Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Multi-Point Search Based Identification under Severe Numerical Conditions

Lianming Sun * **Ryoka Uto** * **Xinyu Liu** * **Akira Sano** **

* *Department of Information Systems Engineering, The University of Kitakyushu, 1-1 Hibikino, Wakamatsu-ku, Kitakyushu, Japan*
e-mail: sun@kitakyu-u.ac.jp, {a9mca004, z8dcb001}@eng.kitakyu-u.ac.jp
** *Department of System Design, Keio University, 3-14-1 Hiyoshi, Kouhoku-ku, Yokohama, Japan*
e-mail: sano@sd.keio.ac.jp

Abstract: It is necessary to perform the system identification under severe numerical conditions in many practical applications. When less external test signals are available for parameter estimation from experimental data, the identification performance often suffers from numerical problems in the optimization procedure due to the less independent informative components, the influence of complicated noise, or the local minima problem. In this paper, a multi-point search based identification algorithm is investigated for system identification under severe numerical conditions. It introduces the output over-sampling scheme to collect the experimental input-output data, and extracts the information in time and space domains to complement information criterion for numerical optimization. Furthermore, the multi-point search is utilized to decrease the influence of local minima. The numerical simulation examples illustrate that the identification performance has been improved in the proposed algorithm.

Keywords: System identification, output over-sampling, cyclo-stationarity, multi-point search.

1. INTRODUCTION

System identification is an important data-driven approach to construct a mathematical model for a practical system, and a large number of parametric and nonparametric identification approaches have been developed over several decades, see Ljung (1999); Söderström and Stoica (1989). Most of parametric identification approaches utilize information criterion functions to evaluate how well the parametric model approximates the dominant dynamic characteristics of the system, then the model parameters are estimated by optimizing the information criteria through numerical optimization algorithms. For example, the information criterion in prediction error method (PEM) is a quadratic function of the prediction error, which is often a nonlinear function of the model parameters. The numerical optimization uses Gauss-Newton or Levenberg-Marquardt methods, which are gradient based search algorithms to estimate the parameters.

The informativeness of the experimental data and optimization algorithm are the crucial factors in identification algorithms. In order to provide sufficient information, many conventional methods design external test signals to complement the information deficiency, see Van den Hof and Schrama (1995); Eckhard et al. (2013). However, some systems are subject to rigid restrictions for the system safety, operating performance or economical reasons. For example, the unstable industrial processes are required to be stabilized by the feedback controllers, and the large external test signals are not allowed since the external signals interfere the process operation. As a result, the information extracted from the experimental data may be too weak to deal with the influence of disturbance and model uncertainty. Moreover, the numerical conditions are so poor that the numerical optimization becomes unstable and fails to obtain satisfactory parameter estimates, see Wang et al. (2003); Sun et al. (2018).

It has been illustrated that the output over-sampling scheme, where the sampling rate for the output signals is multiple times higher than the input rate of input, can offer informative data for blind channel identification, see Moulines et al. (1995), and closed-loop identification, see Sun et al. (2001) when the external signals are not available. They use the space or the time domain information of cyclo-stationarity extracted from the data in output over-sampling to complement the information deficiency, thus make the system be possibly identifiable. However, when the system operating restrictions lead to severe numerical conditions, or the system is disturbed by a complicated noise, the numerical optimization will suffer from severely ill-conditioned problem, and the identification performance degrades dramatically, see Wang et al. (2003); Sun et al. (2018). In order to deal with the ill-conditioned problem, regularization terms are introduced into the optimization algorithm to stabilize the numerical computation, as shown in Pillonetto (2018); Chen (2018), but the determination of regularization terms are expertise, see Nelles (2001). In Sun and Sano (2017), some cyclo-stationarity based spatial information is extracted from the experimental data, and is used to develop a temporal-spatial algorithm that regularizes the ill-conditioned Hessian matrix. The temporal-spatial algorithm has been successfully used in the closed-loop identification of an unstable magnetic levitation system, see Sun et al. (2018).

Generally the gradient based local search is used for numerical optimization in system identification. Under severe numerical conditions where the attraction region of the global optimal estimates is very narrow, the numerical optimization is easily influenced by the initial values and the estimation of noise model, accordingly, some metaheuristics such as iterated local search are developed to reduce the affection of initial values, see Talbi (2009). On the other hand, some global search algorithms such as genetic algorithms (GA), ant colony optimization (ACO), particle swarm optimization (PSO) use multi-point scheme for better global convergence in scheduling problems. Inspired the idea of multi-point search, this paper investigates a new identification algorithm that has better convergence performance under the severe numerical conditions. Compared with the existing methods, the proposed algorithm considers the following issues: it betters the numerical conditions by combining the cyclo-stationary information in time and space domains; the influence of local minima is decreased by updating multiple estimates parallelly. Furthermore, the search efficiency is improved by merging the estimates that are close to each other, while the random walk scheme moves the estimates from the local minima toward to a new attraction region.

