International Journal of Systems Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsys20

AFD-based ILC designs in frequency domain for linear discrete-time systems

Wen-Yuan Fu, Xiao-Dong Li & Tao Qian

To cite this article: Wen-Yuan Fu, Xiao-Dong Li & Tao Qian (2020) AFD-based ILC designs in frequency domain for linear discrete-time systems, International Journal of Systems Science, 51:16, 3393-3407, DOI: 10.1080/00207721.2020.1815097

To link to this article: https://doi.org/10.1080/00207721.2020.1815097

Published online: 11 Sep 2020.

INTERNATIONAL JOURNAL OF SYSTEMS SCIENCE
2020, VOL. 51, NO. 16, 3393–3407
https://doi.org/10.1080/00207721.2020.1815097

Taylor & Francis
Taylor & Francis Group

AFD-based ILC designs in frequency domain for linear discrete-time systems

Wen-Yuan Fu[a,b], Xiao-Dong Li[a,c] and Tao Qian[d]

[a]School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, People's Republic of China; [b]College of Information Science and Engineering, Huaqiao Univesity, Xiamen, People's Republic of China; [c]School of Information Science, Xinhua College of Sun Yet-sen University, Dongguan, People's Republic of China; [d]Institute of Systems Engineering, Macau University of Science and Technology, Macau, People's Republic of China

ABSTRACT

The existing frequency-domain-based iterative learning control (ILC) methods are highly dependent on the mathematical models of the controlled systems. For linear discrete-time single-input single-output (SISO) systems with unknown mathematical models, this paper tries to present fully data-driven ILC designs in frequency domain. With the help of support vector machine (SVM), the input-output data of the linear discrete-time SISO system at the first repetition is utilised to constitute an adaptive Fourier decomposition (AFD) model. Then, based on the AFD model, a P-type ILC law and an extended D-type ILC law with data-driven determining techniques for learning gains are presented. It is noted that comparing with the conventional D-type ILC law, the newly proposed extended D-type ILC law exhibits superior tracking characteristic due to involving the frequency information during the ILC process. A numerical example is utilised to illustrate the effectiveness of the proposed ILC algorithms with the data-driven determining techniques for learning gains.

ARTICLE HISTORY
Received 7 May 2019
Accepted 20 August 2020

KEYWORDS
Adaptive Fourier decomposition; data-driven; frequency domain; iterative learning control; support vector machine

1. Introduction

Iterative Learning Control (ILC) is a control methodology which focuses on improving the tracking performance of dynamic system that performs the same task repetitively over a fixed time interval. It allows a controller to learn from previous tasks and determine a feedforward control input which forces the system output to track a given reference trajectory. Recent three decades have witnessed a renascent interest in ILC designs (Meng & Moore, 2020; Shen & Zhang, 2017; Xiong et al., 2016; Xu, 2011; Zhang et al., 2010) with extensive applications in industrial robots (Cheah et al., 1994), manufacturing (Zhao et al., 2015), traffic systems (Yu et al., 2018), and chemical processes (Mezghani et al., 2002), etc.

Hitherto, the existing ILC algorithms have been designed mainly in time domain, such as contraction mapping-based ILC (Meng & Moore, 2020; Sun et al., 2017), adaptive ILC (Li et al., 2016; Shao & Xiang, 2019; Xu, 2016), norm-optimal ILC (Yu et al., 2018), point-to-point ILC (Freeman, 2012), terminal ILC (Chi et al., 2014), and hybrid ILC (Ouyang et al., 2006). Many of the aforementioned ILC methods are represented in a proportional-plus-integral-plus-derivative (PID)-type formation (Kurek & Zaremba, 1993; Shen & Zhang, 2017; Xu, 2011). To estimate or determine the learning gains existed in these PID-type ILC laws, relevant parameter information of the controlled system are normally required. In recent years, although generalised Lipschitz condition on controlled system was required, data-driven ILC had attracted considerable attention (Bu et al., 2018; Yu et al., 2020) due to its effectiveness in removing the need of model knowledge of plants. Therefore, to extend the applications of the conventional PID-type ILC algorithms, a fully data-driven technique to determine the learning gains in the PID-type ILC algorithms, which doesn't require relevant model information of the controlled system, is expected.

Compared with the fruitful ILC results in time domain, the frequency-domain-based ILC works are very limited (Ruan & Li, 2014; Ye & Wang, 2006; Zhang, Wang, & Ye, 2009; Zhang, Wang, Ye, Zhou, et al., 2009; Zhang et al., 2010). In Ruan and Li (2014),

CONTACT Xiao-Dong Li lixd@mail.sysu.edu.cn School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510006, People's Republic of China

the average energy of the ILC tracking errors over a finite time interval was converted into a quarter of a summation of fundamental spectrums plus harmonic spectrums. Then, by analysing the frequency spectrum features of the ILC tracking errors, sufficient and necessary conditions for monotone convergence of the ILC tracking errors were presented. In Ye and Wang (2006), for linear continuous-time single-input single-output (SISO) systems, a negative learning gain was used to increase the learnable frequency range in P-type ILC law through the analysis on contraction condition of the ILC tracking errors in frequency domain. And in Zhang, Wang, Ye, Zhou, et al. (2009), a multriate cyclic pseudo-downsampled P-type ILC scheme was proposed and applied to the control of joint 3 which moved in a horizontal plane on a four-axis, closed-loop DC servo industrial SCARA robot. Experimental results demonstrated that the proposed P-type ILC scheme could produce a good learning transient for the desired trajectories with high frequency components with/without initial state errors. Furthermore, the convergence and robustness of the cyclic pseudo-downsampled ILC scheme in Zhang, Wang, Ye, Zhou, et al. (2009) were examined analytically, and proved mathematically in Zhang et al. (2010). In Zhang, Wang, and Ye (2009), for linear discrete-time SISO systems, a cutoff frequency phase-in ILC scheme utilising P-type ILC law was proposed to deal with initial position offset and to improve ILC tracking accuracy. It is worth noting that to achieve good learning transient and high tracking accuracy, all the existing frequency-domain-based ILC designs (Ruan & Li, 2014; Ye & Wang, 2006; Zhang, Wang, & Ye, 2009; Zhang, Wang, Ye, Zhou, et al., 2009; Zhang et al., 2010) are highly dependent on the mathematical models of the controlled systems, and thus lose the most important feature of ILC systems. From an engineering point of view, a frequency-domain-based control technique is sometimes preferable as it may exhibit the spectrum characteristics of system signals and has the lower computation complexity on convolution operation of time-domain signals. Therefore, there is a need to exploit data-driven ILC schemes in frequency domain without utilising accurate model knowledge on the controlled systems.

To address the unknown model knowledge of the controlled systems in frequency domain, it should be noticed that in recent years, a novel Adaptive Fourier Decomposition (AFD) series was proposed as a universal approximator to dynamical systems in frequency domain (Mi & Qian, 2012; Mo et al., 2015; Qian, 2010; Qian et al., 2011). In AFD theory, a frequency-domain-based signal is decomposed as an infinite linear combination of some selected parameterised kernel functions. The breakthrough of AFD is that an easy way to select the poles for the kernel functions one-by-one is found (Qian, 2010; Qian et al., 2011). Owing to the outstanding approximating ability of AFD series to nonlinear functions in frequency-domain, AFD has made much success in some applications such as stock index trend analysis (Zhang et al., 2014), image denoising (Wang et al., 2016), and system identification (Mi & Qian, 2012; Mo et al., 2015), etc. Numerical examples in Mi and Qian (2012) demonstrated that the AFD-based frequency-domain identification algorithm could exhibit a better performance than the finite impulse response (FIR) filter and the Laguerre models of Mäkilä (1991).

The objective of this paper is to exploit fully data-driven ILC algorithms based on AFD for linear discrete-time SISO systems in frequency domain. In the AFD-based ILC algorithms, the input-output data of the linear discrete-time SISO system at the first repetition/cycle is utilised to constitute an AFD model, which can well approximate the transfer function of the linear discrete-time SISO system in statistical learning sense. During the approximation process, a SVM technique is employed to determine the coefficients of the AFD series. Then, based on the AFD model of the linear discrete-time SISO system, a fully data-driven technique to select the learning gain in the P-type ILC law is presented. Furthermore, a novel extended D-type ILC law, which utilises the frequency information of the ILC system, is proposed. To the best of authors' knowledge, this study is the first attempt to apply the AFD theory to ILC designs in frequency domain. Notably, several data-driven ILC techniques based on model approximation with basis functions were ever presented in Bolder and Oomen (2015) and Bolder et al. (2014), but they were designed in time domain. The main features of the paper and its contribution relative to the related works are summarised as follows: (1) A novel AFD series with SVM application is used to approximate the transfer function of the linear discrete-time SISO system, which operates over a finite time interval. (2) Different from other frequency-domain-based ILC designs, the proposed ILC algorithms with determining techniques for learning gains are fully dependent on the input-output data

of the linear discrete-time SISO system, and don't require any model information of the controlled system. (3) The extended D-type ILC law involves the frequency information during the ILC process, and thus exhibits more superior tracking characteristic to the conventional D-type ILC law.

The remainder of this paper is organised as follows: Preliminaries and problem formulation are exhibited in Section 2. Section 3 presents the AFD-based ILC designs in frequency domain. The simulation example is illustrated in Section 4. Finally, Section 5 concludes this paper.

Notations

$\mathbb{H}$ Complex Hilbert space
$\mathbb{X}$ Real Hilbert space
$\mathbb{C}$ Complex space
$\mathbb{R}$ Real space
$\mathbb{D}$ Unit disc
$H^2(\mathbb{D})$ Hardy space on $\mathbb{D}$
$\bar{f}$ denotes the complex conjugation of f, and $\langle\cdot,\cdot\rangle$ denotes the inner product of the Hilbert space.

2. Preliminaries and problem formulation

2.1. The related mathematical theory to complex SVM

In this paper, AFD combined with complex SVM is used to design ILC algorithms in frequency domain. Just like real SVM, the complex SVM is firmly grounded in the framework of statistical learning theory, which solves the minimised upper bound of the generalisation errors embodying the sum of the approximating errors of input-output complex data with a confidence interval (Mo et al., 2015; Shawe-taylor & Cristianini, 2002).

Let S be an input-output complex data set, which is typically given by

$$S = \{(U^{(1)}, Y^{(1)}), (U^{(2)}, Y^{(2)}), \ldots, (U^{(l)}, Y^{(l)})\} \quad (1)$$

where $U^{(r)} \in \mathbb{H}$ and $Y^{(r)} \in \mathbb{C}$, $(r = 1, 2, 3, \ldots, l)$ represent the input data and the output data in S, respectively. Suppose that $\tilde{G}(\cdot)$ is a bounded linear function on the complex Hilbert space $\mathbb{H}$. According to Riesz representation theorem, $\tilde{G}(\cdot)$ is incurred by an element w of $\mathbb{H}$. Let $w = w_{Re} + jw_{\mathrm{Im}} \in \mathbb{H}$, $(w_{Re}, w_{\mathrm{Im}} \in \mathbb{R})$. Then, for any $U = U_{Re} + jU_{\mathrm{Im}} \in \mathbb{H}$, $(U_{Re}, U_{\mathrm{Im}} \in \mathbb{R})$, there is $\tilde{G}(U) = \langle U, w\rangle$. Furthermore,

$$\tilde{G}(U) = \langle U_{Re} + jU_{\mathrm{Im}}, w_{Re} + jw_{\mathrm{Im}}\rangle = \tilde{G}_{Re}(U) + j\tilde{G}_{\mathrm{Im}}(U)$$

where $\tilde{G}_{Re}(U) = \langle U_{Re}, w_{Re}\rangle + \langle U_{\mathrm{Im}}, w_{\mathrm{Im}}\rangle$ and $\tilde{G}_{\mathrm{Im}}(U) = \langle U_{Re}, -w_{\mathrm{Im}}\rangle + \langle U_{\mathrm{Im}}, w_{Re}\rangle$.

