



## Brief paper

Approximation of nonnegative systems by moving averages of fixed order<sup>☆</sup>Lorenzo Finesso<sup>a</sup>, Peter Spreij<sup>b,c,\*</sup><sup>a</sup> IEIIT-CNR, c/o DEI, Università di Padova, via Gradenigo 6/A, 35127 Padova, Italy<sup>b</sup> Korteweg–de Vries Institute for Mathematics, Universiteit van Amsterdam, PO Box 94248, 1090GE Amsterdam, The Netherlands<sup>c</sup> IMAPP, Radboud University, Nijmegen, The Netherlands

## ARTICLE INFO

## Article history:

Received 7 September 2018

Received in revised form 15 April 2019

Accepted 18 April 2019

Available online 27 May 2019

## Keywords:

Moving average

Positive system

Informational divergence

Estimation

Consistency

## ABSTRACT

We pose the approximation problem for scalar nonnegative input/output systems via impulse response convolutions of finite order, i.e. finite order moving averages, based on repeated observations of input/output signal pairs. The problem is converted into a nonnegative matrix factorization with special structure for which we use Csiszár's I-divergence as the criterion of optimality. Conditions are given, on the input/output data, that guarantee the existence and uniqueness of the minimum. We propose an algorithm of the alternating minimization type for I-divergence minimization, and present its asymptotic behaviour. For the case of noisy observations we give the large sample properties of the statistical version of the minimization problem for different observation regimes. Numerical experiments confirm the asymptotic results and exhibit fast convergence of the proposed algorithm.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Nonnegative systems (also known as positive systems) have attracted a lot of attention in the engineering literature both from the theoretical and the practical point of view. See Benvenuti and Farina (2004) for a survey, and for instance Anderson, Deistler, Farina, and Benvenuti (1996), Farina and Benvenuti (1995), Gurvits, Shorten, and Mason (2007) and Shu, Lam, Gao, Du, and Wu (2008) for theoretical contributions. Possible applications of nonnegative systems are e.g. in the fields of image processing, emission tomography, industrial processes, charge routing networks, compartmental systems, storage systems. For these we refer for instance to Dewasurendra, Bauer, and Premaratne (2007), Farina and Rinaldi (2000), O'Sullivan and Benac (2007), Snyder, Schulz, and O'Sullivan (1992), Vardi, Shepp, and Kaufman (1985) and references therein.

In this paper we pose the problem of the time-domain approximation of nonnegative input/output systems by finite (nonnegative) impulse response convolutions of fixed order  $q$  (positive

moving averages  $MA(q)$ ), based on input/output observations. We propose an iterative algorithm to find the best approximation, and study its asymptotical behaviour. The contributions of the present paper are theoretical, possible applications of the algorithm are in the afore mentioned applied fields. The paper complements Finesso and Spreij (2015), where the order of the convolution is not fixed, but varies with the sample size. Our treatment allows for  $m > 1$  input/output pairs. This setting leads easily to a statistical analysis when the output is observed with noise. We then study large sample properties of the resulting parameter estimators when (1) the number of input/output pairs  $m$  grows unboundedly but the time horizon is fixed, (2) the number of observations, the time horizon,  $N$ , tends to infinity, but  $m$  is fixed and (3) a mixture of the previous two cases. It is noted that the last two cases are not meaningful when the order of the convolution is not fixed, as in Finesso and Spreij (2015). Indeed, fixing the order  $q$  is the main difference with our earlier contributions. Algorithms similar to ours for the case  $m = 1$  have been studied in Snyder et al. (1992) and Vardi et al. (1985). Following the choice made in those early contributions our criterion of optimality will be Csiszár's I-divergence, which as argued in Csiszár (1991) (see also Snyder et al. (1992)), is the best choice for approximation problems under nonnegativity constraints.

We emphasize that our approach to the approximation of a given input/output system by a linear time invariant system is different from the usual identification or realization of (nonnegative) linear systems, see Benvenuti and Farina (2004) for a survey,

<sup>☆</sup> This work was partially funded under the STM-2017 CNR Fund, project number CUPB53C1 7001610005. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Alessandro Chiuso under the direction of Editor Torsten Söderström.

\* Corresponding author at: Korteweg–de Vries Institute for Mathematics, Universiteit van Amsterdam, PO Box 94248, 1090GE Amsterdam, The Netherlands.

E-mail addresses: [finesso@cnr.it](mailto:finesso@cnr.it) (L. Finesso), [spreij@uva.nl](mailto:spreij@uva.nl) (P. Spreij).

and for instance Nagy and Matolcsi (2005) and Nagy, Matolcsi, and Szilvási (2007), and the earlier mentioned Anderson et al. (1996), Farina and Benvenuti (1995), Gurvits et al. (2007) and Shu et al. (2008). From the mathematical point of view, the techniques that we have used in Finesso and Spreij (2006) to analyse a nonnegative matrix factorization algorithm have been shown to be very useful in the present context as well, as demonstrated in Finesso and Spreij (2015), and provided several benefits over the analyses contained in Snyder et al. (1992). We will provide explicit conditions for the existence and uniqueness of the minimizer of the criterion in terms of the data. The algorithm that minimizes the informational divergence criterion is of the same alternating minimization type as in Finesso and Spreij (2015), and the optimality conditions (the Pythagorean relations) are shown satisfied at each step. As demonstrated in Finesso and Spreij (2015), these are at the core of a proof of convergence which is more transparent than other proofs in the literature, e.g. Cover (1984), Snyder et al. (1992) and Vardi et al. (1985).

A brief summary of the paper follows. In Section 2 we state the problem and formulate conditions for strict convexity of the objective function, and hence for the existence and uniqueness of the solution. In Section 3 the original problem is lifted into a higher dimensional setting, thus making it amenable to alternating minimization. The optimality properties (Pythagoras rules) of the ensuing partial minimization problems are recalled here. Then we derive the iterative minimization algorithm combining the solutions of the partial minimizations, we present its first properties and the important result on the convergence of the algorithm. In Section 4, taking advantage of the repeated input/output measurements setup or the possibility of a growing time horizon, we give a concise treatment of a statistical version of the approximation problem, focusing on its large sample properties. In the last Section 5 we present numerical experiments that confirm the asymptotic results and exhibit fast convergence properties of the algorithm.

As mentioned above, the present paper is a follow up to Finesso and Spreij (2015), although it treats a different problem. Accordingly, to avoid unnecessary repetitions, in several places we only highlight differences and for proofs, unless extra or different arguments are needed, we refer to Finesso and Spreij (2015). The main differences with Finesso and Spreij (2015) are the following. First, the order  $q$  of the moving average is now fixed, which results in a different optimization algorithm. Secondly, the fixed order allows for an asymptotic analysis of the properties of the estimators under observation regimes, including those where the time horizon tends to infinity, that cannot be treated in the setup of Finesso and Spreij (2015).

## 2. Problem statement and preliminary results

A discrete time, causal, convolutional moving average system  $S_h$  of order  $q$  maps input sequences  $(u_t)_{t \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  into output sequences  $(y_t)_{t \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ , and is completely characterized by an impulse response vector  $h = (h_t)_{t \in \{0, \dots, q\}}$ , such that

$$y_t = S_h u_t = \sum_{k=0}^t h_k u_{t-k}, \quad t \in \mathbb{N},$$

where  $h_k$  is set to zero for  $k > q$ . Alternatively, one can also write

$$y_t = S_h u_t = \sum_{k=0}^{t \wedge q} h_k u_{t-k}, \quad t \in \mathbb{N}, \quad (1)$$

where we write  $t \wedge q$  for  $\min\{t, q\}$ .

