Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Convex Relaxation Approach to the Identification of the Wiener-Hammerstein Model

Kin Cheong Sou, Alexandre Megretski and Luca Daniel

Abstract— **In this paper, an input/output system identification technique for the Wiener-Hammerstein model and its feedback extension is proposed. In the proposed framework, the identification of the nonlinearity is non-parametric. The identification problem can be formulated as a non-convex quadratic program (QP). A convex semidefinite programming (SDP) relaxation is then formulated and solved to obtain a sub-optimal solution to the original non-convex QP. The convex relaxation turns out to be tight in most cases. Combined with the use of local search, high quality solutions to the Wiener-Hammerstein identification can frequently be found. As an application example, randomly generated Wiener-Hammerstein models are identified.** [1]

I. Introduction

Classical treatments of the Wiener-Hammerstein system identification problem can be found, for example, in [1], [2], [3]. Many more recent treatments of the problem can be found, for example, in [4], [5], [6]. In those references, however, the identification of the nonlinearity is parametric (i.e. the nonlinearity is assumed to be of some form such as piecewise linear or polynomial functions). Therefore, those previous results can be restrictive in application. Non-parametric identification of block oriented models, on the other hand, are more flexible in terms of modeling power. Reference [7] proposed an algorithm for the non-parametric identification of the Wiener system under the assumption that the input is Gaussian noise. The authors of [8], assuming that the LTI block is known, reduced the identification problem of the Wiener system to a least squares problem. [9] proposed an unbiased identification algorithm based on maximum likelihood estimation.

In a sense, the idea of the system identification scheme proposed in this paper has been explored under the banner of model validation [10], [11], [12], [13], [14], [15], [16]. In this problem, a model with a given block diagram is to be invalidated by proving that it is inconsistent with some input/output measurement obtained from experiment. The invalidation is typically performed through the finding of some infeasibility certificate of some constraint set. Conversely, the finding of a feasibility certificate will prove the consistency of a model with the given input/output measurement data. This forms the basis of the block diagram oriented system identification schemes such as [17], [18], [19]. In particular, [19] proposed a very general approach for the identification of the Wiener system assuming only the monotonicity of the nonlinearity. [19] sets up a convex QP based on the idea of enforcing an input/output functional relationship of the nonlinearity. The algorithm proposed in this paper can be considered as an extension of the idea in [19]. In fact, the formulation of the optimization problem in this paper also centers around some sector bound property of the nonlinearity. However, because of the more complicated Wiener-Hammerstein structure, the resultant optimization problem is more involved. In fact, it is a non-convex QP. Nevertheless, with the proposed SDP relaxation, it will be demonstrated that the non-convex QP formulated in this paper is not necessarily hard to solve.

A. Feedback Wiener-Hammerstein system

In this paper, the unknown system in the input/output system identification problem is assumed to be from a specific class – either of the Wiener-Hammerstein form, or the Wiener-Hammerstein with feedback in Figure 1.

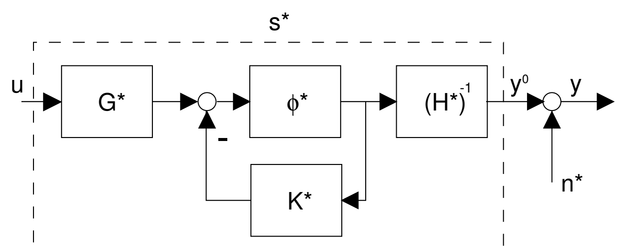

Fig. 1. The Wiener-Hammerstein system with feedback. S^* denotes the unknown system. $K \equiv 0$ corresponds to the Wiener-Hammerstein system without feedback. The output measurement y is assumed to be corrupted by some noise n^*.

The following assumptions are made in Figure 1.

1) The signals u, y, y^0 and n^* are causal and of finite length N.
2) G^*, H^* and K^* are assumed to be single-input-single-output (SISO) FIR systems. In addition, H^* and K^* are assumed to be positive-real passive. That is,
$$\begin{array}{l}\text{Re}\left\{H^*\left(e^{j\omega}\right)\right\} > 0, \quad \forall\,\omega \in [0, 2\pi) \\ \text{Re}\left\{K^*\left(e^{j\omega}\right)\right\} > 0, \quad \forall\,\omega \in [0, 2\pi)\end{array} \tag{1}$$
3) Nonlinearity ϕ^* is assumed to be scalar valued and memoryless, and it is assumed to satisfy a certain sector bound criterion in incremental sense. That is, there exists a scalar $0 < \beta < \infty$ such that for all $a, b \in \mathbb{R}$,
$$\Big(\phi^*(b) - \phi^*(a)\Big)\Big(\phi^*(b) - \phi^*(a) - \beta b + \beta a\Big) \leq 0. \tag{2}$$

[1]The first author is now affiliated with the Department of Automatic Control, Lund Institute of Technology, Sweden. The rest of the authors are affiliated with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. The research work was completed before the first author's transition from MIT to Lund.

Condition (2) means that the nonlinearity ϕ^* is monotonically non-decreasing and its derivative has an upper bound. Further details can be found in [20].

B. Organization of the paper

The rest of the paper is organized as follows: in The main ideas of the problem formulation and solution procedure will be explained in Section II and Section III respectively, through a special setup in which there is no output measurement noise or feedback. Then in Section IV the identification setup with output measurement noise is considered. Differences in the analysis and algorithm due to the noise will be highlighted. After that, the full feedback Wiener-Hammerstein system identification problem will be considered in Section V. Application examples will be presented in Section VI.

II. Identification of the Wiener-Hammerstein System – No Measurement Noise

The first problem to be considered in this paper is the identification of the Wiener-Hammerstein system without the feedback or the output measurement noise. The identification problem will be formulated as two equivalent optimization problems in Subsections II-A and II-C respectively. The solution technique for the optimization problems will be described in Section III.

A. System identification problem formulation

Problem data. The problem data is the input signal u and the output measurement signal y of the true (but unknown) system S^* in Figure 1. For ease of exposition, a signal will also be denoted as the vector of its non-zero values (e.g. vector $\mathbf{u}$ for the signal v).

System identification model and decision variables. It is natural to choose a model with the same structure as the true but unknown system (i.e. the Wiener-Hammerstein structure in Figure 2). In Figure 2 the G and H are FIR systems, and ϕ is a scalar memoryless nonlinearity (i.e. a nonlinear function). Obviously, the model is specified when G, H and ϕ are specified.

