Identification of Wiener–Hammerstein benchmark model via rank minimization

Younghee Han*, Raymond A. de Callafon

Department of Mechanical and Aerospace Engineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0401, USA

ARTICLE INFO

ABSTRACT

In this paper, a Wiener–Hammerstein system identification problem is formulated as a semidefinite programming (SDP) problem which provides a sub-optimal solution for a rank minimization problem. In the proposed identification method, the first linear dynamic system, the static nonlinear function, and the second linear dynamic system are parameterized as an FIR model, a polynomial function, and a rational transfer function respectively. Subsequently the optimization problem is formulated by using the over-parameterization technique and an iterative approach is proposed to update two unmeasurable intermediate signals. For the modeling of static nonlinearity, the monotonically non-deceasing condition was applied to limit the number of possible selections for intermediate signals. At each step of iteration, the over-parametrized parameters are estimated and then system parameters are separated by using a singular value decomposition (SVD). The proposed method is applied to the benchmark problem and the estimation result shows the effectiveness of the proposed algorithm.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Wiener–Hammerstein systems are dynamical systems characterized by a series connection of three parts: a linear dynamical system, a static nonlinearity and another linear dynamical system, as shown in Fig. 1. Early works on Wiener–Hammerstein system identification can be found in Billings and Fakhouri (1982), and Hunter and Korenberg (1986). In this early research, the correlation analysis-based identification method under Gaussian excitation has been proposed. Chen and Fassois (1992) introduced a time-domain identification method based on the Maximum Likelihood principle. Boutayed and Darouach (1995) presented a simple technique for recursive identification of the Wiener–Hammerstein model with extension to the multi-input single-output (MISO) case. More recent work can be found in Crama and Schoukens (2001), Vörös (2007), Paduart, Lauwers, Pintelon, and Schoukens (2009) and Ase, Katayama, and Tanaka (2009). Paduart et al. (2009) proposed an identification method using the polynomial nonlinear state space (PNLSS) approach. Ase et al. (2009) presented a method iteratively identifying the linear system and the Hammerstein system by minimizing the square norm of output prediction error and by using the orthogonal decomposition subspace method (ORT).

Recently, a system identification method was introduced based on the sector bound property of static nonlinearity using Quadratic Programming (QP) and semidefinite programming (SDP) relaxation in Zhang, Iouditski, and Ljung (2006) and Sou, Megretski, and Daniel (2008). In Sou et al. (2008), the identification problem was formulated as a non-convex QP. A convex SDP relaxation is then formulated and solved to obtain a sub-optimal solution to the original non-convex QP. However, the formulation of the problem is based on the existence of the inverse of the second dynamic system, which cannot be generally guaranteed.

In this paper, the SDP relaxation approach by Sou et al. (2008) is extended by using rank minimization to propose a Wiener–Hammerstein system identification method which does not require invertibility of any sub-systems. The order of a finite dimensional model can be expressed as the rank of a matrix that is filled with input and output measurement. If the set of feasible models is described by convex constraints, then choosing the simplest model can often be expressed as a rank minimization problem (Fazel, Hindi, & Boyd, 2004). Based on this idea, in this paper a Wiener–Hammerstein system identification problem is formulated as a rank minimization problem and the non-convex rank minimization problem is then formulated as a convex problem via SDP relaxation.

The objective of this paper is to formulate a procedure that allows the characterization and identification of the three parts in a Wiener–Hammerstein system individually based on the finite number of input $u(t)$ and the output $y(t)$ observations. In this paper, this is accomplished by reconstructing unmeasurable

* Corresponding author.
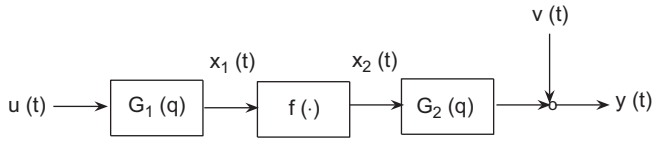*E-mail address:* y3han@ucsd.edu (Y. Han).

Fig. 1. Wiener–Hammerstein system consisting of the cascade of a linear dynamic block $G_1(q)$, a static non-linear block $f(\cdot)$ and another linear dynamic block $G_2(q)$.

intermediate signals $x_1(t)$ and $x_2(t)$ that satisfy the conditions on the finite dimensional dynamical representation of the linear systems $G_1(q)$ and $G_2(q)$, and the memoryless static nonlinearity $f(\cdot)$. Once the intermediate signals $x_1(t)$ and $x_2(t)$ are reconstructed, the identification and characterization of the three parts becomes trivial.

2. Problem description

The system to be modeled is a Wiener–Hammerstein system as shown in Fig. 1. The purpose of this study is to propose a method to identify the unknown linear dynamic systems, $G_1(q)$ and $G_2(q)$, and a static nonlinear function $f(\cdot)$ from a finite number of observations of the time domain data $u(t)$ and $y(t)$. In this paper, each block is parametrized separately. System parameters for each block will be estimated simultaneously by finding feasible models consistent with the input and output data, and by satisfying the following basic properties of the Wiener–Hammerstein system:

Condition 1.

