UC Berkeley

Building Efficiency and Sustainability in the Tropics (SinBerBEST)

Title

Improved System Identification with Renormalization Group

Permalink

<https://escholarship.org/uc/item/2426w8k9>

Author

Qing-Guo, Wang

Publication Date

2013-06-12

Peer reviewed

Improved System Identification with Renormalization Group

Qing-Guo Wang, *Member, IEEE*, Chao Yu and Yong Zhang

Abstract—This paper proposes an improved system identification method with Renormalization Group. Renormalization Group is applied to a fine data set to obtain a coarse data set. The least-squares algorithm is performed on the coarse data set. The theoretical analysis under certain conditions shows that the parameter estimation error could be reduced. The proposed method is illustrated with examples.

I. INTRODUCTION

System identification is concerned with building mathematical models of dynamical systems from measured data [1]. The ordinary least-squares (OLS) method has since been the dominant algorithm for parameter estimation due to its simplicity in concept and convenience in implementation. Given a data set, one can always get an estimate of OLS. It is known that the OLS estimate could be biased for a regression model with noise. And it is very challenging to study the properties of OLS estimate analytically.

With regard to the asymptotic properties of the least square estimates, it is known that the OLS estimate will converge to its real value when the system is with white noise. Otherwise, when the system is disturbed by correlated noise, the OLS estimate could be biased. Griliches [2] gives the expression of the bias for ARAR(1,1) models. Phillips [3] analyzes the bias for simple first order ARARX models. Stocker [4] presents an expression of the bias for a common model but only gives very simple examples for illustration. He argues that for more complicated models (e.g. models with higher order or with several exogenous variables), it is not practicable any more to get fully parameterized bias formulas. These would become very extensive even if the order of the model or the number of exogenous variables increases only slightly. Zheng [5] proposes a bias-eliminated least squares (BELS) algorithm to identify system parameters. The instrumental variable (IV) method [6] is also a very efficient way to avoid correlated noise and reduce the estimate bias.

In this paper, we present an improved system identification method with Renormalization Group (RG). Technically, the proposed method groups the given data set into a RG data set.

*This research is funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

Q.-G. Wang is with the Department of Electrical and Computer Engineering, National University of Singapore, 119260, Singapore (email: elewqg@nus.edu.sg)

C. Yu is with the Department of Electrical and Computer Engineering, National University of Singapore, 117576, Singapore (email: yuchao@nus.edu.sg)

Y. Zhang is with the Qiming Venture Partners, Shanghai, 200121, China (email: yong.zhang@qimingvc.com)

Performing the least squares algorithm, the proposed method obtains an estimate based on the given data and another estimate based on the RG data. Through comparisons of the two estimates, we find the proposed method could get a better estimate under certain conditions. The contributions of this article can be summarized as follows.

- The proposed method forms a data set based on the system inputs and outputs, and produces a new data set from the given data set with the members of the former different from those of the latter, whereas none of the existing methods has a similar idea.
- We present theoretical analysis and simulation results for an academic model to illustrate the effectiveness of our method.

The rest of this paper is organized as follows. Section II states our problem and motivation. Section III details the theoretical analysis. Section IV presents the simulation examples. Section V concludes the paper.

II. PROBLEM STATEMENT AND MOTIVATION

In this section, we first describe the system and present some details of OLS estimation. Then we state the problem and give our motivation. In the end, we illustrate the idea of the proposed method by a simple example.

A. System description and OLS estimation

Consider a linear dynamic model

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_n y_{t-n} + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_m u_{t-m} + \varepsilon_t, \quad (1)$$

where u_t is input, y_t is output and ε_t is noise.

We rewrite (1) as

$$y_t = \varphi_t^T \theta + \varepsilon_t, \quad (2)$$

where

$$\theta^T = [\alpha_1, \dots, \alpha_n, \beta_0, \dots, \beta_m], \quad (3)$$

and

$$\varphi_t^T = [y_{t-1}, \dots, y_{t-n}, u_t, \dots, u_{t-m}]. \quad (4)$$

The parameter vector θ is chosen to minimize the loss function

$$J = \frac{1}{2} \sum_{t=1}^N (y_t - \varphi_t^T \theta)^2. \quad (5)$$

The OLS estimate is given by

$$\hat{\theta} = (\Phi^T \Phi)^{-1} (\Phi^T Y), \quad (6)$$

where $Y = [y_1, \dots, y_N]^T$, $\Phi = [\varphi_1, \dots, \varphi_N]^T$ and $\mathcal{E} = [\varepsilon_1, \dots, \varepsilon_N]^T$. The estimation error is given by

$$\begin{aligned}\Delta\theta &= \hat{\theta} - \theta \\ &= (\Phi^T \Phi)^{-1} (\Phi^T Y) - \theta \\ &= (\Phi^T \Phi)^{-1} (\Phi^T (\Phi \theta + \mathcal{E})) - \theta \\ &= (\Phi^T \Phi)^{-1} (\Phi^T \mathcal{E}).\end{aligned}\quad (7)$$

