A Sufficient Condition to Guarantee the Quasiconvexity of the Hammerstein Identification Problem

M. Rasouli, D. T. Westwick, W. D. Rosehart

Abstract:
The Hammerstein identification problem is studied using a prediction error method in a separable least squares framework. Thus, the identification is recast as an optimization over the parameters used to describe the nonlinearity. Under certain conditions the identification problem is quasiconvex. First, the identification problem is shown to be quasiconvex under certain assumptions, including the use of an IID input. Next, the IID requirement is relaxed, and a sufficient condition for quasiconvexity is derived. The results are illustrated using a series of simulations.

Keywords: Hammerstein, Identification, Quasiconvex

1. INTRODUCTION

The Hammerstein model, a memoryless nonlinearity followed by a dynamic linear element, is one of the simplest nonlinear dynamic system models. Despite its simplicity, it has been used to model a variety of nonlinear systems including heat exchangers (Eskinat et al., 1991), and stretch reflexes (Dempsey and Westwick, 2004).

The identification of Hammerstein models is a well studied problem. A thorough review is beyond the scope of this paper. The earliest contributions were an iterative method, proposed by Narendra and Gallman (1966), and a non-iterative approach (Chang and Luus, 1971) that was based on identifying an over-parametrized linear regression model.

The convergence of the iterative algorithm was extensively studied by Bai and Li (2004), who showed that the appropriate use of a normalization step can be used to guarantee convergence. In particular, they showed that this approach addresses the famous counter-example due to Söderström and Stoica (1989). Furthermore, Bai and Li (2004) discussed the convexity of the identification problem, and proved that all of the estimation problems encountered in the iterative algorithm are convex, although they did not address the convexity of the overall identification problem and the proof was limited to IID inputs.

Separable least squares (SLS) methods have been used to identify certain classes of Hammerstein models (Westwick and Kearney, 2001; Bai, 2002; Dempsey and Westwick, 2004). In these methods, the parameters are divided into two subsets: those that appear linearly in the output, and those that do not. The linear parameters are solved in closed-form via a least squares regression that is dependent on the values of the nonlinear parameters. An optimization is performed over the nonlinear parameters, where the linear parameters are treated as functions of the nonlinear parameters (due to the dependence of the regressors on the nonlinear parameters). This approach suggests that the convexity of the identification problem need only be assessed with respect to the so-called nonlinear parameters, since the linear parameters have been omitted.

Unimodality is a useful property of quasiconvex functions in practical problems (Dattorro, 2005). This guarantees existence of a global minimum over any convex set in the domain of such functions. In this paper, the sum of squared errors objective function for identification of the Hammerstein model is studied and sufficient conditions on the input that guarantee its quasiconvexity are derived.

The outline of this paper is as follows: Section 2 will introduce the Hammerstein model used throughout the paper and list the assumptions made regarding the model’s inputs and elements. The problem formulation and identification of the Hammerstein model will be described in Sec. 3. This will be followed, in Sec. 4, by proving the quasiconvexity of the problem for IID inputs. Section 5 will provide sufficient conditions under which the problem is quasiconvex for correlated inputs. Simulation results for clarifying the presented materials are shown in Sec. 6. Section 7 will conclude the paper.

1.1 Notation

This paper deals with discrete time models. Thus, the time index, $t$, is an integer, and $q$ is the forward shift operator. Thus $q^{-1}x(t) = x(t-1)$. Bold lower and uppercase roman letters will denote vectors and matrices, respectively. The vector $\mathbf{x}$ will be understood to contain elements of the signal, $x(t)$. Similarly, $G(q)\mathbf{x}$ denotes a vector that contains the signal $G(q)x(t)$.