For the input-output complex data set S in (1), some results bound on the generalisation error of complex SVM to complex function $\tilde{G}(\cdot)$ are presented as follows:

Definition 2.1: Consider the regression analysis of the input-output complex data set S in (1) with the set $\wp$ of complex linear functions on $\mathbb{H}$. For $\tilde{G}(\cdot) \in \wp$, an approximating error $\hat{\theta}$ of SVM, and a loss margin χ, $(0 < \chi \le \hat{\theta})$, margin slack variables ξ_r and ς_r, $(r = 1, 2, 3, \ldots, l)$ are defined as

$$\begin{cases} \xi_r = \xi_r((U^{(r)}, Y^{(r)}), \tilde{G}(\cdot), \hat{\theta}/2, \chi/2) \\ \quad = \max\{0, |\tilde{G}_{Re}(U^{(r)}) - Re(Y^{(r)})| - (\hat{\theta} - \chi)/2\} \\ \varsigma_r = \varsigma_r((U^{(r)}, Y^{(r)}), \tilde{G}(\cdot), \hat{\theta}/2, \chi/2) \\ \quad = \max\{0, |\tilde{G}_{\mathrm{Im}}(U^{(r)}) - \mathrm{Im}(Y^{(r)})| - (\hat{\theta} - \chi)/2\} \end{cases} \quad (2)$$

where $\tilde{G}_{Re}(U^{(r)}) = \langle U_{Re}^{(r)}, w_{Re}\rangle + \langle U_{\mathrm{Im}}^{(r)}, w_{\mathrm{Im}}\rangle$, $\tilde{G}_{\mathrm{Im}}(U^{(r)}) = \langle U_{Re}^{(r)}, -w_{\mathrm{Im}}\rangle + \langle U_{\mathrm{Im}}^{(r)}, w_{Re}\rangle$, and $U^{(r)} = U_{Re}^{(r)} + jU_{\mathrm{Im}}^{(r)}$, $(U_{Re}^{(r)}, U_{\mathrm{Im}}^{(r)} \in \mathbb{R})$.

As a consequence, the margin slack vectors of input-output complex data set S with the complex function $\tilde{G}(\cdot) \in \wp$ are given by

$$\begin{cases} \xi = \xi(S, \tilde{G}(\cdot), \hat{\theta}/2, \chi/2) = (\xi_1, \xi_2, \ldots, \xi_l)^T \\ \varsigma = \varsigma(S, \tilde{G}(\cdot), \hat{\theta}/2, \chi/2) = (\varsigma_1, \varsigma_2, \ldots, \varsigma_l)^T \end{cases} \quad (3)$$

Lemma 2.1 (Mo et al., 2015): *Let $\wp$ be the set of complex linear functions on $\mathbb{H}$. Fix $\hat{\theta} \in R^+$ and a probability distribution on the space $\mathbb{H} \times \mathbb{C}$. If we restrict the inputs to the ball $B(0, \hat{r}) = \{U \in \mathbb{H} : |U| \le \hat{r}\}$, then there exists a constant c such that with probability at least 1–δ over the randomly drawn input–output complex data set S of size l and for all χ, $0 < \chi \le \hat{\theta}$, the probability that a function $\tilde{G}(\cdot) \in \wp$ with its representation w in $\mathbb{H}$ has error larger than $\hat{\theta}$ on a randomly*

chosen input within $B(0,\hat{r})$ *is bounded by*

$$q(l,\delta,\chi) = \frac{c}{l}\left(\ln^2(l)\cdot\frac{\begin{array}{c}4||w||_2^2\hat{r}^2+||\xi||_2^2\ln(2/\chi)\\+||\varsigma||_2^2\ln(2/\chi)\end{array}}{\chi^2}+2\ln\frac{2}{\delta}\right) \tag{4}$$

with ξ, ς *defined in (3). In other words, with the notation* $err_P(\tilde{G}(\cdot),\hat{\theta}) = P\{(U,Y)\in\mathbb{H}\times\mathbb{C}:|\tilde{G}(U)-Y|\geq\hat{\theta}\}$, *there holds* $P^l\{S: err_P(\tilde{G}(\cdot),\hat{\theta})\leq q(l,\delta,\gamma)\}\geq 1-\delta$, *where* P^l *is the product probability induced by* P *over* $(\mathbb{H}\times\mathbb{C})^l$.

Lemma 2.2: *(Struble, 2013): Let* Y *be a set and* $K: Y\times Y\to\mathbb{R}$ *be a function. If* K *is a kernel function, then there exists a Reproducing Kernel Hilbert Space (RKHS) of functions on* Y *such that* K *is the reproducing kernel of real Hilbert space* $\mathbb{X}$.

According to Lemma 2.2 and Riesz representation theorem, for all $y\in Y$, there exists a unique element $K(\cdot,y)$ of $\mathbb{X}$ with the reproducing property $f(y)=\langle f,K(\cdot,y)\rangle_{\mathbb{X}}$, where $f\in\mathbb{X}$. Meanwhile, for any $z\in Y$, supposed that $f=K(\cdot,z)$, then $K(y,z)=f(y)=\langle f,K(\cdot,y)\rangle_{\mathbb{X}}=\langle K(\cdot,z),K(\cdot,y)\rangle_{\mathbb{X}}$. Thus, we can define the reproducing kernel of $\mathbb{X}$ as a function $K: Y\times Y\to\mathbb{R}$ by $K(y,z)=\langle K(\cdot,z),K(\cdot,y)\rangle_{\mathbb{X}}$ for any $y,z\in Y$. It is easily obtained that the reproducing kernel $K(y,z)$ is positive definite and symmetric.

The regression analysis of complex SVM is essentially to formulate an unknown function $\tilde{G}(\cdot)$ with the input-output complex data set S via minimising the objective function $q(l,\delta,\chi)$ in probability sense. Thus, a complex support vector regression algorithm should be used to minimise the objective function $q(l,\delta,\chi)$ in (4) to improve the approximating ability of complex SVM.

2.2. ILC and the AFD approximation with SVM

Consider the linear discrete-time SISO system which operates repeatedly over the discrete-time instants $\{0,1,2,\ldots,N\}$,

$$\begin{cases}x_k(n+1)=Ax_k(n)+Bu_k(n),\\ y_k(n)=Cx_k(n),\end{cases} \tag{5}$$

where $k=0,\ 1,\ 2,\cdots\cdots$ denotes the repetition times, and $n\in\{0,1,2,\ldots,N\}$ is the discrete-time index. $x_k(n)\in\mathbb{R}^n$, $u_k(n)\in\mathbb{R}^1$, and $y_k(n)\in\mathbb{R}^1$ represent system state, system input, and system output, respectively. For the linear discrete-time system (5), if it is obtained by sampling a linear continuous-time system with sampling period T, then the running time of the output $y_k(n)$ is NT in one repetition.

As the iterative initial state satisfies $x_k(0)=\mathbf{0}\in\mathbb{R}^n$ for $k=0,\ 1,\ 2,\cdots$, it is known that the linear discrete-time system (5) can be converted into the following form in frequency domain,

$$Y(z^{-1})=G(z^{-1})\cdot U(z^{-1}), z\in\mathbb{D} \tag{6}$$

where $G(z^{-1})=C(z^{-1}I-A)^{-1}B$ is the transfer function of the linear discrete-time system (5).

For the linear discrete-time system (5) or (6), given a desired output trajectory $Y_d(z^{-1})$ at $\omega\in[0,\pi/T)$ for $z^{-1}=e^{j\omega T}$, the task of ILC design in frequency domain is to iteratively determine a control input sequence $\{U_k(e^{j\omega T})\}$ such that the system output $\{Y_k(e^{j\omega T})\}$ can well track the desired trajectory $Y_d(e^{j\omega T})$ at $\omega\in[0,\pi/T)$ as the repetition times k goes to infinity.

Regarding the transfer function $G(z^{-1})$, $z\in\mathbb{D}$, it is continuous on $|z|=1$. We have the following Lemma 2.3.

Lemma 2.3: *Let* $G(z^{-1})$, $z\in\mathbb{D}$ *be the transfer function of the linear discrete-time system (5). As the frequency-domain-based input–output data set* $V=\{(\vartheta_r,G(e^{j\vartheta_r})),r=0,1,\ldots,\hat{N}-1\}$ *is obtained by sampling the frequency response of the linear discrete-time system (5) with impulse input signal* $\delta(n)$, *where* $\vartheta_r=\frac{2\pi r}{\hat{N}}$, *and* $\hat{N}$ *denotes the number of sampling points on the boundary of the unit disc* $\mathbb{D}$. *If*

$$\hat{G}(z^{-1},\hat{N})=\frac{1}{\hat{N}}\sum_{r=0}^{\hat{N}-1}\frac{G(e^{j\vartheta_r})e^{-j\vartheta_r}}{e^{-j\vartheta_r}-z} \tag{7}$$

then, $G(z^{-1})=\lim_{\hat{N}\to+\infty}\hat{G}(z^{-1},\hat{N})$.

The proof of Lemma 2.3 is provided in Appendix 1.

From Lemma 2.3, we know that the transfer function $G(z^{-1})$ can be approximated by $\hat{G}(z^{-1},\hat{N})$ as the number $\hat{N}$ of sampling points is large enough. In Mi and Qian (2012), it was proved that for $\vartheta\in[0,2\pi)$ and

$\vartheta + \Delta\vartheta \in [0, 2\pi)$, there is

$$|G(z^{-1}) - \hat{G}(z^{-1}, \hat{N})| < \sup_{\Delta\vartheta \le \left|\frac{2\pi}{\hat{N}}\right|} |G(e^{j(\vartheta+\Delta\vartheta)}) - G(e^{j\vartheta})|$$

On the other hand, it is known that the ILC techniques are often used in the dynamical systems without accurate model knowledge. For the transfer function $G(z^{-1}) = C(z^{-1}I - A)^{-1}B$ of the linear discrete-time system (5), as the matrices A, B, and C in (5) are unknown, $G(z^{-1})$ cannot be available. An input-output data set-based approximating technique is thus needed. For this purpose, using the SVM and AFD theory, we have the following Theorem 1 addressing the input-output data set-based approximation to $G(z^{-1})$ in probability sense.

