Optimality analysis of the Two-Stage Algorithm for Hammerstein system identification

Qinghua Zhang, Jiandong Wang, Lennart Ljung

Division of Automatic Control

E-mail: zhang@irisa.fr, jiandong@pku.edu.cn, ljung@isy.liu.se

13th May 2009

Report no.: LiTH-ISY-R-2909

Accepted for publication in 15th IFAC Symposium on System Identification, Saint-Malo, France

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Abstract

The Two-Stage Algorithm (TSA) has been extensively used and adapted for the identification of Hammerstein systems. It essentially deals with the parameter estimation problem for a linear regression parameterized in a bilinear form. This paper is motivated by a somewhat contradictory fact: though the optimality of the TSA has been established only in the case of some special weighting matrices, this algorithm usually uses the unweighted case, i.e., no weighting matrix or an identity one is used. We find out that the unweighted TSA indeed gives the optimal solution of the weighted nonlinear least-squares optimization problem for a particular weighting matrix. This provides a theoretical justification of the unweighted TSA, and leads to a generalization of the obtained result to the case of colored noise with noise whitening. Numerical examples of identification of Hammerstein systems are presented to validate the theoretical analysis.

Keywords: System identification, Hammerstein system, two-stage algorithm

Optimality analysis of the Two-Stage Algorithm for Hammerstein system identification ⋆

Jiandong Wang * **Qinghua Zhang** **,1 **Lennart Ljung** ***

* *College of Engineering, Peking University, Beijing, China 100871*
** *INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France*
*** *Dept. of Electrical Eng., SE-581 83, Linköping University, Sweden*

Abstract: The Two-Stage Algorithm (TSA) has been extensively used and adapted for the identification of Hammerstein systems. It essentially deals with the parameter estimation problem for a linear regression parameterized in a bilinear form. This paper is motivated by a somewhat contradictory fact: though the optimality of the TSA has been established in [2] only in the case of some special weighting matrices, this algorithm usually uses the unweighted case, i.e., no weighting matrix or an identity one is used. We find out that the unweighted TSA indeed gives the optimal solution of the weighted nonlinear least-squares optimization problem for a particular weighting matrix. This provides a theoretical justification of the unweighted TSA, and leads to a generalization of the obtained result to the case of colored noise with noise whitening. Numerical examples of identification of Hammerstein systems are presented to validate the theoretical analysis.

Keywords:
System identification, Hammerstein system, two-stage algorithm

1. INTRODUCTION

Within the class of block-oriented nonlinear systems, a Hammerstein system is composed of a static nonlinearity block followed by a linear dynamic block. In the quite vast literature on Hammerstein systems, most known identification methods are covered by the ten methods summarized in Section 3.9 of [11]. Another quite complete survey given in Chapter 1 of [12] classifies most existing methods into four groups. The present paper focuses on the so-called Two-Stage Algorithm (TSA), also known as the over-parameterized method. It has been extensively used and adapted in the identification of block-oriented nonlinear systems, e.g., for Hammerstein-Wiener systems [2][14][15][1], and for Hammerstein/Wiener systems [5][13][9][10].

The TSA is essentially based on a particular formulation of the Hammerstein system in the form of a linear regression parameterized in a *bilinear form*, also referred to as *bilinear equation* (see, e.g., [6][3]). More specifically, let $\Psi(t) \in \mathbb{R}^{n\times m}$ be a matrix filled with $l = nm$ regressors, the linear regression

$$y(t) = b^T \Psi(t)\, a + v(t) \tag{1}$$

bilinearly parameterized by $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$ can be used to formulate a Hammerstein system with $y(t) \in \mathbb{R}$ and $v(t) \in \mathbb{R}$ corresponding respectively to the output and to the additive noise of the system (more details will be given in the next section). The estimation of the bilinear parameters a and b is usually formulated as a nonlinear Least Squares (LS) problem. The TSA uses a relaxation approach to solve this problem, by first over-parameterizing the bilinearly parameterized model (1) with linear parameters before reducing the estimated linear parameters back to the bilinear parameters a and b.

The parameter estimation problem for bilinear equations in the form of (1) is receiving an increasing attention. Cohen & Tomasi [6] made some preliminary remarks on the problem of solving systems of bilinear equations. Bai [2] studied the TSA regarding its optimality in the sense of a weighted LS criterion. Goethals *et al.* [8] applied the equivalence of the TSA in [2] to an LS support vector machine context for identification of Hammerstein systems. Bai & Liu [3] compared the normalized iterative method, the TSA, and the application of some numerical search method for nonlinear LS solution. Abrahamsson, Kay & Stoica [1] proposed a new method to estimate the parameters of a general bilinear equation based on a better approximation of a weighting matrix occuring in the LS problem, with applications to submarine detection and Hammerstein-Wiener model identification. Despite these works, the study of the problem is far from mature, just like stated in [1]: “Bilinear systems of equations and models, however, are still not very well understood even though they are fairly common and despite the fact that they could be considered the next logical step after linear models.” As a matter of fact, the TSA has not been very well understood, for some important questions have not been answered and need further investigations.

