Decentralized Gauss-Newton method for nonlinear least squares on wide area network

Lanchao Liu*, Qing Ling†, and Zhu Han*

*Electrical and Computer Engineering Department, University of Houston, U.S.A
†Department of Automation, University of Science and Technology of China, China
E-mail: lliu10@uh.edu

Abstract. This paper presents a decentralized approach of Gauss-Newton (GN) method for nonlinear least squares (NLLS) on wide area network (WAN). In a multi-agent system, a centralized GN for NLLS requires the global GN Hessian matrix available at a central computing unit, which may incur large communication overhead. In the proposed decentralized alternative, each agent only needs local GN Hessian matrix to update iterates with the cooperation of neighbors. The detail formulation of decentralized NLLS on WAN is given, and the iteration at each agent is defined. The convergence property of the decentralized approach is analyzed, and numerical results validate the effectiveness of the proposed algorithm.

1. Introduction

The significant importance of the nonlinear least squares (NLLS) in applications of state estimation in power system [1], signal detection in wireless networks [2], and target tracking in mobile networks [3] has been appreciated for decades. The Gauss-Newton (GN) method, which can be seen as a modification of the Newton’s method, is widely used to solve the NLLS [4]. The GN method finds the minimizer of the NLLS in an iterative fashion, and obtains the solution with provable local optimality and convergence rate. In this work, a decentralized GN method for NLLS on WAN is presented. In particular, only local GN Hessian matrix is used and limited communication is performed between neighboring agents. The decentralized optimization enjoys the advantage of scalability to network size, robustness to dynamic topologies and privacy preservation in data-sensitive applications [5–8]. A detailed formulation of the decentralized optimization problem for NLLS on WAN is provided, and the updating rule at each agent is explicitly given. Finally, we investigate the convergence property of the proposed algorithm, which turns out the convergence rate is related to the number of agents as well as the minimum node degree in the network. Numerical tests validate the performance of the proposed algorithm.

The contributions of this work are threefold. Firstly, we do not assume any specific structure for the global Hessian matrix, and proposed a decentralized GN method for NLLS use only local Hessian matrix. Whereas the localization application in [3] has a block-wise Jacobian matrix which is convenient to decompose, and needs global Hessian through network-wide consensus. [1] proposes a generalized gossip-based GN method, which still requires global Hessian through Gossip exchange. Secondly, we proved the local superlinearly convergence property of the proposed algorithm. Finally, we validated the proposed method through numerical simulations.
2. Problem Formulation

2.1. Centralized Nonlinear Least Squares

Consider an unknown variable \( \tilde{x} \in \mathbb{R}^n \) in a network, and \( m \) observations are obtained through a vector-valued function \( h(\tilde{x}) = (h_1(\tilde{x}), \ldots, h_m(\tilde{x})) : \mathbb{R}^n \to \mathbb{R}^m \). Each entry in function \( h(\tilde{x}) \) is a real value function and not necessarily convex. Let \( z = h(\tilde{x}) + e \), where \( e \) stands for measurement errors, which are assumed zero mean and known covariance \( R \in \mathbb{R}^{m \times m} \). The unknown variable \( \tilde{x} \in \mathbb{R}^n \) can be estimated by the NLLS as follows:

\[
\text{minimize} \quad (z - h(\tilde{x}))^T R^{-1} (z - h(\tilde{x})).
\] (1)

The GN method can be adopted to solve (1) given that all the observations and functions are available at a central computing node. Specifically, define \( r(\tilde{x}) = R^{-1/2} (z - h(\tilde{x})) \) and its Jacobian \( J(\tilde{x}) = \partial r(\tilde{x}) / \partial \tilde{x} \). Let \( F(\tilde{x}) = ||r(\tilde{x})||^2 \), and problem (1) can be solved iteratively as 
\[
\tilde{x}^{k+1} = \tilde{x}^k + \alpha^k d^k,
\]
where the descent direction \( d^k \) at each iteration can be obtained by solving 
\[
J^T(\tilde{x}^k)J(\tilde{x}^k) d = J^T(\tilde{x}^k)r(\tilde{x}^k). \] (2)

The GN method can solve the problem (1) at a superlinear convergence rate with order at least two under Assumption 1. Note that majority of NLLS problems are non-convex, and in this paper we only consider the local convergence property of the algorithm. We assume Assumption 1 holds throughout the paper.

**Assumption 1.** Consider a function \( F(\tilde{x}) \), suppose the following assumptions hold.

(i) The function \( F(\tilde{x}) \) is continuous, differentiable and bounded below.

(ii) There exists a vector \( \tilde{x}^* \) such that the greatest lower bound can be achieved.

