Approximation of Nonnegative Systems by Finite Impulse Response Convolutions

Lorenzo Finesso and Peter Spreij

***Abstract*— We pose the deterministic, nonparametric, approximation problem for scalar nonnegative input/output systems via finite impulse response convolutions, 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 study its asymptotic behavior. For the case of noisy observations, we give the large sample properties of the statistical version of the minimization problem. Numerical experiments confirm the asymptotic results and exhibit the fast convergence of the proposed algorithm.**

***Index Terms*— Positive systems, FIR approximation, I-divergence, alternating minimization.**

I. Introduction

INVERSE problems are at the core of system modeling and identification. Since the publication of [21] they have been the subject of a vast technical literature in applied mathematics, engineering, and specialized applied fields. The focus of this paper is on the subclass of problems for which the models are linear and time (or space) invariant. Even within this much narrower field the literature is very rich, with many of the contributions leaning towards specific computational aspects of interest for specialized applications.

The paper has three goals: to pose the problem of the time-domain approximation of nonnegative input/output systems by finite (nonnegative) impulse response convolutions when input/output observations are available, to propose an iterative algorithm to find the best approximation, and to study the asymptotical behavior of the algorithm. The frequency domain properties of nonnegative impulse response systems have been treated in [14] and [15]. Contrary to previous contributions our treatment allows for $m > 1$ input/output pairs. An advantage of allowing multiple input/output pairs is that this setting leads easily to a statistical analysis. The algorithm for the case $m = 1$ has been studied in [20] and [23]. Following the choice made in those early contributions the criterion of optimality will be Csiszár's I-divergence, which as argued in [5] (see also [20]), is the best choice for approximation problems under nonnegativity constraints.

We emphasize that here we pursue a nonparametric approach to the approximation of a given input/output system by a linear time invariant system. The point of view is different from the usual identification or realization of (nonnegative) linear systems, see [2] for a survey, and for instance [1], [9], [12], [16], [17], [19]. From the mathematical point of view the techniques that have been used in [11] to analyse a nonnegative matrix factorization algorithm are perfectly suited to deal with the present approximation problem and provide several benefits over the analyses contained in [20]. We 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 alternating minimization type, and the optimality conditions (the Pythagorean relations) are satisfied at each step. Exploiting this, we are able to present a proof of convergence which is more transparent than other proofs in the literature, see [3], [20], and [23].

Although the contributions of the paper are theoretical, possible applications of the algorithm are in the field of image processing and emission tomography. For these we refer for instance to [18], [20], [23], and the references therein. Design of nonnegative impulse response systems for industrial processes can be found in for instance [8]. Other fields of applications are charge routing networks, compartmental systems, storage systems, see [10].

A brief summary of the paper follows. In Section II 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 III 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 established here. In Section IV we derive the iterative minimization algorithm combining the solutions of the partial minimizations and we present its first properties. Section V is devoted to the convergence analysis of the algorithm. The Pythagoras rules facilitate compact and transparent proofs. In Section VI, taking advantage of the repeated input/output measurements setup, we give a concise treatment of a statistical version of the approximation

Manuscript received June 26, 2013; revised December 8, 2014; accepted April 29, 2015. Date of publication June 10, 2015; date of current version July 10, 2015.

L. Finesso is with the Institute of Electronics, Computer, and Telecommunication Engineering, National Research Council–Institute of Electronics, Computer and Telecommunication, Padua 35131, Italy (e-mail: lorenzo.finesso@ieiit.cnr.it).

P. Spreij is with the Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, Amsterdam 1012WX, The Netherlands (e-mail: spreij@uva.nl).

Communicated by G. Matz, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2015.2443786

problem, focusing on its large sample properties. In the last Section VII we present numerical experiments that confirm the asymptotic results and exhibit the fast convergence properties of the algorithm.

II. Problem Statement and Preliminary Results

A discrete time, causal, convolutional system $\mathcal{S}_h$ 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 sequence $(h_t)_{t\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, such that

$$y_t = \mathcal{S}_h u_t = \sum_{k=0}^{t} h_k u_{t-k}, \qquad t \in \mathbb{N}. \tag{II.1}$$

Rewriting equation (II.1), for $t = 0, \dots, N$, in matrix form, one gets the system of equations

$$\begin{pmatrix} y_0 \\ \vdots \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} h_0 & 0 & \cdots & \cdots & 0 \\ h_1 & h_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ h_N & \cdots & \cdots & h_1 & h_0 \end{pmatrix} \begin{pmatrix} u_0 \\ \vdots \\ \vdots \\ u_N \end{pmatrix}, \tag{II.2}$$

compactly written as

$$y = T(h)u, \tag{II.3}$$

having introduced the notations $u = (u_0, \dots, u_N)^\top$, $y = (y_0, \dots, y_N)^\top$ and $T(h)$ for the matrix in (II.2). For m input sequences u^j, with corresponding output sequences y^j, where $j = 1, \dots, m$, equation (II.3) becomes

$$Y = T(h)U, \tag{II.4}$$

where $Y = (y^1, \dots, y^m) \in \mathbb{R}^{(N+1)\times m}$ and $U = (u^1, \dots, u^m) \in \mathbb{R}^{(N+1)\times m}$.