FIR systems G and H in Figure 2 are characterized by their impulse responses of length N_g and N_h respectively. That is,

$$\begin{aligned} \mathbf{g} &:= \begin{bmatrix} g_0 & g_1 & \dots & g_{N_g-1} \end{bmatrix}', \\ \mathbf{h} &:= \begin{bmatrix} h_0 & h_1 & \dots & h_{N_h-1} \end{bmatrix}'. \end{aligned} \tag{3}$$

The identification of the nonlinearity ϕ is non-parametric. That is, ϕ is specified only by some samples of its input/output pair. The values of ϕ other than those given by the samples can be obtained using an interpolation scheme (e.g. linear interpolation). In addition, the samples will be restricted to those computable by the FIR impulse response $\mathbf{g}$ and $\mathbf{h}$. Therefore, $\mathbf{g}$ and $\mathbf{h}$ are the decision variables sufficient to specify ϕ as well as the full model in Figure 2.

Treatment of the passivity constraint. A sufficient condition for the stability of the identified model is that the FIR system H in Figure 2 is positive real passive (see [21], Chapter 3). Ideally the positive real constraint should be enforced. However, it turns out to be inconsistent with the solution technique proposed. Therefore, in all subsequent sections the stability requirement will not be dealt with explicitly. In Subsection III-C this issue will be revisited, and a post-processing algorithm will be given to enforce the passivity of H (and hence the stable of the final model).

System identification problem formulation – a feasibility problem. Consider the Wiener-Hammerstein model in Figure 2 in which the output and the input are constrained to be the given data $(\mathbf{u}, \mathbf{y})$. Let's investigate the possible choices of the decision variables $\mathbf{g}$ and $\mathbf{h}$ so that there exist signals $\mathbf{v} \in \mathbb{R}^N$ and $\mathbf{w} \in \mathbb{R}^N$ with the property that $(\mathbf{u}, \mathbf{v})$, $(\mathbf{v}, \mathbf{w})$, $(\mathbf{y}, \mathbf{w})$ are valid input/output pairs of the blocks G, ϕ and H respectively.

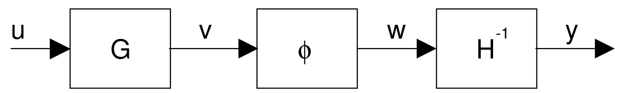

Fig. 2. A feasibility problem to determine the impulse responses of the FIR systems G and H. Here $\mathbf{u}$ and $\mathbf{y}$ are the given input and output measurements generated by the true (but unknown) system. The signals $\mathbf{v}$ and $\mathbf{w}$ are the outputs of G and H, respectively. $\mathbf{v}$ and $\mathbf{w}$ are chosen so that they define a function ϕ satisfying sector bound constraint eq. (6).

The pairs $(\mathbf{u}, \mathbf{v})$ and $(\mathbf{y}, \mathbf{w})$ satisfy the following convolution relationship.

$$\begin{aligned} \mathbf{v} &= \mathbf{U}\mathbf{g}, \\ \mathbf{w} &= \mathbf{Y}\mathbf{h}, \end{aligned} \tag{4}$$

where $\mathbf{U} \in \mathbb{R}^{N \times N_g}$ and $\mathbf{Y} \in \mathbb{R}^{N \times N_h}$ are defined as

$$\mathbf{U} := \begin{bmatrix} u[0] & 0 & \dots & 0 \\ u[1] & u[0] & \ddots & \vdots \\ & & \ddots & 0 \\ \vdots & \vdots & & u[0] \\ & & & \vdots \\ u[N-1] & u[N-2] & \dots & u[N-N_g] \end{bmatrix}_{N \times N_g}, \tag{5}$$

and $\mathbf{Y}$ is defined in a fashion analogous to eq. (5).

For the pair $(\mathbf{v}, \mathbf{w})$, in principle, the only constraint imposed is that there exists some function ϕ such that $\mathbf{w}_i = \phi(\mathbf{v}_i)$, $\forall\, i = 0, 1, \dots, N-1$. However, to maximally reduce the redundancy of the possible choices of $(\mathbf{v}, \mathbf{w})$, an additional constraint is enforced: ϕ should satisfy the sector bound of the form of eq. (2). That is,

$$(\phi(b) - \phi(a))\,(\phi(b) - \phi(a) - \beta b + \beta a) \le 0, \quad \forall\, a, b \in \mathbb{R}. \tag{6}$$

Constraint eq. (6) imposed on the function $\phi : \mathbb{R} \mapsto \mathbb{R}$ is equivalent to a constraint on the generating pair $(\mathbf{v}, \mathbf{w})$ as

$$(\mathbf{w}_i - \mathbf{w}_j)\,(\mathbf{w}_i - \mathbf{w}_j - \beta \mathbf{v}_i + \beta \mathbf{v}_j) \le 0, \quad \forall\, i > j. \tag{7}$$

The equivalence of eq. (6) and eq. (7) is shown in [20].

In summary, the Wiener-Hammerstein system identification problem in the noiseless case can be defined as

Definition 2.1: **[Wiener-Hammerstein system identification problem – noiseless case]**

Given the input/output measurement $(\mathbf{u}, \mathbf{y}) \in \mathbb{R}^N \times \mathbb{R}^N$ of an unknown Wiener-Hammerstein system and positive integers N_g and N_h, find decision vectors $\mathbf{g} \in \mathbb{R}^{N_g}$ and $\mathbf{h} \in \mathbb{R}^{N_h}$ such that there exist signals $\mathbf{v} \in \mathbb{R}^N$ and $\mathbf{w} \in \mathbb{R}^N$ satisfying eq. (4), eq. (7). ■

Typically there are infinitely many solutions of the problem in Definition 2.1, the corresponding normalization issue will be discussed in Subsection II-B.

Comparison with the model validation techniques. The principles of the identification problem in Definition 2.1 and that of the problem of model validation (e.g. [10]) are very similar. Both problems call for satisfiability certificate of the input/output relationships of the blocks in the model structures. Definition 2.1 seeks a feasibility certificate while model validation seeks an infeasibility certificate. However, there are two major distinctions between the proposed identification setup and the model validation setup. First, for the model validation problem, proving the *existence* of the infeasibility certificate is sufficient. For example, in [10], [15] the question of whether an infeasibility certificate exists is answered by a structured singular value bounding problem. The Wiener-Hammerstein identification problem in Definition 2.1, on the other hand, requires the computation of all signals presented in the model. This computation can potentially be expensive. The second distinction of the proposed identification setup from the model validation setup is that the feasibility problem in Definition 2.1 will lead to a *non-convex* quadratic program, while most of the previously considered model validation setups lead to the formulation of convex problems. The convexity properties of the optimization problems also lead to a distinction in the solution approaches. The published model validation results are mostly based on rigorous analysis, while the approach adopted in this paper will be more experimental – some observations will be substantiated by numerical experiments only.