Throughout the paper we consider a time horizon  $N$  for which we assume  $N \geq q$ , a standing assumption. Hence we have to replace (1) by

$$y_t = S_h u_t = \sum_{k=0}^{t \wedge q} h_k u_{t-k}, \quad 0 \leq t \leq N. \quad (2)$$

The special case where  $q = N$  has been treated in Finesso and Spreij (2015) and  $N < q$  yields a redundancy, as the parameters  $h_{N+1}, \dots, h_q$  do not play a role in (2). Rewriting Eq. (2) in matrix form, one gets the system of equations

$$\begin{bmatrix} y_0 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} h_0 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & & & \vdots \\ h_q & & \ddots & \ddots & & \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & h_q & \cdots & h_0 \end{bmatrix} \begin{bmatrix} u_0 \\ \vdots \\ u_N \end{bmatrix}, \quad (3)$$

compactly written as

$$y = T(h)u, \quad (4)$$

having introduced the notations  $u = (u_0, \dots, u_N)^\top$ ,  $y = (y_0, \dots, y_N)^\top$  and  $T(h) \in \mathbb{R}^{(N+1) \times (N+1)}$  for the matrix in (3). For  $m$  input sequences  $u^j$ , with corresponding output sequences  $y^j$ , where  $j = 1, \dots, m$ , Eq. (4) becomes

$$Y = T(h)U, \quad (5)$$

where  $Y = (y^1, \dots, y^m) \in \mathbb{R}^{(N+1) \times m}$  and  $U = (u^1, \dots, u^m) \in \mathbb{R}^{(N+1) \times m}$ . Elements of  $Y$  and  $U$  are denoted  $Y_{ij}$  and  $U_{ij}$ , instead of  $y_i^j$  and  $u_i^j$ .

In many practical contexts the inputs and outputs  $U$  and  $Y$  are directly measured *data*, while  $h$  is not known or, more generally, a causal convolutional system  $S_h$  is not known to exist such that  $Y = T(h)U$ . In either of these cases a problem of interest is to find an  $h$  that solves the best approximation problem

$$Y \approx T(h)U, \quad (6)$$

according to a specified loss criterion. In this paper we concentrate on this problem, under the extra condition that (6) is the approximate representation of the behaviour of a positive system, i.e. all quantities in (6) are nonnegative real numbers. The goal is the determination of the *best* nonnegative vector  $h = (h_0, \dots, h_q)^\top$ , where the loss criterion, chosen to measure the discrepancy between the left and the right hand side in (6), is the *I-divergence* between nonnegative matrices. See Csiszár (1991) for a justification from first principles.

For given nonnegative vectors, matrices, tensors  $M$  and  $N$  of the same size, indexed by some variable  $\alpha$ , the *I-divergence* between them is defined as

$$\mathcal{I}(M \parallel N) := \sum_{\alpha} \left( M_{\alpha} \log \frac{M_{\alpha}}{N_{\alpha}} - M_{\alpha} + N_{\alpha} \right) \leq \infty. \quad (7)$$

In definition (7) we also adopt the usual conventions  $\frac{0}{0} = 0$ ,  $0 \log 0 = 0$  and  $\frac{p}{0} = \infty$  for  $p > 0$ .

**Problem 1.** For given  $Y \geq 0$  and  $U \geq 0$ , find a nonnegative vector  $h = (h_0, \dots, h_q)^\top \in \mathcal{H} := \mathbb{R}_+^{q+1}$  such that  $F : \mathcal{H} \rightarrow [0, \infty]$ ,

$$F(h) := \mathcal{I}(Y \parallel T(h)U)$$

is minimized over  $\mathcal{H}$ .

**Problem 1** is well posed if there exists at least one  $h \in \mathbb{R}_+^{q+1}$  such that  $F(h)$  is finite. Under a rather weak condition on the data  $(U, Y)$ , the loss  $F(h)$  is strictly convex (and hence **Problem 1** is well posed), a property that simplifies the study of the existence and uniqueness of the solution of **Problem 1**.

**Condition 1.** For all  $i \in \{0, \dots, N\}$  there exists  $j \in \{1, \dots, m\}$  such that  $U_{0j} > 0$  and  $Y_{ij} > 0$ .

This condition holds e.g. under the (stronger) assumption that for some experiment  $j$ , with initial input  $U_{0j} > 0$ , the output trajectory  $Y_{ij}$  is strictly positive for all  $i$ . **Lemma 2** below is similar to **Finesso and Spreij (2015, Lemma II.6)**, but its proof uses different arguments.

**Lemma 2.** Under **Condition 1** the loss  $F(h)$  is strictly convex on its effective domain  $\{h \in \mathcal{H} : F(h) < \infty\}$ .

**Proof.** We exploit strict concavity of the logarithm. It is sufficient to prove strict concavity of  $h \mapsto \sum_{ij} Y_{ij} \log(T(h)U)_{ij}$ . Note that all mappings  $h \mapsto Y_{ij} \log(T(h)U)_{ij}$  are concave. Hence it is sufficient to show that at least one of them is strictly concave. Fix  $i$  and choose  $j = j(i)$  such that  $Y_{ij} > 0$  and  $U_{0j} > 0$ . We show that for at least on pair  $(i, j)$  one has strict concavity of  $h \mapsto \log(T(h)U)_{ij}$ . Choose different vectors  $h^0, h^1 \in \mathbb{R}_+^{q+1}$  and let  $\bar{h} = (1-t)h^0 + th^1$  for  $t \in (0, 1)$ . We have to show that there is an  $i$  such that  $(T(\bar{h})U)_{ij}$  is not equal to one of the  $(T(h^0)U)_{ij}$  and  $(T(h^1)U)_{ij}$ . Suppose on the contrary that  $(T(\bar{h})U)_{ij} = (T(h^0)U)_{ij}$  for all  $i$ . It is sufficient to restrict our attention to  $i \leq q$ , since we assumed  $N \geq q$ . This is then equivalent to  $\sum_{l=0}^q v_l U_{l-i,j} = 0$  for all  $i$ , where  $v_l = h_l^1 - h_l^0$ . This gives a linear system of  $q+1$  equations in the  $v_l$  in which the coefficient matrix is lower triangular with the  $U_{0,j(i)}$  on the diagonal. But these diagonal elements are all strictly positive, hence the  $v_l$  are all zero, which contradicts  $h^0 \neq h^1$ .  $\square$

**Remark 3.** In solving **Problem 1**, minimizers  $h^*$  at the boundary of  $\mathcal{H} = \mathbb{R}_+^{q+1}$ , i.e. with some zero components, are the rule rather than an exception when  $q = N$ , see **Finesso and Spreij (2015, Remark 10)**. But, if  $N$  is much larger than  $q$ , it has been observed that one often has interior solutions. See Section 5 for an illustration of this remark.

We now state the existence and uniqueness result. The statement and its proof are verbatim the same as for Proposition 7 in **Finesso and Spreij (2015)**. An important ingredient of the proof is that the search for a minimizer can be confined to a suitable compact set, on which the divergence is finite.

**Proposition 4.** Assume **Condition 1** is satisfied, then **Problem 1** admits a unique solution.

**Remark 5.** Suppose that given the input sequences, the outputs are obtained by a true convolutional system  $Y = T(h^*)U$  for some  $h^* \in \mathcal{H}$ . It follows from **Proposition 4** that under **Condition 1**, the minimizer of  $h \mapsto F(h)$  is  $h^*$  and  $F(h^*) = 0$ . Note too that under the same **Condition 1** the system of equations  $T(h)U = T(h^*)U$  has the unique solution  $h = h^*$ . If for the general case one wants to check whether a proposed vector  $h^*$  is a minimizer, it is by the convexity result of **Lemma 2** sufficient to check the Kuhn–Tucker conditions (see e.g. **Zangwill & Mond, 1969, Theorem 2.19**).

