1} - z^k\|_2^2 + (\mu - 1)M_f^2\|x^{k+1} - x^*\|_2^2 \\
\geq\ & \|cM_+(z^{k+1} - z^k)\|_2^2 + (\mu - 1)\|\nabla f(x^{k+1}) - \nabla f(x^*)\|_2^2 \\
\geq\ & (1 - \tfrac{1}{\mu})\|M_-(\beta^{k+1} - \beta^*)\|_2^2 \\
\geq\ & (1 - \tfrac{1}{\mu})\tilde{\sigma}_{\min}^2(M_-)\|\beta^{k+1} - \beta^*\|_2^2,
\end{aligned}
\tag{19}
$$

where the first inequality follows from Lipschitz continuity of $\nabla f(x)$; the second inequality follows from (12) and the basic inequality $\|a + b\|_2^2 \geq (1 - \mu)\|a\|_2^2 + (1 - \tfrac{1}{\mu})\|b\|_2^2, \forall \mu > 0$; the third inequality holds since we choose β^0 in the column space of M_-^T such that β^{k+1} and β^* also lie in the column space of M_-^T, as discussed in Section 2.

From (18) and (19), we know δ in (9) satisfies (17), and equivalently, (16). Assumptions 1 and 2 guarantee that $m_f > 0$ and $\tilde{\sigma}_{\min}(M_-) > 0$, such that $\delta > 0$. Combining (15) and (16) further leads to (10).

Above we show that the sequence $\{[c^{\frac{1}{2}}z^{kT}, c^{-\frac{1}{2}}\beta^{kT}]^T\}$ is Q-linear convergent to $\{[c^{\frac{1}{2}}z^{*T}, c^{-\frac{1}{2}}\beta^{*T}]^T\}$ with rate $\sim (1 + \delta)^{-\frac{1}{2}}$. Note that it does not mean Q-linear convergence of either the primal sequence $\{z^k\}$ or dual sequence $\{\beta^k\}$. However, from (15) we know that (11) holds, i.e., the primal sequence $\{x^k\}$ is R-linear convergent to x^*. $\square$

4. TUNING OF CONVERGENCE RATE

This section discusses how the convergence rate is influenced by the network topology, property of the local objective functions, and algorithm parameter c.

In (9), the value of δ is related with an arbitrary $\mu > 1$, the algorithm parameter c, strongly convexity constant m_f of $f(x)$ and Lipschitz continuity constant M_f of $\nabla f(x)$, as well as $\sigma_{\max}(M_+)$ and $\tilde{\sigma}_{\min}(M_-)$. For convenience, we define the condition number of $f(x)$ as $\kappa_f = \frac{M_f}{m_f}$ and the condition

number of G_d and G_u as $\kappa_G = \frac{\sigma_{\max}(M_+)}{\tilde{\sigma}_{\min}(M_-)} = \sqrt{\frac{\sigma_{\max}(L_+)}{\tilde{\sigma}_{\min}(L_-)}}$. Recall that larger $\tilde{\sigma}_{\min}(L_-)$ and $\sigma_{\max}(L_+)$ mean better connectedness, but the latter is a weaker measure of network connectivity [16, 17]. Roughly speaking, larger κ_G means worse connectedness.

To maximize δ, we choose c in (9) as

$$c = \frac{2\mu^{\frac{1}{2}} M_f}{\sigma_{\max}(M_+)\tilde{\sigma}_{\min}(M_-)} \tag{20}$$

such that

$$\delta = \min\left\{ \frac{(\mu-1)\tilde{\sigma}_{\min}^2(M_-)}{\mu\sigma_{\max}^2(M_+)}, \frac{m_f\tilde{\sigma}_{\min}(M_-)}{\mu^{\frac{1}{2}} M_f \sigma_{\max}(M_+)} \right\}. \tag{21}$$