The rest of the paper is organized as follows. In the next section, the main description of the identification problem, the system models and signals are summarized. In Section 3, the information complementation by introducing the cyclo-stationary in time and space domains extracted in the output over-sampling are investigated, then the multi-point search based identification algorithm is illustrated in Section 4, and the results of numerical simulation are shown in Section 5. Finally, the conclusion and the future research work are given in Section 6.

2. PROBLEM STATEMENT

The system model and signals are discussed in this section.

2.1 Model description

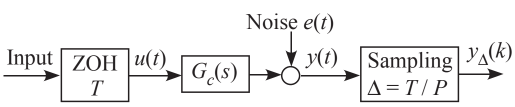

Fig. 1. Illustration of a system model

Consider an nth order linear system $G_c(s)$ as shown in Fig.1, where the input is added to the system through a zero-order holder (ZOH), whose holding period is T. Correspondingly, the signal $u(t)$ is a piece-wise signal. The noise $e(t)$ is often approximated as a stationary stochastic process, and $y(t)$ is the system output, which is sampled at an sampling interval $\Delta = T/P$. When $P = 1$, the experimental data of input and output are obtained at the same sampling rate, which are denoted as $u(KT)$, $y(KT)$, $K = 0, 1, 2, \cdots$. The data $\{u(KT),\ y(KT)\}$ are used in many conventional identification methods. With respect to the interval T, the system model can be described by a discrete-time model as follows

$$y(KT)=G(z^{-1})u(KT)+e(KT),\ \text{for}\ K=0,1,2,\cdots \quad (1)$$

where $G(z^{-1})$ is an nth order discrete-time model of $G_c(s)$, z^{-1} is a backwards shift operator corresponding to interval T. For $P \geq 2$, i.e., the output is sampled at a higher rate than the input signal, then the experimental data of $u(t)$, $y(t)$ at the instant $t = k\Delta$ are recorded as $u_\Delta(k)$, $y_\Delta(k)$, respectively. When P is an integer $P \geq 2$, the system can also be described by the following discrete-time model

$$y_\Delta(k) = G_\Delta(q^{-1})u_\Delta(k) + e_\Delta(k),\ k = 0, 1, 2, \cdots \quad (2)$$

where $G_\Delta(q^{-1})$ is a discrete-time transfer function model with respect to the interval Δ, while q^{-1} is the corresponding backwards shift operator, $q^{-P} = z^{-1}$. It has been illustrated that the model $G_\Delta(q^{-1})$ determines the discrete-time model $G(z^{-1})$ uniquely. Hence, in the output over-sampling scheme, the identification problem can be solved by modelling $G_\Delta(q^{-1})$ from the experimental data of $u_\Delta(k)$, $y_\Delta(k)$.

$e_\Delta(k)$ is considered as the sample of $e(t)$ at $k\Delta$, and the stochastic process is approximated by

$$e_\Delta(k) = H_\Delta(q^{-1})w_\Delta(k), \quad (3)$$

where $H_\Delta(q^{-1})$ is an n_Hth order stable transfer function with minimum phase, and $w_\Delta(k)$ is a white i.i.d stochastic process with $\mathcal{N}(0, \sigma^2_{w_\Delta})$.

2.2 Signals in over-sampling

It is noticed that the input $u(t)$ holds the same value $u_\Delta(KP) = u(KT)$ within the interval $KP\Delta$, $(KP+1)\Delta$, $\cdots$, $((K+1)P-1)\Delta$, while the system's transient response contains some information different from the noise terms. Fig.2 illustrates the data examples of $P = 2$ and $P = 3$ in Sun et al. (2018).

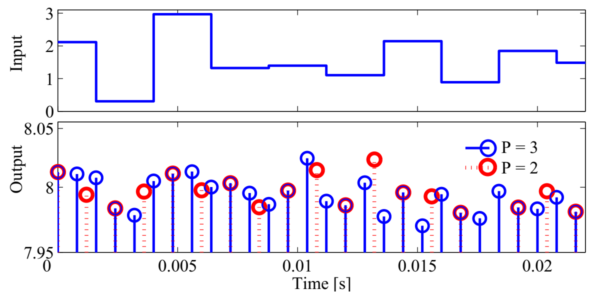

Fig. 2. Illustration of experimental input-output data. From the top: input $u(t)$; Output $y_\Delta(k)$ for $P = 2$ (dashed) and $P = 3$ (solid).

Suppose $x_1(k)$ and $x_2(k)$, $k = 0, 1, \cdots$, are the data of stochastic processes. Let their correlation function be defined as follows

$$\mathcal{R}_{x_1,x_2}(k,\tau) := E\left\{x_1(k+\tau)x_2(k)\right\}, \quad (4)$$

where $E\{\cdot\}$ indicates the expectation, and τ is the shift time. Then, for the stationary noise term $e_\Delta(k)$

at an arbitrary instant $k\Delta$, the auto-correlation function $\mathcal{R}_{e_\Delta,e_\Delta}(k,\tau)$ is a constant for a fixed shift time τ. For $u(KT)$, $K=0,1,\cdots$, if it can be approximated as a wide-sense stationary process, i.e.,