Theorem 2.1: *For the transfer function $G(z^{-1})$ of the linear discrete system (5), and the function $\hat{G}(z^{-1}, \hat{N})$ given in (7), assume that the data set $S = \{(U^{(1)}, Y^{(1)}), (U^{(2)}, Y^{(2)}), \ldots, (U^{(l)}, Y^{(l)})\}$, $l \le \hat{N}$ with $Y^{(m)} = \hat{G}(1/U^{(m)}, \hat{N}), m = 1, 2, \ldots, l$ is obtained by randomly taking $U^{(m)} \in \mathbb{D}$. Let $\tilde{G}(z^{-1}) = \sum_{m=1}^{l} \frac{\bar{\phi}_m}{1-U^{(m)}\bar{z}}$, where $\phi_m = (\alpha_m - \alpha_m^*) - j(\beta_m - \beta_m^*)$ with $\alpha_m, \alpha_m^*, \beta_m, \beta_m^*$ determined by (B.12 ~ 13) in Appendix 2. Then, the probability, in which the inequality $|G(z^{-1}) - \tilde{G}(z^{-1})| < \theta$ is derived with probability at least $1 - \delta$, is larger than $1 - q(l, \delta, \chi)$, where $q(l, \delta, \chi)$ is the same as in (4) of Lemma 2.1 and $\theta = \hat{\theta} + \sup_{\Delta\vartheta \le \left|\frac{2\pi}{\hat{N}}\right|} |G(e^{j(\vartheta+\Delta\vartheta)}) - G(e^{j\vartheta})|$ with $\hat{\theta}$ the error of approximating $\hat{G}(z^{-1}, \hat{N})$ by $\tilde{G}(z^{-1})$ and $\vartheta, \vartheta + \Delta\vartheta \in [0, 2\pi)$.*

According to the AFD theory on $H^2(\mathbb{D})$, the estimated transfer function $\tilde{G}(z^{-1})$ can be represented as a linear combination of parameterised Szegö kernels (Mo et al., 2015). That is $\tilde{G}(z^{-1}) = \sum_{m=1}^{l} \breve{C}_m \cdot K(U^{(m)}, z)$, where $z \in \mathbb{D}$ and $\breve{C}_m$, $(m = 1, 2, \cdots, l)$ is the coefficient of parameterised Szegö kernel $K(U^{(m)}, z)$. In Theorem 1, there are $\breve{C}_m = \bar{\phi}_m$ and $K(U^{(m)}, z) = \frac{1}{1-U^{(m)}\bar{z}}$, $(m = 1, 2, \cdots, l)$.

In Theorem 1, regarding the approximating error θ to transfer function $G(z^{-1})$ by using the AFD series $\tilde{G}(z^{-1})$, from the notation $\theta = \hat{\theta} + \sup_{\Delta\vartheta \le \left|\frac{2\pi}{\hat{N}}\right|} |G(e^{j(\vartheta+\Delta\vartheta)}) - G(e^{j\vartheta})|$, where $\vartheta \in [0, 2\pi)$ and $\vartheta + \Delta\vartheta \in [0, 2\pi)$, it is noted that θ in fact includes an approximating error to $\hat{G}(z^{-1}, \hat{N})$ by using the AFD series $\tilde{G}(z^{-1})$ and a residual error by using $\hat{G}(z^{-1}, \hat{N})$ to approximate $G(z^{-1})$ in Lemma 2.3. Theorem 1 demonstrates theoretically that the AFD series $\tilde{G}(z^{-1}) = \sum_{m=1}^{l} \frac{\bar{\phi}_m}{1-U^{(m)}\bar{z}}$ can well approximate the unknown transfer function $G(z^{-1})$ in probability sense. Based on the SVM technique and AFD theory, Appendix 2 presents a detailed proof of Theorem 2.1 by utilising Lemmas 2.1–2.3. It is worth noting that the proof process actually provides an approximating algorithm to transfer function $G(z^{-1})$ by using the data set $S = \{(U^{(1)}, Y^{(1)}), (U^{(2)}, Y^{(2)}), \ldots, (U^{(l)}, Y^{(l)})\}$.

Lemma 2.3 and Theorem 2.1 actually constitute a two-step approximation process of the AFD series $\tilde{G}(z^{-1})$ to $G(z^{-1})$. According to Lemma 2.3, $G(z^{-1})$ is firstly approximated by $\hat{G}(z^{-1}, \hat{N})$ as the number $\hat{N}$ of sampling points is large enough due to $G(z^{-1}) = \lim_{\hat{N}\to+\infty} \hat{G}(z^{-1}, \hat{N})$. Then, the data set $S = \{(U^{(1)}, Y^{(1)}), (U^{(2)}, Y^{(2)}), \ldots, (U^{(l)}, Y^{(l)})\}$, $(l \le \hat{N})$ with $Y^{(m)} = \hat{G}(1/U^{(m)}, \hat{N}), m = 1, 2, \ldots, l$ is obtained by randomly taking $U^{(m)} \in \mathbb{D}$. Resultantly, the AFD series $\tilde{G}(z^{-1}) = \sum_{m=1}^{l} \frac{\bar{\phi}_m}{1-U^{(m)}\bar{z}}$ becomes the estimation of $G(z^{-1})$.

Remark 2.1: To achieve the above approximation of $G(z^{-1})$ by using the AFD series $\tilde{G}(z^{-1})$, according to Mi and Qian (2012) and Mo et al. (2015), it is essentially required that the poles of $G(z^{-1})$ are located within the Unit disc $\mathbb{D}$.

Remark 2.2: Actually, function approximation/identification issue by using AFD series was ever addressed in Mi and Qian (2012) and Mo et al. (2015). For subsequent data-driven ILC applications of AFD approximation, with the results in Mi et al. (2012), we cannot obtain the convergence of ILC tracking error in statistical learning sense because the coefficients of parameterised Szegö kernels in AFD series are decided according to the maximal selection principle. While in Mo et al. (2015), the bound of approximation error with AFD series is not provided such that the resultant ILC tracking error cannot be well analysed if the approximation procedure in Mo et al. (2015) is employed.

3. AFD-based ILC designs in frequency domain

To clearly exhibit the AFD-based ILC approach, a simplified design procedure of ILC in frequency domain is depicted in Figure 1.

According to Figure 1, for the system $G(z^{-1})$ in (6), at the initial iteration $k = 0$, input $u_0(n) = \delta(n)$ to get the response signal of $G(z^{-1})$ such that a frequency-domain-based data set $V = \{(\vartheta_r, G(e^{j\vartheta_r})), r = 0, 1, \ldots, \hat{N} - 1\}$ is obtained by sampling, where $\vartheta_r = \frac{2\pi r}{\hat{N}}$. Then, using the proposed AFD technique with SVM, the unknown transfer function $G(z^{-1})$ is well approximated by $\tilde{G}(z^{-1})$ in statistical learning sense. From the estimated transfer function $\tilde{G}(z^{-1})$, the learning gains in the designed ILC laws can be conveniently determined. In the following, two types of ILC laws in frequency domain with determination of learning gains are investigated.

3.1. P-type ILC design with determination of learning gain

For the linear discrete-time system (5), the following P-type ILC law is applied at $n \in \{0, 1, 2, \ldots, N-1\}$,

$$u_{k+1}(n) = u_k(n) + \Gamma \cdot e_k(n+1) \tag{8}$$

where Γ stands for the learning gain, and for the desired output trajectory $y_d(n)$, the ILC tracking error is defined as

$$e_k(n) = y_d(n) - y_k(n) \tag{9}$$

It is noted that the P-type ILC law (8) is not new (Kurek & Zaremba, 1993), but a novel determining scheme for the learning gain Γ is presented in this paper.

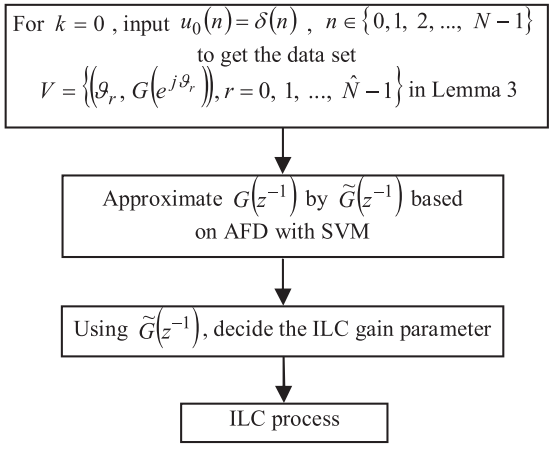

Figure 1. Simplified design procedure of ILC.

Correspondingly, the P-type ILC law (8) in frequency domain is written as

$$U_{k+1}(e^{j\omega T}) = U_k(e^{j\omega T}) + \Gamma \cdot e^{j\omega T} E_k(e^{j\omega T}), \quad \omega \in [0, \pi/T) \tag{10}$$

with the ILC tracking error in frequency domain

$$E_k(e^{j\omega T}) = Y_d(e^{j\omega T}) - Y_k(e^{j\omega T}) \tag{11}$$

Then, the learning gain Γ in the P-type ILC law (8) or (10) can be determined according to the following Theorem 2.

Theorem 3.1: *For the linear discrete-time system (5), the transfer function $G(z^{-1})$ is approximated by $\tilde{G}(z^{-1}) = \sum_{m=1}^{l} \frac{\bar{\phi}_m}{1 - U^{(m)}\bar{z}}$ in Theorem 2.1, and the P-type ILC law (8) is used. If the learning gain Γ makes that*

$$\left| 1 - \Gamma e^{j\omega T} \sum_{m=1}^{l} \frac{\bar{\phi}_m}{1 - U^{(m)} e^{j\omega T}} \right| + |\Gamma|\theta < 1, \quad \omega \in [0, \pi/T) \tag{12}$$

then the probability, in which $\lim_{k \to +\infty} E_k(e^{j\omega T}) = 0$ is derived with probability at least $1 - \delta$, is larger than $1 - q(l, \delta, \chi)$, where $q(l, \delta, \chi)$ is defined in (4) of Lemma 2.1, and $\theta \in \mathbb{R}^+$ denotes the approximating error of transfer function $G(z^{-1})$.

Proof: The ILC tracking error at the $(k+1)$-th trial for the linear discrete-time system (5) is represented as

$$\begin{aligned} E_{k+1}(e^{j\omega T}) &= Y_d(e^{j\omega T}) - Y_{k+1}(e^{j\omega T}) \\ &= Y_d(e^{j\omega T}) - Y_k(e^{j\omega T}) \\ &\quad - [Y_{k+1}(e^{j\omega T}) - Y_k(e^{j\omega T})] \\ &= E_k(e^{j\omega T}) - [Y_{k+1}(e^{j\omega T}) - Y_k(e^{j\omega T})] \end{aligned} \tag{13}$$

According to (6), for $z^{-1} = e^{j\omega T}$, we have

$$Y_k(e^{j\omega T}) = G(e^{j\omega T}) U_k(e^{j\omega T}) \tag{14}$$

From (10), there is

$$\begin{aligned} &Y_{k+1}(e^{j\omega T}) - Y_k(e^{j\omega T}) \\ &= G(e^{j\omega T}) \\ &\quad \times [U_{k+1}(e^{j\omega T}) - U_k(e^{j\omega T})] \\ &= \Gamma e^{j\omega T} G(e^{j\omega T}) E_k(e^{j\omega T}) \end{aligned} \tag{15}$$

Substituting (15) into (13), there is

$$\begin{aligned} E_{k+1}(e^{j\omega T}) &= [1 - \Gamma e^{j\omega T} G(e^{j\omega T})] E_k(e^{j\omega T}) \\ &= [1 - \Gamma e^{j\omega T} \tilde{G}(e^{j\omega T}) - \Gamma e^{j\omega T} (G(e^{j\omega T}) \\ &\quad - \tilde{G}(e^{j\omega T}))] E_k(e^{j\omega T}) \end{aligned} \tag{16}$$

where $\tilde{G}(e^{j\omega T})$ is the estimation to the transfer function $G(e^{j\omega T})$.