I. The static nonlinear function has no memory: The current output $x_2(t)$ only depends on the current input $x_1(t)$.
II. The first linear dynamic system has a finite, but unknown, McMillan degree n_1: $x_1(t) = \phi_1^T(t)\theta_1$, where $\phi_1^T(t) = [u(t)\ldots u(t-n_b)\, x_1(t-1)\ldots x_1(t-n_a)]$, θ_1 is the first linear system parameter, and $n_1 \leq \max(n_b-1, n_a)$.
III. The second linear dynamic system has a finite, but unknown, McMillan degree n_2: $y(t) = \phi_2^T(t)\theta_2$, where $\phi_2^T(t) = [x_2(t)\ldots x_2(t-n_d)\, y(t-1)\ldots y(t-n_c)]$, θ_2 is the second linear system parameter, and $n_2 \leq \max(n_d-1, n_c)$.

The intermediate signals $x_1(t)$ and $x_2(t)$ in Fig. 1 are not measurable, and the properties in Condition 1 are used to formulate a procedure to reconstruct $x_1(t)$ and $x_2(t)$. The unknown signals $x_1(t)$ and $x_2(t)$ will be parametrized, and the estimation of the unknown coefficients will be formulated as an SDP problem. Let $\hat{x}_1(t)$ be the reconstructed signal of $x_1(t)$, $\hat{x}_2(t)$ be the reconstructed signal of $x_2(t)$, and $\hat{y}(t)$ be the model output. The SDP problem will be formulated in such a way that $\hat{x}_1(t)$ and $\hat{x}_2(t)$ are related via a memoryless static nonlinearity, $u(t)$ and $\hat{x}_1(t)$ are related via a linear dynamical system with the smallest McMillan degree, and $\|y-\hat{y}\|_2$ is minimized under Condition 1. Once $\hat{x}_1(t)$ and $\hat{x}_2(t)$ have been reconstructed, the identification of $G_1(q)$ from $u(t)$ to $\hat{x}_1(t)$, and the identification of $G_2(q)$ from $\hat{x}_2(t)$ to $y(t)$ can be solved with the standard Prediction Error (PE) identification method in Ljung (1999) and the identification of $f(\cdot)$ from $\hat{x}_1(t)$ to $\hat{x}_2(t)$ can be solved via the Least Squares (LS) method. The proposed identification method deals with Wiener–Hammerstein systems in this paper, but the idea of constraining rank for signal reconstruction can be extended to Hammerstein, Wiener, or Hammerstein–Wiener systems.

3. System parametrization

3.1. Modeling of the first dynamic system

In order to formulate the parameter estimation problem, a finite impulse response (FIR) model is used to model the first dynamic system G_1. Let $h_k, k = 0, 1, \ldots$ be a causal sequence of unit impulse responses for $G_1(q)$. The relationship between the input $u(t)$ and the intermediate signal $x_1(t)$ can be described by the convolution as

$$x_1(t) = \sum_{k=0}^{\infty} h_k u(t-k).$$

Due to Condition 1 (finite McMillan degree), the Hankel matrix defined as

$$H = \begin{bmatrix} h_1 & \cdots & h_{N/2} \\ h_2 & \cdots & h_{N/2+1} \\ \vdots & \ddots & \vdots \\ h_{N/2} & \cdots & \ldots h_{N-1} \end{bmatrix} \tag{1}$$

has a $rank(H) \leq n_1$. The order of the linear dynamical system is determined by the $rank(H)$ as H is simply the product of the extended observability and controllability matrices (Goethals, Pelckmans, Suykens, & De Moor, 2005). Let

$$\hat{x}_1 = [\hat{x}_1(1)\ \hat{x}_1(2)\ \ldots\ \hat{x}_1(N)]^T$$

and

$$U = \begin{bmatrix} u(1) & u(0) & \cdots & u(2-N) \\ u(2) & u(1) & \cdots & u(1-N) \\ \vdots & \vdots & \vdots & \\ u(N) & u(N-1) & \cdots & u(1) \end{bmatrix} \tag{2}$$

With the system parameter

$$h = [h_0\ h_1 \ldots h_{N-1}]^T \tag{3}$$

to be estimated, $\hat{x}_1$ can be written as

$$\hat{x}_1 = Uh \tag{4}$$

The finite order sequence of $h_k, k = 0, 1, \ldots, N-1$, for a lower order model for G_1 can be estimated by minimizing the rank of H in (1) (Fazel et al., 2004). The rank minimization of H is used only to minimize the order of G_1. The FIR approximation of G_1 is used to formulate a convex optimization problem to estimate system parameters in Section 4. Once $\hat{x}_1(t)$ has been reconstructed, the identification of $G_1(q)$ from $u(t)$ to $\hat{x}_1(t)$ can be solved with the standard Prediction Error (PE) identification method in Ljung (1999). Based on (4), the error is defined by

$$e(t) = x_1(t) - \hat{x}_1(t) = x_1(t) - \sum_{k=0}^{L-1} h_k u(t-k)$$

Thus,

$$\|e(t)\|_2^2 = \sum_{t=1}^{N} \left[x_1(t) - \sum_{k=0}^{L-1} h_k u(t-k) \right]^2 = \sum_{t=1}^{N} x_1^2(t) - 2\sum_{k=0}^{L-1} h_k R_{x_1 u}(k) + \sum_{k=0}^{L-1}\sum_{l=0}^{L-1} h_k h_l R_{uu}(k-l)$$

where

$$R_{x_1 u}(k) = \sum_{t=1}^{N} x_1(t)u(t-k)$$

$$R_{uu}(k) = \sum_{t=1-k-l}^{N-k-l} u(t)u(t+k)$$

If L tends toward infinity, the h_k obtained by minimizing the *rank*(H) will satisfy $|h_k| \ll 1$ for $k \geq L$, resulting in $R_{x_1 u}(k) \to \sum_{l=0}^{L-1} h_l R_{uu}(l-k)$. Then,

$$\lim_{N \to \infty, L \to \infty} \|e(t)\|_2^2 = 0 \qquad (5)$$

As a result, the estimate $\hat{x}_1(t)$ in (4) will converge to $x_1(t)$ provided that $N \to \infty$ and $L \to \infty$.