⋆ This research was partially supported by the National Natural Science Foundation of China under grants No.60704031 and No.10832006.). Paper titles should be written in uppercase and lowercase letters, not all uppercase.

[1] Corresponding author: Telephone: +33(2)9984-7463; Fax: +33(2)9984-7171; Email: zhang@irisa.fr

In this paper, we will study a somewhat contradictory fact: though the optimality of the TSA has been established in [2] only in the case of some special weighting matrices, this algorithm, adopted in some other works, usually uses the unweighted LS solution in its first stage. That is, the identity matrix, which usually does not belong to the above special weighting matrices, is used. The contribution of this paper is to remove this "contradiction" by justifying the use of unweighted LS solution in this case, and to generalize the obtained result to the case of colored noise with noise whitening.

The rest of the paper is organized as follows. Section 2 describes the bilinear equation for Hammerstein systems. Some optimality results of the TSA are obtained in Section 3. Numerical examples of identification of Hammerstein systems are presented in Section 4 to validate the theoretical analysis. Section 5 concludes the paper.

2. BILINEAR EQUATION FORMULATION OF HAMMERSTEIN SYSTEMS

It will be shown in this section that, when appropriately parameterized, Hammerstein system identification can be formulated in the form of (1), which is a linear regression with bilinear parameters. In order to focus on the main issues of this paper, let us consider a single-input and single-output discrete-time Hammerstein system with a finite impulse response (FIR) linear part, as illustrated in Fig. 1 where $x(t)$ and $y(t)$ are respectively the input and the noise-corrupted output of the Hammerstein system. The case of infinite impulse response linear part can be addressed similarly using the separable LS technique; the details are omitted due to space limitation.

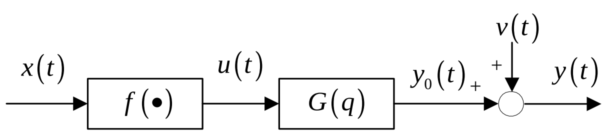

Fig. 1. A discrete-time Hammerstein model

More precisely, the FIR linear part $G(q)$ in Fig. 1 is parameterized by $b_1, \ldots, b_n$ such that

$$y(t) = \sum_{i=1}^{n} b_i u(t-i) + v(t). \tag{2}$$

The additive noise $v(t)$ is assumed to be white for the time being, whereas the case of colored noise is discussed later in Section 3.3 and Example 2. The static nonlinearity $f(\cdot)$ is assumed to be a linear combination of known basis functions $\phi_i(\cdot)$,

$$u(t) = \sum_{j=1}^{m} a_j \phi_j\left(x(t)\right). \tag{3}$$

Here the orders n and m are assumed to be known *a prior*. Substituting $u(t)$ in (3) into (2) yields a linear regression model parameterized in a bilinear form,

$$y(t) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_i a_j \phi_j\left(x(t-i)\right) + v\left(t\right). \tag{4}$$

In order to more clearly show that (4) is in the bilinear form (1), rewrite (4) as

$$\begin{aligned} y\left(t\right) = & \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}^T \begin{bmatrix} \phi_1\left(x\left(t-1\right)\right) & \cdots & \phi_m\left(x\left(t-1\right)\right) \\ \phi_1\left(x\left(t-2\right)\right) & \cdots & \phi_m\left(x\left(t-2\right)\right) \\ \vdots & \ddots & \vdots \\ \phi_1\left(x\left(t-n\right)\right) & \cdots & \phi_m\left(x\left(t-n\right)\right) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \\ & + v\left(t\right) \\ =: & \; b^T \phi\left(t\right) a + v\left(t\right). \end{aligned} \tag{5}$$

Let us introduce a notation useful for a different formulation of the bilinearly parametrized linear regression. For any matrix $M = \left[\, M_1 \; M_2 \; \cdots \; M_m \,\right]$, the overlined notation $\overline{M}$ denotes the column vector obtained by stacking the columns $M_1, M_2, ..., M_m$, namely,

$$\overline{M} = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_m \end{bmatrix}.$$

Hence the overlined notation $\overline{M}$ indicates a vectorization of the matrix M. With this notation, (5) can also be written as

$$\begin{aligned} y(t) = & \begin{bmatrix} \phi_1\left(x\left(t-1\right)\right) \\ \phi_1\left(x\left(t-2\right)\right) \\ \vdots \\ \phi_1\left(x\left(t-n\right)\right) \\ \vdots \\ \phi_m\left(x\left(t-1\right)\right) \\ \phi_m\left(x\left(t-2\right)\right) \\ \vdots \\ \phi_m\left(x\left(t-n\right)\right) \end{bmatrix}^T \begin{bmatrix} b_1 a_1 \\ b_2 a_1 \\ \vdots \\ b_n a_1 \\ \vdots \\ b_1 a_m \\ b_2 a_m \\ \vdots \\ b_n a_m \end{bmatrix} + v\left(t\right) \\ = & \; \overline{\phi\left(t\right)}^T \overline{ba^T} + v\left(t\right). \end{aligned} \tag{6}$$