Since $|G(e^{j\omega T}) - \tilde{G}(e^{j\omega T})| < \theta$ from Theorem 2.1, (16) can be converted to

$$\begin{aligned} |E_{k+1}(e^{j\omega T})| &< (|1 - \Gamma e^{j\omega T} \tilde{G}(e^{j\omega T})| + |\Gamma|\theta) \\ &\quad \cdot |E_k(e^{j\omega T})| \end{aligned} \tag{17}$$

By Theorem 2.1, $G(e^{j\omega T})$ is estimated as

$$\tilde{G}(e^{j\omega T}) = \sum_{m=1}^{l} \frac{\bar{\phi}_m}{1 - U^{(m)} e^{j\omega T}} \tag{18}$$

Substituting (18) into (17), we have

$$\begin{aligned} &|E_{k+1}(e^{j\omega T})| \\ &< \left(\left| 1 - \Gamma e^{j\omega T} \sum_{m=1}^{l} \frac{\bar{\phi}_m}{1 - U^{(m)} e^{j\omega T}} \right| + |\Gamma|\theta \right) \\ &\quad \cdot |E_k(e^{j\omega T})| \end{aligned} \tag{19}$$

It is concluded that

$$\begin{aligned} &|E_k(e^{j\omega T})| \\ &< \left(\left| 1 - \Gamma e^{j\omega T} \sum_{m=1}^{l} \frac{\bar{\phi}_m}{1 - U^{(m)} e^{j\omega T}} \right| + |\Gamma|\theta \right)^{k-1} \\ &\quad \cdot |E_1(e^{j\omega T})| \end{aligned} \tag{20}$$

Consequently, from the condition (12), if $\left|1 - \Gamma e^{j\omega T} \sum_{m=1}^{l} \frac{\bar{\phi}_m}{1 - U^{(m)} e^{j\omega T}}\right| + |\Gamma|\theta < 1$, we have $\lim_{k \to +\infty} E_k(e^{j\omega T}) = 0$.

It is noted that the derivation of $\lim_{k \to +\infty} E_k(e^{j\omega T}) = 0$ is with a certain probability because the approximation result $|G(e^{j\omega T}) - \tilde{G}(e^{j\omega T})| < \theta$ is obtained in probability sense. From Theorem 2.1, we know that

$$P^l\{P\{|G(e^{j\omega T}) - \tilde{G}(e^{j\omega T})| \geq \theta\} \leq q(l, \delta, \chi)\} \geq 1 - \delta$$

where $P(\cdot)$ accounts for the event probability, and $P^l(\cdot)$ denotes the product probability induced by randomly sampling from the data set S of the ILC system. Subsequently,

$$\begin{aligned} &P^l\{1 - P\{|G(e^{j\omega T}) - \tilde{G}(e^{j\omega T})| < \theta\} \\ &\leq q(l, \delta, \chi)\} \geq 1 - \delta \end{aligned}$$

that is, $P^l\{P\{|G(e^{j\omega T}) - \tilde{G}(e^{j\omega T})| < \theta\} > 1 - q(l, \delta, \chi)\} \geq 1 - \delta$. ■

Therefore, the probability, in which $\lim_{k \to +\infty} E_k(e^{j\omega T}) = 0$ is derived with probability at least $1 - \delta$, is larger than $1 - q(l, \delta, \chi)$. This completes the proof.

Remark 3.1: To design the P-type ILC law (8) for the linear discrete-time system (5), a challenging work in applications is to determine the learning gain Γ. To guarantee the convergence of the ILC tracking error $e_k(n)$, it is normally required that the learning gain Γ should meet the condition $\rho(I - CB\Gamma) < 1$, where $\rho(\cdot)$ represents the spectral radius of matrix (Kurek & Zaremba, 1993). As a result, the knowledge on B and C of the system (5) is in fact required. In this paper, the presented convergence condition (12) in Theorem 1 is based on the estimated transfer function $\tilde{G}(z^{-1})$, and does not require any model information of the system (5). A feasible value of the learning gain Γ can be obtained by solving the inequality (12). Therefore, Theorem 1 with a data-driven determining technique for the learning gain Γ extends the applications of the P-type ILC law (8) to the 'fully-blind' linear discrete-time systems. Furthermore, following a similar procedure to the case of the P-type ILC law (8), some data-driven determining techniques for learning gains in other PID-type ILC laws, i.e. the D-type ILC law, can be easily obtained.

3.2. Relation between the feasible range of Γ in the P-type ILC law (8) and the approximating error θ

In Theorem 1, the condition (12) of determining learning gain $\Gamma \in \mathbb{R}$ is closely related to the approximating error θ of the transfer function $G(e^{j\omega T})$. Suppose that $\tilde{G}(e^{j\omega T}) = \kappa(\omega) \cdot e^{j\nu(\omega)}$, where $\nu(\omega) \in [0, 2\pi)$ and $\kappa(\omega) \in \mathbb{R}^+$. From (12), we get

$$\begin{aligned}\theta &< \frac{1}{|\Gamma|} \cdot (1 - |1 - \Gamma e^{j\omega T}\tilde{G}(e^{j\omega T})|) \\ &= \frac{1}{|\Gamma|} \cdot \left(1 - \sqrt{\begin{matrix}1 + \Gamma^2\kappa^2(\omega) - 2\Gamma\kappa(\omega) \\ \cdot \cos(\nu(\omega) + \omega T)\end{matrix}}\right)\end{aligned} \tag{21}$$

Let $f(\Gamma) = \frac{1}{|\Gamma|}\left(1 - \sqrt{\begin{matrix}1 + \Gamma^2\kappa^2(\omega) - 2\Gamma\kappa(\omega) \\ \cdot \cos(\nu(\omega) + \omega T)\end{matrix}}\right)$ and $g(\Gamma) = -\sqrt{1 + \Gamma^2\kappa^2(\omega) - 2\Gamma\kappa(\omega) \cdot \cos(\nu(\omega) + \omega T)} + 1 - \Gamma\kappa(\omega) \cdot \cos(\nu(\omega) + \omega T)$.

In Appendix 3, it is proved that $g(\Gamma) < 0$ as $0 < \Gamma < \frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}$ or $\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)} < \Gamma < 0$.

If $\Gamma > 0$, then we have $\theta < f(\Gamma) = -\frac{\Gamma\kappa^2(\omega) - 2\kappa(\omega)\cdot\cos(\nu(\omega)+\omega T)}{1+\sqrt{1+\Gamma^2\kappa^2(\omega)-2\Gamma\kappa(\omega)\cdot\cos(\nu(\omega)+\omega T)}}$ from (21). Resultantly, $0 < \Gamma < \frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}$ is obtained from $\theta > 0$. Subsequently, it can be derived that for each $\omega \in [0, \pi/T)$, the function $f(\Gamma)$ decreases monotonically on $0 < \Gamma < \frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}$ due to

$$\begin{aligned}&\frac{\partial f(\Gamma)}{\partial \Gamma} \\ &= \frac{g(\Gamma)}{\Gamma^2\sqrt{1 + \Gamma^2\kappa^2(\omega) - 2\Gamma\kappa(\omega) \cdot \cos(\nu(\omega) + \omega T)}} \\ &< 0.\end{aligned}$$

On the other hand, if $\Gamma < 0$, then we have $\theta < f(\Gamma) = \frac{\Gamma\kappa^2(\omega) - 2\kappa(\omega)\cdot\cos(\nu(\omega)+\omega T)}{1+\sqrt{1+\Gamma^2\kappa^2(\omega)-2\Gamma\kappa(\omega)\cdot\cos(\nu(\omega)+\omega T)}}$ from (21). Thus, $\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)} < \Gamma < 0$ is gotten from $\theta > 0$. Subsequently, it is obtained that for each $\omega \in [0, \pi/T)$, $f(\Gamma)$ increases monotonically on $\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)} < \Gamma < 0$ due to

$$\begin{aligned}&\frac{\partial f(\Gamma)}{\partial \Gamma} \\ &= \frac{-g(\Gamma)}{\Gamma^2\sqrt{1 + \Gamma^2\kappa^2(\omega) - 2\Gamma\kappa(\omega) \cdot \cos(\nu(\omega) + \omega T)}} \\ &> 0\end{aligned}$$

Based on the properties of $f(\Gamma)$ mentioned above, the smaller the approximating error θ is, the larger the feasible range of Γ can be selected.

3.3. Extended D-type ILC design with determination of learning gain

It is known that for the linear discrete-time system (5), a D-type ILC law was ever proposed in Cheah et al. (1994) as follows:

$$u_{k+1}(n) = u_k(n) + L \cdot [e_k(n+1) - e_k(n)] \tag{22}$$

In this paper, an extended D-type ILC law is further suggested for $n \in \{0, 1, 2, \ldots, N-1\}$,

$$\begin{aligned}u_{k+1}(n) = u_k(n) + L \cdot [e_k(n+1) - e_k(n) - \frac{T}{\pi} \\ \times \int_0^{\frac{\pi}{T}} E_k(e^{j\omega T})e^{j\omega nT}(e^{j\omega T} - 1 - j\omega T)d\omega]\end{aligned} \tag{23}$$

where $E_k(e^{j\omega T})$ is the frequency domain form of $e_k(n)$, and L is the learning gain.

The extended D-type ILC law (23) can be transferred to the following form in frequency domain

$$\begin{aligned}U_{k+1}(e^{j\omega T}) = U_k(e^{j\omega T}) + L \cdot [e^{j\omega T}E_k(e^{j\omega T}) - E_k(e^{j\omega T}) \\ - \frac{T}{\pi}\sum_{s=0}^{N-1} e^{-j\omega sT}\int_0^{\frac{\pi}{T}} E_k(e^{j\omega T})e^{j\omega sT} \\ \times (e^{j\omega T} - 1 - j\omega T)d\omega]\end{aligned} \tag{24}$$

Applying the Discrete Time Fourier Transform (DTFT) Theorem in Tu and Zhang (2008) to (24), we get

$$\begin{aligned}U_{k+1}(e^{j\omega T}) &= U_k(e^{j\omega T}) + L \cdot [e^{j\omega T}E_k(e^{j\omega T}) - E_k(e^{j\omega T}) \\ &\quad - E_k(e^{j\omega T}) \cdot (e^{j\omega T} - j\omega T - 1)] \\ &= U_k(e^{j\omega T}) + L \cdot j\omega T E_k(e^{j\omega T})\end{aligned} \tag{25}$$

According to (25), (13), and (14), we have

$$E_{k+1}(e^{j\omega T}) = (1 - L \cdot j\omega T G(e^{j\omega T}))E_k(e^{j\omega T}) \tag{26}$$

Theorem 3.2: *For the linear discrete-time system (5), the transfer function $G(z^{-1})$ is approximated by $\tilde{G}(z^{-1}) = \sum_{m=1}^{l} \frac{\bar{\phi}_m}{1 - U^{(m)}\bar{z}}$ in Theorem 2.1, and the extended D-type ILC law (23) is used. If the learning gain L makes that*

$$\begin{aligned}&\left|1 - j\omega TL\sum_{m=1}^{l}\frac{\bar{\phi}_m}{1 - U^{(m)}e^{j\omega T}}\right| + \omega T\theta|L| < 1, \\ &\times \omega \in [0, \pi/T)\end{aligned} \tag{27}$$

then the probability, in which $\lim_{k\to+\infty} E_k(e^{j\omega T}) = 0$ is derived with probability at least $1 - \delta$, is larger than $1 -$

$q(l,\delta,\chi)$, where $q(l,\delta,\chi)$ is defined in (4) of Lemma 2.1, and $\theta \in \mathbb{R}^+$ denotes the approximating error of transfer function $G(e^{j\omega T})$.