⋆ This work is supported by NSERC (Canada).
2. HAMMERSTEIN MODEL

A Hammerstein model comprises a memoryless nonlinearity followed by a linear filter as shown in Fig. 1. In

\[ u(t) \xrightarrow{\sum_{i=1}^{M} c_i \phi_i (u(t))} x(t, c) \xrightarrow{\sum_{j=1}^{K} h_j G_j(q)x(t, c)} y(t, c, h) \]

Fig. 1. Block diagram of the Hammerstein model consisting of static nonlinearity followed by a linear dynamic system.

In this study, the nonlinearity will be represented by a basis expansion of known maximum degree, \( M \). Thus,

\[ x(t, c) = \sum_{i=1}^{M} c_i \phi_i (u(t)) \] (1)

where \( u(t) \) is the input signal, \( x(t) \) is the intermediate signal, \( c \) is a vector containing the expansion coefficients, \( c_k \), and the \( \phi_k(\cdot) \) are the basis functions. Such expansions, including polynomials and various linear and cubic splines, have been widely used in nonlinear block structured models, such as the Hammerstein cascade.

Similarly, the linear element will be modeled using an expansion

\[ y(t, c, h) = \sum_{j=1}^{K} h_j G_j(q)x(t, c) \] (2)

where the \( G_j(q) \) are a bank of linear time-invariant filters. Common choices for these filters include delays, \( G_j(q) = q^{-j} \), or discrete Laguerre or Kautz filters (Heuberger et al., 2005).

The model may be represented by the parameters of the nonlinearity and linear filter, \( [c^T h^T]^T \). Note, however, that this representation contains a redundant degree of freedom, in that the models \( [c^T h^T]^T \) and \( [K c^T \frac{1}{M} h^T]^T \), where \( K \) is any non-zero real number, will produce identical outputs, given the same input. Thus, it is customary to normalize one of the two parameter vectors.

The system identification problem is to estimate \( c \) and \( h \), to within a constant gain, given measurements of the input, and the possibly noise corrupted output, \( y(t) \).

2.1 Assumptions

In this sub-section, the assumptions used in the initial quasiconvexity proof are detailed. Section 5 will explore the degree to which some of these assumptions may be relaxed.

A1 The input is a realization of an independent identically distributed (IID) random process unless otherwise stated, e.g. Section 5.

A2 The probability density function of the input is non-zero in at least enough points that the matrix:

\[ \Phi = [\phi_1(u) \phi_2(u) \ldots \phi_M(u)] \]

has full column rank.

A3 The input is chosen such that the intermediate signal has a non-zero variance.

A4 The true system is within the model class described by (1) and (2).

A5 The output is corrupted by an additive, Gaussian IID noise process of finite variance.

A1 is relatively standard in the system identification literature. If A1 holds, then \( x(t) \) will be an IID sequence. In Section 5, this will be relaxed so that we will only require the over parametrized model regressors to be full rank.

A2 would permit the estimation of the nonlinearity, given access to the intermediate signal.

A3 is needed to avoid the pathological case where the intermediate signal is constant. This could occur, for example, if the input PDF was only non-zero inside a dead-zone or saturation region.

A4 and A5 ensure that the residuals are independent of the input. A4 requires that there is no undermodelling, and is relatively standard in the analysis of identification algorithms.

3. PROBLEM FORMULATION AND IDENTIFICATION

Since the measurement noise is assumed to be Gaussian and IID (A5), and there is no undermodelling (A4), minimizing the sum of squared prediction errors produces the maximum likelihood estimate of the model parameters.

The prediction errors and cost function are given by:

\[ \epsilon(c, h) = y_0(c_0, h_0) - y(c, h) \] (4)

\[ V_N(c, h) = \frac{1}{2N} \sum_{t=1}^{N} \epsilon(c, h) \] (5)

where \( c_0, h_0 \), represent the true nonlinear block parameters, and true linear block coefficients. The identification is performed in a prediction error framework (i.e. the error is a function of the data \( u \) and \( y_0 \)), but it will be analyzed in terms of the true model \( (c_0, h_0) \).