3.2. Modeling of the static nonlinearity

It is well known that the static nonlinear function can be approximated as a linear combination of a finite set of basis functions as

$$f(u(t)) \approx \sum_{m=1}^{M} \lambda_m \xi_m(u(t)) \qquad (6)$$

where λ_m are weighting parameters to be estimated and $\xi_m(\cdot)$ are basis functions. A user can choose any basis function if it is linear in parameters as shown in (6). In this paper, a n_f th order polynomial function is used to model static nonlinearity. With the polynomial basis functions, $\hat{x}_2(t)$ is defined by

$$\hat{x}_2(t) = \lambda_0 + \lambda_1 \hat{x}_1(t) + \lambda_2 \hat{x}_1^2(t) + \cdots + \lambda_{n_f} \hat{x}_1^{n_f}(t) \qquad (7)$$

Weierstrass's Theorem guarantees that the polynomial approximation $\hat{x}_2(t)$ in (7) will converge to $x_2(t)$ as n_f tends toward to infinity for an arbitrary interval.

Weierstrass's Theorem: If $f(x_1)$ is a given continuous function for an arbitrary interval $i \leq x_1 \leq j$, and ϵ is a small magnitude positive constant, there is a polynomial $\hat{f}(x_1)$ such that

$$|f(x_1) - \hat{f}(x_1)| < \epsilon \quad \forall x \in [i,j]$$

There could be many possible combinations of $(\hat{x}_1(t), \hat{x}_2(t))$ that satisfy Condition 1 and (7). In order to limit the number of possible selections of $(\hat{x}_1(t), \hat{x}_2(t))$, in this paper, a monotonically non-decreasing static nonlinearity with the maximum slope of 1 is considered as follows:

Condition 2.

I. The static nonlinear function is monotonically non-decreasing with the maximum slope of 1: $(\hat{x}_2(i) - \hat{x}_2(j))(\hat{x}_2(i) - \hat{x}_2(j) - \hat{x}_1(i) + \hat{x}_1(j)) \leq 0 \forall i > j$.

In Condition 2,

$$\hat{x}_2(i) - \hat{x}_2(j) \geq 0 \Rightarrow \hat{x}_2(i) - \hat{x}_2(j) \leq \hat{x}_1(i) - \hat{x}_1(j)$$

or

$$\hat{x}_2(i) - \hat{x}_2(j) \leq 0 \Rightarrow \hat{x}_2(i) - \hat{x}_2(j) \geq \hat{x}_1(i) - \hat{x}_1(j).$$

In both cases,

$$\hat{x}_1(i) - \hat{x}_1(j) = 0 \Rightarrow \hat{x}_2(i) - \hat{x}_2(j) = 0$$

or

$$\hat{x}_1(i) - \hat{x}_1(j) \neq 0 \Rightarrow \frac{\hat{x}_2(i) - \hat{x}_2(j)}{\hat{x}_1(i) - \hat{x}_1(j)} \leq 1.$$

The maximum slop of the static nonlinearity is a user-chosen value, thus can be adjusted by a user. In this paper, we assumed that $\lambda_0 = 0$ and $\lambda_1 = 1$. The assumptions are not necessary for the proposed method, but chosen for notational brevity and normalization. Without loss of generality, this monotonicity assumption on the unknown static nonlinearity combined with the

assumptions ($\lambda_0 = 0$, $\lambda_1 = 1$, and the maximum slope ≤ 1) guarantees a solution for an FIR linear system for G_1 and serves as a normalization condition on the static nonlinearity. Based on (4) and (7), $\hat{x}_2 = [\hat{x}_2(1) \ldots \hat{x}_2(N)]^T$ can be calculated as

$$\hat{x}_2 = Uh + X_1 \lambda \qquad (8)$$

where

$$\lambda = [\lambda_2 \ldots \lambda_{n_f}]^T \qquad (9)$$

and

$$X_1 = \begin{bmatrix} \hat{x}_1^2(1) & \cdots & \hat{x}_1^{n_f}(1) \\ \vdots & \vdots & \vdots \\ \hat{x}_1^2(N) & \cdots & \hat{x}_1^{n_f}(N) \end{bmatrix} \qquad (10)$$

An iterative approach will be used to update the higher order nonlinear terms of $\hat{x}_1$ in (10) that are included in the description of $\hat{x}_2$ in (8).