With different weights w_t assigned to the measurements [1], the criterion (5) becomes

$$J = \frac{1}{2} \sum_{t=1}^N w_t (y_t - \varphi_t^T \theta)^2.$$

Then the resulting estimate is given by

$$\hat{\theta}_W = (\Phi^T W \Phi)^{-1} (\Phi^T W Y), \quad (8)$$

where $W = \text{diag}(w_1, \dots, w_N)$ is the weighting matrix. The estimation error is given by

$$\begin{aligned}\Delta\theta_W &= \hat{\theta}_W - \theta \\ &= (\Phi^T W \Phi)^{-1} (\Phi^T W \mathcal{E}).\end{aligned}$$

B. Idea of the proposed method

This paper studies the estimation error of OLS analytically and find some way to reduce it. We present an improved identification method using Renormalization Group (RG). RG was first proposed to study the critical phenomena in the quantum field in 1971 by Kenneth G. Wilson, who won the Nobel Prize for physics in 1982 [7], due to this great contribution. RG designs some Renormalization Group transformation (RGT) to relate macroscopic physics quantity to microscopic one and invokes “scale invariance” to solve the problem.

Considering the system (2), we first form a data $S_t^T = (y_t, \varphi_t^T)$, $t = 1, 2, \dots, N$ and construct a data set $\mathcal{S} = \{S_1, S_2, \dots, S_N\}$. Then we perform a RGT on $\mathcal{S}$ to obtain a transformed data set $\mathcal{S}_R = \{S_{R1}, S_{R2}, \dots, S_{RK}\}$, with $S_{Rj}^T = (y_{Rj}, \varphi_{Rj}^T)$, $j = 1, 2, \dots, K$. A RGT on $\mathcal{S}$ is to group a number of data points of $\mathcal{S}$ into one data point in $\mathcal{S}_R$ in a systematic way. There are different ways to do this grouping. One example is to group all the data of $\mathcal{S}$ using K -means clustering method [8]. We define $r = \frac{K}{N}$ as the transformation ratio. For each group, the coarse data may be determined by simple linear superposition of all the fine data that belong to this group. And due to this, the noise might be reduced. Then the weighted least squares algorithm can be performed on the coarse data. For easy reference, we call the data set, $\mathcal{S}$, as the fine data set and least squares estimate based on $\mathcal{S}$ as the OLS estimate. The data set $\mathcal{S}_R$ obtained from a RGT on the fine data set is called as the coarse data set and the resulting estimate as the Renormalization Group weighted least squares (RGWLS) estimate. In this paper, through both theoretical analysis and simulation examples, we will show that the estimation error of RGWLS could be smaller than that of OLS estimate under certain conditions.

To quickly see the idea of the proposed method, we consider a simple model

$$y_t = 0.8y_{t-1} + 0.5u_t + \gamma(e_t - 0.8e_{t-1}),$$

where $y_0 = 0$, u_t is a unit step signal and e_t is a white noise which has the properties:

- $E[e_t] = 0$,
- $E[e_t^2] = \sigma_e^2$,
- $E[e_t e_s] = 0$ for all $t \neq s$.

Let $\sigma_e^2 = 1$ and $\gamma = 0.0625$. The model is simulated over 30 steps. We have

$$\begin{aligned}\theta^T &= [0.8, 0.5], \\ \varphi_t^T &= [y_{t-1}, u_t], \quad t = 1, 2, \dots, 30, \\ Y &= [y_1, \dots, y_{30}]^T, \\ \Phi &= [\varphi_1, \dots, \varphi_{30}]^T.\end{aligned}$$

Then the OLS estimate can be obtained based on (6). Next we carry out the RG method. We first form the fine data point as $S_t^T = (y_t, \varphi_t) = (y_t, y_{t-1}, u_t) = (y_t, y_{t-1}, 1)$ and the fine data set as $\mathcal{S} = \{S_1, S_2, \dots, S_{30}\}$. Then a RGT is performed on the fine data set $\mathcal{S}$ to get the coarse data set $\mathcal{S}_R$. In this case, it is done by K -means clustering method based on Euclidean distance measure with $K = 5$, and for each group, the coarse data is obtained by the linear superposition of all the fine data in this group. We have

$$\begin{aligned}S_{R1} &= S_1, \\ S_{R2} &= S_2, \\ S_{R3} &= S_3 + S_4, \\ S_{R4} &= S_5 + \dots + S_8, \\ S_{R5} &= S_9 + \dots + S_{30}.\end{aligned}$$

The spatial distribution of different groups of the fine data is shown in Fig. 1a and the system response is shown in Fig. 1b. It is obvious that y_{Rj} , φ_{Rj} and ε_{Rj} satisfy

(a) Spatial distribution of fine data (b) System response

Fig. 1. Data grouping

$$y_{Rj} = \varphi_{Rj}^T \theta + \varepsilon_{Rj}, \quad j = 1, 2, 3, 4, 5.$$

Choosing the weighting matrix as $W = \text{diag}(1, 1, \frac{1}{4}, \frac{1}{16}, \frac{1}{484})$, the RGWLS estimate can be obtained based on (8). Employing the Monte Carlo method, the simulation is repeated 100 times. Then we get $E(\|\Delta\theta\|_2) = 0.38$ and $E(\|\Delta\theta_R\|_2) = 0.19$, where $\|\cdot\|_2$ denotes the l^2 -norm. Therefore, the RGWLS estimate is better than the OLS estimate.