### 3. The algorithm

To solve **Problem 1** we propose an alternating minimization algorithm, based on a variation of the lifting technique pioneered by **Csiszár and Tusnády (1984)**. The same approach was previously adopted in **Finesso and Spreij (2015)** for the solution of

**Problem 1** under the condition  $q = N$ . The results of this section are in spirit the same as the corresponding ones in **Finesso and Spreij (2015, Section III)** and can be derived in an analogous way. Proofs are therefore omitted.

This leads to the following algorithm, almost identical to Algorithm 19 in **Finesso and Spreij (2015)**, with minor differences only, see also **Remark 6**.

**Algorithm 1.** Initialize at a strictly positive vector  $h^0$  and define recursively for  $t \geq 0$

$$h^{t+1} = I(h^t),$$

where the map  $I$  acts on the components of  $h^t$  as follows. For  $k = 0, \dots, q$ ,

$$h_k^{t+1} = I_k(h^t) := \frac{h_k^t}{\sum_{l=0}^{N-k} U_l} \sum_{j=1}^m \sum_{i=k}^N \frac{Y_{ij} U_{i-k,j}}{\sum_{p=0}^{i \wedge q} h_p^t U_{i-p,j}}. \quad (8)$$

If the data satisfy  $U_{0j} > 0$ , as is the case under **Condition 1**, any  $h^0 > 0$  componentwise is sufficient for  $F(h^0) < \infty$ .

**Remark 6.** If  $q = N$ , **Algorithm 1** is exactly the same as Algorithm 19 in **Finesso and Spreij (2015)**. If  $q < N$ , one can add artificial parameters  $h_k = 0$  for  $k = q+1, \dots, N$ . Starting the algorithm in  $h_k^0 = 0$  for those  $k$ , we see that all iterated values  $h_k^t$  are zero as well. The extension of the algorithm with these iterates then also yields the algorithm of **Finesso and Spreij (2015)**, with the modification of the zero initial values for  $k = q+1, \dots, N$ . Note that, although **Algorithm 1** can thus be viewed as a special case of **Finesso and Spreij (2015, Algorithm 19)**, it requires separate derivation being a consequence of the partial minimization problems, the second one of which is a constrained version of analogue problem in **Finesso and Spreij (2015)**. The constraints in effect are  $h_k = 0$  for  $k = q+1, \dots, N$ . Curiously enough the solution of the second minimization problem coincides with the  $h_k^*$  in the non-constrained problem in **Finesso and Spreij (2015)**.

Here are a few properties, paralleling those in **Finesso and Spreij (2015)**. Positivity of the initial values is preserved by the iterations; the algorithm decreases the divergence  $\mathcal{I}(Y \parallel T(h^t)U)$  at each step; the recursion enjoys a stability property, if  $h^t$  is such that  $F$  is increasing (decreasing) in the  $k$ th coordinate of  $h^t$ , then  $h_k^{t+1} < h_k^t$  ( $h_k^{t+1} > h_k^t$ ); the vectors  $h^t$  belong to a certain compact set, in fact a simplex.

**Remark 7.** **Algorithm 1** has multiplicative update rules for the  $h_k^t$  and all iterates remain positive. In principle the algorithm risks to get trapped if some component  $h_k^t$  is (nearly) zero. But **Theorem 8** below guarantees that the algorithm converges to the minimizing  $h$ , and hence will not get trapped elsewhere. This is in contrast with many other algorithms with a multiplicative update rule. For further discussion on this issue see **Lin (2007)**.

The next result, **Theorem 8**, concerns the asymptotic behaviour of **Algorithm 1**. The proof is omitted being much like that of **Theorem 25** in **Finesso and Spreij (2015)**, relying heavily on the optimality results for the partial minimizations. A very simple example closes the section.

**Theorem 8.** The sequence of iterates  $h^t$  converges to a limit  $h^\infty$  which minimizes  $h \rightarrow \mathcal{I}(Y \parallel T(h)U)$ .

**Example 9.** Suppose  $q = 0$  and  $N \geq 0$ . This is an instance in which **Problem 1** has an explicit solution,  $h_0^* = \frac{\sum_{ij} Y_{ij}}{\sum_{ij} U_{ij}}$ . Starting with  $h_0^0 > 0$ , **Algorithm 1** produces  $h_0^1 = h_0^*$ , so it reaches

the minimizing value in one step. When  $q > 0$  there is no termination of the algorithm in finitely many steps that achieves the minimizing vector  $h^*$ , but an explicit solution for  $q = 1$ ,  $N = 1$  is available, see [Finesso and Spreij \(2015, Example II.9\)](#). Depending on the data, there are boundary solutions in the latter case.

#### 4. Statistics

In the previous sections we focused on the minimization of  $F(h) = \mathcal{I}(Y \parallel T(h)U)$  over  $h \in \mathbb{R}_+^{q+1}$ , where  $Y$  and  $U$  were given matrices and we presented an algorithm that asymptotically yields the minimizer. In the present section we concentrate on a statistical version of the minimization problem and its large sample properties under different observation regimes. Specifically we study the asymptotics when the number of input sequences grows and/or the time horizon tends to infinity. The latter is possible in the present context, unlike in [Finesso and Spreij \(2015\)](#), because of the fixed dimension of the parameter.

Recall that  $Y, U \in \mathbb{R}^{(N+1) \times m}$ , but are now random quantities. For each fixed  $m, N$ , [Algorithm 1](#) can be used to find the optimizing  $\hat{h}^{N,m}$ , which now becomes a random vector in  $\mathbb{R}_+^{q+1}$ . We will give limit results on consistency and asymptotic normality for the  $\hat{h}^{N,m}$  in three cases. First for  $m \rightarrow \infty$ , and the columns  $U^j$  of  $U$  ( $j = 1, \dots, m$ ) form an i.i.d. sample. Then for  $N \rightarrow \infty$ , and the rows  $U_i$  of  $U$  ( $i = 0, \dots, N$ ) form an i.i.d. sample. Finally for  $N, m \rightarrow \infty$ , and all  $U_{ij}$  ( $i = 0, \dots, N, j = 1, \dots, m$ ) form an i.i.d. sample.

**Assumption 10.** We assume throughout this section that, for  $i = 0, \dots, N$ , and  $j = 1, \dots, m$ , the following *true* relationships hold

$$Y_{ij} = (T(h^*)U^j)_i \delta_{ij}, \quad (9)$$

where  $h^*$  is an interior point of  $\mathbb{R}_+^{q+1}$ , and  $\delta_{ij}$  (multiplicative noise) are nonnegative, i.i.d. random variables, independent of all  $U_{ij}$ , and with  $\mathbb{E} \delta_{ij} = 1$ .

Further assumptions will be detailed in the subsections below.

##### 4.1. Asymptotics for $m \rightarrow \infty$ , $N$ fixed

For matrices  $Y, U$  one can write  $\mathcal{I}(Y \parallel T(h)U) = \sum_{j=1}^m \mathcal{I}(Y^j \parallel T(h)U^j)$ , with the  $Y^j$  and  $U^j$  the columns of the matrices  $Y$  and  $U$  respectively. In this section we assume, next to [Assumption 10](#), that the pairs  $(Y^j, U^j)$  are i.i.d. Let  $(y, u)$  be a pair of random vectors that has the same distribution as each of the  $(Y^j, U^j)$ . Elements of  $y$  (and  $u$ ) are denoted  $y_i$  (and  $u_i$ ). Here is the first result, basically the same as [Finesso and Spreij \(2015, Lemma 27\)](#).