For an input-output data set $\left\{x\left(t\right), y\left(t\right)\right\}_{t=1}^{N}$, define

$$Y = \begin{bmatrix} y\left(1\right) \\ y\left(2\right) \\ \vdots \\ y\left(N\right) \end{bmatrix}, \Phi = \begin{bmatrix} \overline{\phi\left(1\right)}^T \\ \overline{\phi\left(2\right)}^T \\ \vdots \\ \overline{\phi\left(N\right)}^T \end{bmatrix}, V = \begin{bmatrix} v\left(1\right) \\ v\left(2\right) \\ \vdots \\ v\left(N\right) \end{bmatrix}, \tag{7}$$

then (6) leads to

$$Y = \Phi\overline{ba^T} + V. \tag{8}$$

It is clear that these new notations are such that $Y \in \mathbb{R}^N$, $\Phi \in \mathbb{R}^{N\times l}$ and $V \in \mathbb{R}^N$ with $l = mn$.

Now let us introduce a standard assumption to remove the scale ambiguity between a and b and to make the parameterization unique.

Assumption 1. The first non-zero entry of b is positive, and

$$\|b\|_2^2 = b^T b = 1. \tag{9}$$

With the above formulation, the identification of the Hammerstein system amounts to the estimation of the bilinear parameters a and b from the data set $\left\{x\left(t\right), y\left(t\right)\right\}_{t=1}^{N}$.

3. THE TSA AND OPTIMALITY RESULTS

In this section we analyze some optimality properties of the TSA after a recall of the algorithm.

3.1 Weighted nonlinear LS criteria and the TSA

For the estimation of the bilinear parameters a and b in (8), a traditional approach is to consider the weighting nonlinear LS criterion

$$\min_{b,a} L(b,a,W) = \min_{b,a} \frac{1}{2}\left(\Phi\overline{ba^T} - Y\right)^T W \left(\Phi\overline{ba^T} - Y\right), \tag{10}$$

where $W \in \mathbb{R}^{N\times N}$ is some chosen symmetric weighting matrix. To ensure the uniqueness of its solution, this LS problem should be solved under the constraints stated in Assumption 1. Since the TSA described later in this section provides a solution naturally satisfying these constraints, we will not always explicitly mention these constraints when the nonlinear LS problem (10) is referred to. The following Assumption 2 is also for ensuring the uniqueness of this nonlinear LS solution. Remark that this assumption implies that Φ has full column rank, a condition related to the excitation property of the system input.

Assumption 2. The matrices $\Phi \in \mathbb{R}^{N\times l}$ and $W \in \mathbb{R}^{N\times N}$ are such that $\Phi^T W \Phi$ has full rank.

Let us slightly reformulate the nonlinear LS problem (10) in order to make the connection with the TSA. By introducing the notation

$$\Theta := ba^T,$$

the LS problem (10) is equivalent to the following one with a rank constraint,

$$\min_{\Theta} L(\Theta,W) = \min_{\Theta} \frac{1}{2}\left(\Phi\overline{\Theta} - Y\right)^T W \left(\Phi\overline{\Theta} - Y\right) \tag{11a}$$

$$\text{s.t. } \operatorname{rank}(\Theta) = 1. \tag{11b}$$

Some constraints equivalent to those of Assumption 1 should also be added for a unique solution of this LS problem.

Remark that now in (11) the single parameter vector $\theta := \overline{\Theta}$ shows up quadratically in the weighted LS problem; however, the associated matrix $\Theta \in \mathbb{R}^{n\times m}$ has a rank constraint, which is usually difficult to be taken into account in optimization problems. In this paper we are interested in a solution known as the TSA. The following description of the TSA closely follows that in [2].

The TSA:

(1) Choose a weighting matrix $W \in \mathbb{R}^{N\times N}$ and use it to estimate the parameter vector $\theta = \overline{\Theta}$ through the solution of the *unconstrained* weighted linear LS problem (11a), namely,

$$\hat{\theta}(W) = \left(\Phi^T W \Phi\right)^{-1} \Phi^T W Y. \tag{12}$$

(2) Build the matrix $\hat{\Theta}(W) \in \mathbb{R}^{n\times m}$ from the vector $\hat{\theta}(W) \in \mathbb{R}^l$ such that

$$\overline{\hat{\Theta}(W)} = \hat{\theta}(W).$$

Let σ_1 be the largest singular value of this matrix, and u_1 and v_1 be its left and right singular vectors associated with σ_1, then the bilinear parameters a and b are estimated by

$$\hat{\hat{a}}(W) = s_1 u_1, \tag{13}$$

$$\hat{\hat{b}}(W) = s_1 \sigma_1 v_1, \tag{14}$$

where $s_1 = \pm 1$ is the sign of the first non-zero entry of u_1.