Convention 1: In expressions containing elements of U the first index is allowed to run out of range, posing $U_{ij} := 0$ for all $i < 0$.

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 $\mathcal{S}_h$ is not known to exist such that $Y = T(h)U$. In either of these cases an interesting problem is to find h such that the approximate relation

$$Y \approx T(h)U \tag{II.5}$$

is the best possible with respect to a specified loss criterion.

In the paper we concentrate on this problem, under the extra condition that (II.5) is the approximate representation of the behavior of a positive system, i.e. all quantities in (II.5) are nonnegative real numbers. The goal is the determination of the *best* nonnegative sequence $h = (h_0, \dots, h_N)^\top$, where the loss criterion, chosen to measure the discrepancy between the left and the right hand side in (II.5), is the *I-divergence* between nonnegative matrices. See [5] for a justification from first principles.

For given nonnegative matrices M and N of the same size, M is said to be absolutely continuous with respect to N, denoted $M \ll N$, if elementwise $M_{ij} = 0$ for all (i, j) such that $N_{ij} = 0$. The *I-divergence* between the nonnegative matrices of the same size M, and N is defined as

$$\mathcal{I}(M||N) := \sum_{ij} \left(M_{ij} \log \frac{M_{ij}}{N_{ij}} - M_{ij} + N_{ij} \right), \tag{II.6}$$

if $M \ll N$, otherwise set $\mathcal{I}(M||N) := +\infty$.

In definition (II.6) we also adopt the usual conventions $\frac{0}{0} = 0$ and $0 \log 0 = 0$. This leads to

Problem 2: For given $Y \geq 0$ and $U \geq 0$, find a nonnegative vector $h = (h_0, \dots, h_N)^\top \in \mathcal{H} := \mathbb{R}_+^{N+1}$ such that

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

is minimized over $\mathcal{H}$.

Remark 3: In Problem 2 one can assume, without loss of generality, that $S := \sum_{ij} Y_{ij} = 1$. Indeed, for any $S > 0$, put $\widetilde{Y}_{ij} = Y_{ij}/S$ and $\widetilde{U}_{ij} = U_{ij}/S$. It then holds that $\mathcal{I}(Y||T(h)U) = S\mathcal{I}(\widetilde{Y}||T(h)\widetilde{U})$, and since S does not depend on h the two problems have the same minimizers. This property will be useful in Section V.

Problem 2 is well posed if there exists at least one $h \in \mathbb{R}_+^{N+1}$ such that $F(h)$ is finite. From definition (II.6) it follows that $F(h)$ is finite if and only if $Y \ll T(h)U$, or equivalently iff $(T(h)U)_{ij} > 0$ for all (i, j) such that $Y_{ij} > 0$. Since

$$(T(h)U)_{ij} = \sum_{k=0}^{i} h_k U_{i-k,j}, \tag{II.7}$$

the following condition characterizes the data (U, Y) that produce a well posed Problem 2.

Condition 4: For all (i, j) such that $Y_{ij} > 0$ there exists $\ell \leq i$ such that $U_{\ell j} > 0$.

Condition 4 is rather weak. In terms of the data sequences it states that if $y_i^j > 0$ then $u_\ell^j > 0$ for some $\ell \leq i$, i.e. if the present output is strictly positive then the present or at least one of the past inputs must be strictly positive. This condition is always satisfied if the data (U, Y) are produced by linear, causal systems.

We prove below that, under a stronger condition on the data (U, Y), the loss $F(h)$ is strictly convex, a property that simplifies the study of the existence and uniqueness of the solution of Problem 2.

Condition 5: For all $i \in \{0, \dots, N\}$ there exists $j \in \{1, \dots, m\}$ such that $U_{0j} > 0$ and $Y_{ij} > 0$.