III. ASYMPTOTIC ANALYSIS

It is known that the bias of the OLS estimate is related to the number of data N if N is finite. If the system is disturbed by colored noise, it has been proved to be difficult to

investigate finite-sample bias analytically [9]. In this section, we discuss the properties of the OLS and RGWLS estimates as $N \rightarrow \infty$, analytically.

With (7) being multiplied and divided by N , we have

$$\lim_{N \rightarrow \infty} \Delta\theta = \lim_{N \rightarrow \infty} \left[\left(\frac{\Phi^T \Phi}{N} \right)^{-1} \left(\frac{\Phi^T \mathcal{E}}{N} \right) \right]. \quad (9)$$

Property 1 [10]: If the limits $\lim_{N \rightarrow \infty} H(N)$ and $\lim_{N \rightarrow \infty} G(N)$ both exist, and $\lim_{N \rightarrow \infty} G(N) \neq 0$, then

$$\lim_{N \rightarrow \infty} [G^{-1}(N)H(N)] = \left[\lim_{N \rightarrow \infty} G(N) \right]^{-1} \left[\lim_{N \rightarrow \infty} H(N) \right].$$

With the above property, (9) is equivalent to

$$\begin{aligned} \lim_{N \rightarrow \infty} \Delta\theta &= \left[\lim_{N \rightarrow \infty} \left(\frac{\Phi^T \Phi}{N} \right) \right]^{-1} \left[\lim_{N \rightarrow \infty} \left(\frac{\Phi^T \mathcal{E}}{N} \right) \right] \\ &= \left[\lim_{N \rightarrow \infty} \frac{\sum_{t=1}^N (\varphi_t \varphi_t^T)}{N} \right]^{-1} \left[\lim_{N \rightarrow \infty} \frac{\sum_{t=1}^N (\varphi_t \varepsilon_t)}{N} \right]. \end{aligned} \quad (10)$$

When the system (1) is asymptotically stable with ε_t a stationary stochastic process that is independent of the input signal, Söderström [6] show that the sums $\frac{\sum_{t=1}^N (\varphi_t \varphi_t^T)}{N}$ and $\frac{\sum_{t=1}^N (\varphi_t \varepsilon_t)}{N}$ tend to the corresponding expected values with probability one as the number of data points, N , tends to infinity. Then (10) becomes

$$\lim_{N \rightarrow \infty} \Delta\theta = [E(\varphi_t \varphi_t^T)]^{-1} [E(\varphi_t \varepsilon_t)]. \quad (11)$$

As can be seen from (11), when $E(\varphi_t \varphi_t^T)$ is non-singular and $E(\varphi_t \varepsilon_t) = 0$, $\hat{\theta}$ will converge to θ . Otherwise, the OLS estimate could be biased.

A. The asymptotic properties of OLS estimate

It can be seen from (11) that when ε_t is white noise, the OLS estimate will converge to its real value. However, when ε_t is colored noise, the OLS estimate could be biased. It is known that $\Delta\theta$ is related to model structure, input signal and noise properties. But these factors cannot be reflected in (11) explicitly. Therefore, for formulation of the estimation error and easy illustration of the proposed method, we consider a simple ARMA(1,1) model, which is given by

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t \quad |\alpha_1| < 1, \quad (12)$$

where $\varepsilon_t = e_t + \lambda_1 e_{t-1}$. According to (3) and (4), we have $\theta = \alpha_1$ and $\varphi_t = y_{t-1}$. It follows from (11) that

$$\lim_{N \rightarrow \infty} \Delta\alpha_1 = [E(y_{t-1}^2)]^{-1} [E(y_{t-1} \varepsilon_t)]. \quad (13)$$

Introducing the lag operator $Lz_t = z_{t-1}$ into (12), we can obtain

$$y_t = \frac{\varepsilon_t}{(1 - \alpha_1 L)} = \sum_{p=0}^{\infty} \alpha_1^p \varepsilon_{t-p}.$$