Notes on Notations.

- The solution of the nonlinear LS problem (10) under Assumptions 1 and 2 will be noted as $\hat{a}(W)$ and $\hat{b}(W)$.
- The solution Θ of the *unconstrained* weighted linear LS problem (11a), vectorized as $\theta = \overline{\Theta}$, will be noted as $\hat{\theta}(W)$.
- The result of the TSA will be noted as $\hat{\hat{a}}(W)$ and $\hat{\hat{b}}(W)$.

In these notations the weighting matrix W is explicitly indicated, as its different choices will take an important role in this paper.

The TSA is based on a relaxation approach. It first solves the unconstrained LS problem (11a) by omitting the rank constraint (11b) and then subsequently projects the solution onto the class of bilinear equations via the singular-value decomposition (SVD). By omitting the rank constraint (11b), the solution set in (12) becomes broader than that of the optimization problem (11). For instance, as noticed by Goethals *et al.* [8], if $\theta_{i,j} := b_i a_j$ is the solution of (11a), so is $\theta'_{i,j} := b_i a_j + \beta_i \alpha_j$ for any set of variables α_j, $j = 1,\cdots,m$ such that $\sum_{j=1}^m \alpha_j \phi_j(x(t)) =$ constant, $\forall t \in \mathbb{R}$, and any set of variables β_i, $i = 1,\cdots,n$ such that $\sum_{i=1}^n \beta_i = 0$. However, the matrix

$$\begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \cdots & \theta_{1,m} \\ \theta_{2,1} & \theta_{2,2} & \cdots & \theta_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n,1} & \theta_{n,2} & \cdots & \theta_{n,m} \end{bmatrix}$$

with $\theta_{i,j} := b_i a_j$ satisfies the rank constraint (11b), while the corresponding matrix consisting of $\theta'_{i,j}$ does not. Though in principle the TSA can provide the estimates of a and b for any weighting matrix W, it is not obvious if these estimates are the solution of the original LS problem (10) or if they minimize another error criterion. Hence, the significance of the solution given by the TSA is the main question this paper attempts to answer.

This question has already been partly answered by Theorem 2.2 of [2] which is restated as follows.

Theorem 1. If the weighting matrix W, used *both* in the nonlinear LS problem (10) and in the TSA, satisfies

$$\Phi^T W \Phi = \alpha I_l, \tag{15}$$

where α is any positive scalar and I_l is the $l \times l$ identity matrix, then the TSA produces the optimal solution of the nonlinear LS problem (10), or more precisely (using notations introduced at the end of Section 3.1),

$$\hat{a}(W) = \hat{\hat{a}}(W)$$

$$\hat{b}(W) = \hat{\hat{b}}(W).$$

3.2 The Unweighted TSA

Theorem 1 answers the optimality question about the TSA only in the case of a special class of weighting matrices W satisfying (15). For such special weighting matrices, the TSA is thus perfectly justified. However, in other works the TSA is typically applied with the unweighted LS solution in its first stage, corresponding to the use of the identity matrix for the weighting matrix. It is clear that usually the identity weighting matrix does not satisfy condition (15). There thus appears to be a contradiction between the established optimality of the algorithm and the usual practice. The purpose of this subsection is to remove this "contradiction" by justifying the use of the unweighted LS solution in the TSA.

For the ease of presentation, the TSA using the unweighted LS solution in its first stage will be referred to as the Unweighted TSA in the sequel. Hence, the first stage of the algorithm in this case consists in computing the unweighted LS solution

$$\hat{\theta}(I_N) = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y, \tag{16}$$

while the second stage is the same as (13) and (14).

The use of the unweighted LS solution is well in line with the assumed additive white noise of the considered system. As a matter of fact, the use of any non diagonal weighting matrix would color the noise, which is usually not a desired effect. Though each stage of the Unweighted TSA is optimal for its own purpose, it is not obvious if the overall algorithm results in a solution which is optimal in any sense. The following result provides an answer to this question.

Theorem 2. The Unweighted TSA produces the optimal solution of the nonlinear LS problem (10), if the weighting matrix W of this latter is chosen to be

$$\check{W} = \alpha(\Phi\Phi^T)^+ \tag{17}$$

where α is any positive scalar and $(\Phi\Phi^T)^+$ is the Moore-Penrose pseudo-inverse of the matrix $(\Phi\Phi^T)$. In other words, for such a special $\check{W}$,

$$\hat{\hat{a}}(I_N) = \hat{a}(\check{W})$$
$$\hat{\hat{b}}(I_N) = \hat{b}(\check{W})$$

where the hatted notations have been defined in Notes on Notations at Section 3.1. □

The proof of this theorem will need the following classical result which can be found in textbooks on matrices, for instance in [4].

Lemma 1. For any two matrices $A \in \mathbb{R}^{m\times l}$ and $B \in \mathbb{R}^{l\times n}$, both of rank l, the equality

$$(AB)^+ = B^T(BB^T)^{-1}(A^TA)^{-1}A^T \tag{18}$$

holds, where $(AB)^+$ is the Moore-Penrose pseudo-inverse of the matrix (AB).

