International Journal of Control
Vol. ??, No. ?, Month?? 2008, 1–17

Taylor & Francis
Taylor & Francis Group

Dual-mode predictive control algorithm for constrained Hammerstein systems

Hai-Tao Zhang[ab]*, Han-Xiong Li[b] and Guanrong Chen[c]

[a]*The Key Laboratory of Image Processing and Intelligent Control, Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, P.R. China;* [b]*Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong;* [c]*Department of Electronic Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong*

(*Received 21 April 2007; final version received 28 December 2007*)

In the process industry, there exist many systems which can be approximated by a Hammerstein model. Moreover, these systems are usually subjected to input magnitude constraints. In this paper, a multi-channel identification algorithm (MCIA) is proposed, in which the coefficient parameters are identified by least squares estimation (LSE) together with a singular value decomposition (SVD) technique. Compared with traditional single-channel identification algorithms, the present method can enhance the approximation accuracy remarkably, and provide consistent estimates even in the presence of coloured output noises under relatively weak assumptions on the persistent excitation (PE) condition of the inputs. Then, to facilitate the following controller design, this MCIA is converted into a two stage single-channel identification algorithm (TS-SCIA), which preserves most of the advantages of MCIA. With this TS-SCIA as the inner model, a dual-mode non-linear model predictive control (NMPC) algorithm is developed. In detail, over a finite horizon, an optimal input profile found by solving a open-loop optimal control problem drives the non-linear system state into the terminal invariant set; afterwards a linear output-feedback controller steers the state to the origin asymptotically. In contrast to the traditional algorithms, the present method has a maximal stable region, a better steady-state performance and a lower computational complexity. Finally, simulation results on a heat exchanger are presented to show the efficiency of both the identification and the control algorithms.

Keywords: Hammerstein systems; singular value decomposition; model predictive control

Nomenclature

$\mathbf{H}_2^{m\times n}(\mathbf{T})$: Hardy space of $(m\times n)$ transfer matrices whose elements are in the Hardy space $\mathbf{H}_2(\mathbf{T})$
$\mathbb{R}^{n\times m}$: $n\times m$-dimensional real space
$\mathbb{R}^+$: one-dimensional positive real space
$\gamma(k)$: stochastic output noise
$u(k)$: input of the Hammerstein system
$v(k)$: intermediate variable of the Hammerstein system
$y(k)$: output of the Hammerstein system
S: number of the sampling input/output data pairs
$\mathbf{N}(\cdot)$: memoryless non-linear block of the Hammerstein system
$G(z^{-1})$: dynamic linear block of the Hammerstein system
z^{-1}: one-step backward shifting operator
N: Laguerre series truncation length
r: non-linear basis series truncation length
$g_i(\cdot)(i=1,\ldots,r)$: non-linear basis of the non-linear block of the Hammerstein system
a_i $(i=1,\ldots,r)$: coefficients of $g_i(\cdot)$
$L_l(\cdot)(l=1,\ldots,N)$: Laguerre series basis
$c_l(l=1,\ldots,N)$: Laguerre series coefficients
Θ_{ac}: coefficient matrix of the Hammerstein system
$\varpi_j(j=1,\ldots,\mathrm{rank}(\hat{\Theta}_{ac}))$: singular values of $\hat{\Theta}_{ac}$
NC: number of identification channels
$L^{(j)}(k)(j=1,\ldots,N_1)$: Laguerre state vector of the jth channel
$\hat{f}(\cdot)$: non-linear block of the single-channel identification model
$\hat{a}$: coefficient parameters of the non-linear block of the single-channel identification model

*Corresponding author. Email: zht@mail.hust.edu.cn

ISSN 0020–7179 print/ISSN 1366–5820 online

DOI: 10.1080/00207170701885453
http://www.informaworld.com

$\hat{c}$: coefficient parameters of the linear block of the single-channel identification model
$v(k), \bar{v}(k)$: the nominal and actual intermediate variables of the single-channel identification model, respectively
$\bar{u}$: upper bound of the input magnitude
σ: upper bound of the inversion error
$\hat{f}^{-1}_{zeroin}(\cdot)$: inversion of the single-channel model's non-linear block using the zeroin method
K: state-feedback gain
$\bar{K}$: extended state-feedback gain
$L(k)$: Laguerre state
$\hat{L}(k)$: estimation of $L(k)$
$D(k)$: perturbation signal vector
S_L: initial ellipsoidal invariant set of $\hat{L}(k)$
M: prediction horizon
$S_{I,M}$: extended ellipsoidal invariant set of $\hat{L}(k)$
$\hat{x}(k)$: extended system state estimation
S_x: feasible stable ellipsoidal invariant set of $\hat{x}(k)$
$\mathbf{E}(\cdot)$: mathematical expectation
$\varsigma(\cdot)$: standard deviation

1. Introduction

In industrial processes, most dynamical systems can be better represented by non-linear models, which are able to describe the systems over large operation ranges, rather than by linear ones that are only able to approximate the systems around given operation points (Minesh and Matthew 1999; Gómez and Baeyens 2004). One of the most frequently studied classes of non-linear models is the Hammerstein model (Eskinat and Johnson 1991; Minesh and Matthew 1999), which consists of the cascade connection of a static (memoryless) non-linear block followed by a dynamic linear block. Under certain considerations such as fading memory assumption (Boyd and Chua 1985), the Hammerstein approximation could be a good representation. Thus, this model structure has been successfully applied to chemical processes: heat exchanger (Eskinat and Johnson 1991); distillation (Huner and Korenberg 1986; Eskinat and Johnson 1991; Bhandari and Rollins 2004); biological processes (Garcia et al. 1989; Hasiewicz 1999); signal processing (Stapleton and Bass 1985; Bai and Fu 2002), and communications (Greblicki 1996; Bai and Fu 2002). In recent years, identification and control of Hammerstein systems has become one of the most needed and yet very difficult tasks in the field of the process industry.

In model predictive control (MPC) framework (Garcia et al. 1989; Henson 1998), the input is calculated by on-line minimisation of a performance index based on model predictions. It is well known that the control quality relies on the accuracy of the model. In recent years, extensive efforts were devoted to the modelling of Hammerstein non-linearities (Greblicki and Pawlak 1989; Eskinat and Johnson 1991; Bai 1998; Greblicki 2002; Gómez and E.Baeyens 2004; Hasiewicz and Mzyk 2004). For example, Bai (1998) studied single input/single output (SISO) systems subject to external white noise. Gómez and Baeyens (2004) designed a non-iterative identification with guaranteed consistent estimation even in the presence of coloured output noise. Both of their works use only one channel to identify the system; therefore, owing to the singular value decomposition (SVD) nature of their methods, the identification errors usually cannot be minimised. A basic reason is that the error is determined by the second largest singular value (for SISO system) or the $(n+1)$st largest singular value (for MIMO system with n inputs) of the estimated coefficients matrix. For a SISO system, if the sampling set is not big enough or the persistent excitation (PE) conditions are not fulfilled, the second largest singular value cannot be neglected, making the identification accuracy unsatisfactory or even unacceptable.

On the other hand, the research on the control of Hammerstein systems is still on the midway so far. Most of the existent control algorithms have some of the following disadvantages:

- reliance on prior knowledge;
- insufficiently large closed-loop stable regions;
- limited capacity of handling input constraints.

In detail, Haddad and Chellaboina (2001) suggested a design that can guarantee global asymptotic closed-loop stability for non-linear passive systems by embedding a non-linear dynamic compensator with a suitable input non-linearity, which requires the memoryless non-linear block to be partially known or measurable without considering input constraints. Patwardhan et al. (1998) used a partial least square (PLS) framework to decompose the modelling problem into a series of univariate problems in the latent subspace while preserving optimality of the constraints. In this way, they can extend the SISO formulation into a constrained MIMO scenario. In this approach, however, the computational complexity is prohibitive, and the reliance on prior knowledge cannot be eliminated. Knohl et al. (2003) slightly alleviated this reliance by an artificial neural network (ANN) inverse compensation, which makes the control scheme more flexible, but its stable region is still small. Zhu et al. (1991) and Fruzzetti et al. (1997) developed generalised predictive control (GPC) and MPC algorithms respectively by taking input constraints into account. These schemes still cannot

ensure a large stable region in general, and require prior knowledge of the real plant such as order, structure, partial coefficients, etc. Bolemen et al. (2001) extended their own work (Bolemen et al. 2000), which preserves the convex property of the optimisation problem, but does not consider input constraints. In order to enlarge the asymptotically stable region for constrained non-linear systems, Chen and Allgöwer (1998b) developed a quasi-infinite horizon non-linear model predictive control (NMPC) algorithms based on a dual-mode (or two-step) technique, which has opened a new avenue in this fascinating field. In Chen and Allgöwer's (1998b) paper there are three important investigations made by Lin and Saberi (1993), Kouvartakis et al. (2000) and Ding and Xi (2006). More precisely, Kouvartakis et al. (2000) proposed a new approach that deployed a fixed state-feedback law with the assistance of extra degrees of freedom through the use of perturbations, which led to a significant reduction in computational cost. More generally, for linear systems with actuator rate saturation, Lin and Saberi (1993) designed both state-feedback and output-feedback control laws that achieve semi-global asymptotic stabilisation based on the assumption of detectability of the system. For input saturated Hammerstein systems, Ding and Xi (2006) designed a two-step MPC by solving non-linear algebraic equation group and desaturation. The stable region is enlarged and its domain of attraction is designed applying semi-global stabilisation techniques. Unfortunately, this work is still based on the measurability of the state of the linear block.

Based on the above analysis, two important tasks are formulated as follows:

- Task one: Develop a better identification algorithm to separate the non-linear/linear blocks of the Hammerstein system more effectively so that some mature linear control theories can be used to facilitate the non-linear control algorithm design.
- Task two: Develop a more efficient control algorithm for constrained Hammerstein systems.

Bearing these tasks in mind, we propose a NMPC algorithm based on a two stage single-channel identification algorithm (TS-SCIA). More precisely:

- A multi-channel identification algorithm (MCIA) is developed for Hammerstein systems which eliminates requirement of prior knowledge about the plant and minimises the identification errors. The MCIA is then converted to a TS-SCIA thereby facilitating the controller design. A sufficient condition for the convergence and approximation capability is given for the new algorithm.
- A dual-mode NMPC algorithm is developed by taking the above mentioned two stage single-channel identification model (TS-SCIM) as the internal model. The closed-loop stable region is maximised by using ellipsoidal invariant set theory together with linear matrix inequality (LMI) techniques.

This paper is organised as follows. The research problems are first formulated in the next section. Then, in § 3, the MCIA and its conversion, TS-SCIA, are proposed, where their approximation capabilities are also analysed. Some technical issues about the dual-mode NMPC algorithm are presented in § 4. Afterwards, simulations are reported in § 5. Finally, concluding remarks are given in § 6.

2. Problems formulation

A Hammerstein system is schematically represented in Figure 1 (Gómez and Baeyens 2004). The model consists of memoryless non-linear block $\mathbf{N}(\cdot)$ in cascade with a dynamic linear block $G(z^{-1}) \in \mathbf{H}_2^{m\times n}(\mathbf{T})$. The measured output $y(k)$ is subject to an unknown additive noise input $\gamma(k)$.

The input/output relationship is then given by

$$y(k) = G(z^{-1})\mathbf{N}(u(k)) + \gamma(k), \tag{1}$$

where $u(k) \in \mathbf{D} \subset \mathbb{R}^n$, $v(k) = \mathbf{N}(u(k)) \in \mathbb{R}^n$, $y(k) \in \mathbb{R}^m$, $\gamma(k) \in \mathbb{R}^m$ are the input, intermediate variable, output and output noise vector at time k, respectively, in which $\mathbf{D}$ is the permission region or constraints of the input, e.g., for SISO system $|u(k)| \le \bar{u}$. Output noise $\gamma(k)$ can be a white or coloured noise sequence induced by measurement or external disturbances, and input signal $u(k)$ can be random or stationary.