Proof: From (26), it yields

$$\begin{aligned} &E_{k+1}(e^{j\omega T}) \\ &= [1 - j\omega TL\tilde{G}(e^{j\omega T}) - j\omega TL(G(e^{j\omega T}) - \tilde{G}(e^{j\omega T}))] \\ &\quad \cdot E_k(e^{j\omega T}) \end{aligned} \tag{28}$$

Considering (18) and $|G(e^{j\omega T}) - \tilde{G}(e^{j\omega T})| < \theta$ in Theorem 2.1, the following can be further derived from (28),

$$\begin{aligned} &|E_{k+1}(e^{j\omega T})| \\ &< (|1 - j\omega TL\tilde{G}(e^{j\omega T})| + \omega T\theta|L|) \\ &\quad \cdot |E_k(e^{j\omega T})| \\ &< \left(\left|1 - j\omega TL\sum_{m=1}^{l}\frac{\bar{\phi}_m}{1 - U^{(m)}e^{j\omega T}}\right| + \omega T\theta|L|\right) \\ &\quad \cdot |E_k(e^{j\omega T})| \\ &< \left(\left|1 - j\omega TL\sum_{m=1}^{l}\frac{\bar{\phi}_m}{1 - U^{(m)}e^{j\omega T}}\right| + \omega T\theta|L|\right)^k \\ &\quad \cdot |E_1(e^{j\omega T})| \end{aligned} \tag{29}$$

Consequently, according to (27), we have $\lim_{k\to+\infty} E_k(e^{j\omega T}) = 0$.

Similar to the subsequent proof process of Theorem 2, we can further show that the probability, in which $\lim_{k\to+\infty} E_k(e^{j\omega T}) = 0$ is derived with probability at least $1 - \delta$, is larger than $1 - q(l,\delta,\chi)$. This completes the proof.

Remark 3.2: It is worth noting that the D-type ILC law (22) is very noisy from numerical differentiation, and the bound of ultimate ILC tracking errors is proportional to the noise level. Also, the convergence rate of the D-type ILC law (22) is slower at lower and higher frequency ranges (Heinzinger et al., 1992). Thus the effectiveness of the conventional D-type ILC law (22) is often decreased by the noise and variety of frequencies involved in control signals, which can be well demonstrated by the following D-type ILC law (22) in frequency form,

$$\begin{aligned} &U_{k+1}(e^{j\omega T}) \\ &= U_k(e^{j\omega T}) + L \cdot [e^{j\omega T}E_k(e^{j\omega T}) - E_k(e^{j\omega T})] \\ &= U_k(e^{j\omega T}) \\ &\quad + L \cdot \left[E_k(e^{j\omega T})\sum_{i=0}^{+\infty}\frac{(j\omega T)^i}{i!} - E_k(e^{j\omega T})\right] \\ &= U_k(e^{j\omega T}) \\ &\quad + L \cdot \left[E_k(e^{j\omega T})\sum_{i=2}^{+\infty}\frac{(j\omega T)^i}{i!} + j\omega TE_k(e^{j\omega T})\right] \end{aligned} \tag{30}$$

According to (30), it is the harmonic frequency components $E_k(e^{j\omega T})\sum_{i=2}^{+\infty}\frac{(j\omega T)^i}{i!}$ of error signal that deteriorate the performance of the conventional D-type ILC law. While from the extended D-type ILC law (25) in frequency form, it can be seen that the original harmonic frequency components $E_k(e^{j\omega T})\sum_{i=2}^{+\infty}\frac{(j\omega T)^i}{i!}$ have been fully excluded. Consequently, as illustrated in Figure 3 of the simulation example in next section, compared with the results of the conventional D-type ILC law (22), the convergence rate and accuracy degree of the ILC tracking are greatly improved by using the extended D-type ILC law (23). ■

3.4. Relation between the feasible range of L in the extended D-type ILC law (23) and the approximating error θ

In Theorem 3.1, the condition (27) of determining learning gain L in the extended D-type ILC law (23) is closely related to the approximating error θ of the transfer function $G(e^{j\omega T})$. Similar to the analysis in Section 3.2, from (27), we can infer that $\theta < \frac{1}{\omega T|L|}(1 - |1 - j\omega TL\tilde{G}(e^{j\omega T})|)$. Assuming that $\tilde{G}(e^{j\omega T}) = \frac{\mu(\omega)}{\omega T}\cdot e^{j\left(\nu(\omega)-\frac{\pi}{2}\right)}$, where $\nu(\omega) \in \left[\frac{\pi}{2}, \frac{5\pi}{2}\right)$ and $\mu(\omega) \in \mathbb{R}^+$, we get

$$\theta < \frac{1}{\omega T|L|}\left(1 - \sqrt{1 + L^2\mu^2(\omega) - 2L\mu(\omega)\cos\nu(\omega)}\right)$$

Let $h(L) = \frac{1}{\omega T|L|}\left(1 - \sqrt{\begin{matrix}1+L^2\mu^2(\omega)\\ -2L\mu(\omega)\cos\nu(\omega)\end{matrix}}\right)$. If $L > 0$, then $\theta < h(L) = -\frac{L\mu^2(\omega) - 2\mu(\omega)\cos\nu(\omega)}{\omega T\left(1+\sqrt{1+L^2\mu^2(\omega)-2L\mu(\omega)\cos\nu(\omega)}\right)}$, where $0 < L < \frac{2\cos\nu(\omega)}{\mu(\omega)}$ can be obtained by $\theta > 0$. It can be derived that for each $\omega \in [0, \pi/T)$, $h(L)$ decreases monotonically on $0 < L < \frac{2\cos\nu(\omega)}{\mu(\omega)}$. Furthermore, if $L < 0$, we get $\frac{2\cos\nu(\omega)}{\mu(\omega)} < L < 0$, and for each $\omega \in [0, \pi/T)$, $h(L)$ increases monotonically on

$\frac{2\cos\nu(\omega)}{\mu(\omega)} < L < 0$. As a result, it is concluded that the smaller the approximating error θ is, the larger the feasible range of the learning gain L can be chosen for the extended D-type ILC law (23).

4. Simulation example

In this section, an example is presented to show the effectiveness of the proposed AFD-based ILC designs in frequency domain for linear discrete-time SISO systems.

Example 4.1: Consider the ILC issue of the following linear discrete-time SISO system (Zhang et al., 2005)

$$G(z^{-1}) = \frac{0.02277z^{-1}}{z^{-2} - 1.659z^{-1} + 0.683} \tag{31}$$

The desired output trajectory of the system (31) is $y_d(n) = 1 - e^{-0.4n}$ for $n = 0, 1, 2, .., 100$. In the ILC process of the system (31), the iterative initial output is $y_k(0) = 0$ for $k = 0, 1, 2, ...$ The accuracy of ILC is evaluated by the following average error power index $FEE^{(k)}$ in frequency domain,

$$FEE^{(k)} = \frac{1}{p}\sum_{i=1}^{p} |E_k(e^{j\omega_i T})|^2 \tag{32}$$

where $p = 232$ is set as the maximum number of i in ω_i, and ω_i is evenly sampled at $[0, \pi/T)$. With the same parameter settings of $T = 1s$, $l = 640$, $\chi = 0.001$, and $\delta = 0.06$, the simulation consists of the following three parts:

(1) The transfer function $G(z^{-1})$ is approximated by $\tilde{G}(z^{-1})$. Input $u_0(n) = \delta(n)$ to the system (31), and then obtain a frequency-domain-based data set $V = \{(\vartheta_r, G(e^{j\vartheta_r})), r = 0, 1, \ldots, \hat{N} - 1\}$ with $\vartheta_r = \frac{2\pi r}{\hat{N}}$ and $\hat{N} = 640$. Based on the data set V and a SVM software program, $\tilde{G}(e^{j\omega T})$ is obtained by following the procedure introduced in Section 2. As the approximating error is set as $\theta = 0.04$, Figure 2 presents the curve of the approximating function $\tilde{G}(e^{j\omega T})$ at $\omega \in [0, \pi)$. As a comparison, the curve of the transfer function $G(e^{j\omega T})$ is also given in Figure 2. Clearly, the effect of the AFD-based approximation method to transfer function $G(e^{j\omega T})$ is very promising.

(2) Apply the P-type ILC law (8) and the extended D-type ILC law (23) to the system (31). As the approximating error of transfer function $G(z^{-1})$ is set as $\theta = 0.04$, with the approximating function $\tilde{G}(e^{j\omega T})$, $\omega \in [0, \pi)$ obtained in (1), the learning gains in the P-type ILC law (8) and the extended D-type ILC law (23) can be determined by a searching algorithm to satisfy the convergent conditions (12) and (27), respectively. In this example, they are taken as $\Gamma = 1.31$ and $L = 1.31$. With the initial control input $u_0(n) = 1.8 \cdot \sin(n)$ for $n = 0, 1, 2, .., 99$, Figure 3 presents the profiles of the ILC tracking error index $FEE^{(k)}$ by using the P-type ILC law (8) and the extended D-type ILC law (23), respectively. And the actual tracking performance of the system output $y_k(n)$ to the desired

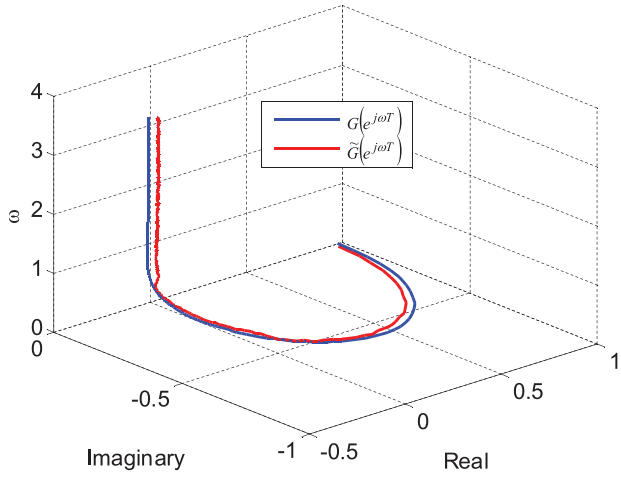

Figure 2. Curves of the approximating function $\tilde{G}(e^{j\omega T})$ and the transfer function $G(e^{j\omega T})$.

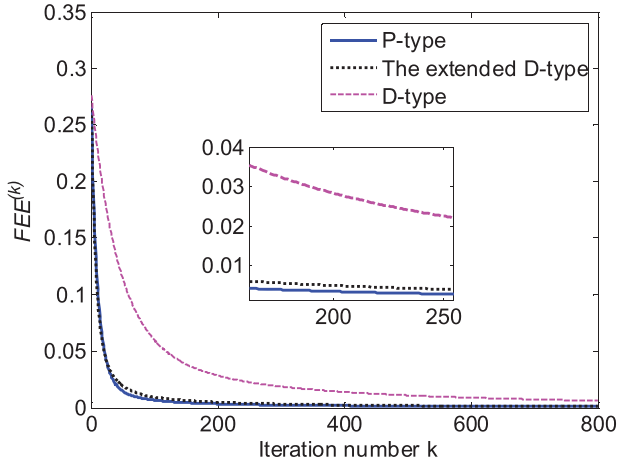

Figure 3. Profiles of the ILC tracking error index $FEE^{(k)}$ at different iterations by using the P-type ILC law (8), the D-type ILC law (22), and the extended D-type ILC law (23), respectively.