Proof of Theorem 2. Assumption 2 implies that Φ has full column rank. The application of Lemma 1 with $A = \Phi$ and $B = \Phi^T$ then leads to

$$\Phi^T\check{W}\Phi = \Phi^T(\alpha(\Phi\Phi^T)^+)\Phi \tag{19}$$
$$= \Phi^T(\alpha\Phi(\Phi^T\Phi)^{-1}(\Phi^T\Phi)^{-1}\Phi^T)\Phi$$
$$= \alpha I_l. \tag{20}$$

According the Theorem 1, when such a weighting matrix $\check{W}$ is used in the TSA, it produces the optimal solution of the nonlinear LS problem (10) defined with the same weighting matrix $\check{W}$, or more formally,

$$\hat{a}(\check{W}) = \hat{\hat{a}}(\check{W}) \tag{21a}$$
$$\hat{b}(\check{W}) = \hat{\hat{b}}(\check{W}). \tag{21b}$$

Now let us examine the first stage of the TSA. When it uses the weighting matrix $\check{W} = \alpha(\Phi\Phi^T)^+$, the unconstrained LS solution is, following (12) and then (20),

$$\hat{\theta}(\check{W}) = (\Phi^T\check{W}\Phi)^{-1}\Phi^T\check{W}Y$$
$$= (\alpha I_l)^{-1}\Phi^T(\alpha(\Phi\Phi^T)^+)Y$$

Again apply Lemma 1 with $A = \Phi$ and $B = \Phi^T$, then

$$\hat{\theta}(\check{W}) = (\alpha I_l)^{-1}\Phi^T(\alpha\Phi(\Phi^T\Phi)^{-1}(\Phi^T\Phi)^{-1}\Phi^T)Y$$
$$= (\Phi^T\Phi)^{-1}\Phi^T Y.$$

The last result is nothing but the unweighted LS solution $\hat{\theta}(I_N)$ as defined in equation (16). It then follows that $\hat{\theta}(\check{W}) = \hat{\theta}(I_N)$. This result indicates that the first stage of the TSA produces the same result with the weighting matrix $\check{W}$ or I_N. As the second stage does not involve the weighting matrix, the TSA produces the same result with these two weighting matrices, or more formally,

$$\hat{\hat{a}}(I_N) = \hat{\hat{a}}(\check{W})$$
$$\hat{\hat{b}}(I_N) = \hat{\hat{b}}(\check{W}).$$

This result and (21) then lead to

$$\hat{\hat{a}}(I_N) = \hat{a}(\check{W})$$
$$\hat{\hat{b}}(I_N) = \hat{b}(\check{W}),$$

which complete the proof. □

The above result shows that using the unweighted LS solution in the TSA does solves the nonlinear LS problem (10), provided this nonlinear LS problem is formulated with a particular weighting matrix $\check{W}$ which is different from the identity matrix. It is thus necessary to distinguish the weighting matrix used in the formulation of the nonlinear LS problem (10) from the one used in the first stage of the TSA. This remark will be particularly important for the colored noise case considered in the next section.

3.3 Dealing with colored noises

If the additive noise $v(t)$ in (2) is colored instead of being white, the Unweighted TSA still yields unbiased and consistent estimates, because the regressor $\overline{\phi(t)}$ in (6) is uncorrelated with $v(t)$ for an FIR transfer funtion $G(q)$. However, the estimates are less efficient. In other words, there is a choice of the weighting matrix W for the TSA leading to lower estimation uncertainties.

In order to better deal with colored noises, as remarked previously, it is important to distinguish two weighting matrices, one is used in the first stage of the TSA, the other in the weighted nonlinear LS problem (10). As a matter of fact, the TSA using an arbitrary weighting matrix in its first stage (12) solves the weighted nonlinear LS problem (10) which is formulated with another weighting matrix, as stated in the following theorem.

Theorem 3. The TSA using any symmetric positive definite weighting matrix $W \in \mathbb{R}^{N\times N}$ in its first stage (12) produces the optimal solution of the nonlinear LS problem (10), if in this latter the weighting matrix W is replaced by

$$\tilde{W} = \alpha W^{\frac{1}{2}}(W^{\frac{1}{2}}\Phi\Phi^T W^{\frac{1}{2}})^{+}W^{\frac{1}{2}}.$$

Here α is any positive scalar, $W^{\frac{1}{2}}$ is the symmetric matrix square root of W, and M^{+} denotes the Moore-Penrose pseudo-inverse of the matrix M. More formally, for such a special $\tilde{W}$,

$$\hat{\hat{a}}(W) = \hat{a}(\tilde{W})$$
$$\hat{\hat{b}}(W) = \hat{b}(\tilde{W}).$$

See again Notes on Notations at Section 3.1 for the hatted notations.