3.3. Modeling of the second dynamic system

Since the output of the Wiener–Hammerstein system is measured, a rational transfer function is used to model the second dynamic system G_2. The simulation output $\hat{y}(t)$ is defined as

$$\hat{y}(t) = G_2(q)\hat{x}_2(t) = \frac{D(q)}{C(q)}\hat{x}_2(t)$$

where

$$C(q) = 1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}$$

$$D(q) = d_0 + d_1 q^{-1} + \cdots + d_{n_d} q^{-n_d} \qquad (11)$$

Using the system parameters in (11), the linear difference equation between the output $\hat{y}(t)$ and the intermediate signal $\hat{x}_2(t)$ is defined as

$$\hat{y}(t) = -\sum_{k=1}^{n_c} c_k y(t-k) + \sum_{k=0}^{n_d} d_k \hat{x}_2(t-k)$$

Let

$$\Gamma(t) = \sum_{k=0}^{n_d} d_k \hat{x}_2(t-k) = \sum_{k=0}^{n_d} d_k U(t-k, :)h + \sum_{k=0}^{n_d} d_k X_1(t-k, :)\lambda \qquad (12)$$

and

$$T = Uhd^T + X_1 \lambda d^T = \begin{bmatrix} T_{1,1} & \cdots & T_{1,n_d+1} \\ \vdots & \vdots & \vdots \\ T_{N,1} & \cdots & T_{N,n_d+1} \end{bmatrix} \qquad (13)$$

using U in (2), h in (3), λ in (9) and X_1 in (10), where

$$d = [d_0 \ldots d_{n_d}]^T \qquad (14)$$

and the notations $(k,:)$ and $(:,k)$ are used to denote the kth row and the kth column in a matrix respectively. Then $\Gamma(t)$ in (12) can be rewritten as

$$\Gamma(t) = \sum_{k=1}^{min(t,n_d)} T_{t-k+1,k}$$

Using the given parameterization, the simulated output vector $\hat{y} = [\hat{y}(1) \ldots \hat{y}(N)]^T$ can be written as

$$\hat{y} = Yc + \Gamma \qquad (15)$$

where

$$c = [c_1 \ldots c_{n_c}]^T \qquad (16)$$

$$Y = \begin{bmatrix} -y(1-1) & \ldots & -y(1-n_c) \\ \vdots & \vdots & \vdots \\ -y(N-1) & \ldots & -y(N-n_c) \end{bmatrix} \tag{17}$$

and

$$\Gamma = [\Gamma(1)\ldots\Gamma(N)]^T \tag{18}$$

With the system parametrization in Section 3 based on Condition 1, system parameters for a lower order model for G_1, consistent with the input and output measurement data, can be estimated by minimizing $\|y-\hat{y}\|_2$, with $\hat{y}$ in (15) and $rank(H)$, with H in (1) simultaneously under the constraints developed based on Condition 2.

4. Problem formulation

4.1. Optimization problem

With the system parametrization and constraints explained in Section 3, the optimization problem to obtain system parameters can be written as

Optimization Problem 1.

Consider variables

$h_{N\times 1}$ in (3)

$\lambda_{n_f-1\times 1}$ in (9)

$c_{n_c\times 1}$ in (16)

$d_{n_d+1\times 1}$ in (14)

and define

$\hat{x}_1 = Uh$ in (4)

$\hat{x}_2 = Uh + X_1\lambda$ in (8)

$\hat{y} = Yc + \Gamma$ in (15)

minimize

$w_1 rank(H) + w_2 \|y-\hat{y}\|_2$ with H in (1)

subject to

$(\hat{x}_2(i)-\hat{x}_2(j))(\hat{x}_2(i)-\hat{x}_2(j)-\hat{x}_1(i)+\hat{x}_1(j)) \leq 0$ $\forall i > j.$

In Optimization Problem 1, w_1 and w_2 are weighting factors. Optimization Problem 1 is a non-convex quadratic programming (QP) problem. Semidefinite programming (SDP) relaxation is a standard approach to solve non-convex QP problems. An SDP relaxation procedure converts a non-convex optimization problem to a convex optimization problem by defining a feasible convex set, which is easier to solve and whose solution is close to the solution of the original non-convex optimization problem.

4.2. Semidefinite programming relaxation

In order to convert the non-convex Optimization Problem 1 to a convex optimization problem, the over-parametrization technique is used in this paper. Let us define a system parameter matrix Θ that includes system parameters h, λ and b. Let us define the parameter θ as

$$\theta = [h^T \ \lambda^T \ d^T]^T \tag{19}$$

and then define the over-parametrized parameter matrix Θ as

$$\Theta = \theta \cdot \theta' = \begin{bmatrix} h_0 h_0 & \ldots & \lambda_{n_f} h_0 & \ldots & d_{n_d} h_0 \\ h_0 h_1 & \ldots & \lambda_{n_f} h_1 & \ldots & d_{n_d} h_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_0 \lambda_{nf} & \ldots & \lambda_{n_f}\lambda_{n_f} & \ldots & d_{n_d}\lambda_{n_f} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_0 d_{n_d} & \ldots & \lambda_{n_f} d_{n_d} & \ldots & d_{n_d} d_{n_d} \end{bmatrix} \tag{20}$$

An arbitrary gain may be distributed among the static nonlinearity and the two linear dynamic systems. In order to avoid an ambiguous gain, the scaling of the first dynamic system will be fixed by setting $\sum_{k=0} h_k^2 = 1$. The scaling of the static nonlinear function is fixed as explained in Section 3.2. Let us define U_1 and U_2 as

$$U_1 = [U \ zeros(N,n_f-1)] \tag{21}$$

and

$$U_2 = [U \ X_1] \tag{22}$$

respectively with U in (2) and X_1 in (10) where $zeros(N,n_f-1)$ is an N-by-(n_f-1) matrix of zeros. Then, with U_1 in (21) and U_2 in (22), let

$$\delta U_1 = \begin{bmatrix} U_1(2, :)-U_1(1, :) \\ U_1(3, :)-U_1(2, :) \\ \vdots \\ U_1(N, :)-U_1(N-1, :) \\ \vdots \\ U_1(N, :)-U_1(1, :) \end{bmatrix} \tag{23}$$