(iii) For \( \delta > 0 \), let \( S_\delta \) denote the sphere \( \{\tilde{x} : ||\tilde{x} - \tilde{x}^*||^2 \leq \delta\} \). The Hessian matrix \( J^T(\tilde{x})J(\tilde{x}) \) is invertible in the sphere \( S_\delta \). For some \( L > 0 \), \( M > 0 \), \( \delta > 0 \), and for all \( \tilde{x} \) and \( \tilde{y} \) in \( S_\delta \), we have \( ||J^T(\tilde{x})J(\tilde{x}) - J^T(\tilde{y})J(\tilde{y})|| \leq L||\tilde{x} - \tilde{y}|| \) and \( \sigma_{\min}(J^T(\tilde{x})J(\tilde{x})) \geq \frac{1}{M} \geq 0 \).

2.2. Decentralized Nonlinear Least Squares

In a multi-agent system consisted of \( N \) networked agents, each agent is engaged in its own monitoring and controlling task in the network, and at the same time, each agent is cooperating with other agents in the context of estimating the global system states \( \tilde{x} \). Suppose these \( N \) agents are loosely coupled; there is very little, if any, central coordination and control among those agents, and each agent is able to exchange information with its neighbors. The system states \( \tilde{x} \) can be obtained by solving the following optimization problem:

\[
\text{minimize} \quad f(\tilde{x}) = \sum_{i=1}^{N} (z_i - h_i(\tilde{x}_i))^T R_i^{-1} (z_i - h_i(\tilde{x}_i)),
\]

subject to \( \tilde{x}_i = \ldots = \tilde{x}_N \).

(3)

where \( z_i \) is the local observation which is a subset of \( z \), i.e., \( z = (z_1; \ldots ; z_N) \), and \( h_i \) is the local observation function which is a subset of \( h \), i.e., \( h = (h_1; \ldots ; h_N) \). \( R_i^{-1} \) is the covariance matrix of local noise vector \( e_i \). \( \tilde{x}_i \in \mathbb{R}^n \) is the local duplicate of \( \tilde{x} \), and \( \tilde{x} = (\tilde{x}_1; \ldots ; \tilde{x}_N) \in \mathbb{R}^{Nn} \).

3. Proposed Algorithms

For concreteness, the network model of agents is first described. Specifically, consider a graph \( G = (V, E) \). \( V \) represents the set of agents, and \( E \) represents the set of communication links between each pair of agents. An arc \( e \) is associated with an order pair \( (i, j) \) as \( e \sim (i, j) \), which means the information is transmitted from agent \( i \) to agent \( j \). Assume the graph formed by the
agents is connected. By introducing the auxiliary variables $w_{ij} \in \mathbb{R}^n$ associated with each arc $e \sim (i, j) \in \mathcal{E}$, problem (3) can be reformulated as:

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{N} (z_i - h_i(x_i))^\top R^{-1}_i(z_i - h_i(x_i)), \\
\text{subject to} & \quad x_i = w_{ij}, \quad x_j = w_{ij}, \quad \forall (i, j) \in \mathcal{E},
\end{align*}
\]

(4)

where $w_{ij}$ is used to enforce the equality of variables $x_i$ and $x_j$ for agents $i$ and $j$ connected by arc $(i, j)$. We use compact notations in the following for the sake of discussion simplicity. Concatenating $w_{ij}$ in vector $w$, problem (4) can be reformulated as:

\[
\begin{align*}
\text{minimize} & \quad f(x), \quad \text{subject to} \quad A_s x - w = 0, \quad A_d x - w = 0,
\end{align*}
\]

(6)

where $A_s$ and $A_d$ are the extended arc source matrix and extended arc destination matrix for the network graph $G$, respectively. Stacking $A_s$ and $A_d$ form $A = [A_s; A_d] \in \mathbb{R}^{2Mn \times Nn}$. The optimization problem (6) reduces to:

\[
\begin{align*}
\text{minimize} & \quad f(x), \quad \text{subject to} \quad A x + B w = 0,
\end{align*}
\]

(7)

where $B = [-I_{Mn}; -I_{Mn}] \in \mathbb{R}^{2Mn \times Mn}$. The GN method is utilized to solve the optimization problem (7), where the updates of $x$ are implemented in a decentralized fashion. The local update rule at each agent is given in the following proposition.