Then,

$$\begin{aligned} E(y_{t-1}^2) &= E \left[\left(\sum_{p=0}^{\infty} \alpha_1^p \varepsilon_{t-1-p} \right)^2 \right] \\ &= E \left(\sum_p \sum_q \alpha_1^p \alpha_1^q \varepsilon_{t-1-p} \varepsilon_{t-1-q} \right) \\ &= \sum_p \sum_q \alpha_1^p \alpha_1^q E(\varepsilon_{t-1-p} \varepsilon_{t-1-q}) \\ &= \left\{ \sum_p \sum_q \alpha_1^p \alpha_1^q E[(e_{t-1-p} + \lambda_1 e_{t-2-p}) \right. \\ &\quad \times (e_{t-1-q} + \lambda_1 e_{t-2-q})] \Big\} \\ &= \sum_{p=q} \alpha_1^p \alpha_1^q (1 + \lambda_1^2) \sigma_e^2 + 2 \sum_{q=p+1} \alpha_1^p \alpha_1^q \lambda_1 \sigma_e^2 \\ &= \sum_{p=0}^{\infty} \alpha_1^{2p} (1 + \lambda_1^2) \sigma_e^2 + 2 \sum_{p=0}^{\infty} \alpha_1^{2p+1} \lambda_1 \sigma_e^2 \\ &= \frac{(1 + \lambda_1^2)}{1 - \alpha_1^2} \sigma_e^2 + \frac{2\lambda_1 \alpha_1}{1 - \alpha_1^2} \sigma_e^2 \\ &= \frac{(1 + \lambda_1^2 + 2\lambda_1 \alpha_1)}{1 - \alpha_1^2} \sigma_e^2. \end{aligned} \quad (14)$$

We also have

$$\begin{aligned} E(y_{t-1} \varepsilon_t) &= E \left(\sum_{p=0}^{\infty} \alpha_1^p \varepsilon_{t-1-p} \varepsilon_t \right) \\ &= \sum_{p=0}^{\infty} \alpha_1^p E[(e_{t-1-p} + \lambda_1 e_{t-2-p}) \\ &\quad \times (e_t + \lambda_1 e_{t-1})], \end{aligned}$$

where

$$E[(e_{t-1-p} + \lambda_1 e_{t-2-p})(e_t + \lambda_1 e_{t-1})] = \begin{cases} \lambda_1 \sigma_e^2, & \text{for } p = 0; \\ 0, & \text{for } p > 0. \end{cases}$$

Hence

$$E(y_{t-1} \varepsilon_t) = \lambda_1 \sigma_e^2. \quad (15)$$

Substituting (14) and (15) into (13) gives

$$\begin{aligned} \lim_{N \rightarrow \infty} \Delta\alpha_1 &= \left[\frac{(1 + \lambda_1^2 + 2\lambda_1 \alpha_1)}{1 - \alpha_1^2} \sigma_e^2 \right]^{-1} (\lambda_1 \sigma_e^2) \\ &= \frac{\lambda_1 (1 - \alpha_1^2)}{1 + \lambda_1^2 + 2\lambda_1 \alpha_1}. \end{aligned} \quad (16)$$

In order to obtain a neater result, we consider a special case that $\lambda_1 = -\alpha_1$. Then (16) becomes

$$\lim_{N \rightarrow \infty} \Delta\alpha_1 = -\alpha_1. \quad (17)$$

B. The asymptotic properties of RGWLS estimate

RG is performed on the fine data set to obtain the coarse data set. The RGT groups all fine data based on the K -means clustering method. Then in each group, it takes linear superposition of all the fine data as the coarse data. Introduce the notations

$$\begin{aligned} y_{Rj} &= \sum_{t \in j} y_t, \\ \varphi_{Rj} &= \sum_{t \in j} \varphi_t, \\ \varepsilon_{Rj} &= \sum_{t \in j} \varepsilon_t, \quad j = 1, \dots, K, \end{aligned} \quad (18)$$

where y_{Rj} , φ_{Rj} and ε_{Rj} satisfy

$$y_{Rj} = \varphi_{Rj}^T \theta + \varepsilon_{Rj}.$$

The parameter θ is chosen to minimize the loss function

$$J_R = \frac{1}{2} \sum_{j=1}^K \left[w_j (y_{Rj} - \varphi_{Rj}^T \theta_R)^2 \right],$$

where w_j is the chosen weight. The RGWLS estimate is given by

$$\hat{\theta}_R = (\Phi_R^T W \Phi_R)^{-1} (\Phi_R^T W Y_R),$$

where $Y_R = [y_{R1} \dots, y_{RK}]^T$, $\Phi_R = [\varphi_{R1}, \dots, \varphi_{RK}]^T$, $\mathcal{E}_R = [\varepsilon_{R1}, \dots, \varepsilon_{RK}]^T$, and $W = \text{diag}(w_1, \dots, w_K)$. The estimation error of RGWLS is given by

$$\begin{aligned} \Delta\theta_R &= \hat{\theta}_R - \theta \\ &= (\Phi_R^T W \Phi_R)^{-1} (\Phi_R^T W Y_R) - \theta \\ &= (\Phi_R^T W \Phi_R)^{-1} (\Phi_R^T W (\Phi_R \theta + \mathcal{E}_R)) - \theta \\ &= (\Phi_R^T W \Phi_R)^{-1} (\Phi_R^T W \mathcal{E}_R). \end{aligned}$$