and

$$\delta U_2 = \begin{bmatrix} U_2(2, :)-U_2(1, :) \\ U_2(3, :)-U_2(2, :) \\ \vdots \\ U_2(N, :)-U_2(N-1, :) \\ \vdots \\ U_2(N, :)-U_2(1, :) \end{bmatrix} \tag{24}$$

where the notations $(k,:)$ and $(:,k)$ are used to denote the kth row and the kth column in a matrix respectively. With the system parameter matrix Θ, the constraint in Optimization Problem 1 can be rewritten as a linear matrix inequity (LMI) condition as

$$\Delta X_2^T \Delta X_2 - \Delta X_2^T \Delta X_1 \leq 0 \tag{25}$$

where

$$\Delta X_2^T \Delta X_2 = diag(diag(\delta X_1 \tilde{\Theta} \delta X_1^T)) \tag{26}$$

and

$$\Delta X_2^T \Delta X_1 = diag(diag(\delta X_1 \tilde{\tilde{\Theta}} \delta U^T)) \tag{27}$$

Here,

$$\delta X_1 = \begin{bmatrix} X_1(2, :)-X_1(1, :) \\ X_1(3, :)-X_1(2, :) \\ \vdots \\ X_1(N, :)-X_1(N-1, :) \\ \vdots \\ X_1(N, :)-X_1(1, :) \end{bmatrix} \tag{28}$$

with X_1 in (10),

$$\delta U = \begin{bmatrix} U(2,:)-U(1,:) \\ U(3,:)-U(2,:) \\ \vdots \\ U(N,:)-U(N-1,:) \\ \vdots \\ U(N,:)-U(1,:) \end{bmatrix} \quad (29)$$

with U in (2), $\tilde{\Theta} = \Theta(N+1:N+M,N+1:N+M)$, and $\tilde{\tilde{\Theta}} = \Theta(N+1:N+M,1:N)$. Here $diag(x)$ indicates a square matrix with the elements of a vector x on the diagonal, and $diag(X)$ indicates the main diagonal of a matrix X. Using the system parameter matrix Θ, the simulated output $\hat{y}$ in (15) is defined as

$$\hat{y} = Yc + \Gamma$$

where

$$T = U\Theta(1:N,N+n_f:N+n_f+n_d) \\ + X_1\Theta(N+1:N+n_f-1,N+n_f:N+nf+n_d)$$

Here, Θ in (20), satisfying the LMI in (25), is a rank 1 matrix for noiseless cases. However, in order to account for the noise effect, the condition can be relaxed to a rank inequality condition as $rank(\Theta) \le \gamma$, where γ is a positive constant. An optimization problem with rank inequality conditions is hard to solve. One simple and effective way, applicable when the matrix is symmetric positive semidefinite, is to use its trace in place of its rank. The motivation for the use of its trace is that if the matrix Θ is a symmetric and positive semidefinite, its singular values are the same as its eigenvalues. Therefore, the nuclear norm reduces to trace, and the nuclear norm is the convex envelope of the rank function on the set of matrices with norms less than one (Fazel, Hindi, & Boyd, 2001). As a result, $tr(\Theta) \le p$, where p is a positive constant, is used instead of the rank inequality condition. The positive constant p can be tuned by investigating estimation results. Due to this SDP relaxation used to formulate Optimization Problem 2 and the user-chosen value p, the constraint I in Condition 2 is relaxed to

$$(\hat{x}_2(i)-\hat{x}_2(j))(\hat{x}_2(i)-\hat{x}_2(j)-(1+r)(\hat{x}_1(i)-\hat{x}_1(j))) \le q \quad \forall i > j$$

where r and q are small magnitude positive constants determined by the user-chosen value p and a noise level. Also, the rank minimization on H in Optimization Problem 1 is eliminated since this condition is absorbed into the $tr(\Theta) \le p$. Finally, the non-convex Optimization Problem 1 is reformulated as an SDP convex optimization problem as

Optimization Problem 2.

Consider variables

$\Theta_{N+M+n_d \times N+M+n_d}$ in (20)

$c_{n_c \times 1}$ in (16)

and define

$\hat{y} = Yc + \Gamma$ in (15)

minimize

$\|y - \hat{y}\|_2$

subject to

$tr(\Theta) \le p$

$\Delta X_2^T \Delta X_2 - \Delta X_2^T \Delta X_1 \le 0$

$\Theta \ge 0$

$$\sum_{k=1}^{N} \Theta(k,k) = 1$$

with $\Delta X_2^T \Delta X_2$ in (26) and $\Delta X_2^T \Delta X_1$ in (27).

In Optimization Problem 2, it is assumed that the user-specified structure variables n_c, n_d, and n_f are known. Once the optimal Θ in (20) is obtained, the optimal θ (19) can be obtained by conducting a Singular Value Decomposition (SVD). The singular vector corresponding to the largest singular value is the optimal solution for θ.

4.3. Iterative approach

Obviously, the proposed identification method requires prior information of $\hat{x}_1$ to obtain X_1 in (10). Let us define a new system parameter ϕ that includes the parameter θ in (19) and the parameter c in (16) as $\phi = [\theta^T \ c^T]^T$. With the initialization $X_1^1 = zeros(N, n_f - 1)$ (this means $\hat{x}_1^1(t) = u(t)$) and the previous parameter estimation $\hat{\phi}^{k-1}$, we propose the following iterative method:

Step 1 Construct the necessary matrices for the optimization problem formulation (δU in (29), δU_1 in (23), δU_2 in (24), Y in (17), Γ in (18), $\Delta X_2^T \Delta X_2$ in (26), $\Delta X_2^T \Delta X_1$ in (27)).