The non-linear block is described as (Gómez and Baeyens 2004)

$$\mathbf{N}(u(k)) = \sum_{i=1}^{r} a_i g_i(u(k)), \tag{2}$$

where $g_i(\cdot)\colon \mathbb{R}^n \to \mathbb{R}^n \quad (i = 1, \ldots, r)$ are known non-linear basis functions, and $a_i \in \mathbb{R}^{n\times n}$ are unknown matrix coefficient parameters. Here, $g(\cdot)$ can be chosen as polynomials, radial basis functions (RBF), wavelets, etc. However, it is not the intention of this paper to give a complete overview of non-linear approximation using basis functions, and the interested reader is referred to the works of Sjoberg et al. (1995) and Juditsky et al. (1995), where a unified overview of non-linear black-box modelling using basis

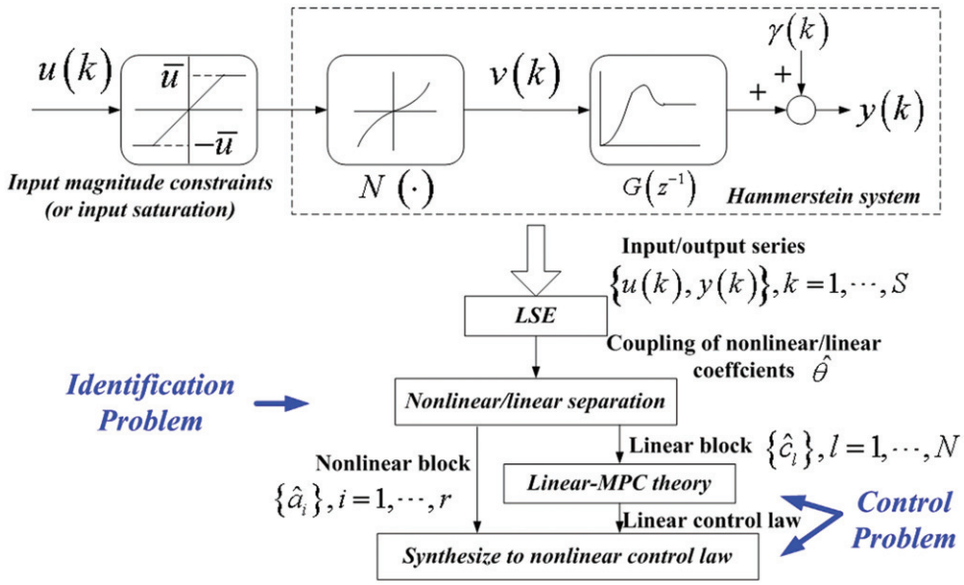

Figure 1. Problem formulation.

functions, as well as the mathematical foundations behind these modelling approaches are presented.

On the other hand, the dynamic linear block is represented by a rational orthonormal basis series as follows (Gómez and Baeyens 2004):

$$G(z^{-1}) = \sum_{l=1}^{N} c_l L_l(z^{-1}), \tag{3}$$

where $c_l \in \mathbb{R}^{m\times n}$ are unknown matrix parameters, $\{L_l(\cdot)\}_{l=1}^{N}$ can be any rational orthonormal basis on the space $\mathbf{H}_2(\mathbf{T})$, and N is the truncation length. Without loss of generality, $\{L_l(\cdot)\}_{l=1}^{N}$ is set as discrete Laguerre series in this paper (see Appendix 1). Certainly, a higher order orthonormal basis such as the Kautz series (Heuberger et al. 1995; Wahlberg and Mäkilä 1996) can also be used in (3), especially to deal with the oscillatory non-linearity. Substituting (2) and (3) into (1) yields

$$\begin{aligned} y(k) &= \left(\sum_{l=1}^{N} c_l L_l(z^{-1})\right)\left(\sum_{i=1}^{r} a_i g_i(u(k))\right) + \gamma(k) \\ &= \sum_{l=1}^{N}\sum_{i=1}^{r} c_l a_i L_l(z^{-1})\,[g_i(u(k))] + \gamma(k). \end{aligned} \tag{4}$$

Define

$$\theta \triangleq [c_1 a_1, \ldots, c_1 a_r, \ldots, c_N a_1, \ldots, c_N a_r]^T, \tag{5}$$

$$\begin{aligned} \phi(k) \triangleq \Big[& L_1(z^{-1})\big(g_1^T(u(k))\big), \ldots, L_1(z^{-1})\big(g_r^T(u(k))\big), \ldots, L_N(z^{-1}) \\ & \times \big(g_1^T(u(k))\big), \ldots, L_N(z^{-1})\big(g_r^T(u(k))\big)\Big]^T. \end{aligned} \tag{6}$$

Note that in real applications, the orthonormal basis $L_l(z^{-1})(g_i(u(k)))(l = 1, \ldots, N; i = 1, \ldots, r)$ in (6) and the system output $y(k)$ in (4) are calculated according to the state-space equations (63) and (64) in Appendix 1 conveniently. Accordingly, Equation (4) can be rewritten in a linear regressor form as (Ljung 1999)

$$y(k) = \theta^T \phi(k) + \gamma(k). \tag{7}$$

Thus, the details of the above operation processes are shown in Figure 2, with Laguerre filter

$$\begin{aligned} G_0(z^{-1}) &= (z^{-1}\sqrt{1-p^2})/(1 - z^{-1}p), \\ G_1(z^{-1}) &= (z^{-1} - p)/(1 - z^{-1}p). \end{aligned} \tag{8}$$

As shown in Figure 3, consider a S-point sampling data set, and carry out non-linear bases and Laguerre operations on the input sequence, one has

$$Y_S = \Phi_S^T \theta + \Upsilon_S, \tag{9}$$

with,

$$\begin{aligned} Y_S &\triangleq [y(1), \ldots, y(S)]^T, \\ \Upsilon_S &\triangleq [\gamma(1), \ldots, \gamma(S)]^T, \\ \Phi_S &\triangleq [\phi(1), \ldots, \phi(S)]^T. \end{aligned} \tag{10}$$

Provided the indicated inverse exists, it is well known that the estimate $\hat{\theta}$ of θ minimising the identification errors $(Y_S - \Phi_S^T\theta)$ can be obtained by least square estimation (LSE) (Ljung 1999) as follows:

$$\hat{\Phi} = \left(\Phi_s \Phi_s^T\right)^{-1} \Phi_s Y_s. \tag{11}$$

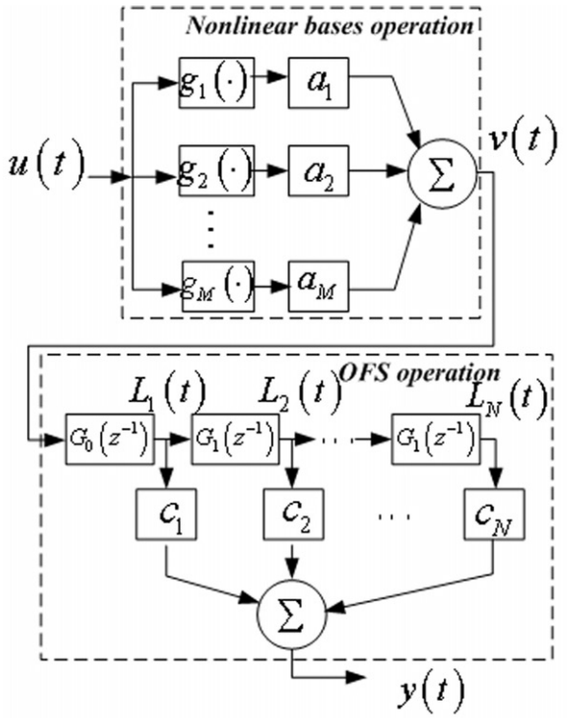

Figure 2. Operation details of each identification channel.

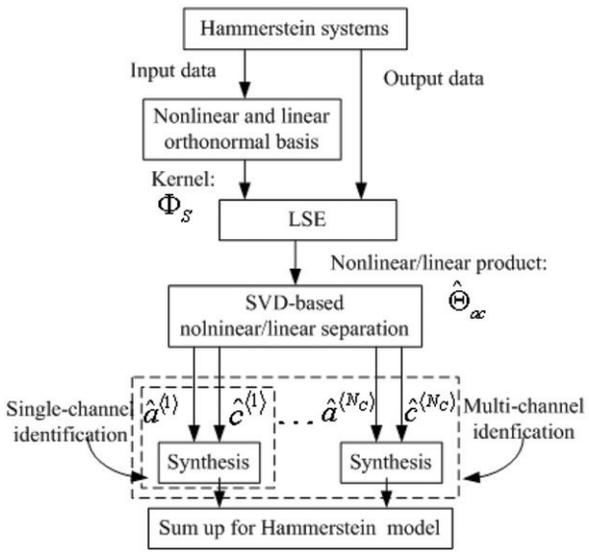

Figure 3. The sketch map of the multi-channel modelling method.

Then, as shown in Figure 1, two unsolved problems concerning identification and control are formulated as follows.

Problem 1 (non-linearity/linearity separation)**:** Separate the non-linear and linear blocks of the Hammerstein system with the sampling input/output series. This operation will extract the unknown parameters $a_i (i = 1, \ldots, r)$ and $c_i (l = 1, \ldots, N)$ from $\hat{\theta}$.

Problem 2 (Simplified non-linear controller development)**:** Based on the above non-linear/linear blocks separation, design a simplified non-linear controller to maximise the closed-loop stable region.

3. Identification algorithm (for Problem 1)

It is clear that the parametrisations (2) and (3) are not unique, because $c_l Q$ and $Q^{-1} a_i$ (Q is an arbitrary compatible nonsingular matrix) lead to the same input/output equation. Thus, in order to have a unique parametrisation, one should normalise the parameter matrices a_i (or c_l) as follows:

$$\|a_i\|_2 = 1 (i = 1, \ldots, r). \tag{12}$$

To solve Problem 1, as shown in Figure 3, first we define

$$\Theta_{ac} \triangleq \begin{bmatrix} a_1^T c_1^T & a_1^T c_2^T & \cdots & a_1^T c_N^T \\ a_2^T c_1^T & a_2^T c_2^T & \cdots & a_2^T c_N^T \\ \vdots & \vdots & \ddots & \vdots \\ a_r^T c_1^T & a_r^T c_2^T & \cdots & a_r^T c_N^T \end{bmatrix} = a c^T, \tag{13}$$

with $a \triangleq [a_1, a_2, \ldots, a_r]^T$, $c \triangleq [c_1^T, \ldots, c_N^T]^T$. It can be easily seen that θ is the block column matrix obtained by stacking the block columns of Θ_{ac} on the top one by one. It is clear that the estimates $\hat{a}$ and $\hat{c}$ are the solutions of the optimisation problem

$$(\hat{a}, \hat{c}) = \arg\min_{a,\, c} \left\{ \left\| \hat{\Theta}_{ac} - a c^T \right\|_2^2 \right\}. \tag{14}$$

This problem can be solved by the standard SVD (Golub and Van Loan 1989), and with the prerequisite (12), the solution of (14) is unique. However, bearing the spectral nature of SVD in mind, one can easily find that the closest estimates of a, c are not a single pair $(\hat{a}, \hat{c})$ but a series of pairs $(\hat{a}^{\langle j \rangle}, \hat{c}^{\langle j \rangle})$ $(j = 1, \ldots, N_C)$ as shown in Figure 3, which solves the optimisation problem

$$\left(\hat{a}^{\langle j \rangle}, \hat{c}^{\langle j \rangle}\right) = \arg\min_{a^{\langle j \rangle},\, c^{\langle j \rangle}} \left[\left\| \hat{\Theta}_{ac} - \sum_{j=1}^{N_C} a^{\langle j \rangle} \left(c^{\langle j \rangle}\right)^T \right\|_2^2 \right]. \tag{15}$$

From now on, the pair $(\hat{a}^{\langle j \rangle}, \hat{c}^{\langle j \rangle})$ is defined as the jth identification channel, with j and N_C being the sequence index and number of the identification channels, respectively. Therefore, in order to separate the non-linear/linear block more effectively, more channels should be used to compensate the separation residuals of the single-channel method (Gómez and Baeyens 2004). To explain it more clearly, we will give a lemma and a theorem as follows. It should be noted that, for SISO scenario, the single-channel

estimates $\hat{a}$ and $\hat{c}$ and the multi-channel estimates $\hat{a}^{(j)}$ and $\hat{c}^{(j)}$ are all column vectors, while for n-input and m-output systems, a (or $a^{(j)}$) $\in \mathbb{R}^{rn\times n}$, c (or $c^{(j)}$) $\in \mathbb{R}^{Nm\times n}$.