**Proposition 1.** Consider iterates $x^k$ and $z^k$ with the initialization $E_s x^0 = 2w^0$, the iterates $x^k$ at each agent $i$ can be iteratively generated by the following recursions for $k > 0$:

\[
x^{k+1}_i = x^k_i + \alpha^k d^k_i,
\]

(8)

where $\alpha^k$ is a positive constant, and $d^k_i$ is the descent direction which can be determined as

\[
(J^\top_i(x^k_i)J_i(x^k_i) + \rho \nu_i I_n) d^k_i = J^\top_i(x^k_i) r_i(x^k_i) + \rho [\nu_i x_i - \frac{1}{2} \sum_{j \in N_i} (x^{k-1}_i + x^{k-1}_j)],
\]

where $r_i(x^k_i) = R^{-1/2}_i(z_i - h_i(x_i))$ and $J_i(x^k_i)$ is its Jacobian $J_i(x^k_i) = \partial r_i(x^k_i)/\partial x_i$. $\nu_i$ is the degree of agent $i$ in the network, and $N_i$ denotes the neighbors of agent $i$.

**Proof.** The proof is omitted here due to the space limitation.

The local convergence property and convergence rate of the proposed decentralized approach are given by the following theorem.

**Theorem 1.** Suppose the Assumption 1 holds, and the start point of each agent $x^0_i$ is in $S_3$. The sequence $\{x^k\}$ generated by the update rule given in Proposition 1 is defined, and converges to $x^* = \{\tilde{x}^*; \ldots; \tilde{x}^*\}$. Furthermore, we have

\[
\|x^{k+1} - x^*\| \leq \frac{ML\sqrt{N}}{2(1 + M\rho \max(\nu_i))} \|x^k - x^*\|^2,
\]

(9)

where $N$ is the number of agents and $\max(\nu_i)$ is the maximum node degree in the network.

**Proof.** The proof is omitted here due to the space limitation.

Remark that Theorem 1 illustrates the local convergence property of the proposed decentralized approach, which converges to the optimal solution superlinearly. The convergence rate is related to the number of agents as well as the minimum node degree in the network.
4. Numerical Results

A bidirectionally connected ring network composed of $N = 100$ agents is considered here, in which each agent connects to exactly two other agents. The unknown system states in the network is $\tilde{x} \in \mathbb{R}^3$. The observation function $h_i(x_i)$ at each agent $i$ is defined as:

$$h_i(x_i) = a_i(x_i(1)^2 + x_i(2)^2) + b_i x_i(2) \sin(x_i(2) - x_i(3)) + c_i x_i(1)x_i(2),$$

where $a_i$, $b_i$, and $c_i$ are i.i.d. random variables follow the standard normal distribution. It is seen that observation function $h_i(x_i)$ is a nonlinear function with a quadratic term, a trigonometric term and a cross product term. The agents in the network work cooperatively to estimate the unknown system states $\tilde{x}$ in a decentralized fashion. The convergence result is depicted in Fig. 1. It is shown that the proposed algorithm is effective in the sense that after a moderate number of iterations, the iterates converge to the optimal values. To investigate the performance of the proposed decentralized approach at each agent, the root-mean-square error (RMSE) of the estimate at each agent is calculated. The best RMSE (agent 2), the worst RMSE (agent 52) and the average RMSE are described in Fig. 2. It can be seen that at each agent, the RMSE decreases as the iteration increases. Furthermore, the convergence rate at each agent is different.

![Figure 1](image1.png)  
**Figure 1.** Convergence performance of the proposed algorithm.

![Figure 2](image2.png)  
**Figure 2.** RMSE performance of the proposed algorithm.

5. Conclusions

We have presented a decentralized approach of GN method for NLLS on WAN. We have given the updating rule at each agent explicitly, and investigated the local convergence of the proposed algorithm. Numerical simulations validated the effectiveness of the proposed algorithm.

References

[1] Li X and Scaglione A 2013 *IEEE Transactions on Signal Processing* **61** 5231–5246
[2] Stoica P, Mose R, Friedlander B and Soderstrom T 1989 *IEEE Transaction on Acoustics, Speech and Signal Processing* **37** 378–392
[3] Bejar B, Belanovic P and Zazo S 2010 2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR) (Pacific Grove, CA)
[4] Bertsekas D P 1999 *Nonlinear Programming* (Nashua, USA: Athena Scientific)
[5] Bertsekas D P and Tsitsiklis J N 1989 *Parallel and Distributed Computing: Numerical Methods* (NJ, USA: Prentice-Hall)
[6] Nedic A and Ozdaglar A 2009 *IEEE Transaction on Automatic Control* **54** 48–61
[7] Ram S, Nedic A and Veeravalli V 2010 *Journal of Optimization Theory and Applications* **147** 516–545
[8] Srivastava K and Nedic A 2011 *IEEE Journal of Selected Topics in Signal Processing* **5** 772–790