$$\mathcal{R}_{u,u}(K,\tau)=\mathcal{R}_{u,u}(K\pm K_1,\tau) \tag{5}$$

holds for arbitrary K and K_1. On the other hand, due to the input's holding property with respect to the interval Δ, the auto-correlation function of $\mathcal{R}_{u_\Delta,u_\Delta}(k,\tau)$ is a periodic function rather than a constant,

$$\mathcal{R}_{u_\Delta,u_\Delta}(k,\tau)=\mathcal{R}_{u_\Delta,u_\Delta}(k\pm P,\tau) \tag{6}$$

whereas the following inequality generally holds that

$$\mathcal{R}_{u_\Delta,u_\Delta}(k,\tau)\neq\mathcal{R}_{u_\Delta,u_\Delta}(k\pm k_1,\tau),\ 1\leq k_1\leq P-1. \tag{7}$$

The same properties also hold for the system output $y_\Delta(k)$ since it is a response of $u_\Delta(k)$ in a causal system. Therefore, if $x_1(k)$ or $x_2(k)$ is $u_\Delta(k)$ or $y_\Delta(k)$, the correlation function $\mathcal{R}_{x_1,x_2}(k,\tau)$ satisfies that

$$\mathcal{R}_{x_1,x_2}(k,\tau)=\mathcal{R}_{x_1,x_2}(k+Pk_1,\tau) \tag{8}$$

for an arbitrary integer k_1, and the property of periodic correlation functions in (8) is called cyclo-stationarity, see Gardner (1994). Furthermore, let the cyclo-stationary correlation function be given by Fourier transform of the periodic correlation functions with respect to k

$$\mathcal{C}_{x_1,x_2}(\alpha,\tau)=\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\mathcal{R}_{x_1,x_2}(k,\tau)e^{-i\alpha k}, \tag{9}$$

where $0\leq\alpha<2\pi$. Then following (8), the cyclo-stationary correlation functions satisfy

$$\begin{aligned}&\mathcal{C}_{x_1,x_2}(\alpha,\tau)\\&=\begin{cases}\dfrac{1}{P}\displaystyle\sum_{k=mP}^{(m+1)P-1}\mathcal{R}_{x_1,x_2}(k,\tau)e^{-i\alpha_p k},\ \alpha=\alpha_p\\ 0,\qquad\qquad\qquad\qquad\qquad\quad \text{others}\end{cases},\end{aligned} \tag{10}$$

where $\alpha_p\in\left\{\alpha_p|\alpha_p=\dfrac{2\pi}{P}p,\ p=0,1,\cdots,P-1\right\}$. On the other hand, if both $x_1(k)$ and $x_2(k)$ are stationary signals, for example, the cyclo-stationary correlation function $\mathcal{C}_{e_\Delta,e_\Delta}(\alpha,\tau)$ becomes to

$$\begin{aligned}&\mathcal{C}_{e_\Delta,e_\Delta}(\alpha,\tau)\\&=\begin{cases}\mathcal{R}_{e_\Delta,e_\Delta}(k,\tau),\ \alpha=0,k\text{ is an arbitrary integer}\\ 0,\qquad\qquad\qquad 0<\alpha<2\pi\end{cases}.\end{aligned} \tag{11}$$

It is seen that the properties on cyclo-stationarity of $u_\Delta(k)$, $y_\Delta(k)$ are quite different from that of the noise term $e_\Delta(k)$ at $\alpha=\alpha_1,\cdots,\alpha_{P-1}$, hence the cyclo-stationarity could offer a possibility to deal with the noise influence.

2.3 Information criterion

Let the prediction error $\varepsilon_\Delta(k,\hat{\boldsymbol{\theta}}_\Delta)$ for the system model be defined as follows:

$$\varepsilon_\Delta(k,\hat{\boldsymbol{\theta}}_\Delta)=\frac{1}{\hat{H}_\Delta(q^{-1})}\Big(y_\Delta(k)-\hat{G}_\Delta(q^{-1})u_\Delta(k)\Big), \tag{12}$$

where $\hat{G}_\Delta(q^{-1})$ and $\hat{H}_\Delta(q^{-1})$ are the estimated models of the system model $G_\Delta(q^{-1})$ and the noise model $H_\Delta(q^{-1})$, and their orders are determined by detecting whether any cyclo-stationary component is contained in $\varepsilon_\Delta(k,\hat{\boldsymbol{\theta}})$ as shown in Sun and Sano (2017). The definition of prediction error can be easily extended to the other model structures such as ARMAX models, BJ models, the model of unstable process, see Forssell and Hjalmarsson (1999); Sun et al. (2018). Then, the information criterion in time domain for estimation of model parameters in the identification algorithm such as PEM is defined as a quadratic function of the prediction error

$$J_{\mathrm{T}}(\hat{\boldsymbol{\theta}}_\Delta)=\frac{1}{2N}\sum_{k=0}^{N-1}\varepsilon_\Delta^2(k,\hat{\boldsymbol{\theta}}_\Delta), \tag{13}$$

when N is the data number, $\hat{\boldsymbol{\theta}}_\Delta$ is the estimate vector of $\hat{G}_\Delta(q^{-1})$ and $\hat{H}_\Delta(q^{-1})$.