Lemma 1 (Golub and Van Loan 1989)**:** *Let* $\text{rank}(\hat{\Theta}_{ac}) = q$, *here* $\hat{\Theta}_{ac}$ *is the estimate of* Θ_{ac} *in* (13), *then the SVD of* $\hat{\Theta}_{ac}$ *is*

$$\hat{\Theta}_{ac} = U_q \Sigma_q V_q^T = \sum_{j=1}^{q} \varpi_j \iota_j \vartheta_j^T, \tag{16}$$

such that the singular value matrix $\Sigma_q = \text{diag}\{\varpi_j\}(j = 1, \ldots, \min(r, N))$ *satisfies*

$$\varpi_1 \geq \cdots \geq \varpi_q > 0 \quad \textit{and} \quad \varpi_l = 0 \ (l > q), \tag{17}$$

where l_j, ϑ_j $(j = 1, 2, \ldots, q)$ *are pairwise orthogonal vectors. As shown in Figure* 3, *if* $\|a^{(j)}\|_2 = 1$, *then for* $\forall NC$ $(1 \leq N_C \leq q)$, *each identification channel can be calculated by*

$$\begin{aligned}\left(a^{(j)}, c^{(j)}\right) &= \arg\min\left[\left\|\hat{\Theta}_{ac} - \sum_{j=1}^{N_C} a^{(j)}\left(c^{(j)}\right)^T\right\|^2\right] \\ &= (\iota_j, \varpi_j\theta_j), \quad j = 1, \ldots, N_C,\end{aligned} \tag{18}$$

with approximation error $e^{(N_C)}$ *given by*

$$e^{(N_C)} \triangleq \left\|\hat{\Theta}_{ac} - \sum_{j=1}^{N_C} \hat{a}^{(j)}(\hat{c}^{(j)})^T\right\|_2^2 = \sum_{j=N_C+1}^{q} \varpi_j^2. \tag{19}$$

It can be seen from Lemma 1 that after the *SVD* operation, $\hat{\Theta}_{ac}$ is decomposed into a series of pairs (or channels) $(a^{(j)}, c^{(j)})$ as shown in Figure 3. These channels should be used to yield a more effective model. The accuracy enhancement will be proven by the following theorem.

Theorem 1: *For the Hammerstein system* (4), *with the identification matrix* $\hat{\Theta}_{ac}$ *calculated by* (11), (14) *and* (15), *if* rank $(\hat{\Theta}_{ac}) = q \geq 1$, *then*, *with the identification pairs* $(\hat{a}^{(j)}, \hat{c}^{(j)})$ $(j = 1, \ldots N_C)$ *obtained by* (16) *and* (18) *and the identification error index defined by* (19), *one has*

$$e^{(1)} > e^{(2)} > \cdots > e^{(q)} = 0. \tag{20}$$

In other words, *the identification error decreases along with the increase of* N_C.

Proof: Under the given conditions, Lemma 1 yields $\text{error}^{(N_C)} = \sum_{j=N_C+1}^{q} \varpi_j^2$ $(1 \leq N_C \leq q-1)$ and $\text{error}^{(q)} = 0$, leading to the conclusion (20). □

In principle, one can select the suitable N_C according to the approximation error tolerance $\bar{e}$ and (19). Even for the extreme case that $\bar{e} = 0$, one can still set $N_C = q$ to eliminate the approximation error, thus such feasible N_C always exists. For simplicity, if rank $(\hat{\Theta}_{ac}) \geq 3$, the general parameter setting $N_C = 2$ or 3 can work well enough.

According to the conclusions of Lemma 1 and Theorem 1, multi-channel identification method should be used instead of the traditional single-channel one to enhance the modelling accuracy. Therefore, multi-channel model $y_m(t) = \sum_{j=1}^{N_C} \hat{G}^{(j)} \times (z^{-1})\hat{\mathbf{N}}^{(j)}(u(t))$ outperforms single-channel model $y_m(t) = \hat{G}(z^{-1})\hat{\mathbf{N}}(u(t))$ in modelling accuracy. Here, the subscript 'm' denotes model output. Then, we hereby design a multi-channel identification algorithm (MCIA) based on Theorem 1 as follows.

3.1 *MCIA*

Multi-channel identification model (MCIM) is composed of N_C parallel channels. As shown in Figure 2, each channel consists of a static non-linear block, which is represented by a series of non-linear basis $\{g_1(\cdot), g_2(\cdot), \ldots, g_r(\cdot)\}$, followed by a dynamic linear block represented by the discrete Laguerre model (Heuberger et al. 1995; Wahlberg and Mäkilä 1996; Wang 2004) in the state-space form (see Appendix 1). Without loss of generality, the bases are generally chosen as polynomial function bases according to the Weierstrass Theorem (Krantz 2004). Thus, each channel of MCIM is represented by

$$\begin{aligned}L^{(j)}(k+1) = AL^{(j)}(k) + B\left[\sum_{i=1}^{r} \hat{a}_i^{(j)} g_i(u(k))\right], \\ (j = 1, \ldots, N_C),\end{aligned} \tag{21}$$

$$y_m^{(j)}(k) = \left(\hat{c}^{(j)}\right)^T L^{(j)}(k) \tag{22}$$

with $y_m^{(j)}(k)$, $L^{(j)}(k) = [L_1^{(j)}(k),\ L_2^{(j)}(k), \ldots L_N^{(j)}(k)]^T$ being the output and the Laguerre state vector of the jth channel, respectively. Finally, the output of the MCIM can be synthesised by

$$y_m(k) = \sum_{j=1}^{N_C} y_m^{(j)}(k). \tag{23}$$

Now, we will give a convergence theorem to support the MCIA.

Theorem 2 (convergence theorem)**:** *For the Hammerstein system*

$$y(k) = \sum_{l=1}^{N} c_l L_l(z^{-1})\left(\sum_{i=1}^{r} a_i g_i(u(k))\right) + \gamma(k), \tag{24}$$

with $\|a_i\|_2 = 1$, *nominal output* $\bar{y}(k) = \sum_{l=1}^{N} c_l L_l(z^{-1}) \times (\sum_{i=1}^{r} a_i g_i(u(k)))$ *and bounded input signal set* **D** (*see* (1)). *Here* $L(z^{-1}) = [L_1(z^{-1}), L_2(z^{-1}), \ldots, L_N(z^{-1})]$ *is the Laguerre state vector. If the regressors* $\phi(k)$ (*see* (6)) *is PE in the sense that for an arbitrary positive integer* k_0 *there exist some integer* N_1 *and positive constants* α_1 *and* α_2 *such that*

$$0 < \alpha_1 I \leq \sum_{k=k_0}^{k_0+N_1} \phi(k)^T \phi(k) \leq \alpha_2 I, \tag{25}$$

and $\phi(k)$ *is unrelated to the external noise* $\gamma(k)$, *then*

$$\sum_{j=1}^{N_C} \hat{a}^{\langle j \rangle} \left(\hat{c}^{\langle j \rangle}\right)^T \xrightarrow{a.s.} \Theta_{ac}, \tag{26}$$

$$y_m(k) \xrightarrow{a.s.} \bar{y}(k), \tag{27}$$

where the symbol '$\xrightarrow{a.s.}$' *denotes* '*converge with probability one as the number of the data points S tends to infinity*', *and the model output* $y_m(k)$ *is determined by* (21–23).

Proof: It is referred to Appendix 2. □

Thus, it can be seen from Theorems 1 and 2 that the increase of the identification channel number will help decrease the identification errors. However, in real industrial processes, owing to its complex structure, MCIA should be converted into a single channel algorithm to facilitate the controller development. This method is a more effective version of SCIA called two-stage single-channel identification algorithm (TS-SCIA), since it preserves most of the advantages of MCIA. In this approach, SVD is first implemented to get the first identification channel coefficient $\hat{a}^{\langle 1 \rangle}$ and $\hat{c}^{\langle 1 \rangle}$, then $\hat{a}^{\langle 1 \rangle}$ is fixed while $\hat{c}^{\langle 1 \rangle}$ is set as the initial value. Afterwards, normalised recursive LSE (Adel et al. 1999) will be used to update $\hat{c}^{\langle 1 \rangle}$ so as to minimise the modelling error.

The advantage of TS-SCIA is that the Laguerre states $\{L(k), k = 1, 2, \ldots\}$ will not be influenced by the updating of $\hat{c}^{\langle 1 \rangle}$. Since $\hat{a}^{\langle 1 \rangle}$ are fixed and A, B (see (21) and Appendix 1) are determined offline, the only parameter required to be tuned online is $\hat{c}^{\langle 1 \rangle}$. Therefore, this approach can yield satisfactory performances even when the convergence conditions of traditional SCIA in the work of Gómez and Baeyens (2004) are not satisfied. Furthermore, this mechanism can facilitate the subsequential controller design. This will be also validated by the case studies discussed in § 6. From now on, for convenience, we will denote $\hat{a}^{\langle 1 \rangle}$, $\hat{c}^{\langle 1 \rangle}$ and $v^{\langle 1 \rangle}$ of TS-SCIA simply by $\hat{a}$, $\hat{c}$ and v, respectively.

4. Dual-mode predictive control algorithm (for Problem 2)

4.1 *Control algorithm*

To address the second problem proposed in § 2, in this section, first the TS-SCIA identification method is used to separate the linear and non-linear blocks of the Hammerstein system, then the intermediate variable control law $v(k)$ is gained based on the linear block dynamics. Afterwards, one can calculate the control law $u(k)$ according to the inverse of $v(k)$. Hence, for SISO Hammerstein systems (4) subject to input magnitude constraints $|u(k)| \leq \bar{u}$ $(\bar{u} \in \mathbb{R}^+)$, suppose the following two assumptions hold

A1 a, c can be precisely identified by TS-SCIA, i.e.,

$$a = \hat{a}, c = \hat{c}. \tag{28}$$

A2 When $|u(k)| \leq \bar{u}$, the inverse of $\mathbf{N}(\cdot)$ exists (see (1), (2)). Moreover,

$$\mathbf{N}\left[\mathbf{N}^{-1}_{zeroin}(v(k))\right] \triangleq \bar{v}(k) = [1 + \delta(v(k))]v(k), \tag{29}$$

with

$$\left|\delta(v(k))\right| \leq \sigma(\sigma \in \mathbb{R}^+), \tag{30}$$

where $\mathbf{N}^{-1}_{zeroin}(\cdot)$ denotes the inverse of $\mathbf{N}(\cdot)$ calculated by the Zeroin algorithm (George et al. 1977). The reason of choosing zeroin algorithm is given as follows. According to the theoretical analysis later in this section, more accurate inversion calculation method of $\mathbf{N}^{-1}(\cdot)$ implies larger size of stable region. Taking into consideration of the high accuracy and high convergent speed of zeroin method (George et al. 1977), we apply it to calculate $\mathbf{N}^{-1}(\cdot)$. In real applications, one can use the function *fzero*() of MATLAB to calculate $\mathbf{N}^{-1}_{zeroin}(\cdot)$ with convenience. Certainly, other kinds of high efficient inverse solving methods can also be applied. Moreover, there are still some other feasible assumptions similar to Assumption A2; interested readers can refer to the works of Ding et al. (2004) and Ding and Xi (2006) for more details.