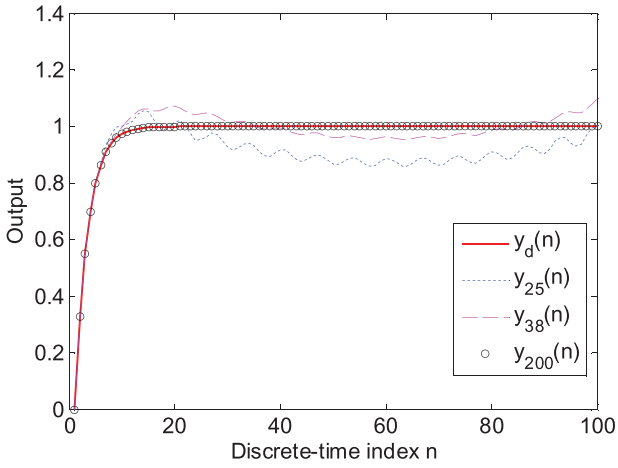

Figure 4. Actual tracking performance of the system output $y_k(n)$ to the desired trajectory $y_d(n)$ at iterations $k = 25$, $k = 38$, and $k = 200$, respectively, with the P-type ILC law (8).

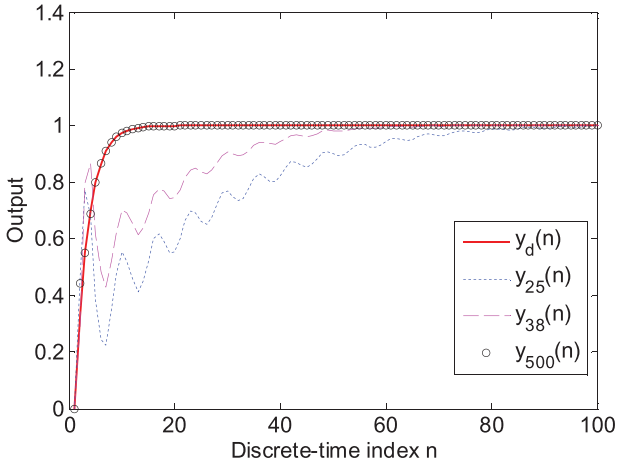

Figure 5. Actual tracking performance of the system output $y_k(n)$ to the desired trajectory $y_d(n)$ at iterations $k = 25$, $k = 38$, and $k = 500$, respectively, with the extended D-type ILC law (23).

trajectory $y_d(n)$ at different iterations is given in Figures 4 and 5, respectively. For further comparison, the ILC tracking error index $FEE^{(k)}$ obtained by using the traditional D-type ILC law (22) with the learning gain value $L = 1.31$ is also exhibited in Figure 3. Clearly, it is observed from Figure 3 that the ILC tracking error index $FEE^{(k)}$ can be uniformly driven to zero by using the three types of ILC laws. The extended D-type ILC law (23) is a little complicated in construction, but it makes the ILC process convergent in less iterations than the D-type ILC law (22) because the original harmonic frequency components of the ILC tracking error are fully filtered by using the extended D-type ILC law (23). Remark 3.2 on the extended D-type ILC law (23) is thus illustrated. In addition, it is worth noting that the P-type ILC law (8) makes the ILC tracking error index $FEE^{(k)}$ to convergence very quickly. However, as indicated in Ye and Wang (2006), compared to the D-type ILC law (22), and furthermore, the extended D-type ILC law (23), the P-type ILC law (8) may introduce a bigger transient in the ILC process because it doesn't capture the direction or trend of tracking errors that occurred in the previous operations.

Table 1. Convergence times and convergence rate for the P-type ILC law (8) and the extended D-type ILC law (23).

	Total times	Convergence times	Convergence rate
The P-type ILC law	300	283	94.33%
The extended D-type ILC law	300	288	96.00%

(3) To further demonstrate the convergence of the proposed AFD-based ILC designs in probabilistic sense, the above simulations are repetitively done 300 times respectively by using the P-type ILC law (8) and the extended D-type ILC law (23). Table 1 presents the convergence times and rates of ILC tracking errors by using the two ILC laws. It is noted that the computational result of SVM program shows $q(l, \delta, \chi) = 0.03$ with the parameter settings at the beginning of this example. Evidently, the convergence rates of ILC tracking errors with the two ILC laws in Table 1 are larger than the value of $(1 - q(l, \delta, \chi)) \cdot (1 - \delta) = 91.18\%$. Theorem 2.1 and Theorem 2.2 are thus illustrated in statistical learning sense.

5. Conclusion

For linear discrete-time SISO systems with unknown mathematical models, two fully data-driven ILC designs in frequency domain have been proposed with verification of some numerical results. In these frequency-domain-based ILC designs, an AFD model combing with the SVM technique is first constituted by utilising the input–output data of the linear discrete-time SISO system at the first cycle such that the unknown linear discrete-time SISO system is well identified in the sense of statistical learning theory. And then, a P-type ILC law and an extended D-type ILC law with data-driven determining techniques for

learning gains are derived. As a result, the high dependence on the controlled system models in conventional frequency-domain-based ILC designs is relaxed. It is worth noting that the proposed extended D-type ILC law involves the frequency information against harmonic frequency distortion during the ILC process, and thus exhibits improved features in convergence rate and tracking accuracy compared to the conventional D-type ILC law. In future work, motivated by the results in Zhang and Luttervelt (2011) and Han et al. (2016), robustness and resilience of the AFD-based ILC designs in frequency domain will be investigated.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported in part by the National Natural Science Foundation of China [grant number 61573385], and in part by the Science and Technology Development Fund of Macao SAR [FDCT079/2016/A2, FDCT0123/2018/A3].

Notes on contributors

Wen-Yuan Fu received a B.S. and an M. Phil. from the School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2005 and 2008, respectively. At present, he is pursuing his Ph.D. from the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His research interests include iterative learning control and digital signal processing.

Xiao-Dong Li received a B.S. from the Department of Mathematics, Shaanxi Normal University, Xian, China, in 1987; an M. Phil. from the Nanjing University of Science and Technology, Nanjing, China, in 1990; and a Ph.D. from the City University of Hong Kong, Hong Kong, in 2007. He is currently a Professor in the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His research interests include 2-D system theory, iterative learning control, and artificial intelligence.

Tao Qian received the M.Sc. and Ph.D. degrees, both in harmonic analysis, from Peking University, Beijing, China, in 1981 and 1984, respectively. From 1984 to 1986, he worked in Institute of Systems Science, the Chinese Academy of Sciences. Then he worked as Research Associate and Research Fellow in Australia till 1992 (Macquarie University, Flinders University of South Australia), and as a faculty teaching member at New England University, Australia, from 1992 to 2000. He started working at University of Macau, Macao, China SAR, from 2000 as Associate Professor. His job in University of Macau was continuing, and he got Full Professorship in 2003, and was Head of Department of Mathematics from 2005 to 2011. He was appointed as Distinguished Professor at University of Macau in April, 2013. In 2019, he worked as Professor at Macau University of Science and Technology and Emeritus Professor at University of Macau. His research interests include harmonic analysis in Euclidean spaces, complex and Clifford analysis and signal and image analysis. He has published over 200 journal and conference papers and two monograph books.

References

Bolder, J., & Oomen, T. (2015). Rational basis functions in iterative learning control – with experimental verification on a motion system. *IEEE Transactions on Control Systems Technology*, *23*(2), 722–729. https://doi.org/10.1109/TCST.2014.2327578

Bolder, J., Oomen, T., Koekebakker, S., & Steinbuch, M. (2014). Using iterative learning control with basis functions to compensate medium deformation in a wide-format inkjet printer. *Mechatronics*, *24*(8), 944–953. https://doi.org/10.1016/j.mechatronics.2014.07.003

Bouboulis, P., & Theodoridis, S. (2010). Extension of Wirtinger's calculus to reproducing kernel Hilbert spaces and the complex kernel LMS. *IEEE Transactions on Signal Processing*, *59*(3), 964–978. https://doi.org/10.1109/TSP.2010.2096420

Bouboulis, P., Theodoridis, S., Mavroforakis, C., & Dalla, L. (2015). Complex support vector machines for regression and quaternary classification. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(6), 1260–1274. https://doi.org/10.1109/TNNLS.2014.2336679

Bu, X., Cui, L., Hou, Z., & Qian, W. (2018). Formation control for a class of nonlinear multiagent systems using model-free adaptive iterative learning. *International Journal of Robust & Nonlinear Control*, *28*(1), 1–11. https://doi.org/10.1002/rnc.3851

Cheah, C. C., Wang, D., & Soh, Y. C. (1994). Convergence and robustness of a discrete-time learning control scheme for constrained manipulators. *Journal of Robotic Systems*, *11*(3), 223–238. https://doi.org/10.1002/rob.4620110308

Chi, R., Hou, Z., Jin, S., & Wang, D. (2014). Improved data-driven optimal TILC using time-varying input signals. *Journal Process Control*, *24*(12), 78–85. https://doi.org/10.1016/j.jprocont.2014.07.007

Freeman, C. T. (2012). Constrained point-to-point iterative learning control with experimental verification. *Control Engineer Practice*, *20*(5), 489–498. https://doi.org/10.1016/j.conengprac.2012.01.003

Han, B., Liu, C. L., & Zhang, W. J. (2016, June 28–30). *A method to measure the resilience of algorithm for operation*

management. 8th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2016, Paper Presentation, Troyes, France.