The estimation of model parameters in PEM are obtained through minimizing the information criterion of (13) by Gauss-Newton or Levenberg-Marquardt methods, which are gradient based search algorithms. The gradient and Hessian matrix are given by

$$\boldsymbol{g}_{\mathrm{T}}(\hat{\boldsymbol{\theta}}_\Delta)=\frac{dJ_{\mathrm{T}}(\hat{\boldsymbol{\theta}}_\Delta)}{d\hat{\boldsymbol{\theta}}_\Delta}=\frac{1}{N}\sum_{k=0}^{N-1}\frac{d\varepsilon(k,\hat{\boldsymbol{\theta}}_\Delta)}{d\hat{\boldsymbol{\theta}}_\Delta}\varepsilon(k,\Delta), \tag{14}$$

$$\boldsymbol{H}_{\mathrm{ess,T}}=\frac{1}{N}\sum_{k=0}^{N-1}\frac{d\varepsilon(k,\hat{\boldsymbol{\theta}}_\Delta)}{d\hat{\boldsymbol{\theta}}_\Delta}\left(\frac{d\varepsilon(k,\hat{\boldsymbol{\theta}}_\Delta)}{d\hat{\boldsymbol{\theta}}_\Delta}\right)^H. \tag{15}$$

Then, the model parameter estimates are updated by

$$\hat{\boldsymbol{\theta}}_\Delta^{(l+1)}=\hat{\boldsymbol{\theta}}_\Delta^{(l)}-\mu\boldsymbol{H}_{\mathrm{ess,T}}^{-1}\boldsymbol{g}_{\mathrm{T}}(\hat{\boldsymbol{\theta}}_\Delta^{(l)}), \tag{16}$$

in Gauss-Newton method, or Levenberg-Marquardt method applied a regularization factor α as follows

$$\hat{\boldsymbol{\theta}}_\Delta^{(l+1)}=\hat{\boldsymbol{\theta}}_\Delta^{(l)}-\mu\big(\boldsymbol{H}_{\mathrm{ess,T}}+\alpha\boldsymbol{I}\big)^{-1}\boldsymbol{g}_{\mathrm{T}}(\hat{\boldsymbol{\theta}}_\Delta^{(l)}), \tag{17}$$

where $0\leq\mu\leq1$ is a step size, $\alpha\geq0$ is the regularization factor, and $\boldsymbol{I}$ is an identity diagonal matrix. Nevertheless, if the numerical conditions are poor, the condition number of Hessian matrix becomes very large and its inverse may cause computation error while the numerical optimization will be largely influenced by the noise term and fall into the local minima. This problem must be remedied in order to obtain an effective model.

3. INFORMATION COMPLEMENTATION

The time domain information can be complemented by the spatial information deduced from the cyclo-stationarity of the input-output data, see Sun and Sano (2017). Consider the noise vector and input-output data vector

$$\boldsymbol{\phi}_e(k)=\begin{bmatrix}e_\Delta(k+\tau)\\e_\Delta(k+\tau-1)\\e_\Delta(k+\tau-2)\\\vdots\\e_\Delta(k+\tau-n_1)\end{bmatrix}, \tag{18}$$

$$\boldsymbol{\phi}(k)=\begin{bmatrix} y_\Delta(k) \\ y_\Delta(k-1) \\ \vdots \\ y_\Delta(k-n_2) \\ -u_\Delta(k-1) \\ \vdots \\ -u_\Delta(k-n_2) \end{bmatrix}, \tag{19}$$

where $n_1 > n_2 + n + n_H$, $n_2 > n$. Following the cyclo-stationary property of $u_\Delta(k)$ and $y_\Delta(k)$, $\mathcal{R}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(k,\tau)$

$$\mathcal{R}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(k,\tau)=E\left\{\boldsymbol{\phi}_e(k+\tau)\boldsymbol{\phi}^T(k)\right\} \tag{20}$$

is periodic in k. Let the following cyclo-stationary correlation matrix $\mathcal{C}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(\alpha,\tau)$

$$\begin{aligned}\mathcal{C}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}&(\alpha,\tau)\\ &=\frac{1}{K_1}\sum_{K=0}^{K_1-1}\left(\sum_{p=0}^{P-1}e^{-ip\alpha}\mathcal{R}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(KP+p,\tau)\right)\end{aligned} \tag{21}$$

be arranged as

$$\boldsymbol{\mathcal{C}}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}=\begin{bmatrix}\mathcal{C}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(\alpha_1,0)\\ \vdots \\ \mathcal{C}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(\alpha_1,n) \\ \vdots \\ \mathcal{C}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(\alpha_{P-1},n)\end{bmatrix} \tag{22}$$

with respect to $\alpha=\alpha_1,\cdots,\alpha_{P-1}$, $\tau=0,\cdots,n$, then its orthogonal vector $\boldsymbol{\psi}=[\psi_0,\ \cdots,\ \psi_{n_2},\ \psi_{n_2+1},\ \cdots,\ \psi_{2n_2}]^T$