For convenience, if $\delta(v(k))$ is denoted by $\delta(\cdot)$ then the controlled plant is described as follows (Ding and Xi 2004):

$$L(k+1) = AL(k) + B\bar{v}(k) = AL(k) + B[1 + \delta(\cdot)]v(k), \tag{31}$$

$$y(k) = cL(k), \tag{32}$$

Since $L(k)$ is an immeasurable state, a state observer $\Gamma \in \mathbb{R}^{N\times 1}$ is used to estimate $L(k)$ as follows:

$$\hat{L}(k+1) = A\hat{L}(k) + B[1+\delta(\cdot)]v(k) + \Gamma ce(k), \quad (33)$$

$$e(k+1) = \Xi e(k), \quad (34)$$

where $\hat{L}(k)$ is the estimation of $L(k)$, $e(k) = L(k) - \hat{L}(k)$ is the state estimation error, and $\Xi = A - \Gamma c$ is Hurwitz.

Then, a *NMPC* law is designed with an additional term $D(k+i\,|\,k)$ as follows:

$$v(k+i\,|\,k) = K\hat{L}(k+i\,|\,k) + ED(k+i\,|\,k), \quad (35)$$

$$u(k\,|\,k) = \mathbf{N}_{zeroin}^{-1}[v(k\,|\,k)]. \quad (36)$$

where $E = [1, 0, \ldots, 0]_{1\times M}$, $\hat{L}(k\,|\,k) = \hat{L}(k)$, $v(k\,|\,k) = v(k)$, $D(k\,|\,k) = D(k) = [d(k), \ldots, d(k+M-1)]^T$ is defined as a perturbation signal vector representing extra degrees of design freedom (Kouvaritakis et al. 2000) and the role of $D(k)$ is merely to ensure the feasibility of the control law (33–36), and $D(k+i\,|\,k)$ is designed such that

$$D(k+i\,|\,k) = TD(k+i-1|k) \quad (i = 1, \ldots, M),$$

where

$$T = \begin{bmatrix} \mathbf{0} & I_{(M-1)\times(M-1)} \\ 0 & \mathbf{0}^T \end{bmatrix}_{M\times M},$$

$M \geq 2$ is called the prediction horizon, and $\mathbf{0}$ is compatible zero column vector.

Then, substituting (35) into (33) yields

$$\begin{aligned}\hat{L}(k+i\,|\,k) = {} & [\Psi + BK\delta(\cdot)]\hat{L}(k+i-1|k) \\ & + [1+\delta(\cdot)]BED(k+i-1|k) \\ & + \Gamma ce(k+i-1|k), \quad (i = 1, \ldots, M),\end{aligned} \quad (37)$$

or

$$\begin{aligned}\hat{x}(k+i\,|\,k) = {} & \Pi\hat{x}(k+i-1|k) + \begin{bmatrix} \delta(\cdot)B\bar{K} \\ 0 \end{bmatrix}\hat{x}(k+i-1|k) \\ & + \begin{bmatrix} \Gamma c \\ 0 \end{bmatrix}e(k+i-1|k), \ (i = 1, \ldots, M),\end{aligned} \quad (38)$$

with

$$\hat{x}(k+i\,|\,k) = \begin{bmatrix} \hat{L}(k+i\,|\,k) \\ D(k+i\,|\,k) \end{bmatrix}, \quad \Pi = \begin{bmatrix} \Psi & BE \\ 0 & T \end{bmatrix}, \quad \bar{K} = [K, E],$$

where $\Psi = A + BK$ is Hurwitz.

Now, in order to stabilise the closed-loop system (38), define two ellipsoidal invariant sets of the extended state estimations $\hat{x}(k)$ and $e(k)$, respectively (Kouvaritakis et al. 2000; Cao et al. 2006)

$$S_x \triangleq \left\{\hat{x}(k)|\hat{x}^T(k)P_x\hat{x}(k) \leq 1\right\}, \quad (39)$$

and

$$S_e \triangleq \left\{e(k)|e^T(k)P_e e(k) \leq \bar{e}\right\}, \quad (0 < \bar{e} \leq 1), \quad (40)$$

where P_x and P_e are both positive-definite symmetric matrices and the perturbation signal vector $D(k)$ (see (35)) is calculated by solving the following optimisation problem

$$\begin{aligned}\min_{D(k)} \quad & J(k) = D^T(k)D(k), \\ \text{s.t.} \quad & \hat{x}^T(k)P_x\hat{x}(k) \leq 1.\end{aligned} \quad (41)$$

To guarantee the feasibility and stability of the control law (33–36), it is required to find the suitable matrices P and P_e assuring the invariance of S_x and S_e, and meanwhile guaranteeing the feasibility of the control law. Hence, the following lemma is derived to provide the sufficient conditions for the invariance and input constraints satisfaction.

Lemma 2 (invariant lemma of dual-mode predictive control)**:** *Consider a closed-loop Hammerstein system* (13) *whose dynamics is determined by the output-feedback control law* (33–36) *and* (41) *and subject to the input magnitude constraints* $|u(k)| \leq \bar{u}$, *the ellipsoidal set* S_x *and* S_e *are invariant in the sense of* (39) *and* (40), *respectively*, *and the control law* (33–36) *and* (41) *is feasible provided that Assumptions A*1, *A*2 *and the following three Assumptions A*3–*A*5 *are all fulfilled.*

A3 *the matrices* Ξ *and* Γ *are both Hurwitz*;

A4 *there exist* $\tau_1 > 1, \tau_2 > 1, 0 < \bar{e} < 1 (\tau_1, \tau_2, \bar{e} \in \mathbb{R})$ *such that*

$$\Xi^T P_e \Xi \leq P_e, \quad (42)$$

$$\eta_1 c^T \Gamma^T E_L^T P_x E_L \Gamma c \leq P_e, \quad (43)$$

$$\tau_1\tau_2\Pi^T P_x\Pi + \tau_1\eta_2\sigma^2\bar{K}^T B^T E_L^T P_x E_L B\bar{K} \leq (1-\bar{e}^2)P_x, \quad (44)$$

where $\eta_1 = 1 + (\tau_1 - 1)^{-1}$, $\eta_2 = 1 + (\tau_2 - 1)^{-1}$ *and*

$$E_L^T = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & 0 & \cdots & \vdots \\ 0 & 0 & 1 & \cdots & 0 \end{bmatrix}_{N\times(N+M)}$$

is the projection matrix such that $E_L^T\hat{x}(k) = \hat{L}(k)$;

A5 *there exist* $\mu > 0$ *and* $\lambda \in (0, \bar{u})$ *such that*

$$|u(k)| = |\mathbf{N}_{Zeroin}^{-1}(v(k))| \leq \mu |v(k)| + \lambda \text{ (local lipschitz condition (Khalil 2002))} \tag{45}$$

and

$$\begin{bmatrix} -((\bar{u}-\lambda)/\mu)^2 & \bar{K} \\ \bar{K}^T & -P_x \end{bmatrix} \leq 0. \tag{46}$$

Proof: It is referred to Appendix 3. □

Based on Lemma 2, we can design the controller structure as shown in Figure 4.

Remark 1: If the control horizon $M = 0$, i.e., the perturbation signal vector is not applied, then the degree of design freedom is minimised and the predictive control law (33–36) reduced to the standard output feedback law (or initial control law) as

$$\begin{aligned} v(k) &= K\hat{L}(k), \\ u(k) &= \mathbf{N}_{zeroin}^{-1}[v(k)], \end{aligned} \tag{47}$$

and the invariant set S_x shrinks to

$$S_x(M=0) \triangleq S_L = \{\hat{L}(k) | \hat{L}^T(k) P_{LL} \hat{L}(k) \leq 1\}. \tag{48}$$

Here, (47) is called the initial control law since it can drive $\hat{L}(k)$ to the origin asymptotically provided that the initial state estimation $\hat{L}(0)$ is inside S_L. However, it will be proven in §4.2 that S_x will be minimised when $M = 0$, which minimises the possibility of $\hat{L}(0) \in S_x$. Fortunately, as will be proven later, the perturbation signal vector $D(k)$ in (35) can enlarge the stable region effectively. The algorithm determined in Lemma 2 is called dual-mode NMPC with more details given as follows.

If the current $\hat{L}(k)$ moves outside of S_L, then the controller enters the first mode, in which the dimension of $\hat{L}(k)$ is extended from N to $(N+M)$ by $D(k)$ (see (38)). Then, $\hat{L}(k)$ will be driven into S_L in no more than M steps, i.e., $\hat{L}(k+M|k) \in S_L$, which will also be proven in §4.2. Once $\hat{L}(k)$ enters S_L, the controller is automatically switched to the second mode, in which the initial control law (47) is feasible and can stabilise the system.

Remark 2: In general, there are two typical control problems (Chen 1998a):

- *Regulator problem*: Suppose the set-point signal $r(k)$ is zero, and the response of the system is caused only by some non-zero initial conditions. The problem is to find an output-feedback gain so that the response will die out at a desired rate.
- *Tracking problem*: Suppose the reference signal $r(t) = a$. The problem is to design an overall system so that $y(t)$ approaches $r(t) = a$ as t approaches infinity.

The above algorithm is focused on the first problem. To solve the second one, (35) can be converted to the following control law:

$$v(k) = \bar{K}\hat{x}(k) + a\rho, \tag{49}$$

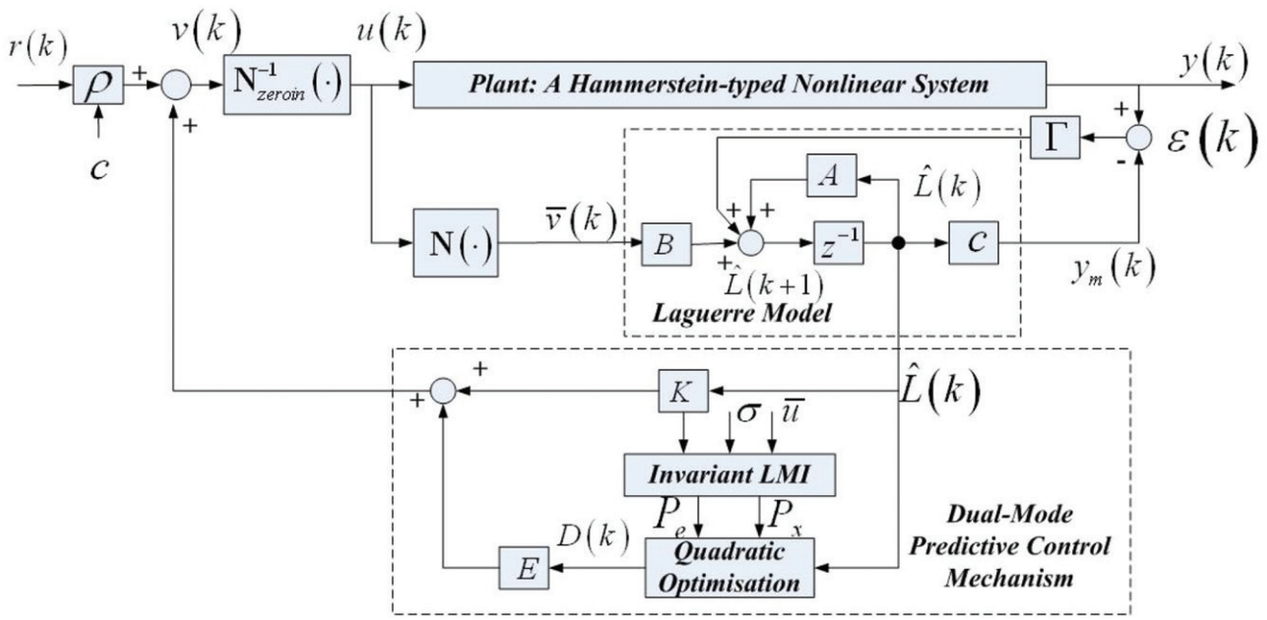

Figure 4. Dual-model predictive controller structure, with $1/\rho = \lim_{z\to 1}(\bar{c}(zI - \Pi)^{-1}\bar{B})$.

with $1/\rho = \lim_{z\to 1}(\bar{c}(zI-\Pi)^{-1}\bar{B})$, $\bar{c} = [c, 0]_{1\times(N+M)}$, $\bar{B} = [B^T, 0]^T_{(N+M)\times 1}$. Moreover, if $(I-\Pi)$ is nonsingular, one can make the coordinate transformation $\hat{x}(k) - x_c \to \hat{x}(k)$, with $x_c = (I-\Pi)^{-1}Ba\rho$, obtaining (38). Even if $(I-\Pi)$ is singular, one can still make some suitable coordinate transformation to obtain (38). In this sense, the above two control problems are equivalent. On the other hand, if the constraints are asymmetric, i.e., $u_{\min} \le u \le u_{\max}, (u_{\min} \ne -u_{\max})$, then one can use another coordinate transformation, $u(k) - (u_{\max} + u_{\min})/2 \to u(k)$, to symmetrize the constraints.