B. Non-uniqueness of solutions and normalization

The system identification problem in Definition 2.1 is feasible with decision vectors $\mathbf{g}^*$ and $\mathbf{h}^*$ (i.e. the impulse responses of the FIR systems in Figure 1). However, there are actually infinitely many solutions. It can be verified that for an infinite set of choices of $c_1 \neq 1$ and $c_2 \neq 1$, the impulse responses $\mathbf{g} = \mathbf{g}^*/c_1$ and $\mathbf{h} = \mathbf{h}^*/c_2$ are also solutions of the problem in Definition 2.1. The non-uniqueness of solutions requires the normalization of $\mathbf{g}$ and $\mathbf{h}$, and the details of the normalization can be found in [20]. Here only the main observations are summarized.

- Normalization of both $\mathbf{g}$ and $\mathbf{h}$ will generally lead to excessive restriction. Therefore, only $\mathbf{h}$ will be normalized in this paper. The particular choice of normalization will be assumed:
$$\mathbf{h}_0 \equiv 1. \tag{8}$$
While the choice of normalization in eq. (8) is somewhat arbitrary, it is not unjustified because $\mathbf{h}_0 = \int_0^{2\pi} \mathrm{Re}\left\{H\left(e^{j\omega}\right)\right\} d\omega > 0.$
- With the normalization, the constant β in sector bound (7) can always be assumed to be one, otherwise it can be absorbed in the part of the decision vector which is not normalized. Therefore, throughout this paper, all sector bound constraints assume values of $\beta = 1$.

C. Formulation of the system ID optimization problem

In this subsection the system identification problem defined in Definition 2.1 will be simplified and put in a format that would facilitate the study of its solution strategy. Some properties of the optimization problem will also be discussed in Subsection II-D.

Definition 2.1 defines a system identification feasibility problem with three constraints given in eq. (4) and eq. (7). The discussion in Subsection II-B concludes that a partial normalization of $\mathbf{h}$ (i.e. eq. (8)) can be assumed. In addition, with the partial normalization, β in eq. (7) can be assumed to be one. Substituting the variables $\mathbf{v}$ and $\mathbf{w}$ using eq. (4), the constraint set eq. (4) and eq. (7) reduces to

$$(\Delta\mathbf{Y}_{ij}\mathbf{h})^2 - (\Delta\mathbf{Y}_{ij}\mathbf{h})(\Delta\mathbf{U}_{ij}\mathbf{g}) \le 0, \quad \forall\, i > j, \tag{9}$$

where

$$\begin{aligned} \Delta\mathbf{U}_{ij} &:= \mathbf{U}_i - \mathbf{U}_j, \\ \Delta\mathbf{Y}_{ij} &:= \mathbf{Y}_i - \mathbf{Y}_j, \end{aligned} \tag{10}$$

and

$$\begin{aligned} \mathbf{U}_i \in \mathbb{R}^{1\times N_g}, &\quad \mathbf{U}_i := \left[\begin{array}{ccc} \mathbf{U}(i,1) & \cdots & \mathbf{U}(i,N_g) \end{array}\right], \\ \mathbf{Y}_i \in \mathbb{R}^{1\times N_h}, &\quad \mathbf{Y}_i := \left[\begin{array}{ccc} \mathbf{Y}(i,1) & \cdots & \mathbf{Y}(i,N_h) \end{array}\right], \end{aligned}$$

with $\mathbf{U}$ and $\mathbf{Y}$ defined in eq. (5).

Conforming to the standard notation in the field of optimization, define the vector of decision variables $x \in \mathbb{R}^{N_g+N_h}$ as

$$x := \left[\begin{array}{c} \mathbf{g} \\ \mathbf{h} \end{array}\right], \tag{11}$$

then corresponding to eq. (8), the partial normalization constraint set will be denoted as

$$\mathcal{X} := \left\{ x = \left[\begin{array}{c} \mathbf{g} \\ \mathbf{h} \end{array}\right] \in \mathbb{R}^{N_g+N_h} \;\middle|\; \mathbf{h}_0 = 1 \right\}. \tag{12}$$

In addition, define matrices $A_{ij} \in \mathbb{R}^{(N_g+N_h)\times(N_g+N_h)}$ as

$$A_{ij} := \left[\begin{array}{cc} (\Delta\mathbf{Y}_{ij})'(\Delta\mathbf{Y}_{ij}) & -\frac{1}{2}(\Delta\mathbf{Y}_{ij})'(\Delta\mathbf{U}_{ij}) \\ -\frac{1}{2}(\Delta\mathbf{U}_{ij})'(\Delta\mathbf{Y}_{ij}) & 0 \end{array}\right]. \tag{13}$$

Then eq. (9) is the same as

$$x' A_{ij} x \le 0, \quad \forall\, N-1 \ge i > j \ge 0. \tag{14}$$

Using the notation A_{ij} defined in eq. (13), the system identification optimization problem can be formulated as follows.

$$\begin{array}{ll} \underset{x\in\mathcal{X}, r\in\mathbb{R}}{\text{minimize}} & r \\ \text{subject to} & x' A_{ij} x \le r, \quad \forall\, i > j \\ & r \ge 0, \end{array} \tag{15}$$

where $\mathcal{X}$ is defined in eq. (12) and A_{ij} are defined in eq. (13). Program (15) and the feasibility problem in Definition 2.1 are equivalent in the following sense: $\hat{x}$ is an optimal of program

(15) if and only if the corresponding $\hat{\mathbf{g}}$ and $\hat{\mathbf{h}}$ (see eq. (11)) is a feasible solution of the problem in Definition 2.1. The equivalence can be explained in the following schematics (with $\hat{x}$ and $\hat{\mathbf{g}}$ and $\hat{\mathbf{h}}$ related by eq. (11)).

$$\begin{aligned} & \hat{\mathbf{g}} \text{ and } \hat{\mathbf{h}} \text{ is a solution according to Definition 2.1.} \\ \iff & \hat{\mathbf{g}} \text{ and } \hat{\mathbf{h}} \text{ satisfies eq. (9).} \\ \iff & \hat{x} \text{ satisfies eq. (14)} \\ \iff & \hat{x} \text{ is an optimal solution of program (15).} \end{aligned} \tag{16}$$

All but the last equivalence have already been discussed. The last equivalence is true only in the noiseless identification case – the normalized FIR system coefficients $\mathbf{g}^*$ and $\mathbf{h}^*$ is an optimal solution of program (15) with an optimal objective value of zero, hence any optimal solution of program (15) satisfies eq. (14).