$$\boldsymbol{\mathcal{C}}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}\boldsymbol{\psi}=0 \tag{23}$$

is a non-zero vector whose parameters are the coefficients of the following polynomials

$$\begin{aligned}\Psi_1(q^{-1})&=\psi_0+\psi_1q^{-1}+\cdots+\psi_{n_2}q^{-n_2}\\ &=A_\Delta(q^{-1})X_\Delta(q^{-1}),\end{aligned} \tag{24}$$

$$\begin{aligned}\Psi_2(q^{-1})&=\psi_{n_1+1}q^{-1}+\cdots+\psi_{2n_2}q^{-n_2}\\ &=B_\Delta(q^{-1})X_\Delta(q^{-1}),\end{aligned} \tag{25}$$

where $A_\Delta(q^{-1})$ and $B_\Delta(q^{-1})$ are the denominator, numerator polynomial of $G_\Delta(q^{-1})$, $X_\Delta(q^{-1})$ is a common factor of $\Psi_1(q^{-1})$ and $\Psi_2(q^{-1})$, as shown in Sun and Sano (2017); Söderström and Stoica (1981); Åström and Söderstrom (1974). (24) and (25) imply that no cyclo-stationary component is contained in $\boldsymbol{\phi}(k)\boldsymbol{\psi}$ if vector $\boldsymbol{\psi}$ satisfies (24) and (25), then $\mathcal{C}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(\alpha_p,\tau)\boldsymbol{\psi}=0$ since $\mathcal{C}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(\alpha_p,\tau)$ are the cyclo-stationary correlation functions of the stationary signals. Furthermore, parameters of $A_\Delta(q^{-1})$ and $B_\Delta(q^{-1})$ can be determined by removing the common factor $X_\Delta(q^{-1})$ from $\Psi_1(q^{-1})$ and $\Psi_2(q^{-1})$.

In identification problem, the true values of $e_\Delta(k)$ are unknown. They are replaced by $y_\Delta(k)-\hat{G}_\Delta(q^{-1})u_\Delta(k)$ or the prediction error $\varepsilon_\Delta(k,\hat{\boldsymbol{\theta}}_\Delta)$.

Let the matrix $\boldsymbol{V}_l$ be defined by

$$\boldsymbol{V}_l=\begin{bmatrix}-\psi_{l,n_2+1} & -\psi_{l,n_2+2} & \cdots & -\psi_{l,2n_2} & 0 & \cdots & 0\\ 0 & \ddots & & & \ddots & & \\ 0 & \cdots & 0 & -\psi_{l,n_2+1} & \cdots & & -\psi_{l,2n_2}\\ \psi_{l,0} & \psi_{l,1} & \cdots & \psi_{l,n_2} & 0 & \cdots & 0\\ 0 & \ddots & & & \ddots & & \\ 0 & \cdots & 0 & \psi_{l,0} & \psi_{l,1} & \cdots & \psi_{l,n_2}\end{bmatrix}^H \tag{26}$$

where $\psi_{l,0},\ \psi_{l,1},\ \cdots,\ \psi_{l,2n_2}$ are the components of the lth orthogonal vector of $\left(\boldsymbol{\mathcal{C}}^H_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}\boldsymbol{\mathcal{C}}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}\right)$. Then, the parameters of $\boldsymbol{\theta}_{G_\Delta}=[a_{\Delta,1},\ \cdots,\ a_{\Delta,n},\ b_{\Delta,1},\ \cdots\ b_{\Delta,n}]^T$ can be estimated by minimizing the following function as shown in Moulines et al. (1995)

$$J_{\mathrm{S}}(\hat{\boldsymbol{\theta}}_{G_\Delta})=\frac{1}{2}\left[1\ \hat{\boldsymbol{\theta}}^T_{G_\Delta}\right]\boldsymbol{\Omega}\left[\begin{matrix}1\\ \hat{\boldsymbol{\theta}}_{G_\Delta}\end{matrix}\right], \tag{27}$$

where $\boldsymbol{\Omega}=\sum_l\left(\boldsymbol{V}_l^H\boldsymbol{V}_l\right)$ are calculated from the orthogonal vectors of cyclo-stationary relation matrix $\mathcal{C}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(\alpha_p,\tau)$.