Remark 3: It is validated by extensive simulations that Assumptions A4 and A5 are not difficult to fulfil. A detailed procedure to calculate P_x and P_e is given here. First, the stable state-feedback gain K (see (35)) and observer gain L (see (33)) are pre-calculated by MATLAB. Then, compute P_e based on (42). Afterwards, select μ (generally in the range $(0, 1)$ and $\lambda \in (0, \bar{u})$ satisfying the local Lipschitz condition (45). Finally, select τ_1, $\tau_2 > 1$ (generally in the range $(1, 1.5)$), and calculate P_x by *LMI* toolbox of MATLAB according to (43), (44), (46), the pre-determined constant $\bar{e}$ and the pre-calculated matrix P_e.

4.2 *Stability analysis*

Rewrite the positive-definite matrix P_x as

$$P_x = \begin{bmatrix} (P_{LL})_{N\times N} & P_{LD} \\ P_{LD} & (P_{DD})_{M\times M} \end{bmatrix},$$

then it follows from (39) that (Kouvaritakis et al. 2000)

$$\hat{L}^T(k)P_{LL}\hat{L}(k) \le 1 - 2\hat{L}^T(k)P_{LD}D(k) - D^T(k)P_{DD}D(k), \tag{50}$$

and

$$\begin{aligned} &-P_{DD}^{-1}P_{LD}^T\hat{L}(k) \\ &\quad = \arg\max_{D(k)}[1 - 2\hat{L}^T(k)P_{LD}D(k) - D^T(k)P_{DD}D(k)]. \end{aligned} \tag{51}$$

Then substituting (51) into (50) yields

$$\hat{L}^T(k)\big(P_{LL} - P_{LD}P_D^{-1}P_{LD}^T\big)\hat{L}(k) \le 1. \tag{52}$$

Hence, the maximised ellipsoidal invariant set of the initial Laguerre state $\hat{L}(k)$ is given as follows:

$$S_{LM} \triangleq \left\{\hat{L}(k) | \hat{L}^T(k)\big(P_{LL} - P_{LD}P_{DD}^{-1}P_{LD}^T\big)\hat{L}(k) \le 1\right\}. \tag{53}$$

Bearing in mind that $P_{LL} - P_{LD}P_D^{-1}P_{DL} = (E_L^TP_x^{-1}E_L)^{-1}$(see Assumption A4 for E_L), it follows immediately from (53) that volume of the ellipsoid set S_{LM} satisfies

$$\mathrm{vol}(S_{LM}) \propto \det\big(E_L^TP_x^{-1}E_L\big), \tag{54}$$

where vol(·) and det(·) denotes volume and matrix determinant and the symbol '$\propto$' means 'proportional to'. A graphical illustration of the benefit of the maximiser $-P_{DD}^{-1}P_{LD}^T\hat{L}(k)$ is given in the reference (Kouvaritakis et al. 2000). Moreover, it will be verified later by case study that the present dual-mode NMPC can effectively increase the value of $E_L^TP_x^{-1}E_L$ with the assistance of perturbation signal vector $D(k)$, and the closed-loop stable region S_{LM} is thus effectively enlarged.

Based on the above mentioned analysis of the size of the invariant set S_{LM}, we give the nominal closed-loop stability theorem as follows.

Theorem 3 (nominal closed-loop stability theorem)**:** *Consider a closed-loop Hammerstein system* (13) *whose dynamics is determined by the output-feedback control law* (33–36) *and* (41) *and subject to the input magnitude constraints* $|u(k)| \le \bar{u}$, *the system is closed-loop asymptotically stable provided that Assumptions* A1–A5 *are fulfilled.*

Proof: Based on Assumptions A1–A5, one has that there exists $D(k+1)$ such that $x(k+1) \in S_x$ for arbitrary $x(k) \in S_x$; then by invariant property, at next sampling time $\tilde{D}(k+1|k) = TD(k)$ (see (36)) provides a feasible choice for $D(k+1)$ (only if $D(k) = 0$, $J(k+1|J(k))$, otherwise $J(k+1| < J(k))$). Thus, the present NMPC law (33–36) and (41) generates a sequence of $D(k+i\,|\,k) = TD(k+i\,|\,k)$ $(i = 1, \ldots, M)$ which converges to zero in M steps and ensures the input magnitudes constraints satisfaction. Certainly, it is obvious that $TD(k)$ need not have the optimal value of $D(k+1)$ at the current time, hence the cost $J(k+1)$ (see (41)) can be reduced further still. Actually, the optimal $D^*(k+1)$ is obtained by solving (41), thus $J^*(k+1) \le J(k+1\,|\,k) < J(k)(D(k) \ne 0)$. Therefore, as the sampling time k increases, the optimisation index function $J(k)$ will decrease monotonously and $D(k)$ will converge to zero in no more than M steps. Given constraints satisfaction, the system state $\hat{L}(k)$ will enter the invariant set S_L in no more than M steps. Afterwards, the initial control law (47) will make the closed-loop system asymptotically stable. This completes the proof. □

It still should be noted that since the above analysis is based on the Assumption A1, Theorem 3 can only guarantee nominal stability.

5. A case study

Plant: Heat exchanger system (Eskinat and Johnson 1991)

$$v(k) = -31.549u(k) + 41.732u^2(k) - 24.201u^3(k) + 68.634u^4(k), \quad (55)$$

$$y(k) = \frac{0.207z^{-1} - 0.1764z^{-2}}{1 - 1.608z^{-1} + 0.6385z^{-2}}v(k) + \gamma(k), \quad (56)$$

the output noise $\gamma(k)$ satisfies

$$\mathbf{E}\{\gamma^2(k)/\mathbb{F}_{k-1}\} \overset{a.s.}{=} 0.5, \quad (57)$$

where $\gamma(k) \in (\Omega, \mathbb{F}, \mathbf{P}), (k = 1, 2, \ldots)$ and $\mathrm{E}(\cdot)$ represents the mathematical expectation.

5.1 *Identification*

In order to examine the effectiveness of MCIA and TS-SCIA for a sufficient broad range of input signal frequency, we select the following input signal containing three different frequencies

$$u(k) = 0.2\cos(0.015k) + 1.3\sin(0.005k) + 0.4\sin(0.01k). \quad (58)$$

The modelling results of SCIA ($N_C = 1$), TS-SCIA ($N_C = 1$) and MCIA ($N_C \geq 2$) based on Laguerre models with truncation length $N = 2$ and $N = 3$ are shown in Figures 5 and 6, respectively. The modelling error $e(k) = [y(k) - y_m(k)]/\max_k[y(k)]$ $(k = 1, \ldots, 600)$ and $e1(k)$, $e2(k)$, $e3(k)$ denote the modelling errors of SCIA, TS-SCIA and MCIA respectively.

The parameters setting is $S = 600$, $p = 0.10$, $r = 5$, the non-linear basis $\{g_1(u), \ldots, g_5(u)\} = \{1, u, \ldots, u^4\}$, and the number of identification channels in MCIA is set as $N_C = 2$ (Figure 5), $N_C = 3$ (Figure 6). Then, calculations shows that rank $(\hat{\Theta}_{ac}) = 2$ (Figure 5), rank $(\hat{\Theta}_{ac}) = 3$ (Figure 6), and the corresponding singular values of $\hat{\Theta}_{ac}$ are $\{19.8779, 2.4032\}$ and $\{2.5818, 0.6212, 0.1742\}$, respectively. Thus, according to Theorem 1, SCIA cannot yield satisfactory modelling performances whereas TS-SCIA and MCIA can. This is also verified by the simulation results in Figures 5 and 6. Further, statistical experiments were also carried out to test the modelling performances. More precisely, we select p and r from the sets $\{0.08, 0.09, 0.10, 0.11, 0.12\}$ and $\{4, 5, 6, 7\}$ respectively. The results is shown in Table 1, where $\varsigma\{|e(k)|\}$ denotes the standard deviation of $|e(k)|$.

From these results, we conclude that

- the identification accuracy increases with the enhancement of N and N_C;
- compared with MCIA, the precision loss of TS-SCIA is sufficiently small while the counterpart of SCIA is not.

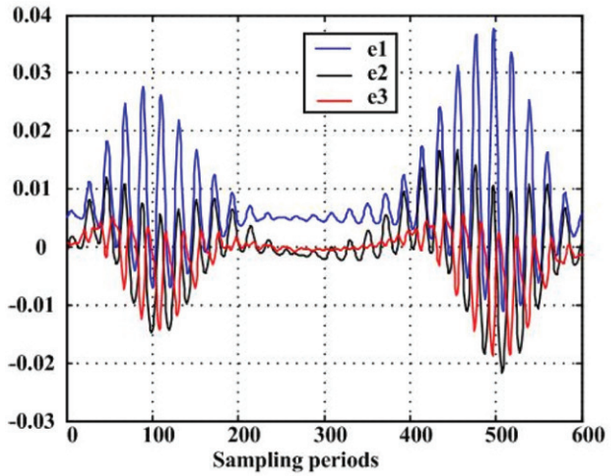

Figure 5. Identification performance ($N = 2$).

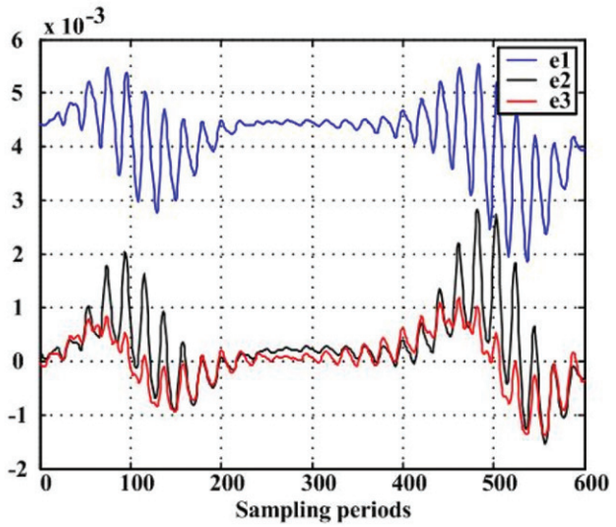

Figure 6. Identification performance ($N = 3$).

Table 1. Statistical modelling performances.