The output of the Hammerstein model can be written as

\[ y(t, c, h) = \sum_{i=1}^{M} \sum_{j=1}^{K} c_i h_j G_j(q) \phi_i(u(t)) \] (6)

Define the following

\[ C_0 = c_0 \otimes I_K \] \hspace{1cm} (7a)

\[ H_0 = h_0 \otimes I_M \] \hspace{1cm} (7b)

where \( I_K \), and \( \otimes \) denote an \( K \times K \) identity matrix, and Kronecker product respectively. Also define \( N \times MK \) matrices \( \Psi_1 \) and \( \Psi_2 \) as

\[ \Psi_1^T = [G_1(q) \phi_1^T(u) \ldots G_1(q) \phi_M^T(u) \ldots G_K(q) \phi_1^T(u) \ldots G_K(q) \phi_M^T(u)] \] (8)

11997
and an $MK \times MK$ matrix

$$
\Psi = \Psi_1^T \Psi_1
$$

(10)

Hence, the true and the estimated output of the system may be written as

$$
y_0(c_0, h_0) = \Psi_1 C_0 h_0 = \Psi_2 H_0 c_0
$$

(11a)

$$
y(c, h) = \Psi_1 Ch = \Psi_2 Hc
$$

(11b)

For the noise free case the prediction error is

$$
\epsilon(c, h) = y_0 - y(c, h) = \Psi_1 [C_0 h_0 - Ch]
$$

Therefore, the sum of squared errors can be written as

$$
J(c, h) = 2NV_N(c, h)
$$

$$
= [h_0^T C_0 - h^T C]^T \Psi_1^T \Psi_1 [C_0 h_0 - Ch]
$$

(13)

In this case, we have chosen to treat the nonlinearity coefficients as nonlinear parameters. For any given choice of nonlinearity parameters, $c$, one could compute $x(t, c)$, and then solve an ordinary least squares problem to obtain the optimal filter weights, $h$, corresponding to the chosen nonlinearity parameters. Thus, the filter weights can be regarded as functions of the nonlinearity coefficients. Hence, the prediction errors and cost function can be treated as functions of just the expansion coefficients.

$$
\epsilon(c) = \epsilon(c, h(c))
$$

Based on this formulation, for any fixed $c$, the optimal $h$ can be found by setting the gradient of (13) with respect to $h$ to zero. Thus

$$
h(c) = (C^T \Psi C)^{-1} C^T \Psi C_0 h_0
$$

(14)

This is the result of solving the least squares regression for $h$, as described above, but expressed in terms of the model parameters. Note that $\Psi$ is positive semidefinite (see (10)), and multiplication from left and right by $C$ does not change its rank. If $\Psi_1$ has full column rank (which can be tested using just the IID data and basis elements) then $\Psi$ will be positive definite, enabling the inversion in (14) and hence the SLS solution. Substituting the optimal $h$, (14), into the objective function (13) gives

$$
J(C) = h_0^T C_0 \Psi C_0 h_0
$$

$$
- h_0^T C_0 \Psi C (C^T \Psi C)^{-1} C^T \Psi C_0 h_0
$$

(15)

Establishing the quasiconvexity of the objective (15) is equivalent to proving the quasiconvexity of the unseparated problem (13). In the next 2 sections, (15) will be shown to be quasiconvex for IID inputs, and for certain classes of correlated inputs.

4. QUASICONVEXITY OF THE HAMMERSTEIN IDENTIFICATION PROBLEM FOR IID INPUTS

Let the matrix $\Phi$ defined in (3) be orthogonal. This could either be due to the use of an appropriate set of basis functions, or the result of a separate orthogonalization step, such as a QR factorization or SVD. Since each column in $\Phi$ is the result of applying a memoryless nonlinearity to the IID input, they are each IID.

Let the $G_k(q)$ be a set of orthogonal filters, such as the Laguerre or Kautz filters (Heuberger et al., 2005).