Introducing the spatial information into the information criterion leads to a new criterion

$$J(\hat{\boldsymbol{\theta}}_\Delta)=J_{\mathrm{T}}(\hat{\boldsymbol{\theta}}_\Delta)+\lambda J_{\mathrm{S}}(\hat{\boldsymbol{\theta}}_{G_\Delta}) \tag{28}$$

to complement the information, where the subscript T and S indicate the information in time and space domain, respectively, λ is a positive coefficient determined by the ratio of smallest singular values of $\boldsymbol{H}_{\mathrm{ess,T}}$ and $\boldsymbol{\Omega}$. Then, the Hessian matrix becomes

$$\boldsymbol{H}_{\mathrm{ess}}=\boldsymbol{H}_{\mathrm{ess,T}}+\lambda\boldsymbol{H}_{\mathrm{ess,S}} \tag{29}$$

where $\boldsymbol{H}_{\mathrm{ess,T}}$ is given in (15), while $\boldsymbol{H}_{\mathrm{ess,S}}=\boldsymbol{\Omega}$. Since $\boldsymbol{\Omega}$ is composed of the normalized orthogonal vectors of $\mathcal{C}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(\alpha_p,\tau)$, the condition number of $\boldsymbol{H}_{\mathrm{ess}}$ is improved largely by introducing the positive-semidefinite matrix $\boldsymbol{\Omega}$ into $\boldsymbol{H}_{\mathrm{ess}}$. Moreover, the new information of $\boldsymbol{g}_{\mathrm{S}}(\hat{\boldsymbol{\theta}}^{(l)}_\Delta)=dJ_{\mathrm{S}}(\hat{\boldsymbol{\theta}}^{(l)}_\Delta)\Big/d\hat{\boldsymbol{\theta}}^{(l)}_\Delta$ is also added into the gradient.

4. MULTI-POINT SEARCH BASED IDENTIFICATION

In order to improve the global convergence performance, the model parameters are estimated from M points parallelly. Assume the M initial values are $\hat{\boldsymbol{\theta}}^{(0)}_{\Delta,m}$ and the multiplicities are set as $\gamma_m=1$, $m=1,\cdots,M$.

4.1 Updating estimate

In the lth iteration, the gradient vector and Hessian matrix of each estimate point are calculated $\boldsymbol{g}^{(l)}_m$ and $\boldsymbol{H}^{(l)}_{\mathrm{ess},m}$, where the subscript m indicates the mth estimate point in the multi-point search. Then the Hessian matrix used for estimate updating is given by

$$\bar{\boldsymbol{H}}^{(l)}_{\mathrm{ess},m}=\begin{cases}\boldsymbol{H}^{(l)}_{\mathrm{ess},m} & l\le l_0\\ \sum_{m=1}^{M}w_m\boldsymbol{H}^{(l)}_{\mathrm{ess},m} & l>l_0\end{cases}, \tag{30}$$

where w_m is a weight coefficient given by

$$w_m = \gamma_m e^{-\sigma^2_{\varepsilon,m}} \Bigg/ \sum_{m=1}^{M} \gamma_m e^{-\sigma^2_{\varepsilon,m}}, \tag{31}$$

here $\sigma^2_{\varepsilon,m}$ is the mean squares of the prediction error with respect to the parameter estimates $\hat{\boldsymbol{\theta}}^{(l)}_{\Delta,m}$. l_0 is such an integer that the inverse of Hessian matrices could be approximated by a common one to reduce the computational complexity. Then the estimates are updated by

$$\hat{\boldsymbol{\theta}}^{(l+1)}_{\Delta,m} = \hat{\boldsymbol{\theta}}^{(l)}_{\Delta,m} - \mu\left(\bar{\boldsymbol{H}}^{(l)}_{\text{ess},m}\right)^{-1} \boldsymbol{g}\left(\hat{\boldsymbol{\theta}}^{(l)}_{\Delta,m}\right),$$
$$\text{for } m = 1, \cdots, M \tag{32}$$

where $\bar{\boldsymbol{H}}^{(l)}_{\text{ess},m}$ and $\boldsymbol{g}\left(\hat{\boldsymbol{\theta}}^{(l)}_{\Delta,m}\right)$ contain the information in both the time and space domains. The larger rate P, the more information can be obtained. However, when the signals are band limited, the information obtained by high sampling rate may do little contribution to identification. Commonly, P is chosen as $2 \sim 4$. On the other hand, for $P = 1$ where only $\{u(KT), y(KT)\}$ can be available, $\bar{\boldsymbol{H}}^{(l)}_{\text{ess},m}$ and $\boldsymbol{g}\left(\hat{\boldsymbol{\theta}}^{(l)}_{\Delta,m}\right)$ are just constructed by the time domain information as in (14), (15). While the multi-point search can be performed work for $P = 1$ without over-sampling, the convergence performance may only be guaranteed when the numerical condition is not too poor.

4.2 Merger and re-generation of estimate points

If some estimates are close to each other, e.g.,

$$\left|\hat{\boldsymbol{\theta}}^{(l+1)}_{\Delta,m_1} - \hat{\boldsymbol{\theta}}^{(l+1)}_{\Delta,m_2}\right| < \delta, \tag{33}$$

where δ is a small positive number. Then,

$$m_0 = \min_{m_1,m_2} \left\{\sigma^2_{\varepsilon,m_1},\ \sigma^2_{\varepsilon,m_2}\right\}, \tag{34}$$

where $\sigma^2_{\varepsilon,m}$ it the mean squares of the prediction error corresponding to the the mth estimate. Correspondingly, the m_1 and m_2th estimates are merged, and a new estimates is initialized by random walk $\boldsymbol{\beta}$ as follows:

$$\hat{\boldsymbol{\theta}}^{(l+1)}_{\Delta,m_1} = \hat{\boldsymbol{\theta}}^{(l+1)}_{\Delta,m_0}, \quad \gamma_{m_1} = \gamma_{m_1} + \gamma_{m_2}, \tag{35}$$
$$\hat{\boldsymbol{\theta}}^{(l+1)}_{\Delta,m_2} = \hat{\boldsymbol{\theta}}^{(l+1)}_{\Delta,m_0} + \boldsymbol{\beta}, \quad \gamma_{m_2} = 1. \tag{36}$$