D. Properties of the system ID optimization problem

The matrices A_{ij} in (13) can be written as

$$A_{ij} = p_{ij}\,(p_{ij})' - q_{ij}\,(q_{ij})',$$

where

$$p_{ij} = \begin{bmatrix} (\Delta\mathbf{Y}_{ij})' \\ -\frac{1}{2}\,(\Delta\mathbf{U}_{ij})' \end{bmatrix} \quad \text{and} \quad q_{ij} = \begin{bmatrix} 0 \\ -\frac{1}{2}\,(\Delta\mathbf{U}_{ij})' \end{bmatrix} \tag{17}$$

From (17), it can be seen that A_{ij} are rank two matrices with one positive and one negative eigenvalues. Therefore, program (15) is a non-convex QP, which is $\mathcal{NP}$ hard.

On the other hand, it can be seen that the absolute value of the positive eigenvalue is (much) greater than that of the negative eigenvalue. This fact suggests that program (15) might be an "easy" $\mathcal{NP}$ hard problem. This hypothesis is indeed justified by the following numerical experiment. Define a proximity function $R : \mathbb{R}^{N_g+N_h} \mapsto \mathbb{R}_+$ as

$$R(x) := \max_{N-1\geq i>j\geq 0} \{0, x'A_{ij}x\}. \tag{18}$$

Then let $\tilde{d} \in \mathbb{R}^{N_g+N_h}$ be such that $\tilde{d}\,(i)$ is a zero mean unit variance Gaussian random variable for all i, and let x^* be the vector corresponding to $\mathbf{g}^*$ and $\mathbf{h}^*$. Then normalize $\tilde{d}$ to d such that $x^* + sd \in \mathcal{X}$ for all $s \in \mathbb{R}$ and $\|d\| = 1$. Consider one dimensional function $\tilde{R} : \mathbb{R} \mapsto \mathbb{R}_+$ such that $\tilde{R}\,(s) := R\,(x^* + sd)$. Plot this function for a range of s (e.g. $s \in [-0.1, 0.1]$). Repeat the process with another randomly generated d for many times and check the shape of the function $\tilde{R}$ (for different d) around $s = 0$. The outcome of the numerical experiment is shown in Figure 3. Such figure suggests that program (15) is almost convex, substantiating the previous notion that program (15) should not be a too difficult problem to solve.

Finally, the following property of the proximity function R defined in eq. (18) will be assumed but not formally proved.

$$\exists K \in \mathbb{R}_+ : \forall\, x \in \mathcal{X},\ \exists \hat{x} \in \operatorname*{argmin}_{\tilde{x}\in\mathcal{X}} R(\tilde{x}) : \|x - \hat{x}\| \leq KR(x), \tag{19}$$

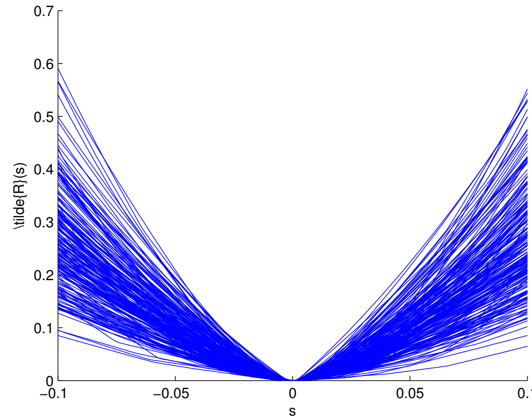

Fig. 3. Plot of $\tilde{R}\,(s)$ in 200 (normalized) randomly generated directions. Note that $\tilde{R}\,(s)$ is not a convex function, but it is almost convex.

III. Solving the Optimization Problem

Subsection II-C concludes with the formulation of program (15), which is a $\mathcal{NP}$ hard non-convex QP. The solution procedure for solving optimization problem (15) can be divided into three steps, which will be discussed in detail in three subsections.

A. Semidefinite programming relaxation

SDP relaxation is a standard attempt to solve non-convex QP's (e.g. [22]). To understand the relaxation, it is noted that in optimization problem (15) the following is true

$$x'A_{ij}x = \operatorname{Tr}(A_{ij}X), \quad X = X' \geq 0,\ \operatorname{rank}(X) = 1. \tag{20}$$

A standard procedure to obtain a SDP relaxation is to drop the rank constraint in (20), which leads to

$$\begin{array}{ll} \underset{X\in\mathcal{X}_{\mathbf{s}},r\in\mathbb{R}}{\text{minimize}} & r \\ \text{subject to} & \operatorname{Tr}(A_{ij}X) \leq r, \quad \forall\, i > j \\ & r \geq 0 \\ & X = X' \geq 0, \end{array} \tag{21}$$

where $\mathcal{X}_{\mathbf{s}}$ is the normalization constraint set for X corresponding to $\mathcal{X}$ for x. Once the relaxation (21) is solved, the singular vector corresponding to the largest singular value of the matrix solution is returned as the best suboptimal solution to (15). It is obvious that the lower the rank of X is, the better the quality of the suboptimal solution will be.