Since the input is IID, the nonlinearity is expanded using an orthogonal basis, and the impulse responses of the basis filters are orthogonal, we have

$$
E [G_i(q) \phi_j(u(t)) G_k(q) \phi_l(u(t))]
$$

$$
= \sum_{\tau_1, \tau_2 = 0}^{\infty} g_i(\tau_1) g_k(\tau_2) E [\phi_j(u(t - \tau_1)) \phi_l(u(t - \tau_2))]
$$

$$
= \frac{1}{N} \sum_{\tau = 0}^{\infty} g_i(\tau) g_k(\tau) \delta_{j,l}
$$

$$
= r_i
$$

where $r_i$ is the norm squared of $g_i(\tau)$.

Define an $M \times M$ matrix

$$
R = \text{diag}[r_1, \cdots, r_M]
$$

(17)

Thus, the matrix $\Psi$ can be written as

$$
\Psi = R \otimes I_K
$$

(18)

Hence,

$$
C^T \Psi C = (c_1^2 r_1 + \cdots + c_M^2 r_M) I_K = c^T R C
$$

(19)

Similarly

$$
C_0^T \Psi C_0 = c_0^T R C_0 I_K
$$

(20)

Substituting (19) and (20) into (15) gives

$$
J(c) = h_0^T h_0 [c^T R C_0 - (c_0^T R C)^2] / c^T R C
$$

(21)

The term $h_0^T h_0$ is a positive scalar related to the true system and does not affect the quasiconvexity of $J(c)$. Also, $c_0^T C_0$ is just a positive bias to the objective function (21). Thus, quasiconvexity of (21) is equivalent to the quasiconvexity of the following function

$$
J_s(c) = (c^T R C)^2 / c^T R C
$$

(22)

The objective function (21) ranges from 0, when $c = c_0$, to $h_0^T h_0 c_0^T R C_0$ for the following case

$$
c_0^T R C = 0
$$

(23)

It can be seen from (21) that $J(c) = J(-c)$, in connection with the gain ambiguity in the Hammerstein model. If $c$ is a local optimum, so is $-c$, and the two local optima will be on opposite sides of the hyperplane defined in (23). Thus, the quasiconvexity need only be proven on one of the half spaces.
Now, we prove the quasiconcavity of (22) in the half space for which $c_0^T Rc > 0$. As long as the initial guess for the nonlinear parameters $c$ is not on the dividing hyperplane (23), quasiconvexity on each of the half spaces would be sufficient to allow one of the equivalent optima to be found using a gradient-based optimization.

**Theorem 4.1.** The objective function (22) is quasiconcave in each of the half spaces $c_0^T Rc > 0$ and $c_0^T Rc < 0$.

One method to show the quasiconcavity of (22) is to prove that the superlevel sets corresponding to this function are convex sets (Boyd and Vandenberghe, 2004).

**Define the superlevel sets**

Choose any arbitrary real valued $\alpha$ in the range of the function, $0 < J_u(c) \leq c_0^T Rc_0$. Consider the following superlevel sets of (22) for the half space $c_0^T Rc > 0$

$$S_\alpha = \{c_0^T Rc \geq \alpha, \quad c_0^T Rc > 0\}$$

(24)

Define the constant vector $d$ as

$$d = \frac{1}{\sqrt{\alpha}}c_0^T R$$

(25)

Then the sets defined in (24) are the same as

$$S_\alpha = \{c_0^T Rc \leq (d^T c)^2, \quad c_0^T Rc > 0\}$$

(26)

**Proof**

We show that for any $c_1$ and $c_2$ in $S_\alpha$, their convex combination, $\theta c_1 + (1 - \theta)c_2$, is in $S_{\alpha}$, where $0 \leq \theta \leq 1$.