**Lemma 11.** Assume the model (9), independence of  $u_i$  and  $\delta_i$ ,  $\mathbb{E} u_i < \infty$ ,  $\mathbb{E} \delta_i = 1$ , and  $\mathbb{E} \delta_i |\log \delta_i| < \infty$ . Then it holds for all  $h \in \mathcal{H}$  that

$$\begin{aligned} \mathbb{E} \mathcal{I}(y \parallel T(h)u) \\ = \mathbb{E} \mathcal{I}(T(h^*)u \parallel T(h)u) + \sum_i \mathbb{E} (T(h^*)u)_i \mathbb{E} \delta_i \log \delta_i. \end{aligned}$$

Minimizing the function  $h \mapsto \mathbb{E} \mathcal{I}(y \parallel T(h)u)$  (referred to below as the limit criterion) is therefore equivalent to minimizing  $h \mapsto \mathbb{E} \mathcal{I}(T(h^*)u \parallel T(h)u)$ .

The following proposition parallels ([Finesso & Spreij, 2015, Proposition 28](#)) with some minor differences in the statement and the proof.

**Proposition 12.** Let  $\mathbb{P}(u_0 > 0) = 1$  and  $\mathbb{E} u_j^2 < \infty$  for all  $j$ . The limit criterion  $h \mapsto \mathbb{E} \mathcal{I}(y \parallel T(h)u)$  is strictly convex on the set where it is finite (and hence on a neighbourhood of  $h^*$ ) and has a unique minimum for  $h = h^*$ .

**Proof.** We show that the Hessian  $H(h) \in \mathbb{R}^{(q+1) \times (q+1)}$  at  $h$  of the limit criterion is strictly positive definite on the set where the limit criterion is finite. A computation shows that the  $kl$ -element of this matrix is equal to (recall the convention  $u_i = 0$  for  $i < 0$ )

$$H(h)_{kl} = \mathbb{E} \sum_{j=0}^N \frac{(T(h^*)u)_j}{(T(h)u)_j^2} u_{j-k} u_{j-l}.$$

Hence, for any vector  $x = (x_0, \dots, x_N)^\top$  one has, using the convolution notation  $(u * x)_j := \sum_k x_k u_{j-k}$ ,

$$x^\top H(h)x = \mathbb{E} \sum_{j=0}^N \frac{(T(h^*)u)_j}{(T(h)u)_j^2} (u * x)_j^2.$$

Suppose that  $x^\top H(h)x = 0$  for some  $x \in \mathbb{R}^{q+1}$ . Then  $\mathbb{E} \frac{(T(h^*)u)_j}{(T(h)u)_j^2} (u * x)_j^2$  has to be zero for all  $j$ , in particular for  $j \in \{0, \dots, q\}$ . Hence  $\frac{(T(h^*)u)_j}{(T(h)u)_j^2} (u * x)_j^2 = 0$  a.s. for  $j \in \{0, \dots, q\}$ . Since  $(T(h^*)u)_j \geq h_j^* u_0$ , which is strictly positive by the assumptions, one can only have  $\frac{(T(h^*)u)_j}{(T(h)u)_j^2} (u * x)_j^2 = 0$  a.s. if  $(u * x)_j = 0$  a.s. for all  $j = 0, \dots, q$ . This gives a system of linear equations  $\bar{U}x = 0$ , where  $\bar{U} \in \mathbb{R}^{(q+1) \times (q+1)}$  is lower triangular with all diagonal elements equal to  $u_0$ . Using  $\mathbb{P}(u_0 > 0) = 1$ , we deduce that  $x = 0$  iff  $x^\top H(h)x = 0$ . From [Lemma 11](#) it follows that the limit criterion has a minimum at  $h = h^*$ , and by strict convexity this must be the unique minimizer.  $\square$

As in the present case  $N$  is fixed, we simply write  $\hat{h}^m$  for the estimators, i.e. the minimizers of  $F_m(h) = \sum_{j=1}^m \mathcal{I}(Y^j \parallel T(h)U^j)$ . The following proposition, basically the same as [Finesso and Spreij \(2015, Proposition 29\)](#), describes the large sample behaviour of the  $\hat{h}^m$  for the number of input sequences  $m \rightarrow \infty$  and the observation horizon  $N$  fixed. We include the proof for the sake of completeness.

**Proposition 13.** Let [Assumption 10](#) be in force, in particular (9), and assume that the random vectors  $U^j$  form an i.i.d. sequence. Let  $\mathbb{P}(U_{0j} > 0) > 0$  and  $\mathbb{E} U_{ij}^2 < \infty$  for all  $i, j$ , moreover assume that  $h^*$  is an interior point. The estimators  $\hat{h}^m$ , defined as the minimizers of the objective function  $\sum_{j=1}^m \mathcal{I}(Y^j \parallel T(h)U^j)$  are consistent. Moreover, this sequence is asymptotically normal, for some positive definite  $\Sigma \in \mathbb{R}^{(q+1) \times (q+1)}$  we have  $\sqrt{m}(\hat{h}^m - h^*) \xrightarrow{d} N(0, \Sigma)$ .

**Proof.** The limit criterion  $h \mapsto \mathbb{E} \mathcal{I}(Y \parallel T(h)U)$  is strictly convex, continuous on the set where it is finite. Therefore from [van der Vaart \(1998, Problem 5.27\)](#) we conclude that the conditions of [van der Vaart \(1998, Theorem 5.7\)](#) are satisfied and consistency follows. To show that the estimators  $\hat{h}^m$  are asymptotically normal with covariance function as given in [van der Vaart \(1998, Theorem 5.23\)](#), we have to show that the Hessian  $H(h^*)$  at  $h^*$  of the limit criterion is strictly positive definite. But this follows from the proof of [Proposition 12](#) taking  $h = h^*$ .  $\square$

##### 4.2. Asymptotics for $N \rightarrow \infty$ , $m$ fixed

The standing assumption is again [Assumption 10](#). Let, as before,  $Y$  and  $U$  be matrices. Write  $\mathcal{I}(Y \parallel T(h)U) = \sum_{i=0}^N \mathcal{I}(Y_i \parallel (T(h)U)_i)$ , with the  $Y_i$  and  $(T(h)U)_i$  the rows of the matrices  $Y$  and  $T(h)U$ .

We would like to have all rows  $Y_i$  mutually independent, but row  $Y_i$  partly uses the same inputs as  $Y_{i+1}$ , namely the rows



$U_{i+1}, \dots, U_{i-q+1}$  (for  $i \geq q$ ). Consider the rows  $U_i$  and  $U_{i+q+1}$ . The elements of these rows that are needed to compute the convolutions  $(T(h)U)_{ij}$  are  $U_{ij}, \dots, U_{i-q,j}$ , whereas for  $(T(h)U)_{i+q+1,j}$  one needs  $U_{i+q+1,j}, \dots, U_{i+1,j}$ . We see that these sets of elements have empty intersection. To have independence of rows  $(T(h)U)_i$  and  $(T(h)U)_{i+q+1}$  we will assume that the rows  $U_i$  are independent. The interpretation is that in the collective experiments, at different times independent row vectors are used as inputs.