Next we analyze the properties of $\lim_{N \rightarrow \infty} \Delta\theta_R$. Let n_j be the number of fine data in the j th group so that

$$n_1 + \dots + n_K = N.$$

Suppose n_j tends to infinity. Similarly with (9) and (10), we have

$$\begin{aligned} \lim_{N \rightarrow \infty} \Delta\theta_R &= \lim_{N \rightarrow \infty} \left[(\Phi_R^T W \Phi_R)^{-1} (\Phi_R^T W \mathcal{E}_R) \right] \\ &= \left(\lim_{N \rightarrow \infty} \Phi_R^T W \Phi_R \right)^{-1} \left(\lim_{N \rightarrow \infty} \Phi_R^T W \mathcal{E}_R \right) \\ &= \left[\lim_{N \rightarrow \infty} \sum_{j=1}^K (w_j \varphi_{Rj} \varphi_{Rj}^T) \right]^{-1} \left[\lim_{N \rightarrow \infty} \sum_{j=1}^K (w_j \varphi_{Rj} \varepsilon_{Rj}) \right] \\ &= \left[\sum_{j=1}^K \lim_{n_j \rightarrow \infty} (w_j \varphi_{Rj} \varphi_{Rj}^T) \right]^{-1} \left[\sum_{j=1}^K \lim_{n_j \rightarrow \infty} (w_j \varphi_{Rj} \varepsilon_{Rj}) \right]. \end{aligned} \quad (19)$$

Let $w_j = 1/n_j^2$. Substituting (18) into (19) gives

$$\begin{aligned} \lim_{N \rightarrow \infty} \Delta\theta_R &= \left[\sum_{j=1}^K \lim_{n_j \rightarrow \infty} \left(\frac{1}{n_j^2} (\sum_{t \in j} \varphi_t) (\sum_{t \in j} \varphi_t^T) \right) \right]^{-1} \\ &\quad \times \left[\sum_{j=1}^K \lim_{n_j \rightarrow \infty} \left(\frac{1}{n_j^2} (\sum_{t \in j} \varphi_t) (\sum_{t \in j} \varepsilon_t^T) \right) \right] \\ &= \left[\sum_{j=1}^K \lim_{n_j \rightarrow \infty} \left(\frac{1}{n_j^2} (\sum_{t \in j} \varphi_t) (\sum_{s \in j} \varphi_s^T) \right) \right]^{-1} \\ &\quad \times \left[\sum_{j=1}^K \lim_{n_j \rightarrow \infty} \left(\frac{1}{n_j^2} (\sum_{t \in j} \varphi_t) (\sum_{s \in j} \varepsilon_s^T) \right) \right]. \end{aligned} \quad (20)$$

As $n_j \rightarrow \infty$, suppose $\sum_{j=1}^K \lim_{n_j \rightarrow \infty} \left(\frac{1}{n_j^2} (\sum_{t \in j} \varphi_t) (\sum_{s \in j} \varphi_s^T) \right)$ and $\sum_{j=1}^K \lim_{n_j \rightarrow \infty} \left(\frac{1}{n_j^2} (\sum_{t \in j} \varphi_t) (\sum_{s \in j} \varepsilon_s^T) \right)$ converge to

$E(\varphi_t \varphi_s^T) |_{t,s \in j}$ and $E(\varphi_t \varepsilon_s) |_{t,s \in j}$, respectively. Then (20) can be rewritten as

$$\begin{aligned} \lim_{N \rightarrow \infty} \Delta\theta_R &= \left[\sum_{j=1}^K \left(E(\varphi_t \varphi_s^T) |_{t,s \in j} \right) \right]^{-1} \\ &\quad \times \left[\sum_{j=1}^K \left(E(\varphi_t \varepsilon_s) |_{t,s \in j} \right) \right]. \end{aligned} \quad (21)$$

Compared with (11), the estimation error of RGWLS (21) is more complicated to be analyzed because of the data grouping.

In order to make a quantitative description of $\lim_{N \rightarrow \infty} \Delta\theta_R$, we still consider the simple model (12). It follows from (21) that

$$\begin{aligned} \lim_{N \rightarrow \infty} \Delta\alpha_{1R} &= \left[\sum_{j=1}^K \left(E(y_{t-1} y_{s-1}) |_{t,s \in j} \right) \right]^{-1} \\ &\quad \times \left[\sum_{j=1}^K \left(E(y_{t-1} \varepsilon_s) |_{t,s \in j} \right) \right]. \end{aligned} \quad (22)$$

In (22),

$$\begin{aligned} E(y_{t-1} y_{s-1}) |_{t,s \in j} &= E \left[\left(\sum_{p=0}^{\infty} \alpha_1^p \varepsilon_{t-p-1} \right) \right. \\ &\quad \left. \times \left(\sum_{q=0}^{\infty} \alpha_1^q \varepsilon_{s-q-1} \right) \right] |_{t,s \in j}. \end{aligned}$$