For the noiseless setup in this section, the minimum value of r is actually zero, attainable by, for example, $x^* := \left[(\mathbf{g}^*)' \quad (\mathbf{h}^*)'\right]'$. Hence, the matrix solution $X^* \equiv x^*x^{*\prime}$ is an optimal solution to relaxation (21). Then by setting the minimum value of r to be zero and instead minimizing the trace of X (to obtain a low rank matrix solution, e.g. [23]), the relaxation of (21) is reformulated as

$$\begin{array}{ll} \underset{X\in\mathcal{X}_{\mathbf{s}}}{\text{minimize}} & \operatorname{Tr}(X) \\ \text{Subject to} & \operatorname{Tr}(A_{ij}X) \leq 0 \\ & X = X' \geq 0 \end{array} \tag{22}$$

The tightness of the relaxation depends upon the nonlinearity in Figure 2, but not too much on the FIR systems G and H. The above observation is made through the following numerical experiment: 300 instances of program (22) were solved. The input/output data was produced by driving 300 randomly generated Wiener-Hammerstein systems with the block diagram in Figure 2. G and H were randomly generated, but the nonlinearity ϕ were fixed. For the first one hundred cases, ϕ was a hyperbolic tangent (i.e. $\phi(v) = \tanh(v)$). For the next one hundred cases, ϕ was a saturated linearity (i.e. $\phi(v) = \operatorname{sgn}(v) \max\{|v|, 1\}$). For the last one hundred cases, ϕ was a cubic nonlinearity (i.e. $\phi(v) = v^3$). It is clear that the cubic nonlinearity does not have a derivative bound, whereas the former two nonlinearities do have such a bound. The results of the tests are shown in Table I. It can be seen that for nonlinearities with strong saturation (i.e. derivative bounds) the SDP relaxation is much tighter.

TABLE I

STATISTICS OF THE RATIO (%) BETWEEN THE SECOND AND THE FIRST LARGEST SINGULAR VALUES OF THE SOLUTION MATRICES FOR THE TEST CASES WITH THREE DIFFERENT PRE-SPECIFIED NONLINEARITIES.

	hyperbolic tangent	saturated linearity	cubic
mean (%)	0.1860	5.842×10^{-10}	1.526
std (%)	0.3223	$2.844 \times 10{-}9$	3.501

While the relaxation (22) provides a reasonably good approximation to the true optimal solution of the original nonconvex problem (15), the approximation should always be refined by some inexpensive procedure such as a linearized local search described in the next subsection.

B. Local search

A local search is the following optimization procedure: given an initial guess $x_0 \in \mathbb{R}^{N_g+N_h}$, generate a sequence $\{x_1, x_2, \ldots, x_m\}$ using the formula

$$x_{k+1} = x_k + s_k \Delta x_k, \quad k = 0, 1, \ldots, m-1$$

where $\Delta x_k \in \mathbb{R}^{N_g+N_h}$ is the search direction and $s_k \in \mathbb{R}$ is the step length defined to minimize some objective function. Given the current iterate x_k, a search direction Δx_k should also be admissible. That is,

$$\Delta x_k \in \mathcal{X}_\Delta(x_k) := \left\{ y \in \mathbb{R}^{N_g+N_h} \,\middle|\, x_k + sy \in \mathcal{X}, \ \forall s \in \mathbb{R} \right\}$$

Then this paper seeks to find $\Delta x_k \in \mathcal{X}_\Delta(x_k)$ such that

$$\max_{i>j} \left\{ 0, (x_k + \Delta x_k)' A_{ij} (x_k + \Delta x_k) \right\} \to \min. \tag{23}$$

Problem (23), however, is as difficult as (15). Nevertheless, if the term $(\Delta x_k)' A_{ij} \Delta x_k$ is ignored, then it leads to

$$\begin{array}{ll} \underset{\Delta x_k, r \in \mathbb{R}}{\text{minimize}} & r \\ \text{subject to} & x_k' A_{ij} x_k + 2 x_k' A_{ij} \Delta x_k \le r, \quad \forall i > j \\ & r \ge 0 \\ & \Delta x_k \in \mathcal{X}_\Delta(x_k). \end{array} \tag{24}$$

Optimization problem (24) is a linear program (LP) with respect to decision variables r and Δx_k.

The treatment of the line search in this paper is standard, see [24] for details.

C. Final optimizations

The main reason for the final optimization is the positive real passivity enforcement of the final model of $\mathbf{h}$. Recall the definition of positive real passivity

$$\operatorname{Re}\left\{ H\left(e^{j\omega}\right) \right\} = h_0 + \ldots + h_{N_h-1} \cos\left((N_h - 1)\,\omega \right) > 0. \tag{25}$$

It can be verified (see [25], for example) that eq. (25) is true if and only if there exists $Q = Q' \in \mathbb{R}^{(N_h-1)\times(N_h-1)}$ such that

$$\left[\begin{array}{cc} Q & \frac{1}{2}\check{\mathbf{h}} \\ \frac{1}{2}\check{\mathbf{h}}' & h_0 \end{array} \right] - \left[\begin{array}{cc} 0 & 0 \\ 0 & Q \end{array} \right] > 0, \tag{26}$$

where

$$\check{\mathbf{h}} := \left[\begin{array}{cccc} h_{N_h-1} & h_{N_h-2} & \cdots & h_1 \end{array} \right]' \in \mathbb{R}^{N_h-1},$$

and inequality (26) means that the left side is a positive definite matrix. Note that (26) is a linear matrix inequality with variables Q, h_0 and $\check{\mathbf{h}}$.

Now suppose $\hat{\mathbf{h}}$ is the identified FIR system impulse response coefficients by the relaxation/local search procedure. Then the passive refinement of $\hat{\mathbf{h}}$ can be found by solving

$$\begin{array}{ll} \underset{\mathbf{h}}{\text{minimize}} & \left\| \mathbf{h} - \hat{\mathbf{h}} \right\|_2 \\ \text{subject to} & (26). \end{array} \tag{27}$$

IV. IDENTIFICATION OF WIENER-HAMMERSTEIN SYSTEM – WITH MEASUREMENT NOISE

The development of this section will be parallel to the combination of Section II and Section III. Differences between the noiseless and the noisy cases will be highlighted.

A. System identification problem formulation

The model to be identified is still of the Wiener-Hammerstein structure in Figure 2 with decision variables $\mathbf{g}$ and $\mathbf{h}$ and ϕ being specified by a lookup table. Because of the output measurement noise, however, the system identification feasibility problem will be different. It is shown in Figure 4.

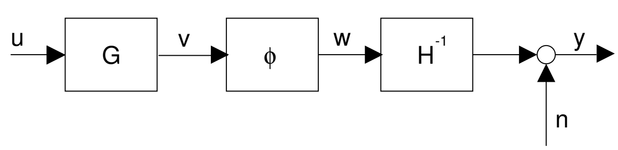

Fig. 4. A feasibility problem to determine the impulse responses of the FIR systems G and H. Here $\mathbf{u}$ and $\mathbf{y}$ are the given input and output measurement generated by the true (but unknown) system. The signals $\mathbf{v}$ and $\mathbf{w}$ are the outputs of G and H, respectively. The signal $\mathbf{n}$ is the noise corrupting the output measurement. In the feasibility problem, $\mathbf{v}$, $\mathbf{w}$ and $\mathbf{n}$ are extra variables chosen so that, together with $\mathbf{g}$ and $\mathbf{h}$, they define a function ϕ satisfying sector bound constraint eq. (6).