The first and second terms in the minimum are monotonically increasing and decreasing with respect to μ when $\mu > 1$, respectively. Hence we choose

$$\frac{1}{\mu} = 1 + \frac{\kappa_G^2}{2\kappa_f^2} - \frac{\kappa_G}{2\kappa_f}\sqrt{\frac{\kappa_G^2}{\kappa_f^2} + 4} < 1$$

to maximize δ and obtain

$$\delta = \frac{1}{2\kappa_f}\sqrt{\frac{1}{\kappa_f^2} + \frac{4}{\kappa_G^2}} - \frac{1}{2\kappa_f^2}. \tag{22}$$

The value of δ in (22) is monotonically decreasing with respect to $\kappa_f \geq 1$ and $\kappa_G > 0$, and its limit as $\kappa_f \to \infty$ or $\kappa_G \to \infty$ is 0. This conclusion suggests that smaller condition number κ_f of $f(x)$ and smaller condition number κ_G of the graph lead to faster convergence. Therefore, when we choose c as recommended in (20), a well-connected network and well-conditioned local objective functions lead to fast convergence in general.

5. SIMULATION RESULTS

In the simulation, we consider a network of $L = 100$ agents. Agent i measures the signal $\tilde{x}_0$ from $y_i = \Phi_i\tilde{x}_0 + e_i$ where $\tilde{x}_0, y_i, e_i \in \mathcal{R}^2$ and $\Phi_i \in \mathcal{R}^{2\times2}$. Elements in $\tilde{x}_0, \Phi_i \sim \mathcal{N}(0,1)$ and elements in $e_i \sim \mathcal{N}(0,0.01)$. The local objective function held by agent i is $f_i(\tilde{x}) = ||y_i - \Phi_i\tilde{x}||_2^2$. We use the same data throughout the simulation such that κ_f is fixed. This way, we can focus our discussion on the impact of c and κ_G on the convergence rate. Convergence of the algorithm is measured by the squared residual $||x^k - x^*||_2^2$. We let K be the first iteration in which the squared residual first reaches a tolerance of 10^{-20}.

The first two simulations use the recommended c in (20) as the algorithm parameter. First, given three different values of κ_G, Fig. 1 depicts the linear convergence of the ADMM.

Second, we randomly generate 200 networks with different values of κ_G and consider the influence of κ_G on the convergence rate. Fig. 2 shows that when $\kappa_G > 2$, i.e., the network connectivity is not strong, K is positively correlated with κ_G; that validates our theoretical analysis. However,
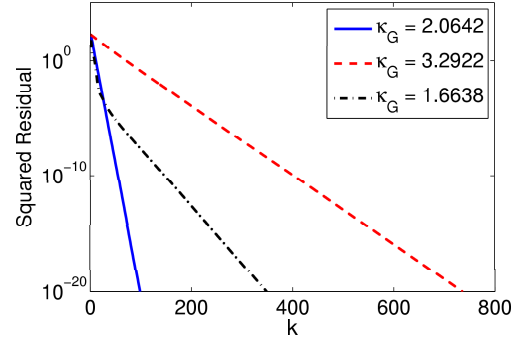

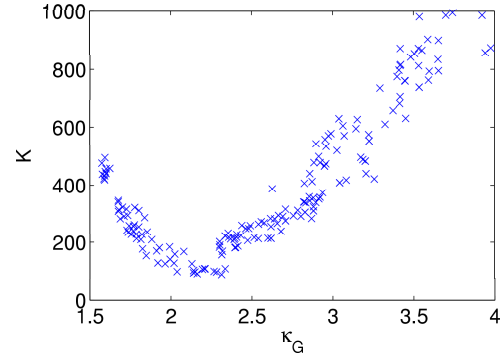**Fig. 1**. Iteration number k vs. squared residual.

Fig. 2. Iteration number of convergence K vs. κ_G.

when the network connectivity is strong (and thus the convergence is fast), the theoretical analysis no longer holds.