It is seen that the multi-point search has larger probability to make some estimates enter into the attraction region of the global optimum than the single-point search. Moreover, the probability is improved with increasing the point number M. Consequently, the convergence performance improves largely compared with the conventional identification algorithms.

4.3 Identification procedure

The procedures of the identification algorithm are summarized as follows.

(1) Select the initial values $\hat{\boldsymbol{\theta}}^{(0)}_{\Delta,m}$ and $\gamma_m = 1$, and let the iteration number $l = 0$.
(2) In the lth iteration, calculate the prediction error $\varepsilon_\Delta(k, \hat{\boldsymbol{\theta}}^{(l)}_{\Delta,m})$ and instrumental vector $\hat{e}_\Delta(k)$.
(3) Estimate the cyclo-stationary correlation matrices $\mathcal{C}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}(\alpha_p, \tau)$ for $p = 1, \cdots, P-1$, $\tau = 0, \cdots, n$, and execute singular value decomposition of $\boldsymbol{\mathcal{C}}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}$, then construct the matrix $\boldsymbol{\Omega}$ by using the orthogonal vectors of $\boldsymbol{\mathcal{C}}_{\boldsymbol{\phi}_e,\boldsymbol{\phi}}$.
(4) Calculate the gradients and Hessian matrices in time and space domains, respectively,
(5) Update the estimates parallelly.
(6) Merge the close estimates, create new points by random walk, and record the optimal estimate.
(7) Let $l = l+1$ then go to step (2) until the the iteration number l exceeds the maximal number l_{tol}, or there is an estimate whose multiplicity is larger than M.
(8) Choose the optimal one that has smallest noise variance from the M estimates.

5. NUMERICAL EXAMPLES

Consider a 3rd order continuous time process $G_c(s)$ whose transfer function is given by

$$G_c(s) = \frac{2s^2 + 28.8s + 368.18}{s^3 + 5.72s^2 + 147.7296s + 720.648}. \tag{37}$$

It is operated by a digital PID controller $C(z^{-1})$ which is followed by a ZOH with holding period $T = 0.18$s. The reference signal is a constant $r(t) = 0.01$, whereas the process is disturbed by a noise term $e_c(t)$ as follows

$$e_c(t) = \frac{0.0001s + 10}{s^2 + s + 0.2344} w_c(t), \tag{38}$$

where $w_c(t)$ is an i.i.d white signal with $\mathcal{N}(0, 1)$. The signal components of $e_c(t)$ are concentrated in the low frequency band, while the reference is a constant. As a result, the numerical conditions of the input-output data in the closed-loop are very poor, and the conventional parameter estimation methods are hard to construct the system model even in the output over-sampling scheme.

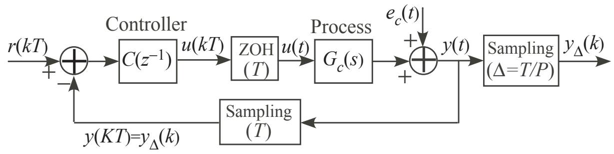

Fig. 3. Process model in closed-loop system

For $P = 1$, i.e., the output is sampled at the same rate of the control input, the data record $\{u(KT), y(KT)\}$ do not have sufficient information for system identification due to the low controller's order and the constant reference signal $r(t)$, see Söderström et al. (1975, 1976). As a result, identification might fail to work just using the data $\{u(KT), y(KT)\}$ sampled at the rate $1/T$. Therefore, the data record $\{u_\Delta(k), y_\Delta(k)\}$ obtained under $P \geq 2$ are used in system identification.

Consider the case of $P = 3$ where the output $y(t)$ is sampled at an interval $\Delta = 0.06$s. The true process model of $G_\Delta(q^{-1})$ with respect to the interval Δ is

$$G_\Delta(q^{-1}) = \frac{0.1483q^{-1} - 0.1477q^{-2} + 0.0638q^{-3}}{1 - 2.2123q^{-1} + 2.0478q^{-2} - 0.7095q^{-3}}. \tag{39}$$

The simulations are performed independently for 10 runs using the input-output data sampled with interval $\Delta = T/3$. In order to improve the numerical conditions, the information in both the time and space domains is utilized in the information criterion $J(\hat{\boldsymbol{\theta}}_\Delta)$. The point number is chosen as $M = 8$, and the initial estimates are selected as random values. The estimated parameters are summarized in Table 1. It is shown that the multi-point search based algorithm has better convergence performance than the conventional single point search algorithms.