	$N = 2$		$N = 3$	
Identification algorithm	$\mathbf{E}(\lvert e(k)\rvert)$ (‰)	$\varsigma[e(k)]$ (‰)	$\mathbf{E}(\lvert e(k)\rvert)$ (‰)	$\varsigma[e(k)]$ (‰)
SCIA	8.84	10.2	3.87	6.4
TS-SCIA	4.92	8.8	0.57	6.4
MCIA	4.13	7.6	0.51	6.3

The effectiveness of TS-SCIA and MCIA is thus validated. Note that the minimal identification is determined by not the SVD in (18) but the LSE $\hat{\theta}$ in (11), thus it is not necessary to discuss the overfitting case in this identification simulation.

5.2 *Control*

The present dual-mode NMPC is performed in the heat exchanger system model (55–57) with the results shown in Figures 7, 8 (regulator problem, $N=2$), Figures 9–11 (Regulator Problem, $N=3$) and Figure 12 (tracking problem, $N=3$), respectively. The correspondence parameter settings are presented in Table 2.

In these numerical examples, the initial state-feedback gain K and state observer gain Γ are optimised offline via DLQR and KALMAN functions of MATLAB 6.5, respectively. The curves of $y(k)$, $u(k)$, $\bar{v}(k)$ and the first element of $D(k)$, i.e. $d(1)$, are shown in Figure 7 ($N=2$) and Figure 8 ($N=3$), respectively. To illustrate the superiority of the proposed dual-mode NMPC, we present the curve of $\hat{L}(k)$, the invariant sets of S_L and S_{LM} in Figure 8 ($N=2$, $M=\{2,8,10\}$) and Figure 10 and 11 ($N=3$, $M=\{0,5,10\}$). One can find that $\hat{L}(0)$ is outside the feasible initial invariant set S_L (refer to (48), see the red ellipse in Figure 10 and the left subfigure of Figure 11). Then the state extension

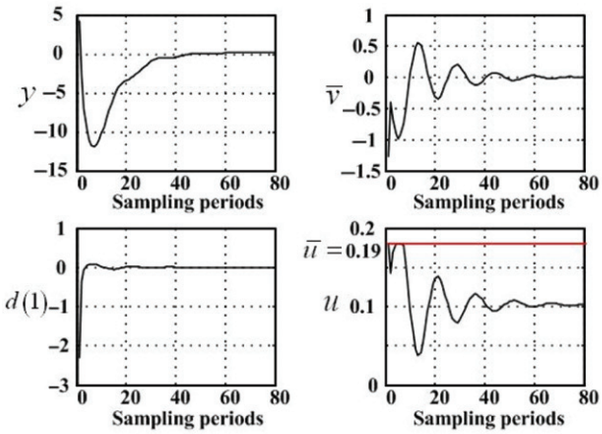

Figure 7. Control performance of regulator problem ($N=2$).

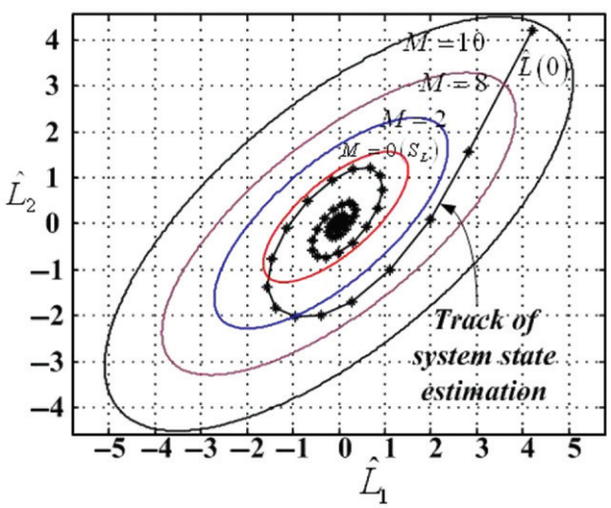

Figure 8. The track of system state and invariant set S_L ($N=2$).

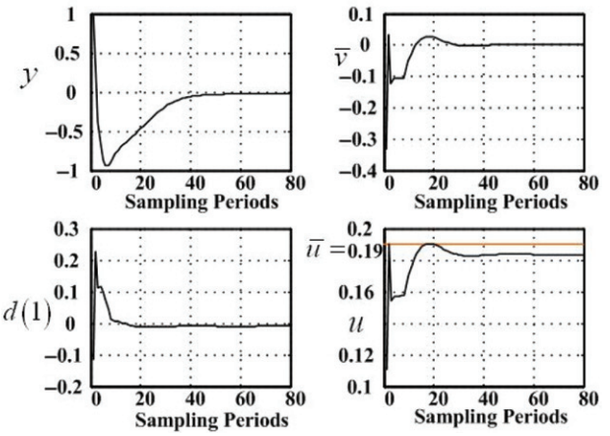

Figure 9. Control performance of regulator problem ($N=3$).

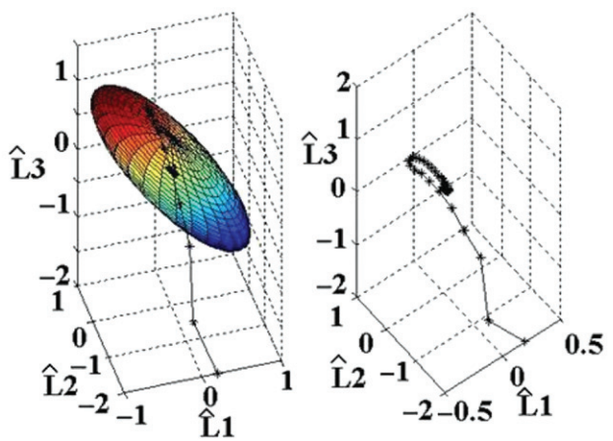

Figure 10. The track of system state and invariant set S_L ($N=3$, $M=10$).

with $M = 10$ is used to enlarge S_L to S_{LM} (refer to (52), see the black ellipse in Figure 8 and the right subfigure of Figure 11) containing $\hat{L}(0)$. After eight (Figure 8) or six steps (Figure 10), $\hat{L}(k)$ enters S_L. Afterwards, the initial control law (47) can stabilise the system and lets the state approach the origin asymptotically. Lemma 2 and Theorem 3 are thus verified. Moreover, the numerical results of Figures 8 and 11 also have verified the conclusion of the ellipsoid volume relation (54), i.e. the size of S_{LM} increases along with the enhancement of the prediction horizon M.

As to the tracking problem (see Figure 12), one should focus on the system state response to the change of the set-point. In this case, $\hat{L}(k)$ moves outside S_L, thus $D(k)$ is activated to enlarge S_L to S_{LM} and then to drive $\hat{L}(k)$ from S_{LM} to S_L in no more than M steps. After 60 sampling periods, the overshooting, modulating time and steady-state error are 2.2%, 15 and 0.3% respectively. Moreover, robustness to the time-delay variations is examined at the 270th sampling period, while the linear block of this plant is changed from (58) to

$$y(k+1) = \frac{0.207z^{-1} - 0.1764z^{-2}}{1 - 1.608z^{-1} + 0.6385z^{-2}} v(k). \qquad (59)$$

Dual-mode NMPC can still yield satisfactory performances, thanks to the capability of the Laguerre series in the inner model. The feasibility and superiority of the proposed control algorithm are thus demonstrated by simulations on both regulator and tracking problems.

It is still worth mentioning that some other simulations also show that the size of S_{LM} increases as σ decreases. In other words, more accurate identification and inverse solving algorithms would help further enlarge the closed-loop stable region. Fortunately, the proposed TS-SCIA can do this job quite well.

To further investigate the proposed dual-mode NMPC, a number of experiments were carried out to yield statistical results. More precisely, $\{\lambda, \mu, \tau\}$ are fixed to $\{0.70, 0.35, 1.12\}$, and N, M and σ are selected from the sets $\{2, 3, 4\}$, $\{8, 9, \ldots, 18\}$ and $\{0.001, 0.002, \ldots, 0.005\}$, respectively. The set-point is the same as Figure 12. In this set-up, 165 experiments were performed. The statistical results, such as expectations and optimal values for the settling time, overshooting, steady-state error and computational time of 400 steps are shown in Table 3. In addition, the corresponding optimal parameters are given. The statistical results further illustrate the advantages of the proposed algorithm regarding transient performance, steady-state performance and robustness to system uncertainties.

Remark 4: The increase of the Laguerre truncation length N can help enhance the modelling and control accuracy at the cost of an increasing

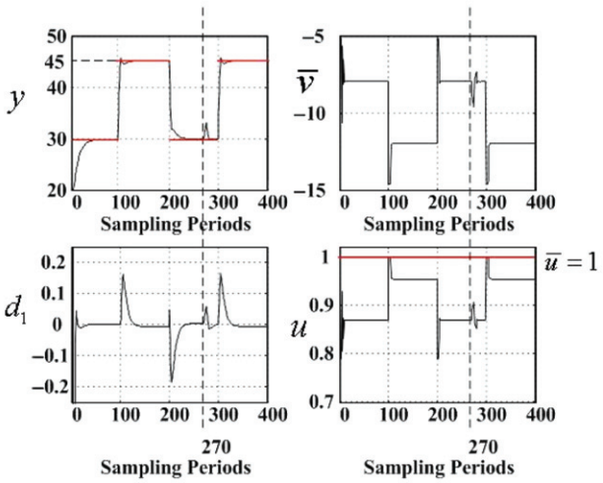

Figure 12. Control performance of tracking problem.

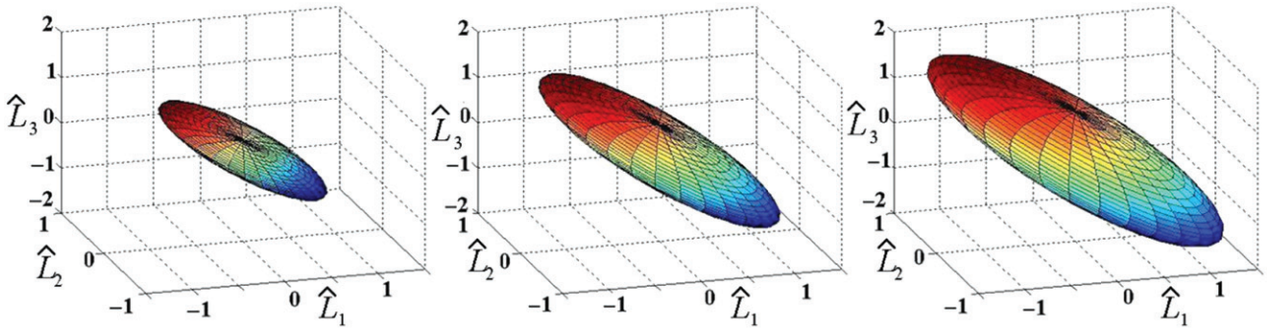

Figure 11. Invariant sets S_L (left) and S_{LM}, $M = 5$ (middle), $M = 3$ (right), $(N = 3)$.

Table 2. Parameter settings.

Parameter	Regulator problem		Tracking problem
	$N=2$ (Figures 7 and 8)	$N=3$ (Figures 9–11)	$N=3$ (Figure 12)
$\hat{L}(0)$	$[4.2, 4.2]^T$	$[0.2, -2, -2]^T$	$[0.2, -2, -2]^T$
K	$[-0.327, -0.570]$	$[-0.328, 0.200, -0.120]$	$[-0.328, 0.200, -0.120]$
Γ	$[0.187, 0.416]^T$	$[0.217, 0.132, 0.276]^T$	$[0.217, 0.132, 0.276]^T$
$\bar{u}$	0.19	0.19	1.0
σ	0.001	0.001	0.001
λ	0.13	0.04	0.17
μ	0.05	0.02	0.35
τ_1	1.05	1.12	1.12
τ_2	1.30	1.25	1.37
$\bar{e}$	0.5	0.5	0.5

Table 3. Statistical control performance of tracking problems (computation platform: 2.8G-CPU and 256M-RAM).