There is an extra signal $\mathbf{n} \in \mathbb{R}^N$ to be determined in the feasibility problem in Figure 4. Define the Toeplitz matrix

$\mathbf{N} \in \mathbb{R}^{N \times N_h}$: similar to $\mathbf{U}$ in eq. (5). Then the constraint set defined in Figure 4 can be given as follows.

$$\mathbf{v} = \mathbf{U}\mathbf{g}, \tag{28a}$$

$$\mathbf{w} = (\mathbf{Y} - \mathbf{N})\,\mathbf{h}, \tag{28b}$$

$$(\mathbf{w}_i - \mathbf{w}_j)(\mathbf{w}_i - \mathbf{w}_j - \mathbf{v}_i + \mathbf{v}_j) \le 0, \quad \forall\, i > j. \tag{28c}$$

Then the Wiener-Hammerstein system identification problem with output measurement noise can be defined as

Definition 4.1: **[Wiener-Hammerstein system identification problem – noisy case]**
Given the input/output measurement $(\mathbf{u}, \mathbf{y}) \in \mathbb{R}^N \times \mathbb{R}^N$ of an unknown Wiener-Hammerstein system and positive integers N_g and N_h, find decision vectors $\mathbf{g} \in \mathbb{R}^{N_g}$ and $\mathbf{h} \in \mathbb{R}^{N_h}$ such that there exist signals $\mathbf{v} \in \mathbb{R}^N$, $\mathbf{w} \in \mathbb{R}^N$ and $\mathbf{n} \in \mathbb{R}^N$ satisfying eq. (28a, 28b, 28c). ■

B. Formulation of the system ID optimization problem

Parallel to the development in Subsection II-C, the feasibility problem in Definition 4.1 will be simplified. However, instead of formulating and solving an equivalent optimization problem as it was in Subsection II-C, a *relaxation* will be formulated due to computation considerations.

Substituting eq. (28a) and eq. (28b) into eq. (28c) yields

$$\begin{aligned} & (\Delta\mathbf{Y}_{ij}\mathbf{h})^2 - (\Delta\mathbf{Y}_{ij}\mathbf{h})(\Delta\mathbf{U}_{ij}\mathbf{g}) \\ \le\ & (\Delta\mathbf{N}_{ij}\mathbf{h})(2\Delta\mathbf{Y}_{ij}\mathbf{h} - \Delta\mathbf{U}_{ij}\mathbf{g}) - (\Delta\mathbf{N}_{ij}\mathbf{h})^2, \quad \forall\, i, j, \end{aligned} \tag{29}$$

where

$$\Delta\mathbf{N}_{ij} := \mathbf{N}_i - \mathbf{N}_j \tag{30}$$

and

$$\mathbf{N}_i \in \mathbb{R}^{1 \times N_h}, \quad \mathbf{N}_i := \left[\begin{array}{ccc} \mathbf{N}(i,1) & \cdots & \mathbf{N}(i,N_h) \end{array}\right].$$

Constraint (29) is difficult to handle because of the terms in the right-hand side with the extra variables of $\mathbf{n}$. Therefore, it is proposed in this paper that the following *relaxed* constraint should be imposed instead. That is,

$$(\Delta\mathbf{Y}_{ij}\mathbf{h})^2 - (\Delta\mathbf{Y}_{ij}\mathbf{h})(\Delta\mathbf{U}_{ij}\mathbf{g}) \le \mathbf{r}_{ij}, \quad \forall\, i > j, \tag{31}$$

with variables $\mathbf{g}$, $\mathbf{h}$ and $\mathbf{r} \in \mathbb{R}_+^{N(N-1)/2}$. Constraint eq. (31) is linear with respect to $\mathbf{r}$, and therefore it is no more difficult to handle than eq. (9) in Subsection II-C. Based on the "robustness principle" that eq. (29) should be satisfied by a noise vector $\mathbf{n}$ (and also $\mathbf{r}$) with the minimum norm (e.g. the infinity norm). Then, using the notations x defined in eq. (11), $\mathcal{X}$ defined in eq. (12) and A_{ij} in eq. (13) in Subsection II-C. The relaxed system identification optimization problems can be given as

$$\begin{array}{ll} \underset{x \in \mathcal{X}, r \in \mathbb{R}}{\text{minimize}} & r \\ \text{subject to} & x' A_{ij} x \le r, \quad \forall\, i > j \\ & r \ge 0. \end{array} \tag{32}$$

Note that program (32) has exactly the same form as program (15), the noiseless case in Subsection II-C. However, in general, the minimum objective value of program (32) will not be zero. Accordingly, the solution procedure described in Section III should be modified. This will be explained in Subsection IV-C.

A question of great concern is how good the relaxed optimization problem (32) is. The following statement, from [20], gives a theoretical solution guideline.

Lemma 4.2: Denote $\mathbf{n}^*$ as the vector of output measurement noise. Let $\hat{\mathbf{g}}$ and $\hat{\mathbf{h}}$ be a solution of program (32) when the matrices A_{ij} are defined with input/output measurement $(\mathbf{u}, \mathbf{y})$ **with** noise $\mathbf{n}^*$. Let $\mathbf{g}^*$ and $\mathbf{h}^*$ be a solution of program (15) when the matrices A_{ij} are defined with input/output measurement $(\mathbf{u}, \mathbf{y})$ **without** noise $\mathbf{n}^*$. Then if the proximity function property in eq. (19) (when A_{ij} are defined with noise) is satisfied, then for $\|\mathbf{n}^*\|$ small enough,

$$\left\| \left(\hat{\mathbf{g}}, \hat{\mathbf{h}}\right) - (\mathbf{g}^*, \mathbf{h}^*) \right\|_2 = O\left(\|\mathbf{n}^*\|_2\right). \tag{33}$$

C. Reformulation of SDP relaxation

The relaxation of the feasibility problem in Definition 4.1 leads to the optimization problem (32), which has exactly the same form as program (15) with only one exception – the minimum of program (32) is not necessarily zero in the presence of output measurement noise. Therefore, all of the solution steps described in Section III apply to the noisy problem (32) with the exception that the feasibility problem (22) is infeasible, and hence it cannot be part of the solution procedure. The following SDP will be solved in place of program (22).

$$\begin{array}{ll} \underset{X \in \mathcal{X}_{\mathbf{s}}, r \in \mathbb{R}}{\text{minimize}} & \mathrm{Tr}(X) + \lambda r \\ \text{Subject to} & \mathrm{Tr}(A_{ij} X) \le r \\ & X = X' \ge 0 \\ & r \ge 0 \end{array} \tag{34}$$

In program (34) $\mathcal{X}_{\mathbf{s}}$ is defined in (22), and A_{ij} are defined in eq. (13). $\lambda > 0$ is a tuning parameter. It turns out that $\lambda = 100$ works pretty well in general.