**Lemma 14.** Assume that the rows  $U_i$  form an i.i.d. sequence, and that all necessary expectations are finite. Consider the random criterion function

$$I_N(h) = \frac{1}{N} \sum_{i=0}^N \mathcal{I}(Y_i \parallel (T(h)U)_i).$$

Then one has, for  $N \rightarrow \infty$  the a.s. convergence

$$I_N(h) \rightarrow \mathbb{E} \mathcal{I}(Y_q \parallel (T(h)U)_q).$$

**Proof.** We split the sum  $\sum_{i=0}^N \mathcal{I}(Y_i \parallel (T(h)U)_i)$  into the  $q+1$  sums

$$\sum_{i=0}^{\lfloor \frac{N}{q+1} \rfloor} \mathcal{I}(Y_{i(q+1)+l} \parallel (T(h)U)_{i(q+1)+l})$$

and a remainder term of at most  $q$  terms  $\mathcal{I}(Y_i \parallel (T(h)U)_i)$ . The remainder term divided by  $N$  tends to zero a.s. For each  $l$  the strong law applies because of the independence properties and we have the a.s. convergence

$$\frac{1}{N} \sum_{i=0}^{\lfloor \frac{N}{q+1} \rfloor} \mathcal{I}(Y_{i(q+1)+l} \parallel (T(h)U)_{i(q+1)+l}) \rightarrow \frac{1}{q+1} \mathbb{E} \mathcal{I}(Y_{q+1+l} \parallel (T(h)U)_{q+1+l}).$$

Since the rows  $(Y_i, (T(h)U)_i)$  have the same distribution for all  $i \geq q$ , one has the identity

$$\mathbb{E} \mathcal{I}(Y_{q+1+l} \parallel (T(h)U)_{q+1+l}) = \mathbb{E} \mathcal{I}(Y_q \parallel (T(h)U)_q).$$

The result follows.  $\square$

**Proposition 15.** Assume the model (9), the rows  $U_i$  form an i.i.d. sequence, and for all  $i, j$ ,  $\mathbb{E} U_{ij} < \infty$ ,  $\mathbb{E} \delta_{ij} = 1$ , and  $\mathbb{E} \delta_{ij} |\log \delta_{ij}| < \infty$ . Then it holds that

$$\mathbb{E} \mathcal{I}(Y_q \parallel (T(h)U)_q) = \mathbb{E} \mathcal{I}(T(h^*)U_q \parallel (T(h)U)_q) + \sum_j (\mathbb{E}(T(h^*)U)_{qj} \mathbb{E}(\delta_{qj} \log \delta_{qj})).$$

Moreover, the divergences  $\mathbb{E} \mathcal{I}(T(h^*)U_i \parallel (T(h)U)_i)$  are identical for all  $i \geq q$  and the limit criterion  $h \mapsto \mathbb{E} \mathcal{I}(Y_q \parallel (T(h)U)_q)$  is strictly convex, and hence continuous, on the set where it is finite (and hence on a neighbourhood of  $h^*$ ). It has a unique minimum for  $h = h^*$ , if  $\mathbb{P}(U_{0j} > 0) > 0$  for at least one  $j$  and  $\mathbb{E} U_{0j}^2 < \infty$  for all  $j$ .

**Proof.** The proof of the first assertion is like the one of Lemma 11. The second assertion follows from the observation that for  $i \geq q$  in the computation of the divergence, one needs  $q+1$  inputs  $U_i, \dots, U_{i-q}$  and these have identical distributions. Strict convexity and uniqueness are proved in the same way as for Proposition 12.  $\square$

Minimizing the function  $h \mapsto \mathbb{E} \mathcal{I}(Y_q \parallel (T(h)U)_q)$  (referred to below as the limit criterion) is thus equivalent to minimizing  $h \mapsto \mathbb{E} \mathcal{I}(T(h^*)U_q \parallel (T(h)U)_q)$ .

As in the present case  $m$  is fixed, we write  $\hat{h}^N$  for the estimators. The following proposition describes the large sample behaviour of the  $\hat{h}^N$  for  $N \rightarrow \infty$ .

**Proposition 16.** Let Assumption 10 be in force, in particular (9), and assume the rows  $U_i$  form an i.i.d. sequence. Let  $\mathbb{P}(U_{0j} > 0) > 0$  for at least one  $j$  and  $\mathbb{E} U_{0j}^2 < \infty$  for all  $j$ , moreover assume that  $h^*$  is an interior point. The estimators  $\hat{h}^N$ , defined as the minimizers of the objective function  $\sum_{i=0}^N \mathcal{I}(Y_i \parallel (T(h)U)_i)$  are consistent. Moreover, this sequence is asymptotically normal, for some positive definite  $\Sigma \in \mathbb{R}^{(q+1) \times (q+1)}$  we have  $\sqrt{N}(\hat{h}^N - h^*) \xrightarrow{d} N(0, \Sigma)$ .

**Proof.** As the convolutions are not independent anymore, we cannot immediately follow the same path as in the proof of Proposition 13. Still, the key to prove the result in the present case is the independence of the rows  $U_i$  and that the  $\delta_{ij}$  are independent.

Recall that a sequence of random variables or vectors  $X_i$  is  $q$ -dependent if for every possible time index  $t$  the (possibly infinite) sequences  $(\dots, X_{t-1}, X_t)$  and  $(X_{t+1+q}, X_{t+2+q}, \dots)$  are independent, and that a  $q$ -dependent sequence is automatically strong mixing. It follows from the assumptions that the  $(Y_k, (T(U))_k)$  are  $q$ -dependent and so are the  $\mathcal{I}(Y_k \parallel (T(h)U)_k)$ , which then trivially become a strong mixing sequence. Hence, one can apply Ibragimov's central limit theorem (Ibragimov, 1975) for strongly mixing stationary sequences to have  $\sqrt{N}(I_N(h) - \mathbb{E} I_N(h))$  converging to a zero mean normal distribution. The asymptotic normality result for the estimators follows by a Taylor argument for M-estimators combined with the laws of large numbers and the CLT result for the  $I_N(h)$  above (see van der Vaart (1998, pages 51 and 72)), or by application of the delta-method. See also van der Vaart (1998, Chapters 5 and 19), in particular the proofs of the general Theorems 5.21 and 5.23, and van der Vaart (1998, Section 5.6) with results on the 'classical case'. To verify the consistency condition in these theorems, one needs strict convexity and continuity of the limit criterion  $h \mapsto \mathbb{E} \mathcal{I}(Y_q \parallel (T(h)U)_q)$  and uniqueness of its minimizer, similar to Proposition 12. From van der Vaart (1998, Problem 5.27) one concludes that the conditions of van der Vaart (1998, Theorem 5.7) are satisfied and consistency follows.  $\square$

### 4.3. Asymptotics for $N, m \rightarrow \infty$

In this section we study the large sample behaviour of the estimators  $h^{N,m}$  when both the time horizon  $N$  and the number of experiments  $m$  may tend to infinity. The model is again (9) and next to Assumption 10 in this section it is additionally assumed that both the  $U_{ij}$  and the  $\delta_{ij}$  are i.i.d. arrays with the relevant expectations finite.