Step 2 Solve Optimization Problem 2 to obtain Θ in (20) and a in (16).

Step 3 Conduct a SVD on Θ in (20) to obtain θ in (19) and define $\hat{x}_1$ in (4).

Step 4 Update X_1^k in (10) using $\hat{x}_1$ estimated in Step 3.

Step 5 Stopping criterion of the algorithm. If $\|\hat{\phi}^k - \hat{\phi}^{k-1}\| / \|\hat{\phi}^{k-1}\| < \varepsilon$, stop. Otherwise, go to Step 1.

Step1 creates the matrices necessary for constructing Optimization Problem 2. Step 2 actually solves Optimization Problem 2 to obtain Θ in (20) and a in (16). Step 3 conducts an SVD to obtain θ in (19). Step 4 updates the prior information to construct Optimization Problem 2. Step 4 formulates a stopping criterion for the algorithm by looking at the relative parameter error.

As long as the classes of models used for the estimation contain the true models for static nonlinearity and for linear dynamic systems, and the assumptions on static nonlinearity are indeed true, $\hat{x}_1(t)$ and $\hat{x}_2(t)$ will converge to $x_1(t)$ and $x_2(t)$ provided N, L and n_f are large enough at each iteration step based on (5) and Weierstrass's Theorem.

5. Benchmark problem

The system to be modeled is an electronic nonlinear system with a Wiener–Hammerstein structure that was built by Vandersteen (1997). The first linear dynamic system G_1 is designed as a third order Chebyshev filter (a pass-band ripple of 0.5 dB and a cut-off frequency of 4.4 kHz). The second linear dynamic system G_2 is designed as a third order inverse Chebyshev filter (a stop-band attenuation of 40 dB starting at 5 kHz). This system has a transmission zero in the frequency band of interest. This can complicate the identification significantly, because the inversion of such a characteristic is difficult.

The proposed iterative identification method is applied to the benchmark problem. In this benchmark, the estimation data are the first part of the measured input $u(t)$ and output $y(t)(t = 1, 2, \ldots, 100\,000)$, and the test data are given by the remaining part of the measured input $u(t)$ and output $y(t)(t = 100\,001, \ldots, 188\,000)$. The goal of the benchmark is to identify a nonlinear model using the estimation data. Next, this model is used to simulate the output $y_{sim}(t)$ of the system on the test set. $n_c = 3, n_d = 3$, and $n_f = 5$ are used in this study. In order to solve the SDP problem (Optimization Problem 2), SEDUMI (Sturm, 1999) and YALMIP (Löfberg, 2004) are used. The estimation results are
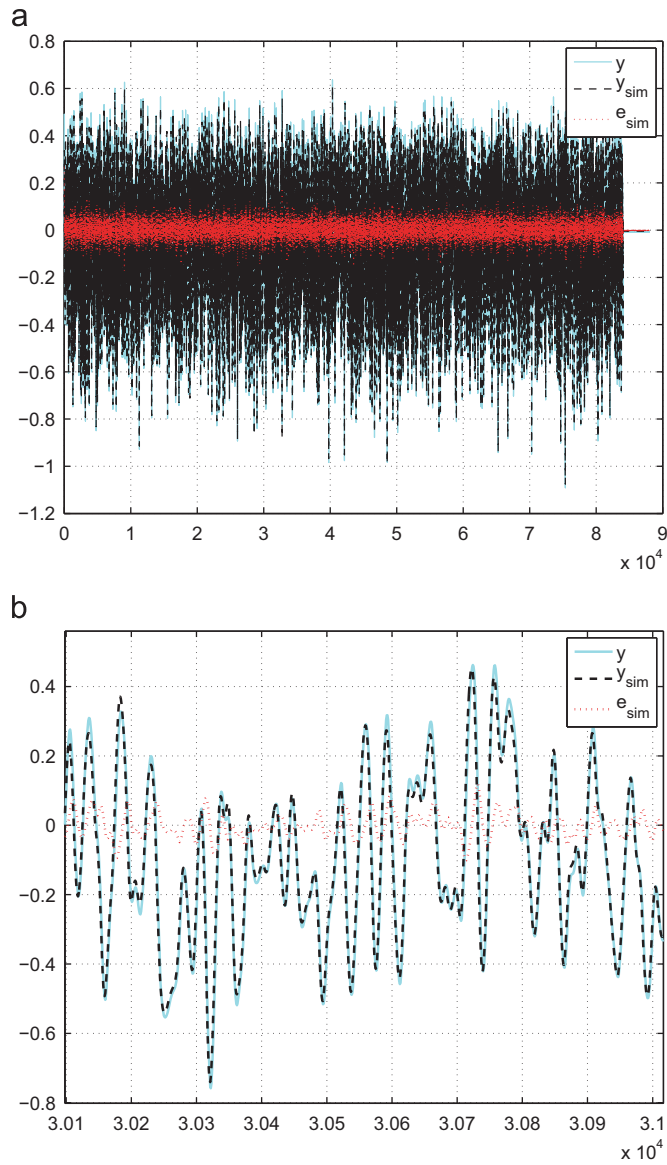
a

b

Fig. 2. Modeled output y_{sim}, test data y, and the simulation error e_{sim} in the time domain (top figure). The magnified figure of the top figure (bottom figure).