V. Identification of Wiener-Hammerstein System – with Feedback and Noise

The setup of the identification feasibility problem is given in Figure 5. In addition to the decision variables $\mathbf{g} \in \mathbb{R}^{N_g}$ and $\mathbf{h} \in \mathbb{R}^{N_h}$ seen in the previous sections, there are decision variables associated with the FIR system K, which is implicitly characterized by the impulse response of the

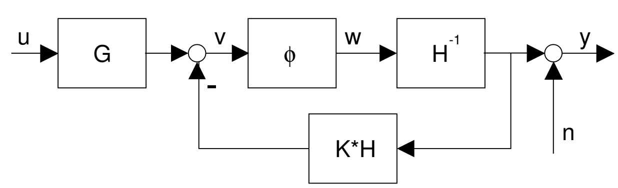

Fig. 5. A feasibility problem to determine the impulse responses of G, H and $K * H$. Here $\mathbf{u}$ and $\mathbf{y}$ are the given input and output measurement generated by the true (but unknown) system. The signals $\mathbf{v}$ and $\mathbf{w}$ are the input and output of the nonlinearity ϕ. The signal $\mathbf{n}$ is the noise corrupting the output measurement. In the feasibility problem, $\mathbf{v}$, $\mathbf{w}$ and $\mathbf{n}$ are extra variables chosen so that, together with $\mathbf{g}$, $\mathbf{h}$ and $\mathbf{k}*\mathbf{h}$, they define a function ϕ satisfying sector bound constraint eq. (6).

product of K and H denoted as $\mathbf{k} * \mathbf{h} \in \mathbb{R}^{N_k+N_h-1}$ and the impulse response of H denoted as $\mathbf{h} \in \mathbb{R}^{N_h}$. Once the vectors $\mathbf{k} * \mathbf{h}$ and $\mathbf{h}$ have been determined, a deconvolution can be applied to retrieve the impulse response of K.

The feasibility problem setup in Figure 5 leads to the following set of constraints.

$$\mathbf{v} = \mathbf{U}\mathbf{g} - \mathbf{Y}\left(\mathbf{k} * \mathbf{h}\right), \tag{35a}$$
$$\mathbf{w} = \left(\mathbf{Y} - \mathbf{N}\right)\mathbf{h}, \tag{35b}$$
$$\left(\mathbf{w}_i - \mathbf{w}_j\right)\left(\mathbf{w}_i - \mathbf{w}_j - \mathbf{v}_i + \mathbf{v}_j\right) \le 0, \quad \forall i > j, \tag{35c}$$

with $\mathbf{U}$, $\mathbf{Y}$ and $\mathbf{N}$ defined in eq. (5) or in some similar fashions. Note that if the following notations are defined

$$\tilde{\mathbf{U}} := \begin{bmatrix} \mathbf{U} & -\mathbf{Y} \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{g}} := \begin{bmatrix} \mathbf{g} \\ \mathbf{k} * \mathbf{h} \end{bmatrix}, \tag{36}$$

then the constraint set eq. (35a,35b,35c) can be written as

$$\mathbf{v} = \tilde{\mathbf{U}}\tilde{\mathbf{g}}, \tag{37a}$$
$$\mathbf{w} = \left(\mathbf{Y} - \mathbf{N}\right)\mathbf{h}, \tag{37b}$$
$$\left(\mathbf{w}_i - \mathbf{w}_j\right)\left(\mathbf{w}_i - \mathbf{w}_j - \mathbf{v}_i + \mathbf{v}_j\right) \le 0, \quad \forall i > j. \tag{37c}$$

As far as the proposed system identification algorithm is concerned, constraint set eq. (37a,37b,37c) has the same form and properties as eq. (28a,28b,28c) in the no feedback case. Therefore, the analysis and algorithm in Section IV can be applied to the feedback Wiener-Hammerstein system identification simply by replacing constraint set eq. (28a,28b,28c) with eq. (37a,37b,37c). Once the optimal values of the decision vectors $\mathbf{g}$, $\mathbf{h}$ and $\mathbf{k}*\mathbf{h}$ have been found, a deconvolution can be applied to obtain the value of $\mathbf{k}$.

VI. Application Examples

A. Identification of randomly generated Wiener-Hammerstein system with feedback

The example given here is the identification of the feedback setup. In this test case, G^*, H^* and K^* are randomly generated positive real passive FIR filters of 4th order. The nonlinearity is $\phi^* = \operatorname{sgn}(x)\{4|x|, 0.1|x| + (4-0.1)\}$. The noise is such that $n[t]$ is uniformly distributed and $n[t] \in [-0.01, 0.01]$ for all t.

For the identification, 86 samples of $(u[t], y[t])$ were used to construct the matrices $\mathbf{U}$ and $\mathbf{Y}$. The identification model has the same structure as in Figure 5, and the orders of the FIR filters are also four. Once the identification is completed, the original test system and the identified model are driven by some test signals (different from the training signals), and the corresponding outputs are recorded. Figure 6 shows the matching of the output of one of the test scenarios. Figure 7 shows the matching of the identified nonlinearity. The identification took about 5 seconds on a PC with a 3GHz CPU and 3GB of RAM.

VII. Conclusion

In this paper, the identification problems of the Wiener-Hammerstein system with and without feedback have been investigated. In the proposed algorithm, the identification of the nonlinearity is non-parametric. The paper formulates the system identification problem as a non-convex QP. Nevertheless, it is demonstrated that the classical SDP relaxation is able to provide very good suboptimal solution to the formulated non-convex QP. Using a local search, high quality solutions of identification problem can often be found. Finally, a numerical example is given to show that the proposed relaxation framework provides an interesting new way to solve the identification problem of the Wiener-Hammerstein system with feedback.