Condition 5 is strictly stronger than Condition 4, but still rather weak. Physically it states that for each time i there exists at least one experiment j with strictly positive initial input U_{0j} and strictly positive output Y_{ij} at time i. 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 6: Under Condition 5 the loss $F(h)$ is strictly convex on its effective domain, i.e. the set $\{h : F(h) < \infty\}$.

Proof: The elements H_{kl} of the Hessian H of the loss $F(h)$ are

$$H_{kl} := \frac{\partial^2 F}{\partial h_k \partial h_l}(h) = \sum_{ij} \frac{Y_{ij}}{(T(h)U)_{ij}^2} U_{i-k,j} U_{i-l,j}.$$

It is enough to show that H is strictly positive definite. Let $x \in \mathbb{R}^{N+1}$, then

$$x^\top H x = \sum_{kl} H_{kl} x_k x_l = \sum_{ij} \frac{Y_{ij}}{(T(h)U)^2_{ij}} (U * x)^2_{ij},$$

where $(U * x)_{ij} = \sum_l x_l U_{i-l,j}$. Let $x^\top H x = 0$. By nonnegativity of the summands, this only happens if $\frac{Y_{ij}}{(T(h)U)^2_{ij}} (U * x)^2_{ij} = 0$ for all i, j. Since $F(h) < \infty$ on its effective domain, we must have $T(h)U_{ij} > 0$ as soon as $Y_{ij} > 0$. Hence $x^\top H x = 0$ iff $Y_{ij}(U * x)_{ij} = 0$ for all i, j, which gives a system of linear equations in x. For every i fixed and summing over j one explicitly obtains $\sum_k (\sum_j Y_{ij} U_{i-k,j}) x_k = 0$. This gives a system of equations in which the matrix of coefficients is lower triangular with $\sum_j Y_{kj} U_{0j}$ as the k-th diagonal element. Hence this system of equations has $x = 0$ as its only solution iff $\sum_j Y_{kj} U_{0j} > 0$ for all k, but the latter constraint is guaranteed by Condition 5, hence the Lemma is proved. □

We are now ready to state the existence and uniqueness result. The proof is deferred to Section IV.

Proposition 7: Assume Condition 5 is satisfied, then Problem 2 admits a unique solution.

Remark 8: Suppose that given the input sequences, the outputs are obtained by true convolutional system $Y = T(h^*)U$ for some $h^* \in \mathcal{H}$. It follows from Proposition 7 that under Condition 5, the minimiser of $h \mapsto F(h)$ is h^* and $F(h^*) = 0$. Note too that under the same Condition 5 the system of equations $T(h)U = T(h^*)U$ has the unique solution $h = h^*$.

We write below the standard Kuhn-Tucker necessary conditions for a vector h to be a minimizer of $F(h)$. Note that, due to the convexity of the divergence $F(\cdot)$ and the concavity of the nonnegativity constraint, the Kuhn-Tucker conditions are sufficient for optimality (see [24, Th. 2.19]). Condition 5, guarantees the strict convexity of $F(\cdot)$ and therefore the uniqueness of the optimizer. We will follow the notational convention that a dot in place of an index denotes summation with respect to the dotted index, e.g. $M_{i\bullet} := \sum_j M_{ij}$.

Denoting $\nabla F(h)_k := \frac{\partial F(h)}{\partial h_k}$, for $k = 0, \ldots, N$, the Kuhn Tucker conditions assert that, if the vector h minimizes $F(h)$ subject to the constraints $h_k \geq 0$, then

$$\nabla F(h)_k = 0 \quad \text{if} \quad h_k > 0, \tag{II.8}$$
$$\nabla F(h)_k \geq 0 \quad \text{if} \quad h_k = 0, \tag{II.9}$$

where the partial derivatives $\nabla F(h)_k$ are explicitly given by

$$\nabla F(h)_k = -\sum_{j=1}^{m} \sum_{i=k}^{N} \frac{Y_{ij} U_{i-k,j}}{\sum_p h_p U_{i-p,j}} + \sum_{l=0}^{N-k} U_{l\bullet}. \tag{II.10}$$

Example 9: To illustrate that the minimizers h may be interior points (all $h_k > 0$) or boundary points (some $h_k = 0$), we consider the following toy example. Let $m = 1$ and $N = 1$, then $T(h)U$ is a two dimensional vector with components $h_0 u_0$ and $h_0 u_1 + h_1 u_0$. The function F is given by

$$F(h) = y_0 \log \frac{y_0}{u_0 h_0} - y_0 + u_0 h_0 + y_1 \log \frac{y_1}{h_0 u_1 + h_1 u_0} - y_1 + h_0 u_1 + h_1 u_0.$$

Condition 4 for well-posedness reads: if $y_0 > 0$, then $u_0 > 0$, and if $y_1 > 0$ then $u_0 > 0$ or $u_1 > 0$. The Condition 5 for strict convexity reads: $y_0 y_1 u_0 > 0$. One checks by immediate inspection of $F(h)$ that strict convexity does not hold if $y_0 = 0$ or $y_1 = 0$.