We look again at the limit criteria of Lemmas 11 and 14. The first limit criterion becomes  $\sum_{i=0}^N \mathbb{E} \mathcal{I}(Y_{ij} \parallel (T(h)U)_{ij})$ , with  $j$  arbitrary, which equals

$$L_N^1 := \sum_{i=0}^{q-1} \mathbb{E} \mathcal{I}(Y_{ij} \parallel (T(h)U)_{ij}) + (N+1-q) \mathbb{E} \mathcal{I}(Y_{qj} \parallel (T(h)U)_{qj})$$

by the assumed identity in distribution. The second limit criterion we can write as  $\sum_{j=1}^m \mathbb{E} \mathcal{I}(Y_{qj} \parallel (T(h)U)_{qj})$ , equal to

$$L_m^2 := m \mathbb{E} \mathcal{I}(Y_{qj} \parallel (T(h)U)_{qj})$$

by the assumed independence for this case. We see that  $\lim_{N \rightarrow \infty} \frac{1}{N} L_N^1 = \lim_{m \rightarrow \infty} \frac{1}{m} L_m^2 = \mathbb{E} \mathcal{I}(Y_{qj} \parallel (T(h)U)_{qj})$ ,  $j$  arbitrary, for instance  $j = 1$ . This motivates the next result.

**Lemma 17.** Consider the random criterion function

$$I_{N,m}(h) = \frac{1}{Nm} \sum_{i=0}^N \sum_{j=1}^m \mathcal{I}(Y_{ij} \parallel (T(h)U)_{ij}).$$

Then one has, for  $N, m \rightarrow \infty$  the convergence in probability

$$I_{N,m}(h) \rightarrow \mathbb{E} \mathcal{I}(Y_{q1} \parallel (T(h)U)_{q1}), \quad (10)$$

which has  $h^*$  as its unique minimizer.

**Proof.** For each  $j$  the random variables  $\mathcal{I}(Y_{ij} \parallel (T(h)U)_{ij})$  are  $q$ -dependent and hence the variance of the sum  $\sum_{i=0}^N \mathcal{I}(Y_{ij} \parallel (T(h)U)_{ij})$  can be shown to be (finite and) of order  $N$ . As the latter sums are i.i.d. for different  $j$  the result is that the variance of the double sum  $\sum_{i=0}^N \sum_{j=1}^m \mathcal{I}(Y_{ij} \parallel (T(h)U)_{ij})$  is of order  $Nm$ . Hence, Chebyshev's inequality gives the result on the convergence. The minimizing property of  $h^*$  follows as in the proof of [Proposition 12](#), using the additive decomposition of the limit in (10) into  $\mathcal{I}((T(h^*)U)_{q1} \parallel (T(h)U)_{q1})$  and a remainder term not involving  $h$ .  $\square$

**Remark 18.** Let  $\rho_{N,m} = \frac{N}{m}$  and  $\rho = \lim_{N,m \rightarrow \infty} \rho_{N,m}$  (assumed to exist). If  $\rho = 0$ , the limit in [Lemma 17](#) coincides with the result for fixed  $N$ , if  $\rho = \infty$  one retrieves the result of [Lemma 14](#), since under the present independence assumptions  $\mathcal{I}(Y_q \parallel (T(h)U)_q) = m\mathcal{I}(Y_{q1} \parallel (T(h)U)_{q1})$ .

**Proposition 19.** Let [Assumption 10](#) be in force, in particular (9), and assume all  $U_{ij}$  form an i.i.d. double array. Let  $\mathbb{P}(U_{ij} > 0) > 0$  and  $\mathbb{E} U_{ij}^2 < \infty$  for all  $i, j$ , moreover assume that  $h^*$  is an interior point. Let  $N, m \rightarrow \infty$ . The estimators  $\hat{h}^{N,m}$ , defined as the minimizers of the objective function  $I_{N,m}(h)$  are consistent. Moreover, this sequence is asymptotically normal, for some positive definite  $\Sigma \in \mathbb{R}^{(q+1) \times (q+1)}$  we have  $\sqrt{Nm}(\hat{h}^{N,m} - h^*) \xrightarrow{d} N(0, \Sigma)$ .

**Proof.** For consistency one needs [Lemma 17](#) and uniqueness of the minimizer of the expectation in (10). The remainder follows as in the proof of [Proposition 16](#), using for any fixed  $j$  the  $q$ -dependence of the  $Y_{ij}$ ,  $i \geq 0$  and the independence, for fixed  $i$  of the  $Y_{ij}$ ,  $j \geq 1$ .  $\square$

#### 4.4. Misspecified models

The standing assumption in this section until now was [Assumption 10](#) that postulated the existence of a 'true' parameter  $h^*$ . In absence of this assumption we have the following counterpart of [Proposition 12](#) under the conditions of Section 4.1. Similar results hold for the situations of Sections 4.2 and 4.3.

**Proposition 20.** Let  $\mathbb{P}(u_0 > 0) = 1$  and  $\mathbb{E} u_j^2 < \infty$  for all  $j$ . The limit criterion  $h \mapsto \mathbb{E} \mathcal{I}(y \parallel T(h)u)$  is strictly convex on the set where it is finite and has a unique minimum. The unique minimizer coincides with  $h^*$  when [Assumption 10](#) holds.

**Proof.** The proof follows the lines of the proof of [Proposition 12](#), but in the computation of the Hessian one has to replace the quantities  $(T(h^*)u)_j$  with  $Y_j$ . The Hessian is then again seen to be positive definite, and the existence of a unique minimum follows. That this minimizer coincides with  $h^*$  under [Assumption 10](#), follows from [Lemma 11](#).  $\square$

Calling the unique minimizer  $h^*$ , one obtains that all previous results on consistency – under the presented conditions – continue to hold with the 'true' parameter replaced with this  $h^*$ . The same is true for the results for asymptotic normality. See Example 5.25 in [van der Vaart \(1998\)](#) for a similar discussion on maximum likelihood estimation for misspecified models. In that situation the 'true' parameter is replaced by the one that minimizes the Kullback–Leibler information between the distribution of the data and the distribution given by the misspecified model. The analogy with our setting is obvious.

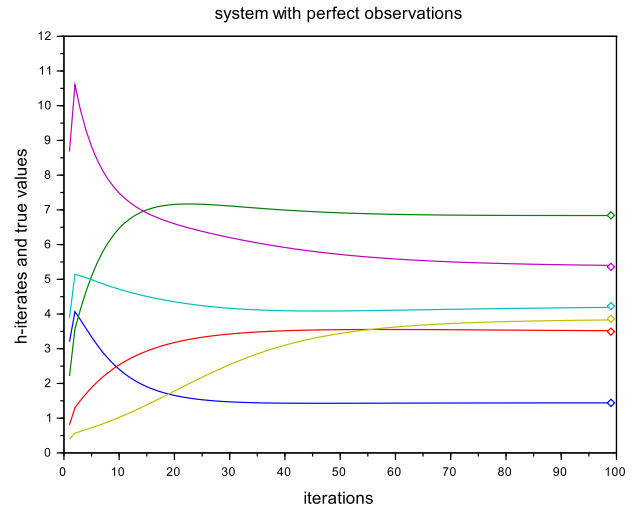


Fig. 1. Noiseless observations,  $m = 5$ ,  $N = 10$ .

## 5. Numerical experiments

In this section we provide the results of a number of numerical experiments that illustrate the behaviour of [Algorithm 1](#). All figures can be found at the end of the paper. We have observed experimentally that usually the iterative algorithm converges very fast in many instances, which is illustrated by the examples. In many cases 50 iterations would have sufficed. For the sake of graph readability in the examples reproduced here the order  $q$  has been limited to 5, leading to a parameter vector  $h$  of length 6. Each of the graphs shows the iterates  $h_k^t$  ( $k = 0, \dots, 5$ ) with the iteration number  $t$  on the horizontal axis, and the 6 values of the impulse response  $h_k^t$  on the vertical axis, different colours representing the different  $k$ 's. As another simplification in the graphs we sometimes omit the first iterates. In [Figs. 1–6](#) the diamonds at the right end of the graph indicate the true  $h^*$  target values. In all cases the  $U_{ij}$  are generated as independent uniform  $U(0.1, 10)$  random variables. The precise features underlying the different experiments are further detailed below. The different experiments highlight the role of the parameters  $m$  and  $N$ , especially when the system is observed with noise. Relatively small values of  $m$  compared to high values of  $N$  give satisfactory results. For the asymptotics of [Proposition 19](#) it is important that only the product  $Nm$  is large.