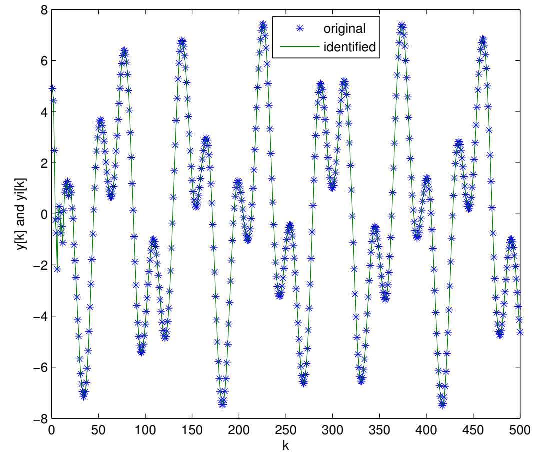

Fig. 6. Matching of output signals by the original (unknown) system and the identified model. $y[k]$ denotes the output by the original system (star). $yi[k]$ denotes the output by the identified model (line). The plots of two output signals almost overlap.

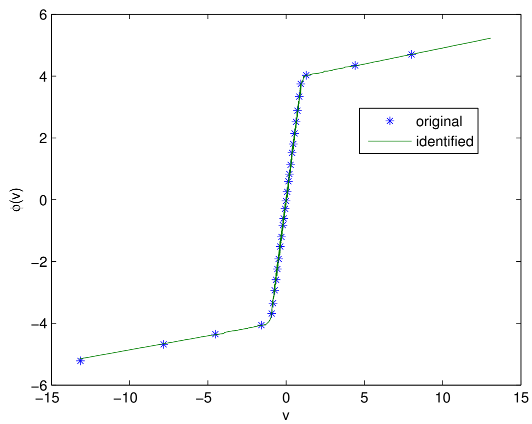

Fig. 7. Matching of the original nonlinearity (star) and the identified nonlinearity (line).

References

[1] L. Ljung, *System Identification, Theory for the User*, 2nd ed. Prentice Hall, 1999.

[2] S. Billings and S. Fakhouri, "Identification of a class of nonlinear systems using correlation analysis," *Proceedings of Institution of Electrical Engineers*, vol. 125, pp. 691–697, July 1978.

[3] R. Haber and L. Keviczky, *Nonlinear System Identification - Input-Output Modeling Approach: Volume 1: Nonlinear System Parameter Identification*. Springer, 1999.
[4] M. Boutayeb and M. Darouach, "Recursive identification method for miso wiener hammerstein model," *IEEE Transactions on Automatic Control*, vol. 40, no. 2, pp. 287–291, February 1995.
[5] M. Kozek and C. Hametner, "Block-oriented identification of Hammerstein/Wiener-models using the RLS-algorithm," *International Journal of Applied Electromagnetics and Mechanics*, vol. 25, pp. 529–535, 2007.
[6] A. Tan and K. Godfrey, "Identification of WienerHammerstein Models Using Linear Interpolation in the Frequency Domain (LIFRED)," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 3, pp. 509–521, June 2002.
[7] W. Greblicki, "Nonparametric identification of wiener systems," IEEE Transcations on Information Theory, vol. 38, no. 5, pp. 1487–1493, September 1992.
[8] K. Hsu, T. Vincent, and K. Poolla, "Identification of Nonlinear Maps in Interconnected Systems," in *the 44th IEEE Conference on Decision and Control*, December 2005, pp. 6430–6435.
[9] A. Hagenblad, L. Ljung, and A. Wills, "Maximum Likelihood Identification of Wiener Models," November 2007, submitted to Automatica.
[10] R. Smith, "Model Validation for Uncertain Systems," Ph.D. dissertation, California Institute of Technology, 1990.
[11] R. Smith and J. Doyle, "Model Validation: A Connection Between Robust Control and Identification," *IEEE Trans. on Automatic Control*, vol. 37, no. 7, pp. 942–952, July 1992.
[12] K. Poolla, P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal, "A Time-Domain Approach to Model Validation," *IEEE Trans. on Automatic Control*, vol. 39, no. 5, pp. 951–959, May 1994.
[13] R. Smith and G. Dullerud, "Modeling and Validation of Nonlinear Feedback Systems," in *Robustness in Identification and Control*. Springer-Verlag, 1999, pp. 87–101.
[14] R. Smith and J. Doyle, "Model Invalidation: A Connection between Robust Control and Identification," in *American Control Conference*, 1989, pp. 1435–1440.
[15] M. Newlin and R. Smith, "A Generalization of the Structured Singular Value and Its Application to Model Validation," *IEEE Trans. on Automatic Control*, vol. 43, no. 7, pp. 901–907, July 1998.
[16] R. Smith and G. Dullerud, "Continuous-Time Control Model Validation Using Finite Experimental Data," *IEEE Trans. on Automatic Control*, vol. 41, no. 8, pp. 1094–1105, August 1996.
[17] T. Zhou and H. Kimura, "Time domain identification for robust control," *Systems & Control Letters*, vol. 20, pp. 167–178, 1993.
[18] R. Kosut, M. Lau, and S. Boyd, "Set-Membership Identification of Systems with Parametric and Nonparametric Uncertainty," *IEEE Trans. on Automatic Control*, vol. 37, no. 7, pp. 929–941, July 1992.
[19] Q. Zhang, A. Iouditski, and L. Ljung, "Identification of wiener systems with monotonous nonlinearity," Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Tech. Rep. LiTH-ISY-R-2787, 2007.
[20] K.C. Sou, "Convex Optimization Methods for Model Reduction," Ph.D. dissertation, Massachusetts Institute of Technology, 2008.
[21] S. Sastry, *Nonlinear Systems*. Springer, 1999.
[22] M. Goemans and D. Williamson, "Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming," *Journal of ACM*, vol. 42, pp. 1115–1145, 1995.
[23] M. Fazel, H. Hindi, and S. Boyd, "Rank Minimization and Applications in System Theory," in *Proceedings of American Control Conference*, Boston, Massachusetts, June 2004, pp. 3273–3278.
[24] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.
[25] B. Alkire and L. Vandenberghe, "Convex optimization problems involving finite autocorrelation sequences," *Mathematical Programming, Series A*, vol. 93, no. 3, pp. 331–359, 2002.