Proof of Theorem 3. The proof of this result is very similar to that of Theorem 2, the only notable difference is that now Lemma 1 is applied with $A = W^{\frac{1}{2}}\Phi$ and $B = \Phi^T W^{\frac{1}{2}}$. Hence here the proof is omitted owing to space limitation. □

For colored noises, the optimal choice of W in the TSA is still unknown to our knowledge. However, the first stage of the TSA ignores the rank constraint (11b) and considers the weighted LS problem (11a) only, which suggests that W could be chosen as the inverse of the noise covariance matrix, according to the principle of the best linear unbiased estimate (BLUE) (see e.g., [16]). That is, we choose W to be the inverse of the noise covariance matrix R of v in (8) which writes, for the finite data sample $\{x(t), y(t)\}_{t=1}^{N}$,

$$R := E\left\{VV^T\right\},$$

where V is defined in (7).

If the TSA takes $W = R^{-1}$, Theorem 3 says that the resulting estimates are the solution of the weighted nonlinear LS problem (10) where W is replaced by

$$\tilde{W} = R^{-\frac{1}{2}}(R^{-\frac{1}{2}}\Phi\Phi^T R^{-\frac{1}{2}})^{+}R^{-\frac{1}{2}}.$$

In practice, the noise $v\,(t)$ and the noise covariance matrix R are normally unknown and have to be estimated. An efficient method for estimating such an inverse noise covariance matrix was recently given by David & Bastin [7], where $v(t)$ is treated as an auto-regressive (AR) process, and the inversion of the large noise covariance matrix R is avoided by using the AR model of the noise. When this approach is adopted in the first stage of the TSA, an unweighted LS estimation $\hat{\theta}(I_N)$ is first estimated, which is used to estimate the noise $v(t)$. Next, the estimated noise sequence is modeled as an AR process in order to compute the first estimate of the noise covariance inverse R_1^{-1}, which is used to compute the weighted LS solution $\hat{\theta}(R_1^{-1})$. A new noise sequence is then estimated with the new model, resulting in a new estimate of the noise covariance inverse R_2^{-1}, and so on. See [7] for more details. The effectiveness of the TSA using this method for noise covariance estimation will be demonstrated in Example 2.

4. EXAMPLES

This section presents two numerical examples to verify the theoretical analysis in Section 3.

Example 1. For the Hammerstein model in Fig. 1,

$$\begin{aligned} y\,(t) &= G\,(q)\,u\,(t) + v\,(t) \\ &= \left(0.4472q^{-1} - 0.8944q^{-2}\right)u\,(t) + v\,(t)\,, \\ u\,(t) &= x\,(t) + 2x^2(t) + 5x^3(t) + 7x^4(t) + x^5\,(t)\,. \end{aligned}$$

The noise $v\,(t)$ is a zero-mean white Gaussian noise with variance σ_v^2. Here $b = \left[\,0.4472 \;\; -0.8944\,\right]^T$ satisfies Assumption 1 so that the parameterization of a and b is unique. The input $x\,(t)$ is generated by passing a fixed realization of uniformly-distributed processes with magnitude range $[-3, 3]$ through the filter $1/(1-0.5q^{-1})$, so that $x\,(t)$ covers sufficient nonlinear ranges of the input nonlinearity. 100 Monte Carlo simulations are implemented, where each simulation takes a different realization of $v\,(t)$ [2].

Two groups of estimates of $a = \left[\,1\;2\;5\;7\;1\,\right]^T$ and $b = \left[\,0.4472\;-0.8944\,\right]^T$ are obtained by the TSA using $\check{W}$ in (17) and by the Unweighted TSA. Another group of estimates is obtained by directly solving the nonlinear weighted LS problem (10) with $\check{W}$ in (17) using the Matlab function "lsqnonlin". The nonlinear LS problem is sensitive to initial estimates; hence the true parameters a and b are used to initiate "lsqnonlin" for the purpose of comparison. Table 1 list the mean and the standard deviations of the estimates in 100 Monte Carlo simulations, for the noise level $\sigma_v^2 = 1$. The means of the associated loss functions $L\,(b, a, W)$ defined in (10) are listed as well at the bottom row in Table 1. In each simulation, 100 data points of $x\,(t)$ and $y\,(t)$ are collected, i.e., $N = 100$.

Table 1. Estimates of a and b and associated loss function $L\,(b, a, W)$ in the bottom row