Table 1. Estimation results for $P = 3$

Parameters	True	Multi-point search	Single-point search
a_1	-2.2123	-2.1693 ± 0.0604	-2.3152 ± 0.1290
a_2	2.0478	1.9835 ± 0.0902	2.2008 ± 0.1917
a_3	-0.7095	-0.6675 ± 0.0591	-0.8065 ± 0.1204
b_1	0.1483	0.1476 ± 0.0014	0.1482 ± 0.0015
b_2	-0.1477	-0.1420 ± 0.0080	-0.1616 ± 0.0178
b_3	0.0638	0.0618 ± 0.0029	0.0634 ± 0.0032

The 8 estimates of the 50th iteration in multi-point search are shown in Table 2. It is seen that from the initial values, the $6 \sim 8$th estimates are close to the true parameters, whereas the others are approaching to the local minima that should be adjusted to escape from the local minima.

Table 2. M estimates of $[a_1, a_2, a_3]$ in the 50th iteration.

Estimate Number	a_1	a_2	a_3
1	-2.4932	2.4651	-0.9718
2	-0.5528	-0.5971	0.3440
3	-2.3680	2.2638	-0.8410
4	-2.6197	2.4974	-0.8647
5	-2.0590	1.8744	-0.6289
6	-2.1805	2.0004	-0.6790
7	-2.2725	2.1381	-0.7686
8	-2.2747	2.1414	-0.7707
True	-2.2123	2.0478	-0.7095

6. CONCLUSIONS

The identification problem under the severe numerical conditions has been investigated. It has been shown that by using the output over-sampling scheme, the cyclostationary properties in both time and space domains can be extracted from the experimental data, and the cyclostationary information can be used in the information criterion to improve the numerical conditions without introducing external test signals. Furthermore, by introducing the multi-point search into the numerical optimization for model parameter estimation, the local minima problem can be mitigated so the performance of global convergence has been improved. The effectiveness of the proposed approach is demonstrated by numerical simulations. The convergence analysis, the approaches to improve the optimization efficiency and to determine the optimal estimate from the multiple points will be investigated in the future research work.

REFERENCES

Åström, K. and Söderstrom, T. (1974). Uniqeness of maximum likelihood estimates of the parameters of an arma model. *IEEE Trans. on Automatic Control*, 19(6), 769–773.

Chen, T. (2018). On kernel design for regularized LTI system identification. *Automatica*, 90, 109 – 122.

Eckhard, D., Bazanella, A., Rojas, C., and Hjalmarson, H. (2013). Input design as a tool to improve the convergence of PEM. *Automatica*, 49, 3282–3291.

Forssell, U. and Hjalmarsson, H. (1999). Maximum likelihood estimation of models with unstable dynamics and nonminimum phase noise zeros. In *Proc. 14th IFAC World Congress*, volume H, 110–115. Beijing.

Gardner, W. (1994). *Cyclostationarity in Communications and Signal Processing*. IEEE Press.

Ljung, L. (1999). *System Identification – Theory for the User*. Prentice Hall, Englewood Cliffs, NJ.

Moulines, E., Duhamel, P., Cardoso, J., and Mayrargue, S. (1995). Subspace methods for the blind identification of multi-channel fir filters. *IEEE Trans. on Signal Processing*, 43, 516–526.

Nelles, O. (2001). *Nonlinear System Identification*. Springer.

Pillonetto, G. (2018). System identification using kernel-based regularization: New insights on stability and consistency issues. *Automatica*, 93, 321 – 332.

Söderström, T., Gustavsson, I., and Ljung, L. (1975). Identifiability conditions for linear systems operating in closed loop. *Int. J. Control*, 21(2), 243–255.

Söderström, T., Gustavsson, I., and Ljung, L. (1976). Identifiability conditions for linear multivariable systems operating under feedback. *IEEE Trans. Automatic Control*, 21(6), 837–840.

Söderström, T. and Stoica, P. (1981). Comparison of some instrumental variable methods – consistency and accuracy aspects. *Automatica*, 17(1), 101–115.

Söderström, T. and Stoica, P. (1989). *System Identification*. Prentice Hall, NJ, USA.

Sun, L., Ohmori, H., and Sano, A. (2001). Output intersampling approach to closed-loop identification. *IEEE Trans. on Automatic Control*, 46(12), 1936–1941.

Sun, L. and Sano, A. (2017). Temporal-spatial information based approach to direct closed-loop identification. *Trans. of the Society on Instrument and Control Engineers*, 53(6), 346–354.

Sun, L., Sano, A., and Liu, X. (2018). Direct closed-loop identification approach to magnetic levitation system. In *Proc. 18th IFAC Symposium on System Identification*, 610–615. Stockholm.

Talbi, E. (2009). *Metaheuristics: from Design to Implementation*. John Wiley & Sons, Inc.

Van den Hof, P. and Schrama, R. (1995). Identification and control - closed-loop issues. *Automatica*, 31(12), 1751–1770.

Wang, J., Chen, T., and Huang, B. (2003). Closed-loop identification via output fast sampling. *Journal of Process Control*, 14, 555–570.