In the first two examples, [Figs. 1 and 2](#), we investigate whether the algorithm is capable of retrieving the true parameter vector  $h^*$ , when the output data are actually generated by  $h^*$ . After that we investigate the behaviour of the algorithm when we have noisy observations of the output. Here we are in the statistical setting of Section 4. The  $\delta_{ij}$  are taken as  $\exp(Z_{ij}/10 - 1/200)$ , where the  $Z_{ij}$  are independent standard normal random variables. Note that indeed  $\mathbb{E} \delta_{ij} = 1$ . [Figs. 3–6](#) illustrate the large sample behaviour of the estimators. We see that for not too large values of  $N$ , already moderate values of  $m$  give good results, this illustrates [Proposition 19](#). For small values of  $m$ , e.g.  $m = 1$ , one needs a relatively large value of  $N$  to have satisfactory results. This is probably partly due to the dependence between rows of  $Y$ . In the last examples the input/output relation generating the outputs is that of an arbitrary positive system. In this case the  $h$  generated by the algorithm is the impulse response of the best convolutional system approximation to the given system. [Figs. 7 and 8](#) also illustrate [Remark 3](#) on boundary solutions.

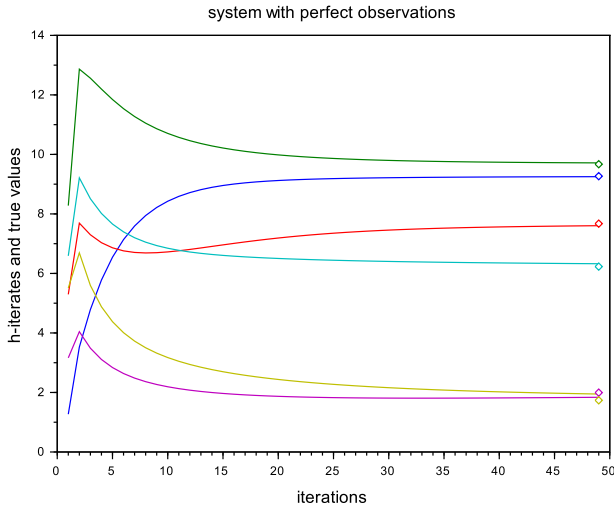


Fig. 2. Noiseless observations,  $m = 10$ ,  $N = 5$ .

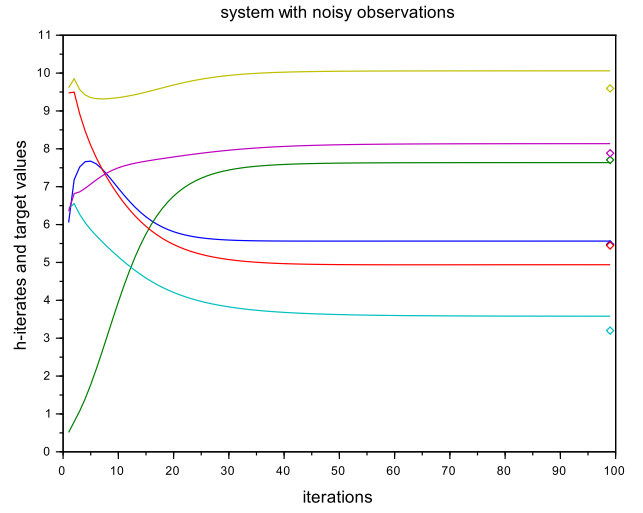


Fig. 5. Noisy observations,  $m = 30$ ,  $N = 5$ .

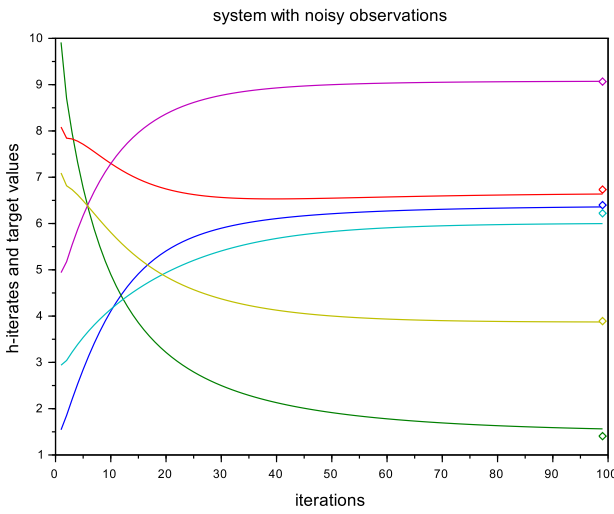


Fig. 3. Noisy observations,  $m = 30$ ,  $N = 20$ .

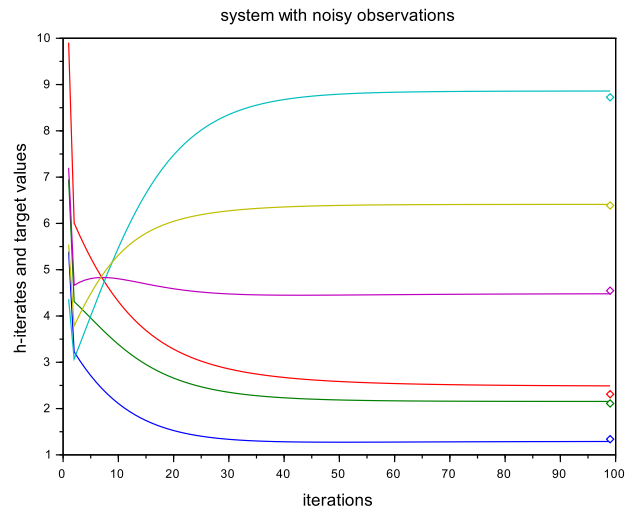


Fig. 6. Noisy observations,  $m = 5$ ,  $N = 30$ .

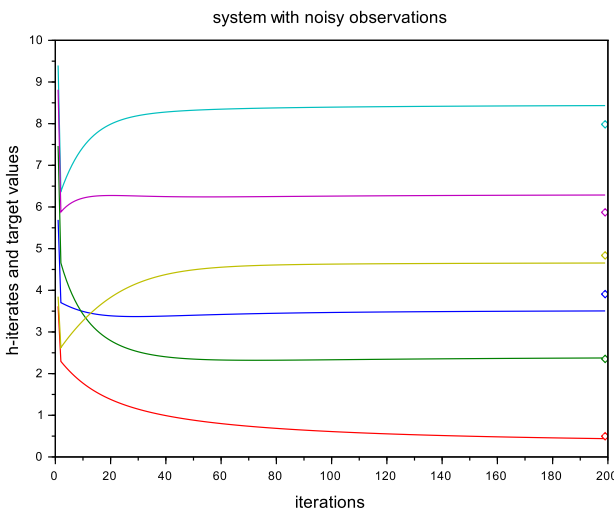


Fig. 4. Noisy observations,  $m = 1$ ,  $N = 100$ .