$$c_1 \in S_\alpha \Rightarrow c_1^T Rc_1 \leq (d^T c_1)^2 \quad (27a)$$

$$c_2 \in S_\alpha \Rightarrow c_2^T Rc_2 \leq (d^T c_2)^2 \quad (27b)$$

The left hand side of the membership in the inequality (26) for convex combination of $c_1$ and $c_2$ is

$$(\theta c_1 + (1 - \theta)c_2)^T R(\theta c_1 + (1 - \theta)c_2)$$

$$= \theta^2 c_1^T Rc_1 + (1 - \theta)^2 c_2^T Rc_2 + 2\theta(1 - \theta)c_1^T Rc_2$$

(28)

Also, the right hand side of (26) for convex combination of $c_1$ and $c_2$ can be written as

$$|d^T (\theta c_1 + (1 - \theta)c_2)|^2 = \theta^2(d^T c_1)^2 + (1 - \theta)^2(d^T c_2)^2 + 2\theta(1 - \theta)(d^T c_1)(d^T c_2)$$

(29)

The first two terms in (28) are less than or equal to the first two terms in (29). This can be seen by multiplying (27a) and (27b) by $\theta^2$ and $(1 - \theta)^2$ respectively.

To show that the remaining term in (28) is less than or equal to (29), multiply both sides of inequalities (27a) and (27b)

$$(c_1^T Rc_1)^2 \leq (d^T c_1)^2$$

(30)

Since $R$ is positive definite, the function $(c^T Rc)^{\frac{1}{2}}$ is a vector norm and hence, the Cauchy-Schwarz inequality can be used for $c_1$ and $c_2$ (Golub and Van Loan, 1989) i.e.:

$$(c_1^T Rc_2)^2 \leq (d^T c_1, d^T c_2)^2$$

(31)

The product $d^T c_1, d^T c_2$ is positive as $d^T c_1$ and $d^T c_2$ have the same sign. Hence,

$$c_1^T Rc \leq d^T c_1, d^T c_2$$

(32)

Therefore, all the terms in (28) are less than or equal to the corresponding terms in (29) and the proof of theorem 4.1 is completed.

It should be mentioned that $c_1^T R(\theta c_1 + (1 - \theta)c_2)^T = \theta c_1^T Rc_1 + (1 - \theta)c_2^T Rc_2 > 0$

This means that the convex combination of any two members of the set (26) remains in the same half space.

**Remark 4.1.** An alternative, but not immediately obvious, method to show the convexity of the sets defined by (26) would be writing those sets as the inverse image of a second order cone under an affine transformation (Boyd and Vandenberghe, 2004).

One final, significant point is that if the output of the Hammerstein model is corrupted by an IID sequence, the proof still holds. The IID noise is orthogonal to the regressors, and hence does not bias the least square problem. This can be seen by adding a white noise term to (12) and substituting it in (13). Due to the presence of white noise, the optimum value of the objective function increases. But, the optimum point remains unchanged in the expected value sense.

5. QUASI CONVEXITY OF THE HAMMERSTEIN IDENTIFICATION PROBLEM FOR CORRELATED INPUTS

In this section, sufficient conditions under which, the Hammerstein identification problem is quasiconvex in two symmetric half spaces are introduced. Basically, the assumption $A_1$ in 2.1 is relaxed. The positive definiteness of $\Psi$, or equivalently $\Psi_1$ or $\Psi_2$ having full column rank, will be shown to be sufficient to guarantee the quasiconvexity of the Hammerstein identification problem. Note that these conditions are also necessary for the solution of the overparameterized problem using the method proposed by Chang and Luus (1971).

Consider the objective function (15) in which (14) was used to express the optimal value of $h$ as a function of $c$. The first term in (15) is just a positive translation and does not affect the quasiconvexity. Due to the leading negative sign in the second term, quasiconvexity of that term is proved in the following theorem.