shown in Figs. 2–5, Table 1. Table 1 shows the mean value (μ), the standard deviation (s), and the root mean square (RMS) value (e_{RMS}) of the simulation error (time domain) for the estimation data and the test data obtained by using the proposed method, and the comparison with the results from Paduart et al. (2009) and Ase et al. (2009). Each value is calculated based on the following equation:

Test data

1. The mean value of the simulation error:

$$\mu = \frac{1}{87\,000} \sum_{t=101\,001}^{188\,000} e_{sim}(t)$$

2. The standard deviation of the simulation error:

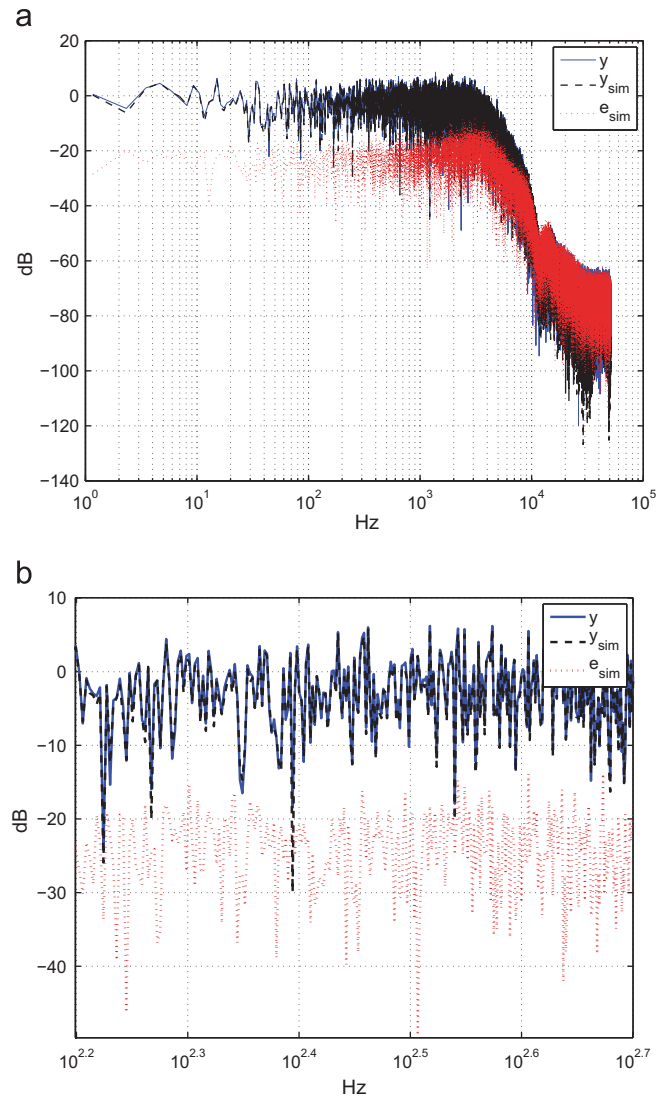$$s = \sqrt{\frac{1}{87\,000} \sum_{t=101\,001}^{188\,000} (e_{sim}(t) - \mu)^2}$$

a

b

Fig. 3. Modeled output y_{sim}, test data y, and the simulation error e_{sim} in the frequency domain (top figure). The magnified figure of the top figure (bottom figure).
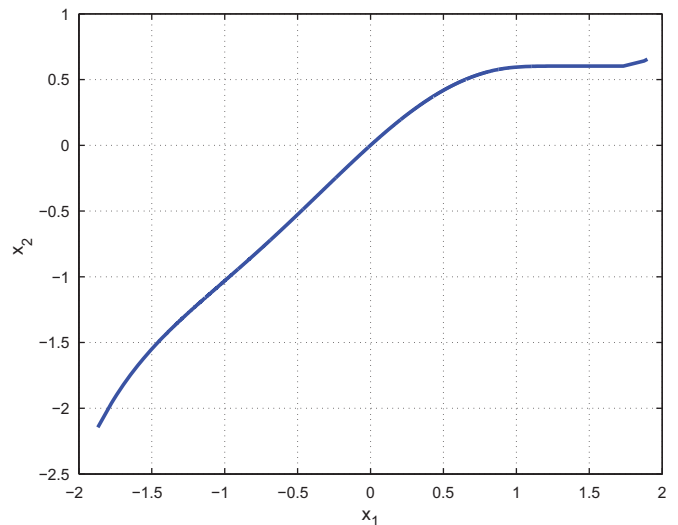

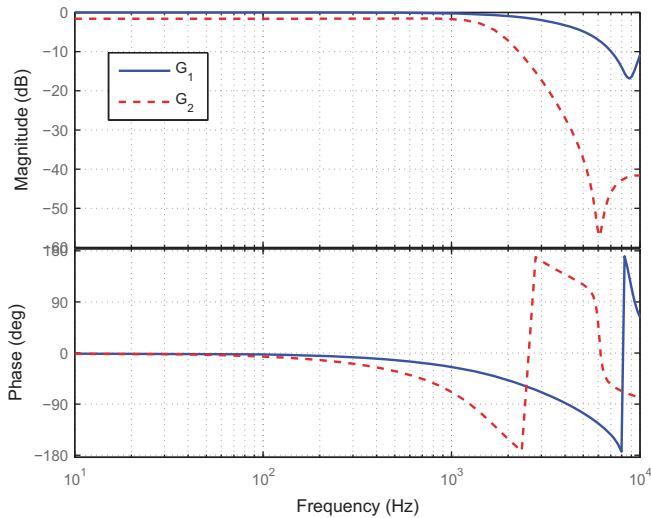**Fig. 4.** Identified static nonlinear function, $f(\cdot)$.