Control indexes	STVSP (steps)	STVTD (steps)	Overshooting (%)	Steady-state error (%)	Computational time of 400 steps (s)
Optimal (N, M, σ)	$(4, 17, 0.003)$	$(3, 12, 0.002)$	$(4, 8, 0.002)$	$(4, 10, 0.001)$	$(2, 8, 0.005)$
Expectation value	16.7	12.4	± 2.78	± 0.41	21.062
Optimal value	7	8	0	± 0.08	11.371

Note: (STVSP, STVTD: the settling time for the variations of set-point and time delay, respectively).

computational complexity. Therefore, a tradeoff must be made between accuracy and computational complexity. Note that the general parameter setting procedure is given in Remark 3.

6. Conclusion

In this paper, a novel multi-channel identification algorithm has been proposed to solve the modelling problem for constrained Hammerstein systems. Under some weak assumptions on the persistent excitation of the input, the algorithm provides consistent estimates even in the presence of coloured output noise, and can eliminate any need for prior knowledge about the system. Moreover, it can effectively reduce the identification errors as compared to the traditional algorithms. To facilitate the controller design, the MCIA is converted to a two-stage identification algorithm called TS-SCIA, which preserve almost all the advantages of the former. In addition, to support these two algorithms, systematical analyses about their convergence and approximation capability has been provided. Based on the TS-SCIA, a novel dual-mode NMPC is developed for process control. This approach is capable of enlarging the closed-loop stable region by providing extra degrees of design freedom. Finally, modelling and control simulations have been performed on a benchmark Hammerstein system, i.e., a heat exchanger model. The statistical results have demonstrated the feasibility and superiority of the proposed identification and control algorithms for a large class of non-linear dynamic systems often encountered in industrial processes.

Acknowledgements

H.-T. Zhang acknowledges the National Natural Science Foundation of China (NSFC) under Grant No. 60704041, and the Natural Scientific Founding Project of Huzhong (Central China) University of Science and Technology under Grant No. 2006Q041B. H. X. Li and G. Chen acknowledge a project from the RGC of HKSAR (CityU 116905).

References

Adel, M.E., Makoudi, M., and Radouane, L. (1999), "Decentralised Adaptive Control of Linear Interconnected Systems Based on Lauguerre Series Representation," *Automatica*, 35, 1873–1881.

Bai, E. (1998), "An Optimal Two-stage Identification Algorithm for Hammerstein–Wiener Nonlinear Systems," *Automatica*, 34, 333–338.

Bai, E., and Fu, M. (2002), "A Blind Approach to Hammerstein Model Identification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 50, 1610–1619.

Bhandari, N., and Rollins, D. (2004), "Continuous-time Hammerstein Nonlinear Modelling Applied to Distillation," *A.I.Ch.E. Journal*, 50, 530–533.

Bolemen, H.H.J., and Van Den Boom, T.T.J. (2000) "Model-based Predictive Control for Hammerstein Systems," in *Proceedings of the 39th IEEE Conference on Decision and Control*, Sydney, Australia, pp. 4963–4968.

Bolemen, H.H.J., Van Den Boom, T.T.J., and Verbruggen, H.B. (2001), "Model-based Predictive Control for Hammerstein-Wiener Systems," *International Journal of Control*, 74, 482–495.

Boyd, E., and Chua, L. (1985), "Fading Memory and the Problem of Approximating Nonlinear Operators with Volterra Series," *IEEE Transactions on Circuits and Systems*, 32, 1150–1161.

Cao, M.L., Wu, Z.M., Ding, B.C., and Wang, C.X. (2006), "On the Stability of Two-Step Predictive Controller Based-on State Observer," *Journal of Systems Engineering and Electronics*, 17, 132–137.

Chen, C.T. (1998a), *Linear System Theory and Design* (3rd ed.), Oxford, UK: Oxford University Press.

Chen, H., and Allgöwer, F. (1998b), "A Quasi-Infinite Horizon Nonlinear Model Predictive Control Scheme with Guaranteed Stability," *Automatica*, 34, 1205–1217.

Ding, B.C., and Xi, Y.G. (2004), "Design and Analysis of the Domain of Attraction for Generalized Predictive Control with Input Nonlinearity," *ACTA Automatica Sinica*, 30, 954–960.

Ding, B.C., and Xi, Y.G. (2006), "A Two-step Predictive Control Design for Input Saturated Hammerstein Systems," *International Journal of Robust and Nonlinear Control*, 16, 353–367.

Ding, B.C., Xi, Y.G., and Li, S.Y. (2004), "On the Stability of Output Feedback Predictive Control for Systems with Input Nonlinearity," *Asian Journal of Control*, 6, 388–397.

Eskinat, E., and Johnson, S.H. (1991), "Use of Hammerstein Models in Identification of Nonlinear Systems," *AI.Ch.E. Journal*, 37, 255–268.

Fruzzetti, K.P., Palazoglu, A., and McDonald, K.A. (1997), "Nonlinear Model Predictive Control Using Hammerstein Models," *Journal of Process Control*, 7, 31–41.

Fu, Y., and Dumont, G.A. (1993), "An Optimum Time Scale for Discrete Laguerre Network," *IEEE Transactions on Automatic Control*, 38, 934–938.

Garcia, C.E., Prett, D.M., and Morari, M. (1989), "Model Predictive Control: Theory and Practice – a Survey," *Automatica*, 25, 335–348.

George, E.F., Malcolm, M.A., and Moler, C.B. (1977), *Computer Methods for Mathematical Computations*, Englewood Cliffs, New Jersey: Prentice–Hall.

Golub, G., and Van Loan, C. (1989), *Matrix Computations* (2nd ed.), Baltimore and London: The John Hopkins University Press.

Gómez, J.C., and Baeyens, E. (2004), "Identification of Block-Oriented Nonlinear Systems Using Orthonormal Basis," *Journal of Process Control*, 14, 685–697.

Greblicki, W. (1996), "Nonlinearity Estimation in Hammerstein Systems Based on Ordered Observations," *IEEE Transactions on Signal Processing*, 44, 1224–1233.

Greblicki, W. (2002), "Stochstic Approximation in Nonparametric Identification of Hammerstein Systems," *IEEE Transactions on Automatic Control*, 47, 1800–1810.

Greblicki, W., and Pawlak, M. (1989), "Nonparametric Identification of Hammerstein Systems," *IEEE Transactions on Information Theory*, 35, 409–418.

Haddad, W.M., and Chellaboina, V.S. (2001), "Nonlinear Control of Hammerstein Systems with Passive Nonlinear Dynamics," *IEEE Transactions on Automatic Control*, 46, 1630–1634.

Hasiewicz, Z. (1999), "Hammerstein System Identification by the Haar Multiresolution Approximation," *International Journal of Adaptive Control and Signal Processing*, 13, 691–717.

Hasiewicz, Z., and Mzyk, G. (2004), "Combined Parametric-Nonparametric Identification of Hammerstein Systems," *IEEE Transactions on Automatic Control*, 49, 1370–1375.

Henson, M.A. (1998), "Nonlinear Model Predictive Control: Current Status and Future Directions," *Computers and Chemical Engineering*, 23, 187–202.

Heuberger, P.S.C., Van den Hof, P.M.J., and Bosgra, O.H. (1995), "A Generalized Orthonormal Basis for Linear Dynamical System," *IEEE Transactions on Automatic Control*, 40, 451–465.

Huner, I.W., and Korenberg, M.J. (1986), "The Identification of Nonlinear Biological Systems Wiener and Hammerstein Cascade Models," *Biological Cybernetics*, 55, 135–144.

Juditsky, A., Hjalmarsson, H., Benveniste, A., Delyon, B., Ljung, J., Sjöberg, J., and Zhang, Q. (1995), "Nonlinear Black-box Modelling in System Identification: Mathematical Foundations," *Automatica*, 31, 1725–1750.

Khalil, H.K. (2002), *Nonlinear Systems*, New Jersey: Prentice Hall.

Knohl, T., Xu, W.M., and Hubelhauen, H. (2003), "Indirect Adaptive Dual Control for Hammerstein Systems using ANN," *Control Engineering Practice*, 11, 277–385.

Kouvaritakis, B., Rossiter, J.A., and Schuurmans, J. (2000), "Efficient Robust Predictive Control," *IEEE Transactions on Automatic Control*, 45, 1545–1549.

Krantz, S.G. (2004), *A Handbook of Real Variables: With Applications to Differential Equations and Fourier Analysis*, Basel, Switzerland: Birkhauser.

Lin, Z.L., and Saberi, A. (1993), "Semi-Global Exponential Stabilization of Linear Systems Subject to 'Input Saturation' via Linear Feedbacks," *Systems and Control Letters*, 21, 225–239.

Ljung, L. (1999), *System Identification: Theory for the User* (2nd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Minesh, A.S., and Matthew, A.F. (1999), "Frequency-based Controller Design for a Class of Nonlinear Systems," *International Journal of Robust and Nonlinear Control*, 9, 825–840.

Patwardhan, R., Lakshminarayanan, S., and Shah, S.L. (1998), "Constrained Nonlinear MPC using Hammerstein," *A.I.Ch.E. Journal*, 44, 1611–1622.

Sjoberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P., Hjalmarsson, H., and Juditsky, A. (1995), "Nonlinear Black-box Modelling in System Identification: A Unified Approach," *Automatica*, 31, 1691–1724.

Stapleton, J., and Bass, S. (1985), "Adaptive Noise Cancellation for a Class of Nonlinear Dynamic Reference Channels," *IEEE Transactions on Circuits and Systems*, 32, 143–150.

Tanguy, N., Morvan, R., Vilbé, P., and Calvez, L.P. (2000), "Online Optimization of the Time Scale in Adaptive Laguerre-based Filters," *IEEE Transactions on Signal Processing*, 48, 1184–1187.

Tanguy, N., Vilbé, P., and Calvez, L.C. (1995), "Optimum Choice of Free Parameter in Orthonormal Approximations," *IEEE Transactions on Automatic Control*, 40, 1811–1813.

Wahlberg, B., and Mäkilä, P.M. (1996), "On Approximation of Stable Linear Dynamic System using Laguerre and Kautz Functions," *Automatica*, 32, 693–708.

Wang, L.P. (2004), "Discrete Model Predictive Controller Design using Laguerre Functions," *Journal of Process Control*, 14, 131–142.

Zhang, H.T., Chen, Z.G., Wang, Y.J., Qin, T., and Li, M. (2006), "Adaptive Predictive Control Algorithm Based on Laguerre Functional Model," *International Journal of Adaptive Control and Signal Processing*, 20, 53–76.

Zhang, H.T., and Li, H.X. (2007), "A General Control Horizon Extension Method for Nonlinear Model Predictive Control," *Industrial & Engineering Chemistry Research*, 46, 91799189.

Zhu, Q.M., Warwick, K., and Douce, J.L. (1991), "Adaptive General Predictive Controller for Nonlinear Systems," *IEEE Proceedings-D*, 138, 33–41.

Appendix 1

Laguerre functional series model

Definition 1 (Heuberger et al. 1995; Wahlberg and Mäkilä 1996; Wang 2004; Zhang et al. 2006; Zhang and Li 2007)**:** A Laguerre function is defined as a functional series

$$\Phi_i(t) \triangleq \sqrt{2p}\frac{e^{pt}}{(i-1)!}\cdot\frac{\mathrm{d}^{i-1}}{\mathrm{d}t^{i-1}}[t^{i-1}\cdot e^{-2pt}] \quad i=1,2,\ldots,\infty, \tag{60}$$

where p is a time-scaling factor and $t\in[0,\infty)$ is a time variable.

Theorem 4 (Wahlberg and Mäkilä 1996; Heuberger et al. 1995)**:** *A Laguerre function series generates a group of self-contained unified orthogonal radixes in the function space* $\mathbf{L}_2(\mathbb{R}^+)$.