**Theorem 5.1.** If the matrix $\Psi$ in (10) is positive definite, then the objective function

$$J_u(C) = h_0^T C^T \Psi C(C^T \Psi C)^{-1} C^T \Psi C_0 h_0$$

(33)

is quasiconcave in each half space, discriminated by the hyperplane

$$h_0^T C^T \Psi C_0 h_0 = 0$$

(34)

**Proof**

Since the matrix $\Psi$ is assumed to be positive definite, it can be decomposed to it’s squared root as $\Psi = \Psi^{\frac{1}{2}} \Psi^{\frac{1}{2}}$ and (33) may be re-written as
\[ J_u(c) = h_0^T \Psi \bar{\Psi} C (C^T \Psi \bar{\Psi} C)^{-1} C^T \Psi \bar{\Psi} C_0 h_0 \] \tag{35} 

Note that because of the special structure of \( C, C_0 \) and \( \Psi, \), \( C_0^T \Psi C = C^T \Psi C_0 \).

Define an \( MK \times 1 \) vector \( g = \Psi \bar{\Psi} C_0 h_0 \) and an \( MK \times K \) matrix \( X(C) = \Psi \bar{\Psi} C \). Re-write (35) in terms of the new variables

\[ J_u(X(C)) = g^T X(C) [X(C)^T X(C)]^{-1} X(C)^T g \] \tag{36} 

Since \( \Psi \) is positive definite, by assumption, the vector \( g \) exists and is a constant vector that depends on the true system \((C_0, h_0)\). Also, \( X(C) \) is an affine transformation of the variable \( C \).

Define \( P(C) = X(C) [X(C)^T X(C)]^{-1} X(C)^T \) so that (36) becomes \( J_u(P(C)) = g^T P(C) g \). The new matrix \( P(C) \) is an orthogonal projection onto the columns of the variable \( X(C) = \Psi \bar{\Psi} C \) because it satisfies two properties, \( P(C) P(C)^T = P(C) \) and \( P(C)^T P(C) = P(C) \). Use these properties to re-write (36)

\[ J_u(P(C)) = g^T P(C) g = g^T P^T (C) P(C) g = ||P(C)||_2 ||g||_2 \] \tag{37} 

The superlevel sets corresponding to (37) for one half space are

\[ S_\alpha = \{ C | \|P(C) g\|_2 \geq \alpha, \ h_0^T C^T \Psi C_0 h_0 > 0 \} \] \tag{38} 

Also, \( h_0^T C^T \Psi C_0 h_0 = h_0^T X(C)^T g > 0 \)

The function (37) is quasiconcave if and only if the superlevel sets of (38) are convex (Boyd and Vandenbergh, 2004).

Choose any two arbitrary \( C_1 \) and \( C_2 \) in (38)

\[ C_1 \in S_\alpha \Rightarrow ||P(C_1) g||_2 \geq \alpha, \ h_0^T X(C_1)^T g > 0 \] \tag{39a} 

\[ C_2 \in S_\alpha \Rightarrow ||P(C_2) g||_2 \geq \alpha, \ h_0^T X(C_2)^T g > 0 \] \tag{39b} 

The convex combination of \( C_1 \) and \( C_2 \) in (38) if the following properties hold

\[ \begin{align*}
  (1) & \quad h_0^T [\Psi \bar{\Psi} (\theta C_1 + (1 - \theta) C_2)] g > 0 \\
  (2) & \quad ||P(\theta C_1 + (1 - \theta) C_2) g||_2 \geq \alpha
\end{align*} \]

The first condition clearly holds. Hence, the convex combination of \( C_1 \) and \( C_2 \) remains in the same half space as \( C_1 \) and \( C_2 \).