Fig. 5. Identified dynamical systems, G_1 and G_2.

Table 1
Characteristics of the simulation error.

Method	Parameters	Estimation data	Test data
The proposed method	μ	0.0011 V	0.0015 V
	s	0.0345 V	0.0345 V
	e_{RMS}	0.0345 V	0.0345 V
PNLSS Paduart (2009)	μ	0.031 mV	0.048 mV
	s	0.359 mV	0.415 mV
	e_{RMS}	0.360 mV	0.418 mV
Ase et al. (2009)	μ	-0.0051 V	-0.0038 V
	s	0.0332 V	0.0333 V
	e_{RMS}	0.0336 V	0.0335 V

3. The root mean square (RMS) value of the error:

$$e_{RMS} = \sqrt{\frac{1}{87\,000} \sum_{t=101\,001}^{188\,000} e_{sim}^2(t)}$$

Estimation data

1. The mean value of the simulation error:

$$\mu = \frac{1}{99\,000} \sum_{t=1001}^{100\,000} e_{sim}(t)$$

2. The standard deviation of the simulation error:

$$s = \sqrt{\frac{1}{99\,000} \sum_{t=1001}^{100\,000} (e_{sim}(t) - \mu)^2}$$

3. The root mean square (RMS) value of the error:

$$e_{RMS} = \sqrt{\frac{1}{99\,000} \sum_{t=1001}^{100\,000} e_{sim}^2(t)}$$

6. Conclusion

An iterative convex optimization algorithm is proposed to identify Wiener–Hammerstein systems. A non-convex rank minimization problem is formulated first, and then the non-convex rank minimization problem is reformulated as a convex optimization problem using an SDP relaxation technique. In the proposed identification method, the first linear dynamic system, the static nonlinear function, and the second linear dynamic system are parameterized as an FIR model, a polynomial function, and a rational transfer function respectively. For the modeling of static nonlinearity, the monotonically non-deceasing condition was applied to limit the number of possible selections for intermediate signals. As two unmeasurable intermediate signals are included in the system description, the over-parameterization technique is used and the parameter estimation problem is solved iteratively. At each step of iteration, the over-parametrized parameters are estimated and then separated by using the singular value decomposition (SVD). The proposed method is applied to the benchmark problem and the estimation result shows the effectiveness of the proposed algorithm.

References

Ase, H., Katayama, T., & Tanaka, H. (2009). A state-space approach to identification of Wiener–Hammerstein benchmark model. In *15th IFAC symposium on system identification* (pp. 1092–1097). Saint-Malo, France.

Billings, S. A., & Fakhouri, S. Y. (1982). Identification of systems containing linear dynamic and static nonlinear elements. *Automatica, 18*, 15–26.

Boutayed, M., & Darouach, M. (1995). Recursive identification method for MISO Wiener–Hammerstein model. *IEEE Transactions on Automatic Control, 40*, 287–291.

Chen, C. H., & Fassois, S. D. (1992). Maximum likelihood identification of stochastic Wiener–Hammerstein-type non-linear systems. *Mechanical Systems and Signal Processing, 6*, 135–153.

Crama, P., & Schoukens, J. (2001). Initial estimates of Wiener and Hammerstein systems using multisine excitation. *IEEE Transactions on Instrumentation and Measurement, 50*, 1791–1795.

Fazel, M., Hindi, H., & Boyd, S. (2001). Rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American control conference* (pp. 4734–4739). Arlington, VA.

Fazel, M., Hindi, H., & Boyd, S. (2004). Rank minimization and applications in system theory. In *Proceedings of the American control conference* (pp. 3273–3278). Boston, Massachusetts.

Goethals, I., Pelckmans, K., Suykens, J., & De Moor, B. (2005). Subspace identification of Hammerstein systems using least squares support vector machines. *IEEE Transactions on Automatic Control, Special Issue on System Identification, 50*, 1509–1519.

Hunter, I. W., & Korenberg, M. J. (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics, 55*, 135–144.

Ljung, L. (1999). *System identification: theory for the user*. New Jersey: Prentice-Hall, Inc, Upper Saddle River.

Löfberg, J. (2004). YALMIP: A toolbox for modeling and optimization in MATLAB. In *IEEE international symposium on CACSD conference* (pp. 284–289). Taipei, Taiwan.

Paduart, J., Lauwers, L., Pintelon, R., & Schoukens, J. (2009). Identification of a Wiener–Hammerstein system using the polynomial nonlinear state space approach. In *15th IFAC symposium on system identification* (pp. 1080–1085). Saint-Malo, France.

Sou, K.C., Megretski, A., & Daniel, L. (2008). Convex relaxation approach to the identification of the Wiener–Hammerstein model. In *47th IEEE conference on decision and control* (pp. 1375–1382). Cancun, Mexico.

Sturm, J. F. (1999). Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software, 11–12*, 625–653.

Vandersteen, G. (1997). Identification of linear and nonlinear systems in an errors-in-variables least squares and total least squares framework. Ph.D. thesis, Vrije Universiteit Brussel.

Vörös, J. (2007). An iterative method for Wiener–Hammerstein systems parameter identification. *Journal of Electrical Engineering, 58*, 114–117.

Zhang, Q., Iouditski, A., & Ljung, L. (2006). Identification of Wiener system with monotonous nonlinearity. In *14th IFAC symposium on system identification* (pp. 166–171). Newcastle, Australia.