Third, we compare the recommended c in (20) and the practically optimal c for networks with different κ_G, as shown in Fig. 3. When the network connectivity is medium ($\kappa_G \simeq 2$), the recommended c is close to the optimal one. The recommended c is smaller than the optimal one when ($\kappa_G < 2$) and larger when ($\kappa_G > 2$). Note that in deriving (9), we introduce several inequalities. Therefore, the recommended c in (20) is not giving the tightest bound. Improving the choice of c will be a future direction of our research.
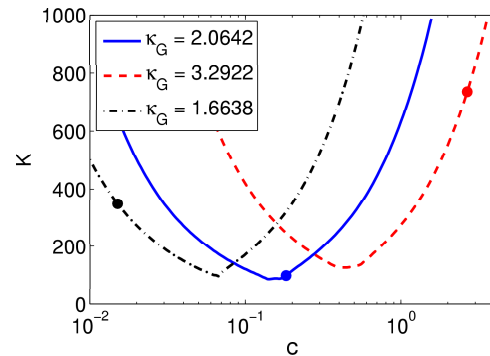

Fig. 3. Iteration number of convergence K vs. c; solid circles are using the recommended c in (20).

6. REFERENCES

[1] A. Dimakis, S. Kar, J. Moura, M. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," Proceedings of the IEEE, vol. 98, pp. 1847–1864, 2010

[2] W. Ren, R. Beard, and E. Atkins, "Information consensus in multivechicle cooperative control: collective group behavier through local interaction," IEEE Control Systems Magazine, vol. 27, pp. 71–82, 2007

[3] Q. Ling and Z. Tian, "Decentralized sparse signal recovery for compressive sleeping wireless sensor networks," IEEE Transactions on Signal Processing, vol. 58, pp. 3816–3827, 2010

[4] V. Kekatos and G. Giannakis, "Distributed robust power system state estimation," IEEE Transactions on Power Systems, In Press

[5] L. Xiao, S. Boyd, and S. Kim, "Distributed average consensus with least-mean-square deviation," Journal of Parallel and Distributed Computing, vol. 67, pp. 33–46, 2007

[6] I. Schizas, A. Ribeiro, and G. Giannakis, "Consensus in ad hoc WSNs with noisy links - Part I: distributed estimation of deterministic signals," IEEE Transactions on Signal Processing, vol. 56, pp. 350–364, 2008

[7] J. Bazerque, G. Mateos, and G. Giannakis, "Group-lasso on splines for spectrum cartograph," IEEE Transactions on Signal Processing, vol. 59, pp. 4648–4663, 2011

[8] J. Predd, S. Kulkarni, and H. Poor, "A collaborative training algorithm for distributed learning," IEEE Transactions on Information Theory, vol. 55, pp. 1856–1871, 2009

[9] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Second Edition, Athena Scientific, 1997

[10] M. Rabbat and R. Nowak, "Quantized incremental algorithms for distributed optimization," IEEE Journal of Selected Areas in Communications, vol. 23, pp. 798–808, 2006

[11] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," IEEE Transactions on Automatic Control, vol. 54, pp. 48–61, 2009

[12] K. Tsianos and M. Rabbat, "Distributed strongly convex optimization," Manuscript

[13] M. Hong and Z. Luo, "On the linear convergence of the alternating direction method of multipliers," Manuscript

[14] W. Deng and W. Yin, "On the global and linear convergence of the generalized alternating direction method of multipliers," Manuscript

[15] T. Erseghe, D. Zennaro, E. Dall'Anese, and L. Vangelista, "Fast consensus by the alternating direction multipliers method," IEEE Transactions on Signal Processing, vol. 59, pp. 5523–5537, 2011

[16] F. Chung, *Spectral Graph Theory*, American Mathematical Society, 1997

[17] Y. Chen and L. Wang, "Sharp bounds for the largest eigenvalue of the signless Laplacian of a graph," Linear Algebra and its Applications, vol. 433, pp. 908–913, 2010