$W = I$ TSA	$W = \check{W}$ as in (17) TSA	$W = \check{W}$ as in (17) lsqnonlin
$\begin{bmatrix} 0.4473 \pm 0.0013 \\ -0.8944 \pm 0.0006 \end{bmatrix}$	$\begin{bmatrix} 0.4473 \pm 0.0013 \\ -0.8944 \pm 0.0006 \end{bmatrix}$	$\begin{bmatrix} 0.4473 \pm 0.0015 \\ -0.8944 \pm 0.0019 \end{bmatrix}$
$\begin{bmatrix} 0.9895 \pm 0.1599 \\ 1.9932 \pm 0.0777 \\ 5.0016 \pm 0.0477 \\ 7.0004 \pm 0.0058 \\ 0.9998 \pm 0.0027 \end{bmatrix}$	$\begin{bmatrix} 0.9895 \pm 0.1599 \\ 1.9932 \pm 0.0777 \\ 5.0016 \pm 0.0477 \\ 7.0004 \pm 0.0058 \\ 0.9998 \pm 0.0027 \end{bmatrix}$	$\begin{bmatrix} 0.9897 \pm 0.1616 \\ 1.9931 \pm 0.0781 \\ 5.0014 \pm 0.0383 \\ 7.0002 \pm 0.0128 \\ 0.9997 \pm 0.0045 \end{bmatrix}$
223.2076	0.0302	0.0302

In Table 1, the two groups of estimates from the TSA using $W = I$ and from that using $W = \check{W}$ as in (17) are exactly the same, which is consistent with Theorem 2. The estimates are also unbiased and consistent, as proved by Theorem 2.1 in [2]. Since "lsqnonlin" takes the luxury of using the true parameters a and b as the initial estimates, the global optima are expected to reach. The resulted estimates are close to the counterparts from the TSA using $W = \check{W}$ as in (17); in fact, the corresponding loss functions $L\,(b, a, W)$ in the last two columns are exactly the same, which validates the conclusion in Theorem 1. On the other hand, $L\,(b, a, W)$ for $W = I$ is much larger than the rest two, which says that the Unweighted TSA does not find the solution of the *unweighted* nonlinear LS optimization problem, namely, (10) with $W = I$. Instead, it provides the solution of the *weighted* nonlinear LS optimization problem (10) with $W = \check{W}$ as in (17), as indicated by the fact that the two groups of estimates from the TSA using $W = I$ and from that using $W = \check{W}$ as in (17) are exactly the same. This is consistent with Theorem 2.

[2] The uniform random generator state in Matlab for $x(t)$ is fixed to 1, while the Gaussian random generator states for $v(t)$ are fixed to $[1, 2, \cdots, 100]$, so that the simulation results can be reproduced.

Example 2. All the configurations are the same as Example 1 except that the noise $v(t)$ is generated as

$$v(t) = \frac{1}{1 - 0.9q^{-1}} e(t),$$

where $e(t)$ is a zero-mean white Gaussian noise with variance σ_e^2. That is, $v(t)$ is colored instead of being white.

Table 2 listed the estimates of a and b obtained by the TSA using estimated inverse covariance matrix, as proposed in Section 3.3. In particular, the order of the noise AR model of $v(t)$ is selected according to the FPE criterion based on F statistical test with 95% confidence level (e.g., see Eq.(4.30) in [16]); in the 100 Monte Carlo simulations, the order 1 is correctly selected for 94 times, and the order 2 is selected for the rest. Some improvements are observed in the first 3 or 4 iterations between the estimation of R^{-1} and that of a and b. As a comparison, another group of estimates of a and b is obtained by the Unweighted TSA. In Table 2, the two groups of estimates appear both unbiased and consistent, but the estimates of a and b associated with $W = \hat{R}^{-1}$ have much smaller uncertainties, in terms of standard deviations, than those obtained by the Unweighted TSA. This observation is in favor of choosing W equal the inverse of the noise covariance matrix.

Table 2. Estimates of a and b for colored noises

$\sigma_e^2 = 1$	$W = I$	$W = \hat{R}^{-1}$
$b = \begin{bmatrix} 0.4472 \\ -0.8944 \end{bmatrix}$	$\begin{bmatrix} 0.4477 \pm 0.0043 \\ -0.8942 \pm 0.0022 \end{bmatrix}$	$\begin{bmatrix} 0.4474 \pm 0.0017 \\ -0.8943 \pm 0.0009 \end{bmatrix}$
$a = \begin{bmatrix} 1.0000 \\ 2.0000 \\ 5.0000 \\ 7.0000 \\ 1.0000 \end{bmatrix}$	$\begin{bmatrix} 0.9771 \pm 0.1911 \\ 1.9720 \pm 0.1232 \\ 4.9980 \pm 0.0601 \\ 7.0015 \pm 0.0091 \\ 0.9999 \pm 0.0034 \end{bmatrix}$	$\begin{bmatrix} 0.9948 \pm 0.1085 \\ 1.9942 \pm 0.0577 \\ 4.9981 \pm 0.0304 \\ 7.0002 \pm 0.0040 \\ 1.0000 \pm 0.0017 \end{bmatrix}$

5. CONCLUSION

This paper revisited the Two-Stage Algorithm (TSA) proposed by Bai [2] for identification of Hammerstein systems, and obtained some optimality results of the TSA. The motivation arises from a somewhat contradictory fact: though the optimality of the TSA has been established in [2] (restated here in Theorem 1) for a special group of weighting matrices satisfying an equality in (17), the TSA, adopted in some other works, usually uses the unweighted LS solution in its first stage. That is, the identity matrix, which usually does not belong to the above special weighting matrices, is used. Theorem 2 found out that the Unweighted TSA indeed gives the solution of (10) or its equivalence (11) for a particular weighting matrix $\check{W}$ in (17). This provides a theoretical justification of using the Unweighted TSA, and leads to Theorem 3 where a generalization is made to the case of colored noise with noise whitening. Finally, these theoretical results were validated via two simulation examples of identification of Hammerstein systems.