Heinzinger, G., Fenwick, D., Paden, B., & Miyazaki, F. (1992). Stability of learning control with disturbances and uncertain initial conditions. *IEEE Transactions on Automatic Control*, *37*(1), 110–114. https://doi.org/10.1109/9.109644

Kurek, J. E., & Zaremba, M. B. (1993). Iterative learning control synthesis based on 2-D system theory. *IEEE Transactions on Automatic Control*, *38*(1), 121–125. https://doi.org/10.1109/9.186321

Li, X. D., Lv, M. M., & Ho, J. K. L. (2016). Adaptive ILC algorithms of nonlinear continuous systems with non-parametric uncertainties for non-repetitive trajectory tracking. *International Journal of Systems Science*, *47*(10), 2279–2289. https://doi.org/10.1080/00207721.2014.992493

Mäkilä, P. M. (1991). Laguerre methods and H_∞ identification of continuous-time systems. *International Journal of Control*, *53*(3), 689–707. https://doi.org/10.1080/00207179108953641

Meng, D., & Moore, K. L. (2020). Contraction mapping-based robust convergence of iterative learning control with uncertain, locally Lipschitz nonlinearity. *IEEE Transactions on Systems Man, Cybernetics-Systems*, *50*(2), 442–454. https://doi.org/10.1109/TSMC.2017.2780131

Mezghani, M., Roux, G., Cabassud, M., Lann, M. V. L., Dahhou, B., & Casamatta, G. (2002). Application of iterative learning control to an exothermic semibatch chemical reactor. *IEEE Transactions Control Systems Technology*, *10*(6), 822–834. https://doi.org/10.1109/TCST.2002.804117

Mi, W., & Qian, T. (2012). Frequency domain identification: An algorithm based on adaptive rational orthogonal system. *Automatica*, *48*(6), 1154–1162. https://doi.org/10.1016/j.automatica.2012.03.002

Mo, Y., Qian, T., & Mi, W. (2015). Sparse representation in Szego kernels through reproducing kernel Hilbert space theory with applications. *International Journal of Wavelets Multi-Resolution & Information Processing*, *13*, 1–22. https://doi.org/10.1142/S0219691315500307

Ouyang, P. R., Zhang, W. J., & Gupta, M. M. (2006). An adaptive switching learning control method for trajectory tracking of robot manipulators. *Mechatronics*, *16*(1), 51–61. https://doi.org/10.1016/j.mechatronics.2005.08.002

Picinbono, B., & Chevalier, P. (2002). Widely linear estimation with complex data. *IEEE Transactions on Signal Processing*, *43*(8), 2030–2033. https://doi.org/10.1109/78.403373

Qian, T. (2010). Intrinsic mono-component decomposition of functions: An advance of Fourier theory. *Mathematical Methods in Applied Science*, *33*(7), 880–891. https://doi.org/10.1002/mma.1214

Qian, T., Zhang, L., & Li, Z. (2011). Algorithm of adaptive Fourier decomposition. *IEEE Transactions on Signal Processing*, *59*(12), 5899–5906. https://doi.org/10.1109/TSP.2011.2168520

Ruan, X., & Li, Z. (2014). Convergence characteristics of PD-type iterative learning control in discrete frequency domain. *Journal of Process Control*, *24*(12), 86–94. https://doi.org/10.1016/j.jprocont.2014.10.001

Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural Computation*, *12*(5), 1207–1245. https://doi.org/10.1162/089976600300015565

Shao, Z., & Xiang, Z. (2019). Adaptive iterative learning control for switched nonlinear continuous-time systems. *International Journal of Systems Science*, *50*(5), 1028–1038. https://doi.org/10.1080/00207721.2019.1587029

Shawe-taylor, J., & Cristianini, N. (2002). On the generalisation of soft margin algorithms. *IEEE Transactions on Information Theory*, *48*(10), 2721–2735. https://doi.org/10.1109/TIT.2002.802647

Shen, D., & Zhang, C. (2017). Learning control for discrete-time nonlinear systems with sensor saturation and measurement noises. *International Journal of Systems Science*, *48*(13), 2764–2778. https://doi.org/10.1080/00207721.2017.1344894

Struble, D. W. (2013). Wavelets adapted to compact domains in reproducing kernel Hilbert spaces. *Applied Mathematical Sciences*, *7*, 893–908. https://doi.org/10.12988/ams.2013.13081

Sun, S. T., Li, X. D., & Zhong, R. X. (2017). An open-closed-loop iterative learning control approach for nonlinear switched systems with application to freeway traffic control. *International Journal of System Science*, *48*(13), 2752–2763. https://doi.org/10.1080/00207721.2017.1346153

Tu, Y., & Zhang, H. (2008). Method for CMF signal processing based on the recursive DTFT algorithm with negative frequency contribution. *IEEE Transactions on Instrumentation and Measurement*, *57*(11), 2647–2654. https://doi.org/10.1109/TIM.2008.925006

Wang, Z., Wan, F., Wong, C. M., & Zhang, L. (2016). Adaptive Fourier decomposition based ECG denoising. *Computers in Biology & Medicine*, *77*, 195–205. https://doi.org/10.1016/j.compbiomed.2016.08.013

Xiong, W., Yu, X., Patel, R., & Yu, W. (2016). Iterative learning control for discrete-time systems with event-triggered transmission strategy and quantization. *Automatica*, *72*, 84–91. https://doi.org/10.1016/j.automatica.2016.05.031

Xu, J. (2016). Adaptive iterative learning control for high-order nonlinear multi-agent systems consensus tracking. *Systems & Controll Letters*, *89*, 16–23. https://doi.org/10.1016/j.sysconle.2015.12.009

Xu, J. X. (2011). A survey on iterative learning control. *International Journal of Control*, *84*(7), 1275–1294. https://doi.org/10.1080/00207179.2011.574236

Ye, Y., & Wang, D. (2006). Learning more frequency components using P-type ILC with negative learning gain. *IEEE Transactions on Industrial Electronics*, *53*, 712–716. https://doi.org/10.1109/TIE.2006.870881

Yu, Q., Hou, Z., & Xu, J. X. (2018). D-type ILC based dynamic modeling and norm optimal ILC for high-speed trains. *IEEE Transactions on Control Systems Technology*, *26*(2), 652–663. https://doi.org/10.1109/TCST.2017.2692730

Yu, X., Hou, Z., Polycarpou, M. M., & Duan, L. (2020). Data-driven iterative learning control for nonlinear discrete-time MIMO systems. *IEEE Transactions on Neural Networks and*

Learning Systems, *99*, 1–13. https://doi.org/10.1109/TNNLS.2020.2980588

Zhang, B., Wang, D., & Ye, Y. (2005). Wavelet transform-based frequency tuning ILC. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, *35*(1), 107–114. https://doi.org/10.1109/TSMCB.2004.841411

Zhang, B., Wang, D., & Ye, Y. (2009). Cutoff-frequency phase-in iterative learning control. *IEEE Transactions on Control Systems Technology*, *17*(3), 681–687. https://doi.org/10.1109/TCST.2008.2000986

Zhang, B., Wang, D., Ye, Y., & Zhou, K. (2010). Stability and robustness analysis of cyclic pseudo-downsampled iterative learning control. *International Journal of Control*, *83*(3), 651–659. https://doi.org/10.1080/00207170903373753

Zhang, B., Wang, D., Ye, Y., Zhou, K., & Wang, Y. (2009). Cyclic pseudo-downsampled iterative learning control for high performance tracking. *Control Engineering Practice*, *17*(8), 957–965. https://doi.org/10.1016/j.conengprac.2009.02.016

Zhang, L., Zhang, D., & Li, W. (2014). Stock index trend analysis based on signal decomposition. *IEICE Transactions on Information & Systems*, *E97-D*(8), 2187–2190. https://doi.org/10.1587/transinf.E97.D.2187

Zhang, W. J., & Luttervelt, C. A. V. (2011). Toward a resilient manufacturing system. *CIRP Annals – Manufacturing Technology*, *60*(1), 469–472. https://doi.org/10.1016/j.cirp.2011.03.041

Zhao, Y. M., Lin, Y., Xi, F. F., & Guo, S. (2015). Calibration-based iterative learning control for path tracking of industrial robots. *IEEE Transactions on Industrial Electronics*, *62*(5), 2921–2929. https://doi.org/10.1109/TIE.2014.2364800

Appendices

Appendix 1. Proof of Lemma 2.3

Proof: From the Cauchy's integral formula, we have $G(z^{-1}) = \frac{1}{2\pi j}\oint_{|\zeta|=1}\frac{G(\zeta^{-1})d\zeta}{\zeta - z}$. Assuming that $\zeta = e^{-j\vartheta}$, we obtain $d\zeta = -j\cdot e^{-j\vartheta}d\vartheta$. As a result, $G(z^{-1}) = -\frac{1}{2\pi j}\int_0^{2\pi}\frac{-je^{-j\vartheta}G(e^{j\vartheta})d\vartheta}{e^{-j\vartheta}-z} = \frac{1}{2\pi}\int_0^{2\pi}\frac{e^{-j\vartheta}G(e^{j\vartheta})d\vartheta}{e^{-j\vartheta}-z}$. Let $[0, 2\pi) \in \mathbb{R}$ be divided equally into $\hat{N}$ sub-intervals $[\vartheta_r, \vartheta_{r+1})$ indexed by $r = 0, 1, \ldots, \hat{N}-1$, where $\vartheta_r = \frac{2\pi r}{\hat{N}}$. From the definition of integral, it yields

$$G(z^{-1}) = \frac{1}{2\pi}\lim_{\breve{\Delta\vartheta}\to 0}\sum_{r=0}^{\hat{N}-1}\left(\frac{e^{-j\vartheta_r}G(e^{j\vartheta_r})}{e^{-j\vartheta_r}-z}\breve{\Delta\vartheta}\right)$$
$$= \lim_{\hat{N}\to+\infty}\frac{1}{\hat{N}}\sum_{r=0}^{\hat{N}-1}\left(\frac{e^{-j\vartheta_r}G(e^{j\vartheta_r})}{e^{-j\vartheta_r}-z}\right)$$

where $\breve{\Delta\vartheta} = \frac{2\pi}{\hat{N}}$. That is $G(z^{-1}) = \lim_{\hat{N}\to+\infty}\hat{G}(z^{-1}, \hat{N})$. This completes the proof.

Appendix 2. Proof of Theorem 2.1

Proof: The proof consists of the following three steps:

(1) From Lemma 2 and the adaptive approximation in Mi and Qian (2012), we obtain that

$$\tilde{G}(z^{-1}) = \langle K(\cdot, z), w\rangle \tag{A1}$$

where $K(\cdot, z)$ denotes the parameterised Szegö kernel and $z \in \mathbb{D}, w \in H^2(\mathbb{D})$. Therefore, there exists a complex number $a_m \in \mathbb{C}^1$ such that

$$Y^{(m)} = \hat{G}(1/U^{(m)}, \hat{N}) = \tilde{G}(1/U^{(m)}) + a_m$$
$$= \left\langle K(\cdot, U^{(m)}), w\right\rangle + a_m, \quad (m = 1, 2, \ldots, l)$$

(2) For $d \in \mathbb{R}^1$ and $\varepsilon \in \mathbb{R}^1$, define a ε-loss function $L_\varepsilon(d)$ as,

$$L_\varepsilon(d) = \begin{cases} 0 |d| \le \varepsilon \\ (|d| - \varepsilon)^2 |d| > \varepsilon \end{cases} \tag{A2}$$

In sequel, for the obtained complex numbers $a_m \in \mathbb{C}^1$, $(m = 1, 2, \ldots, l)$ in step 2, define $L_\varepsilon(a_m) = L_\varepsilon(Re(a_m)) + L_\varepsilon(\mathrm{Im}(a_m))$ as the ε-loss function on $a_m \in \mathbb{C}^1$, where $\varepsilon = (\hat{\theta} - \chi)/2$. According to Lemma 1, there is

$$q(l, \delta, \chi) = \frac{4c\ln^2(l)\hat{r}^2}{l\chi^2}\left(||w||^2 + \frac{\ln(2/\chi)}{4\hat{r}^2}(||\xi||_2^2 + ||\varsigma||_2^2)\right) + \frac{2c}{l}\ln\frac{2}{\delta} \tag{A3}$$

Let

$$S(w) = \frac{1}{2}||w||^2 + b\cdot\left(\sum_{m=1}^{l}\xi_m^2 + \sum_{m=1}^{l}\varsigma_m^2\right) \tag{A4}$$

where $b = \frac{\ln(2/\chi)}{8\hat{r}^2}$. In comparison with (A3) and (A4), the objective of complex support vector regression algorithm is to minimise $q(l, \delta, \chi)$ or $S(w)$ by selecting a suitable $w \in H^2(\mathbb{D})$.