The second condition is verified as follows: \( ||P(C_1) g||_2 \) is the magnitude of the projected true system, under the affine transformation \( \Psi \bar{\Psi} \), onto the subspace \( \Psi \bar{\Psi} C_1 \). Since \( C_1 \in S_\alpha \), this magnitude is greater than or equal to \( \alpha \). Without loss of generality, assume that the angle between the true parameters subspace and the subspace spanned by \( \Psi \bar{\Psi} C_1 , \gamma_1 \), is greater than or equal to the angle between the true parameters subspace and the subspace spanned by \( \bar{\Psi} C_2 , \gamma_2 \). The angle between the convex combination of subspaces \( \Psi \bar{\Psi} C_1 \) and \( \Psi \bar{\Psi} C_2 \) is between \( \gamma_1 \) and \( \gamma_2 \). Thus the magnitude of the projection of the true system onto this combination is greater than or equal to the smallest magnitude, i.e. \( ||P(C_1)||_2 \). Thus,

\[ ||P(\theta C_1 + (1 - \theta) C_2) g||_2 \geq \alpha \]

Therefore, the cost function (36) is quasiconcave in the variable \( X(C) = \Psi \bar{\Psi} C \) which is an affine function of \( C \) and consequently (36) is quasiconcave in \( C \).

Remark 5.1. It should be pointed out that the proof of quasiconcavity was done with respect to \( C \). However, there is a one-to-one relationship between the vectors \( c \) and matrices \( C \), as they are related by a Kronecker product (see (7a)). Thus, quasiconcavity of (33) with respect to \( C \) implies the quasiconcavity of (22) with respect to \( c \), and hence the quasi convexity of the identification problem. Thus, the Hammerstein identification problem is quasiconvex (on the half spaces), provided that \( \Psi \) is positive definite.

6. SIMULATION RESULTS

The static nonlinearity in the simulated Hammerstein model is a second-order Hermite polynomial: \( x[u(t)] = 0.67 r_1[u(t)] + 0.87 r_2[u(t)] \). The linear system is represented by a fourth-order tap delay (finite impulse response) filter, with \( h = [0.7303 - 0.5477 - 0.3651 0.1826]^T \) (Bai and Li, 2004).

The system is first excited by a white Gaussian input of length \( N = 4000 \) and unit variance \( \nu = 1 \). The objective function (21) is computed as a function of nonlinear parameters within the range \([-1.7 1.7]\) and at each step the linear parameters are calculated as functions of nonlinear parameters using least squares. The results are shown in Fig. 2. The dividing line has been plotted using (23). The symmetry of the two half spaces can be verified from the figure. All points on a radial line will have the same cost function, due to the gain ambiguity in the Hammerstein system.

The second simulation has been done for a correlated input. The input was generated by filtering the white Gaussian input using a second order Butterworth filter with a cut off frequency \( f_c = 0.1 \). The surface resulting from the objective function (15) for the contours is depicted in Fig. 3. As expected from (34), since the positive definite matrix \( \Psi \) has changed, the dividing line has rotated. Note that the minima of the objective function are identical along the line from the origin to the true parameters in both Figs. 2 and 3.

To visualize the impact of correlated input on the dividing hyperplane, for the system with two nonlinear parameters, the objective function (15) is evaluated in terms of the angle between \( c_1 \) and \( c_2 \). The results are shown in Fig. 4. This figure shows that the minima occur at the angle corresponding to the true parameters, \( c_0 = [0.6 \ 0.8]^T \), and \( c_0 = [-0.6 \ 0.8]^T \). In the case of correlated input, the minima still happen at the same points as for the IID input, but the maxima, the positions corresponding to the dividing hyperplane, are twisted.
Fig. 2. Contours of the sum of squared error objective function versus the nonlinear parameters for IDD input. The solid black line is the dividing line of the two half spaces.

Fig. 3. Contours of the sum of squared error objective function versus the nonlinear parameters for correlated input. The solid black line is the dividing line of the two half spaces.

7. CONCLUSION

The Hammerstein identification problem was formulated as a separable optimization. The optimization was shown to have two equivalent solutions, due to a sign ambiguity. Within each half space, the objective function was quasiconvex. Thus, for all practical purposes, this formulation of the identification problem is quasiconvex. The quasiconvexity under correlated inputs was considered. A full rank condition on the over parametrized model regressors was shown to be sufficient to guarantee quasiconvexity of the identification problem.

REFERENCES