When $t = s$,

$$\begin{aligned} E(y_{t-1} y_{s-1}) |_{t,s \in j} &= \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \alpha_1^p \alpha_1^q E(\varepsilon_{t-p-1} \varepsilon_{t-q-1}) |_{t \in j} \\ &= \left\{ \sum_p \sum_q \alpha_1^p \alpha_1^q E[(e_{t-p-1} + \lambda_1 e_{t-p-2}) \right. \\ &\quad \times (e_{t-q-1} + \lambda_1 e_{t-q-2})] \Big\} |_{t \in j} \\ &= \left\{ \sum_{p=q} \sum \alpha_1^{2p} (1 + \lambda_1^2) \sigma_e^2 + \sum_{p=q+1} \sum \alpha_1^{2q+1} \lambda_1 \sigma_e^2 \right. \\ &\quad \left. + \sum_{q=p+1} \sum \alpha_1^{2p+1} \lambda_1 \sigma_e^2 \right\} |_{t \in j} \\ &= \frac{1 + \lambda_1^2 + 2\lambda_1 \alpha_1}{1 - \alpha_1^2} \sigma_e^2 |_{t \in j}; \end{aligned} \quad (23)$$

when $t < s$,

$$\begin{aligned}
E(y_{t-1}y_{s-1})|_{t,s \in j} &= \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \alpha_1^p \alpha_1^q E(\varepsilon_{t-p-1} \varepsilon_{s-q-1})|_{t,s \in j} \\
&= \left\{ \sum_p \sum_q \alpha_1^p \alpha_1^q E \left[(e_{t-p-1} + \lambda_1 e_{t-p-2}) \right. \right. \\
&\quad \left. \left. \times (e_{s-q-1} + \lambda_1 e_{s-q-2}) \right] \right\} |_{t,s \in j} \\
&= \left\{ \sum_{q=s-t+p} \sum \alpha_1^{2p+s-t} (1 + \lambda_1^2) \sigma_e^2 \right. \\
&\quad \left. + \sum_{q=s-t+p-1} \sum \alpha_1^{2p+s-t-1} \lambda_1 \sigma_e^2 \right. \\
&\quad \left. + \sum_{q=s-t+p+1} \sum \alpha_1^{2p+s-t+1} \lambda_1 \sigma_e^2 \right\} |_{t,s \in j} \\
&= \frac{\alpha_1^{s-t} (1 + \lambda_1^2) + \alpha_1^{s-t-1} \lambda_1 + \alpha_1^{s-t+1} \lambda_1}{1 - \alpha_1^2} \sigma_e^2 |_{t,s \in j}; \tag{24}
\end{aligned}$$

when $t > s$,

$$\begin{aligned}
E(y_{t-1}y_{s-1})|_{t,s \in j} &= \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \alpha_1^p \alpha_1^q E(\varepsilon_{t-p-1} \varepsilon_{s-q-1})|_{t,s \in j} \\
&= \left\{ \sum_p \sum_q \alpha_1^p \alpha_1^q E \left[(e_{t-p-1} + \lambda_1 e_{t-p-2}) \right. \right. \\
&\quad \left. \left. \times (e_{s-q-1} + \lambda_1 e_{s-q-2}) \right] \right\} |_{t,s \in j} \\
&= \left\{ \sum_{p=t-s+q} \sum \alpha_1^{2q+t-s} (1 + \lambda_1^2) \sigma_e^2 \right. \\
&\quad \left. + \sum_{p=t-s+q-1} \sum \alpha_1^{2q+t-s-1} \lambda_1 \sigma_e^2 \right. \\
&\quad \left. + \sum_{p=t-s+q+1} \sum \alpha_1^{2q+t-s+1} \lambda_1 \sigma_e^2 \right\} |_{t,s \in j} \\
&= \frac{\alpha_1^{t-s} (1 + \lambda_1^2) + \alpha_1^{t-s-1} \lambda_1 + \alpha_1^{t-s+1} \lambda_1}{1 - \alpha_1^2} \sigma_e^2 |_{t,s \in j}. \tag{25}
\end{aligned}$$

Substituting $\lambda_1 = -\alpha_1$ into (23), (24) and (25) gives

$$E(y_{t-1}y_{s-1})|_{t,s \in j} = \begin{cases} \sigma_e^2, & \text{for } t = s; \\ 0, & \text{for } t \neq s. \end{cases}$$

Hence

$$\sum_{j=1}^K \left[E(y_{t-1}y_{s-1})|_{t,s \in j} \right] = K \sigma_e^2. \tag{26}$$

From (22), we also have

$$E(y_{t-1}\varepsilon_s)|_{t,s \in j} = E \left(\sum_{p=0}^{\infty} \alpha_1^p \varepsilon_{t-p-1} \varepsilon_s \right) |_{t,s \in j}.$$