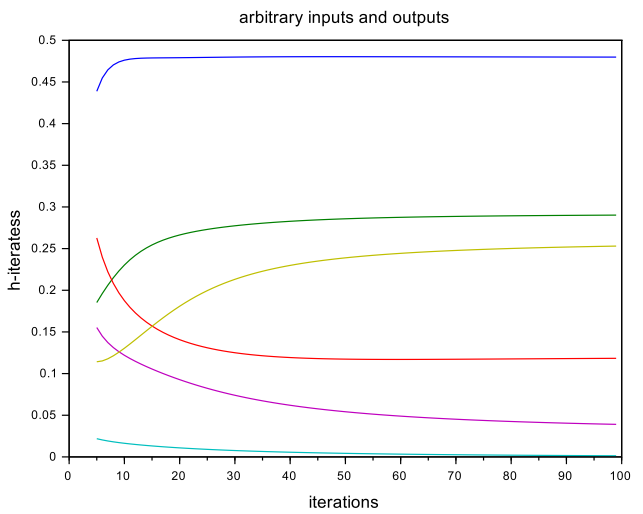


Fig. 7. Arbitrary system,  $m = 10$ ,  $N = 8$ .

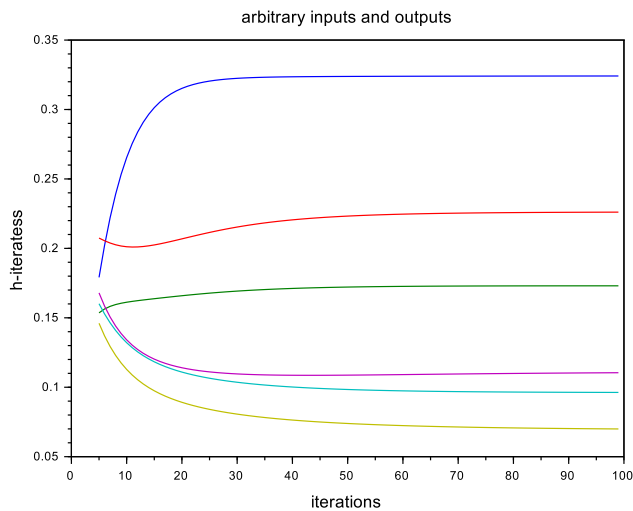


Fig. 8. Arbitrary system,  $m = 10$ ,  $N = 50$ .

## 6. Conclusions

We posed the nonparametric approximation problem for scalar nonnegative input/output systems via impulse response convolutions of finite order, based on multiple observations of input/output signal pairs. The problem is converted into a nonnegative matrix factorization with special structure for which we used Csiszár's  $I$ -divergence as the criterion of optimality. Conditions have been given that guarantee the existence and uniqueness of the minimum. An algorithm whose iterates converge to the unique minimizer has been presented. For the case of noisy observations of a true system we also proved the consistency of the parameter estimators under different large sample regimes (many observation times, many inputs, or a mix of these). Numerical experiments confirm the asymptotic results and often exhibit fast convergence to the minimizer of the objective function.

## References

- Anderson, B. D. O., Deistler, M., Farina, L., & Benvenuti, L. (1996). Nonnegative realization of a linear system with nonnegative impulse response. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 43(2), 134–140.
- Benvenuti, L., & Farina, L. (2004). A tutorial on the positive realization problem. *IEEE Transactions on Automatic Control*, 49(5), 651–664.
- Cover, T. M. (1984). An algorithm for maximizing expected log investment return. *IEEE Transaction on Information Theory*, 30(2, part 2), 369–373.
- Csiszár, I. (1991). Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19(4), 2032–2066.
- Csiszár, I., & Tusnády, G. (1984). Information geometry and alternating minimization procedures. *Statistics & Decisions*, 1(suppl. 1), 205–237.
- Dewasurendra, D. A., Bauer, P. H., & Premaratne, K. (2007). Evidence filtering. *IEEE Transactions on Signal Processing*, 55(12), 5796–5805.
- Farina, L., & Benvenuti, L. (1995). Positive realizations of linear systems. *Systems & Control Letters*, 26(1), 1–9.

- Farina, L., & Rinaldi, S. (2000). *Positive linear systems*. New York: Wiley-Interscience.
- Finesso, L., & Spreij, P. (2006). Nonnegative matrix factorization and  $I$ -divergence alternating minimization. *Linear Algebra and its Applications*, 416(2–3), 270–287.
- Finesso, L., & Spreij, P. (2015). Approximation of nonnegative systems by finite impulse response convolutions. *IEEE Transactions on Information Theory*, 61(8), 4399–4409.
- Gurvits, L., Shorten, R., & Mason, O. (2007). On the stability of switched positive linear systems. *IEEE Transactions on Automatic Control*, 52(6), 1099–1103.
- Ibragimov, I. A. (1975). A note on the central limit theorems for dependent random variables. *Theory of Probability & Its Applications*, 20(1), 135–141.
- Lin, C.-J. (2007). On the convergence of multiplicative update algorithms for nonnegative matrix factorization. *IEEE Transactions on Neural Networks*, 18(6), 1589–1596.
- Nagy, B., & Matolcsi, M. (2005). Minimal positive realizations of transfer functions with nonnegative multiple poles. *IEEE Transactions on Automatic Control*, 50(9), 1447–1450.
- Nagy, B., Matolcsi, M., & Szilvási, M. (2007). Order bound for the realization of a combination of positive filters. *IEEE Transactions on Automatic Control*, 52(4), 724–729.
- O'Sullivan, J. A., & Benac, J. (2007). Alternating minimization algorithms for transmission tomography. *IEEE Transactions on Medical Imaging*, 26(2–3), 283–297.
- Shu, Z., Lam, J., Gao, H., Du, B., & Wu, L. (2008). Positive observers and dynamic output-feedback controllers for interval positive linear systems. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 55(10), 3209–3222.
- Snyder, D. L., Schulz, T. J., & O'Sullivan, J. A. (1992). Deblurring subject to nonnegativity constraints. *IEEE Transactions on Signal Processing*, 40(5), 1143–1150.
- van der Vaart, A. W. (1998). *Cambridge series in statistical and probabilistic mathematics: Vol. 3, Asymptotic statistics*. Cambridge: Cambridge University Press.
- Vardi, Y., Shepp, L. A., & Kaufman, L. (1985). A statistical model for positron emission tomography. *Journal of the American Statistical Association*, 80(389), 8–37, With discussion.
- Zangwill, W. I., & Mond, B. (1969). *Nonlinear programming: A unified approach*. Prentice-Hall.



**Lorenzo Finesso** Laurea in Electrical Engineering (cum laude), U. of Padova, 1979, M.Sc. EE 1987, Ph.D. EE 1990, both at University of Maryland UMCP. Since 1984 he has been a researcher of the Italian National Research Council, formerly with LADSEB, now with IEIIT. His main scientific interests are in the field of realization, approximation, and estimation of finite stochastic systems and of systems with positivity constraints, for which the relative entropy is the natural functional to minimize.



**Peter Spreij** obtained his M.Sc. degree in Mathematics at the Vrije Universiteit (Amsterdam) in 1979 and his PhD with Twente University in 1987. He has been a researcher at CWI (Amsterdam), with the Department of Econometrics (Vrije Universiteit), since 1999 with the Korteweg–de Vries Institute for Mathematics (Universiteit van Amsterdam) and since 2015 also with IMAPP, Radboud University Nijmegen. He has published on a variety of topics in stochastic systems theory, asymptotic statistics, theory of stochastic processes, matrix theory, mathematical finance and information

theory.