The Z-transfer function of the mth Laguerre function is given by (Heuberger et al. 1995)

$$\Gamma_m(z)=\frac{\sqrt{(1-p^2)}}{z-p}\left[\frac{1-pz}{z-p}\right]^m \ (m=1,2,\ldots). \tag{61}$$

The discrete Laguerre function series is defined in the time domain for some $0\le p<1$, $m=1,2,\ldots,\infty$, by the inverse Z-transform of (60)

$$L_m(k)=Z^{-1}\{\Gamma_m(z)\}. \tag{62}$$

By examining the network structure of the Z-transfer functions of Laguerre filters (see Figure 2), one can find that the set of discrete laguerre function series satisfies the following difference equations (Wang 2004)

$$L(k+1)=AL(k)+Bu(k), \tag{63}$$

$$y_m(k)=c^TL(k), \tag{64}$$

where

$$A=\begin{bmatrix} p & 0 & 0 & \cdots & 0\\ \beta & p & 0 & \cdots & 0\\ -p\beta & \beta & p & \cdots & 0\\ p^2\beta & -p\beta & \beta & \cdots & 0\\ \vdots & & & & \\ (-1)^{N-2}p^{N-2}\beta & (-1)^{N-3}p^{N-3}\beta & \cdots & \beta & p \end{bmatrix}_{N\times N},$$

$$B=\begin{bmatrix} \beta^{1/2}\\ (-p)\beta^{1/2}\\ \vdots\\ (-p)^{N-1}\beta^{1/2} \end{bmatrix}_{N\times 1},$$

$\beta=\sqrt{1-p^2}$; $L(k)=[L_1(k),L_2(k),\ldots,L_N(k)]^T$ is the state vector of the Laguerre Series model; $c=[c_1,c_2,\ldots,c_N]^T$ is the Laguerre coefficient vector; and $u(k)$, $y_m(k)$ are the input and output of this model, respectively. Here, the Laguerre filter pole p is the most important parameter, to optimise which Fu and Dumont (1993) and Tanguy et al. (1995) have proposed a suboptimal offline method. The method is based on the minimisation of an upper bound for the modelling error, which has received considerable interest owing to its simplicity, its low computational cost, and its relatively good efficiency. From this method, Tanguy et al. (2000) derive a technique for suboptimal online optimisation of the Laguerre filters parameter, which has the merits of good stability, convergence and numerical robustness. We will follow the method of Tanguy et al. (2000) to set the initial value and then finely adjust p according to the control performances. For another modelling parameter, i.e. Laguerre model truncation length N, there are no theoretical criteria except some empirical ones in the literature.

Appendix 2

Proof of Theorem 2: Since the linear block is stable, $g_i(u(k))(i\,1,2,\ldots,r)$ and is bounded (because $u(k)\in\mathbf{D}$ is bounded and $g_i(\cdot)$ are non-linear basis functions), the model output $y_m(k)$ is bounded. Taking equations (21) and (6) into consideration, one has that $\|\phi(k)\|_2$ is bounded, i.e., $\exists R_L>0$ such that

$$\left\|\phi(k)\right\|_2^2\le R_L. \tag{65}$$

For any $\varepsilon>0$, $\exists\varepsilon_1,\varepsilon_2>0$ such that $\varepsilon=\varepsilon_1+\varepsilon_2$. Let $\varepsilon_3=\varepsilon_1/(\max(r,N)R_L)$ and $\varepsilon_4=\varepsilon_2/R_L$. Because the regressor $\phi(k)$ is *PE* in the sense of (25) and is unrelated to $\gamma(k)$, one has that the *LSE* $\hat{\theta}$ is strongly consistent in sense that $\hat{\theta}\to\theta$ with probability one as $S\to\infty$ (denoted $\hat{\theta}\xrightarrow{a.s.}\theta$) (Ljung 1999); in other words, $\forall\varepsilon_4>0$, $\exists N_\varepsilon>1$ such that

$$\|\hat{\theta}-\theta\|_2^2\le\varepsilon_4, \tag{66}$$

with probability one for $S > N_\varepsilon$. The convergence of the estimation $\hat{\theta}$ implies that

$$\hat{\Theta}_{ac} \xrightarrow{a.s.} \Theta_{ac} \tag{67}$$

It should be noted that the consistency of the estimation $\hat{\theta}$ holds even in the presence of coloured output noise (Gómez and Baeyens 2004).

Using Lemma 1, one can get the SVD of the matrix $\hat{\Theta}_{ac}$ and, assuming rank $(\hat{\Theta}_{ac}) = q$, it follows from Theorem 1 that $\forall \varepsilon_3 > 0$, there exists $N_C \le q$ such that

$$\sum_{m=N_C+1}^{q} \varpi_m^2 \le \varepsilon_3. \tag{68}$$

In other words,

$$\left\| \sum_{j=1}^{N_C} \hat{a}^{\langle j \rangle} \left(\hat{c}^{\langle j \rangle}\right)^T - \hat{\Theta}_{ac} \right\|_2^2 \le \varepsilon_3, \tag{69}$$

or

$$\sum_{j=1}^{N_C} \hat{a}^{\langle j \rangle} \left(\hat{c}^{\langle j \rangle}\right)^T \to \hat{\Theta}_{ac}. \tag{70}$$

Combining (67) with (70) yields conclusion (26). It should be noted that N_C is the total number of the identification channels.

Define

$$\hat{\theta}^{\langle j \rangle} \triangleq [\hat{c}_1^{\langle j \rangle}\hat{a}_1^{\langle j \rangle}, \hat{c}_1^{\langle j \rangle}\hat{a}_2^{\langle j \rangle}, \ldots, \hat{c}_1^{\langle j \rangle}\hat{a}_r^{\langle j \rangle}, \hat{c}_2^{\langle j \rangle}\hat{a}_1^{\langle j \rangle}, \ldots, \hat{c}_2^{\langle j \rangle}\hat{a}_r^{\langle j \rangle}, \ldots, \hat{c}_N^{\langle j \rangle}\hat{a}_r^{\langle j \rangle}]^T, \tag{71}$$

then $\forall S > N_\varepsilon$, the following inequality holds with probability one:

$$\begin{aligned}
[y_m(k) - \bar{y}(k)]^2 &= \left[\phi^T(k) \left(\sum_{j=1}^{N_C} \hat{\theta}^{\langle j \rangle} - \hat{\theta} + \hat{\theta} - \theta \right) \right]^2 \\
&\le R_L \left\| \sum_{j=1}^{N_C} \hat{\theta}^{\langle j \rangle} - \hat{\theta} \right\|_2^2 + R_L \left\| \hat{\theta} - \theta \right\|_2^2 \\
&\le R_L \max(r, N) \left\| \sum_{j=1}^{N_C} \varpi_j \mu_j \varphi_j - \hat{\Theta}_{ac} \right\|_2^2 + R_L \left\| \hat{\theta} - \theta \right\|_2^2 \\
&= R_L \max(r, N) \varepsilon_3 + R_L \varepsilon_4 = \varepsilon,
\end{aligned} \tag{72}$$

where the definition of matrix 2-norm is given in the reference (Gómez and Baeyens 2004). In other words, the conclusion (27) holds. This completes the proof.

Appendix 3

Lemma 3: *For any $\tau > 1$ and $\eta = 1 + (\tau - 1)^{-1}$, one has the following inequality*

$$(A_1 + A_2)^T P (A_1 + A_2) \le \tau A_1^T P A_1 + \eta A_2^T P A_2, \tag{73}$$

where P is a positive-definite symmetric matrix, and A_1 and A_2 are compatible square matrices.

Proof: For any $\tau > 1$ one has

$$\begin{aligned}
(A_1 + A_2)^T P (A_1 + A_2) &= \tau A_1^T P A_1 + [1 + (\tau - 1)^{-1}] A_2^T P A_2 \\
&\quad - (\tau - 1)[A_1 - (\tau - 1)^{-1} A_2]^T \\
&\quad \times P [A_1 - (\tau - 1)^{-1} A_2] \\
&\le \tau A_1^T P A_1 + [1 + (\tau - 1)^{-1}] A_2^T P A_2.
\end{aligned}$$

□

Proof of Lemma 2: From Lemma 3 and (38), one has $\forall \tau_1, \tau_2 > 0$ and $\eta_1 = 1 + (\tau_1 - 1)^{-1}$, $\eta_2 = 1 + (\tau_2 - 1)^{-1}$, such that

$$\begin{aligned}
&\hat{x}^T(k+i\,|\,k) P_x \hat{x}(k+i\,|\,k) \\
&\le \tau_1 \left(\Pi \hat{x}(k+i-1|k) + \begin{bmatrix} \delta(\cdot) B \bar{K} \\ 0 \end{bmatrix} \hat{x}(k+i-1|k) \right)^T P_x \\
&\quad \times \left(\Pi \hat{x}(k+i-1|k) + \begin{bmatrix} \delta(\cdot) B \bar{K} \\ 0 \end{bmatrix} \hat{x}(k+i-1|k) \right) \\
&\quad + \eta_1 \left(\begin{bmatrix} \Gamma c \\ 0 \end{bmatrix} e(k+i-1|k) \right)^T P_x \left(\begin{bmatrix} \Gamma c \\ 0 \end{bmatrix} e(k+i-1|k) \right) \\
&\le \tau_1 \tau_2 (\Pi \hat{x}(k+i-1|k))^T P_x (\Pi \hat{x}(k+i-1|k)) \\
&\quad + \tau_1 \eta_2 \left(\begin{bmatrix} \delta(\cdot) B \bar{K} \\ 0 \end{bmatrix} \hat{x}(k+i-1|k) \right)^T \\
&\quad \times P_x \left(\begin{bmatrix} \delta(\cdot) B \bar{K} \\ 0 \end{bmatrix} \hat{x}(k+i-1|k) \right) \\
&\quad + \eta_1 e^T(k+i-1|k) \hat{c}^T \Gamma^T E_L^T P_x E_L \Gamma c e(k+i-1|k) \\
&\le \tau_1 \tau_2 \hat{x}^T(k+i-1|k) \Pi^T P_x \Pi \hat{x}(k+i-1|k) \\
&\quad + \tau_1 \eta_2 \hat{x}^T(k+i-1|k) \sigma^2 \bar{K}^T B^T E_L^T P_x E_L B \bar{K} \hat{x}(k+i-1|k) \\
&\quad + \eta_1 e^T(k+i-1|k) c^T \Gamma^T E_L^T P_x E_L \Gamma c e(k+i-1|k).
\end{aligned}$$

Thus if (43) and (44) hold and $\hat{x}^T(k+i-1|k) \times P_x \hat{x}(k+i-1|k) \le 1$, then $\hat{x}^T(k+i\,|\,k) P_x \hat{x}(k+i\,|\,k) \le 1$, i.e. S_x is an invariant set.

As to the state estimation error $e(k)$, one has

$$e^T(k+i\,|\,k) P_e e(k+i\,|\,k) = e^T(k+i-1|k) \Xi^T P_e \Xi e(k+i-1|k). \tag{74}$$

Thus, if (42) holds and $e^T(k+i-1|k) P_e e(k+i-1|k) \le \bar{e}$, then $e^T(k+i\,|\,k) P_e e(k+i\,|\,k) \le \bar{e}$, i.e., S_e is an invariant set.

On the other hand,

$$\begin{aligned}
|v(k)| &= |\bar{K} \hat{x}(k)| = |\bar{K} P_x^{-1/2} P_x^{1/2} \hat{x}(k)| \\
&\le \|\bar{K} P_x^{-1/2}\| \cdot \|P_x^{1/2} \hat{x}(k)\| \\
&\le \|\bar{K} P_x^{-1/2}\|.
\end{aligned} \tag{75}$$

Then, substituting (46) into (75) yields

$$|v(k)| \le (\bar{u} - \lambda_1)/\mu_1, \tag{76}$$

and substituting (76) into the local Lipschitz condition (45) yields $|u(k)| \le \bar{u}$.

This completes the proof. □