When $t = s$,

$$\begin{aligned}
E(y_{t-1}\varepsilon_s)|_{t,s \in j} &= \left\{ \sum_{p=0}^{\infty} \alpha_1^p E \left[(e_{t-p-1} + \lambda_1 e_{t-p-2}) \right. \right. \\
&\quad \left. \left. \times (e_t + \lambda_1 e_{t-1}) \right] \right\} |_{t \in j} \\
&= \lambda_1 \sigma_e^2 |_{t \in j}; \tag{27}
\end{aligned}$$

when $t < s$,

$$\begin{aligned}
E(y_{t-1}\varepsilon_s)|_{t,s \in j} &= \left\{ \sum_{p=0}^{\infty} \alpha_1^p E \left[(e_{t-p-1} + \lambda_1 e_{t-p-2}) \right. \right. \\
&\quad \left. \left. \times (e_s + \lambda_1 e_{s-1}) \right] \right\} |_{t,s \in j} \\
&= 0 |_{t,s \in j}; \tag{28}
\end{aligned}$$

when $t > s$, let $k \geq 2$ be a positive integer, then

$$\begin{aligned}
E(y_{t-1}\varepsilon_s)|_{t,s \in j} &= \left\{ \sum_{p=0}^{\infty} \alpha_1^p E \left[(e_{t-p-1} + \lambda_1 e_{t-p-2}) \right. \right. \\
&\quad \left. \left. \times (e_s + \lambda_1 e_{s-1}) \right] \right\} |_{t,s \in j} \\
&= \left\{ \sum_{\substack{p=0 \\ t=s+1}} (1 + \lambda_1^2) \sigma_e^2 + \sum_{\substack{p=1 \\ t=s+1}} \alpha_1 \lambda_1 \sigma_e^2 \right. \\
&\quad \left. + \sum_{\substack{p=k-1 \\ t=s+k}} \alpha_1^{k-1} (1 + \lambda_1^2) \sigma_e^2 + \sum_{\substack{p=k \\ t=s+k}} \alpha_1^k \lambda_1 \sigma_e^2 \right. \\
&\quad \left. + \sum_{\substack{p=k-2 \\ t=s+k}} \alpha_1^{k-2} \lambda_1 \sigma_e^2 \right\} |_{t,s \in j}. \tag{29}
\end{aligned}$$

Substituting $\lambda_1 = -\alpha_1$ into (27), (28) and (29) gives

$$E(y_{t-1}\varepsilon_s)|_{t,s \in j} = \begin{cases} -\alpha_1 \sigma_e^2, & \text{for } t = s; \\ 0, & \text{for } t < s; \\ \sigma_e^2, & \text{for } t = s + 1. \end{cases}$$

Denoting $f_j = d_j/(n_j - 1)$ where d_j is the number of S_t such that S_t and S_{t+1} belong to the j th group, we have

$$E(y_{t-1}\varepsilon_s)|_{t,s \in j} = -\alpha_1 \sigma_e^2 + f_j \sigma_e^2.$$

Hence

$$\sum_{j=1}^K \left[E(\varphi_t \varepsilon_s)|_{t,s \in j} \right] = -K \alpha_1 \sigma_e^2 + \sum_{j=1}^K f_j \sigma_e^2. \tag{30}$$

Substituting (26) and (30) into (22) gives

$$\lim_{N \rightarrow \infty} \Delta \alpha_{1R} = -\alpha_1 + \frac{\sum_{j=1}^K f_j}{K}. \tag{31}$$

Dividing (31) by (17), we have

$$\frac{\lim_{N \rightarrow \infty} \Delta \alpha_{1R}}{\lim_{N \rightarrow \infty} \Delta \alpha_1} = \frac{-\alpha_1 + \frac{\sum_{j=1}^K f_j}{K}}{-\alpha_1}. \tag{32}$$

With $|\alpha_1| < 1$, it is easy to verify

$$\left| \frac{\lim_{N \rightarrow \infty} \Delta \alpha_{1R}}{\lim_{N \rightarrow \infty} \Delta \alpha_1} \right| < 1,$$

if and only if

$$\left\{ \begin{array}{l} \frac{\sum_{j=1}^K f_j}{2K} < \alpha_1 < 1, \\ 0 < \frac{\sum_{j=1}^K f_j}{K} < 1. \end{array} \right. \quad (33)$$

Equation (33) shows when K is fixed, a group of smaller f_j will result in a larger range for α_1 . But a group of small f_j does not necessarily imply a good RGWLS estimate. We can see from (32) that when f_j tends to zero, the RGWLS estimate will not have any improvement. In contrast, when $\sum_{j=1}^K f_j/K$ approaches to α_1 , $\Delta\alpha_{1R}$ tends to zero and the RGWLS estimate will have much improvement. In addition, when $K = N$, then $n_j = 1$ and $f_j = 0$. It is easy to see that (31) becomes

$$\lim_{N \rightarrow \infty} \Delta\alpha_{1R} = -\alpha_1,$$

which is the same as (17), implying that RGWLS is reduced to OLS.