From (A1), we have

$$L_\varepsilon(Re(a_m)) = L_\varepsilon(Re(Y^{(m)}) - \tilde{G}_{Re}(1/U^{(m)})) = \xi_m^2 \tag{A5}$$

$$L_\varepsilon(\mathrm{Im}(a_m)) = L_\varepsilon(\mathrm{Im}(Y^{(m)}) - \tilde{G}_{\mathrm{Im}}(1/U^{(m)})) = \varsigma_m^2 \tag{A6}$$

where ξ_m and ς_m are defined in (2). Therefore, from (A5) and (A6), we obtain

$$S(w) = \frac{1}{2}||w||^2 + b\cdot\sum_{m=1}^{l}L_\varepsilon(a_m) \tag{A7}$$

(1) According to $Y^{(m)} = \tilde{G}(1/U^{(m)}) + a_m$ and (A5), (A6), we derive that $\xi_m = \max\{|Re(a_m)| - \varepsilon, 0\}$ and $\varsigma_m = \max\{|\mathrm{Im}(a_m)| - \varepsilon, 0\}$. Let

$$\begin{cases} \breve{\xi}_m = \max\{Re(a_m) - \varepsilon, 0\}, Re(a_m) > 0 \\ \breve{\xi}_m^* = \max\{Re(-a_m) - \varepsilon, 0\}, Re(a_m) \le 0 \\ \breve{\varsigma}_m = \max\{\mathrm{Im}(a_m) - \varepsilon, 0\}, \mathrm{Im}(a_m) > 0 \\ \breve{\varsigma}_m^* = \max\{\mathrm{Im}(-a_m) - \varepsilon, 0\}, \mathrm{Im}(a_m) \le 0 \end{cases} \tag{A8}$$

According to (A7) and (A8), the minimisation of (A7) is transferred to minimise $\frac{1}{2}||w||^2 + b\cdot\sum_{m=1}^{l}$

$\left(\breve{\xi}_m^2+\breve{\xi}_m^{*\,2}+\breve{\varsigma}_m^2+\breve{\varsigma}_m^{*\,2}\right)$ subject to

$$\begin{cases}\breve{\xi}_m-(Re(a_m)-\varepsilon)\geq 0\\ \breve{\xi}_m^*-(Re(-a_m)-\varepsilon)\geq 0\\ \breve{\varsigma}_m-(\mathrm{Im}(a_m)-\varepsilon)\geq 0\\ \breve{\varsigma}_m^*-(\mathrm{Im}(-a_m)-\varepsilon)\geq 0\end{cases}\tag{A9}$$

Noting that $a_m=Y^{(m)}-\tilde{G}(1/U^{(m)})=Y^{(m)}-\langle K(\cdot,U^{(m)}),w\rangle$, and utilising the Lagrangian multipliers method with the Karush-Kuhn-Tucker condition in Schölkopf et al. (2000) and Picinbono and Chevalier (2002), it turns

$$\begin{aligned}&\underset{w\in H^2(\mathbb{D})}{Minimize}\left\{\tfrac{1}{2}||w||^2+b\cdot\textstyle\sum_{m=1}^{l}\left(\breve{\xi}_m^2+\breve{\xi}_m^{*\,2}+\breve{\varsigma}_m^2+\breve{\varsigma}_m^{*\,2}\right)\right.\\&-\textstyle\sum_{m=1}^{l}\alpha_m\left[\breve{\xi}_m+Re\left(\langle K(\cdot,U^{(m)}),w\rangle-(Y^{(m)})\right)+\varepsilon\right]\\&-\textstyle\sum_{m=1}^{l}\alpha_m^*\left[\breve{\xi}_m^*+Re\left((Y^{(m)})-\langle K(\cdot,U^{(m)}),w\rangle\right)+\varepsilon\right]\\&-\textstyle\sum_{m=1}^{l}\beta_m\left[\breve{\varsigma}_m+\mathrm{Im}\left(\langle K(\cdot,U^{(m)}),w\rangle-(Y^{(m)})\right)+\varepsilon\right]\\&\left.-\textstyle\sum_{m=1}^{l}\beta_m^*\left[\breve{\varsigma}_m^*+\mathrm{Im}\left((Y^{(m)})-\langle K(\cdot,U^{(m)}),w\rangle\right)+\varepsilon\right]\right.\\&\left.-\textstyle\sum_{m=1}^{l}\left(\gamma_m\cdot\breve{\xi}_m+\gamma_m^*\cdot\breve{\xi}_m^*+\eta_m\cdot\breve{\varsigma}_m+\eta_m^*\cdot\breve{\varsigma}_m^*\right)\right\}\end{aligned}\tag{A10}$$

where $\breve{\xi}_m,\breve{\xi}_m^*,\breve{\varsigma}_m,\breve{\varsigma}_m^*,\alpha_m,\alpha_m^*,\beta_m,\beta_m^*\geq 0$.

According to Bouboulis et al. (2015) and Bouboulis and Theodoridis (2010), for (A10), there exists a unique global optimum

$$w=\sum_{m=1}^{l}\phi_m K(\cdot,U^{(m)})\in H^2(\mathbb{D})\tag{A11}$$

where $\phi_m=(\alpha_m-\alpha_m^*)-j(\beta_m-\beta_m^*)$. Introducing the Wolfe duality theorem in Shawe-taylor and Cristianini (2002) and Schölkopf et al. (2000) and exploiting the Karush-Kuhn-Tucker condition in Schölkopf et al. (2000) and Picinbono and Chevalier (2002), the required Lagrangian multipliers $\alpha_m,\alpha_m^*,\beta_m,\beta_m^*$ can be definitely solved by

$$\begin{aligned}&\underset{\alpha_m,\alpha_m^*,\beta_m,\beta_m^*\in\mathbb{R}}{Maximize}\left\{-\tfrac{1}{2}\textstyle\sum_{m=1}^{l}\sum_{p=1}^{l}[(\alpha_m-\alpha_m^*)(\alpha_p-\alpha_p^*)\right.\\&+(\beta_m-\beta_m^*)(\beta_p-\beta_p^*)]\cdot Re(K(U^{(p)},U^{(m)}))\\&+\textstyle\sum_{m=1}^{l}\sum_{p=1}^{l}(\beta_m-\beta_m^*)(\alpha_p-\alpha_p^*)\cdot\mathrm{Im}(K(U^{(p)},U^{(m)}))\\&+\textstyle\sum_{m=1}^{l}[(\alpha_m-\alpha_m^*)Re(Y^{(m)})+(\beta_m-\beta_m^*)\mathrm{Im}(Y^{(m)})]\\&-\textstyle\sum_{m=1}^{l}(\alpha_m+\alpha_m^*+\beta_m+\beta_m^*)\,\varepsilon\\&\left.+b\textstyle\sum_{m=1}^{l}\left(\breve{\xi}_m^2+\breve{\xi}_m^{*^2}+\breve{\varsigma}_m^2+\breve{\varsigma}_m^{*^2}\right)\right\}\end{aligned}\tag{A12}$$

subject to

$$\begin{cases}\breve{\xi}_m\left(2b\breve{\xi}_m-\alpha_m\right)=0\\ \breve{\xi}_m^*\left(2b\breve{\xi}_m^*-\alpha_m^*\right)=0\\ \breve{\varsigma}_m\left(2b\breve{\varsigma}_m-\beta_m\right)=0\\ \breve{\varsigma}_m^*\left(2b\breve{\varsigma}_m^*-\beta_m^*\right)=0\end{cases}\tag{A13}$$

where $\alpha_m,\alpha_m^*,\beta_m,\beta_m^*\in[0,+\infty)$ and $m=1,2,3,\ldots,l$.

As a result, from (A1) and (A11)

$$\begin{aligned}\tilde{G}(z^{-1})&=\langle K(\cdot,z),w\rangle\\&=\left\langle K(\cdot,z),\sum_{m=1}^{l}\phi_m K(\cdot,U^{(m)})\right\rangle\\&=\sum_{m=1}^{l}\bar{\phi}_m\left\langle K(\cdot,z),K(\cdot,U^{(m)})\right\rangle\end{aligned}\tag{A14}$$

According to Bouboulis et al. (2015), there is

$$\left\langle K(\cdot,z),K(\cdot,U^{(m)})\right\rangle=K(U^{(m)},z)\tag{A15}$$

Combing (A14) with (A15), we get

$$\begin{aligned}\tilde{G}(z^{-1})&=\sum_{m=1}^{l}\bar{\phi}_m K(U^{(m)},z)\\&=\sum_{m=1}^{l}\frac{\bar{\phi}_m}{1-U^{(m)}\bar{z}}\end{aligned}\tag{A16}$$

From Lemma 2.1, the probability, in which the inequality $|\hat{G}(z^{-1})-\tilde{G}(z^{-1})|<\hat{\theta}$ is derived with probability at least $1-\delta$, is larger than $1-q(l,\delta,\chi)$, where $q(l,\delta,\chi)$ is the same as (1) in Lemma 2.1. Meanwhile, we obtain that $|G(z^{-1})-\hat{G}(z^{-1},\hat{N})|<\sup_{\Delta\vartheta\leq\left|\frac{2\pi}{\hat{N}}\right|}|G(e^{j(\vartheta+\Delta\vartheta)})-G(e^{j\vartheta})|$ from Lemma 2.3 and Mi and Qian (2012). As a result, $|G(z^{-1})-\tilde{G}(z^{-1})|<\hat{\theta}+\sup_{\Delta\vartheta\leq\left|\frac{2\pi}{\hat{N}}\right|}|G(e^{j(\vartheta+\Delta\vartheta)})-G(e^{j\vartheta})|=\theta$. And the probability, in which $|G(z^{-1})-\tilde{G}(z^{-1})|<\theta$ is derived with probability at least $1-\delta$, is larger than $1-q(l,\delta,\chi)$. This completes the proof of Theorem 1.

Appendix 3. Proof of $g(\Gamma)<0$ as $0<\Gamma<\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}$ or $\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}<\Gamma<0$ in section 3.2

Proof: For

$$\begin{aligned}g(\Gamma)=&-\sqrt{1+\Gamma^2\kappa^2(\omega)-2\Gamma\kappa(\omega)\cdot\cos(\nu(\omega)+\omega T)}\\&+1-\Gamma\kappa(\omega)\cdot\cos(\nu(\omega)+\omega T)\end{aligned}$$

there is

$$\begin{aligned}&\frac{\partial^2 g(\Gamma)}{\partial\Gamma^2}\\&=-\kappa^2(\omega)\sin^2(\nu(\omega)+\omega T)\\&\quad\times\left(\sqrt{1+\Gamma^2\kappa^2(\omega)-2\Gamma\kappa(\omega)\cdot\cos(\nu(\omega)+\omega T)}\right)^{-\frac{3}{2}}<0\end{aligned}$$

That means that $g'(\Gamma)$ decreases monotonically at $0<\Gamma<\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}$ or $\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}<\Gamma<0$. As $0<\Gamma<\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}$, it yields $g'(\Gamma)<g'(0)=0$, which indicates that $g(\Gamma)$ decreases monotonically. Therefore, we have $g(\Gamma)<g(0)=0$ at $0<\Gamma<\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}$. On the other hand, as $\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}<\Gamma<0$, there is $g'(\Gamma)>g'(0)=0$, which means $g(\Gamma)$ increases monotonically. That is $g(\Gamma)<g(0)=0$ at $\frac{2\cos(\nu(\omega)+\omega T)}{\kappa(\omega)}<\Gamma<0$. This completes the proof.