REFERENCES

[1] R. Abrahamsson, S.M. Kay, & P. Stoica (2007). Estimation of the parameters of a bilinear model with applications to submarine detection and system identification. *Digital Signal Processing*, vol.17, pp.756-773.

[2] E.W. Bai (1998). An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica*, vol.34, no.3, pp.333-338.

[3] E.W. Bai, & Y. Liu (2006). Least squares solutions of bilinear equations. *Systems & Control Letters*, vol.55, pp.466-472.

[4] S. Barnet (1990). *Matrices Methods and Applications*, Oxford Applied Mathematics and Computin Science Series, Clarendon Press.

[5] F. Chang, & R. Luus (1971). A noniterative method for identification using Hammerstein model. *IEEE Trans. Automat. Contr.*, vol.16, no.5, pp.464-468.

[6] S. Cohen, & C. Tomasi (1997). Systems of bilinear equations. *Technical report*, Computer Science Dept., Stanford Univ., Standard, CA.

[7] B. David, & G. Bastin (2001). An estimator of the inverse covariance matrix and its application to ML parameter estimation in dynamical systems. *Automatica*, vol.37, no.1, pp.99-106.

[8] I. Goethals, K. Pelckmans, J.A.K. Suykens, & B. De Moor (2005). Identification of MIMO Hammerstein models using least squares support vector machine. *Automatica*, vol.41, pp.1263-1272.

[9] J.C. Gomez, & E. Baeyens (2004). Identification of block-oriented nonlinear systems using orthonormal bases. *J. Process Control*, vol.14, no.6, pp.685-697.

[10] J.C. Gomez, & E. Baeyens (2005). Subspace-based identification algorithms for Hammerstein and Wiener models. *Eur. J. Control*, vol.11, pp.127-136.

[11] R. Haber, & L. Keviczky (1999). *Nonlinear System Identification: Input-Output Modeling Approach*, Dordrecht: Kluwer Academic Publishers.

[12] A. Janczak (2005). *Identification of Nonlinear Systems Using Neural Networks and Polynomial Models: A Block-Oriented Approach*, New York: Springer-Verlag.

[13] T. McKelvey, & C. Hanner (2003). On identification of Hammerstein systems using excitation with a finite number of levels. *Proc. of the 13th International Symp. on System Identification (SYSID2003)*, pp.57-60.

[14] G. Mzyk (2000). Application of instrumental variable method to the identification of Hammerstein-Wiener systems. *6th International Conf. MMAR*, Miedzyzdroje.

[15] H.J. Palanthandalam-Madapusi, D.S. Bernstein, & A.J. Ridley (2006). Subspace identification of periodically switching Hammerstein-Wiener models for magnetospheric dynamics. *14th IFAC Symp. on System Identification*, Newcastle, Australia, pp.535-540.

[16] T. Söderström, & P. Stoica (1989). *System Identification*, London: Prentice Hall.

Avdelning, Institution
Division, Department

Division of Automatic Control
Department of Electrical Engineering

Datum
Date

2009-05-13

Språk
Language

☐ Svenska/Swedish
☒ Engelska/English
☐ ________________

Rapporttyp
Report category

☐ Licentiatavhandling
☐ Examensarbete
☐ C-uppsats
☐ D-uppsats
☒ Övrig rapport
☐ ________________

URL för elektronisk version

`http://www.control.isy.liu.se`

ISBN
—

ISRN
—

Serietitel och serienummer
Title of series, numbering

ISSN
1400-3902

LiTH-ISY-R-2909

Titel Optimality analysis of the Two-Stage Algorithm for Hammerstein system identification
Title

Författare Qinghua Zhang, Jiandong Wang, Lennart Ljung
Author

Sammanfattning
Abstract

The Two-Stage Algorithm (TSA) has been extensively used and adapted for the identification of Hammerstein systems. It essentially deals with the parameter estimation problem for a linear regression parameterized in a bilinear form. This paper is motivated by a somewhat contradictory fact: though the optimality of the TSA has been established only in the case of some special weighting matrices, this algorithm usually uses the unweighted case, i.e., no weighting matrix or an identity one is used. We find out that the unweighted TSA indeed gives the optimal solution of the weighted nonlinear least-squares optimization problem for a particular weighting matrix. This provides a theoretical justification of the unweighted TSA, and leads to a generalization of the obtained result to the case of colored noise with noise whitening. Numerical examples of identification of Hammerstein systems are presented to validate the theoretical analysis.

Nyckelord
Keywords System identification, Hammerstein system, two-stage algorithm