IV. SIMULATION EXAMPLES

In this part, examples are given to illustrate the proposed method. As $N \rightarrow \infty$, introduce the notation R_{EEI} by

$$R_{EEI} = \frac{\left\| \lim_{N \rightarrow \infty} \Delta\theta_R \right\|_2}{\left\| \lim_{N \rightarrow \infty} \Delta\theta \right\|_2},$$

which describes the size of the ratio between the estimation errors of RGWLS and OLS. Apparently, if $R_{EEI} < 1$, the RGWLS estimate does have improvement. We will discuss how the number of data points, the noise level and the transformation ratio affect the R_{EEI} .

Example 1: Consider an ARMA(1,1) model,

$$y_t = 0.95y_{t-1} + \gamma\varepsilon_t,$$

where $\varepsilon_t = e_t - 0.95e_{t-1}$, $y_0 = 1000$, and γ is a chosen parameter that can adjust the noise level. Simulation results are shown in Table I.

TABLE I
 R_{EEI} OF EXAMPLE 1 WITH DIFFERENT CONDITIONS

γ	$r=1/20$			$r=1/40$		
5	0.056	0.057	0.056	0.029	0.030	0.030
25	0.082	0.088	0.110	0.046	0.053	0.062
50	0.128	0.168	0.241	0.076	0.109	0.151
N	5000	10000	20000	5000	10000	20000

Example 2: Consider a second order system with exogenous input,

$$y_t = 1.5y_{t-1} - 0.7y_{t-2} + u_{t-1} + \gamma\varepsilon_t,$$

where $u_t = 1$, $\varepsilon_t = e_t - 1.5e_{t-1} + 0.7e_{t-2}$ and $y_0 = y_1 = 0$. Simulation results are shown in Table II.

Generally, as can be observed from Table I and Table II:

- for all simulation cases, $R_{EEI} < 1$;

TABLE II
 R_{EEI} OF EXAMPLE 2 WITH DIFFERENT CONDITIONS

γ	$r=1/20$			$r=1/40$		
0.1	0.399	0.472	0.496	0.259	0.347	0.410
0.5	0.598	0.694	0.759	0.505	0.586	0.660
1	0.707	0.868	0.885	0.635	0.789	0.853
N	5000	10000	20000	5000	10000	20000

- for the same γ and r , a larger N leads to a larger R_{EEI} , which is because redundant steady state samples with noise will result in bad estimation results;
- for the same N and r , a smaller γ leads to a smaller R_{EEI} ;
- for the same N and γ , a smaller r results in a smaller R_{EEI} .

We can see that the proposed method can be applied on high order systems, even though it is difficult to make theoretical analysis. And for high order systems, sufficient data is required. In addition, as a matter of future research, we will study how different model structures and input signals influence the effectiveness of the proposed method.

V. CONCLUSIONS

In this paper, an improved system identification method has been proposed based on Renormalization Group. Given a fine data set, it is easy to obtain the OLS estimate. By Renormalization Group, a coarse data set is created from the original fine data set. This new data set enables us to get the RGWLS estimate. Through comparisons, we find that the estimation error of RGWLS estimate can be smaller than that of the OLS estimate. Thus, the proposed method could have improvement. It should be pointed out that the proposed method requires sufficient data. Furthermore, its in-depth theory and additional potential values are to be investigated.

REFERENCES

- [1] L. Lennart, "System identification: theory for the user," *PTR Prentice Hall, Upper Saddle River, NJ*, 1999.
- [2] Z. Griliches, "A note on serial correlation bias in estimates of distributed lags," *Econometrica: journal of the Econometric Society*, pp. 65–73, 1961.
- [3] P. C. B. Phillips and M. R. Wickens, *Exercises in econometrics*. P. Allan, 1978, vol. 2.
- [4] T. Stocker, "On the asymptotic bias of ols in dynamic regression models with autocorrelated errors," *Statistical Papers*, vol. 48, no. 1, pp. 81–93, 2007.
- [5] W. X. Zheng, "On a least-squares-based algorithm for identification of stochastic linear systems," *Signal Processing, IEEE Transactions on*, vol. 46, no. 6, pp. 1631–1638, 1998.
- [6] T. Söderström and P. Stoica, *Instrumental variable methods for system identification*. Springer Berlin et al., 1983.
- [7] K. G. Wilson, "Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture," *Physical Review B*, vol. 4, no. 9, p. 3174, 1971.
- [8] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer-Verlag New York, 2009.
- [9] T. J. Sargent, "Some evidence on the small sample properties of distributed lag estimators in the presence of autocorrelated disturbances," *The Review of Economics and Statistics*, vol. 50, no. 1, pp. 87–95, 1968.
- [10] S. Salas, E. Hille, and G. Etgen, *Calculus: One and several variables*. Wiley, 1990.