In this simple case the minimizing $h^* = (h_0^*, h_1^*)^\top$ can be written explicitly by inspection. Since $F(h) \geq 0$, with equality if and only if $Y = T(h)U$, one gets that, if $y_1 u_0 - y_0 u_1 \geq 0$, then $h_0^* = \frac{y_0}{u_0}$ and $h_1^* = \frac{y_1 u_0 - y_0 u_1}{u_0^2}$ satisfies the constraints $h^* \geq 0$ and attains the minimum $F(h^*) = 0$. On the other hand, if $y_1 u_0 - y_0 u_1 < 0$, then the boundary point $h_0^* = \frac{y_0 + y_1}{u_0 + u_1}$, and $h_1^* = 0$ satisfies the constraints $h^* \geq 0$. Checking that h^* satisfies the Kuhn Tucker conditions guarantees that it is a minimizer. From equation (II.10) one gets $\frac{\partial F}{\partial h_0}(h^*) = 0$, and $\frac{\partial F}{\partial h_1}(h^*) = \frac{u_0}{u_1(y_0 + y_1)}(u_1 y_0 - u_0 y_1) \geq 0$, in agreement with (II.8) and (II.9). See also Remark 10 below for more general considerations.

In solving Problem 2, minimizers h^* at the boundary of $\mathcal{H} = \mathbb{R}_+^{N+1}$, i.e. with some zero components, are the rule rather than an exception. This is illustrated in the following remark.

Remark 10: We analyse here the conditions that produce interior and boundary solutions of Problem 2, limiting the discussion to the case $m = 1$ which is more transparent. If the minimizer h belongs to the interior of the domain $\mathcal{H}$, then it can be found imposing that $\nabla F(h)_k = 0$ for all $k = 0, \ldots, N$, i.e. from equation (II.10),

$$\nabla F(h)_k = -\sum_{i=k}^{N} \frac{y_i u_{i-k}}{\sum_p h_p u_{i-p}} + \sum_{l=0}^{N-k} u_l = 0. \tag{II.11}$$

Assume that $u_0 > 0$. Denoting $t_i := \sum_p h_p u_{i-p}$, the above constraints become

$$\nabla F(h)_k = -\sum_{i=k}^{N} \frac{y_i u_{i-k}}{t_i} + \sum_{i=k}^{N} u_{i-k} = 0. \tag{II.12}$$

For $k = N$ this reduces to

$$-\frac{y_N u_0}{t_N} + u_0 = 0,$$

and one gets $t_N = y_N$. Substitution into equation (II.12) for $k = N - 1$ gives,

$$-\frac{y_{N-1} u_0}{t_{N-1}} + u_0 - \frac{y_N u_1}{t_N} + u_1 = 0,$$

and one gets $t_{N-1} = y_{N-1}$. Completing the recursion one gets the system of equations satisfied by the optimal h,

$$y_i = t_i = \sum_{p=0}^{i} h_p u_{i-p}, \qquad \text{for } i = 0, \ldots, N. \tag{II.13}$$

In other words the only interior solution, if it exists, corresponds to perfect modeling, $Y = T(h)U$. Note that, to find the unknown h, system (II.13) can be rewritten as follows

$$\begin{pmatrix} y_0 \\ \vdots \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} u_0 & 0 & \cdots & \cdots & 0 \\ u_1 & u_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ u_N & \cdots & \cdots & u_1 & u_0 \end{pmatrix} \begin{pmatrix} h_0 \\ \vdots \\ \vdots \\ h_N \end{pmatrix},$$

which is an alternative way of writing (II.2). The computation of the solution by Cramer's rule gives necessary and sufficient conditions on the data (u, y), in terms of a number of determinants, for the existence of a feasible solution, $h \in \mathcal{H}$. If at least one of these conditions is violated, a feasible solution of Problem 2 will necessarily be a boundary point. In this sense for $m = 1$ boundary point solutions are the rule rather than the exception. But also for $m > 1$ boundary solutions often occur, as hinted by the numerical experiments of Section VI.

III. Lifted Version of Problem 2

To solve Problem 2 we propose an alternating minimization algorithm, following the approach adopted for the derivation of the algorithm for nonnegative matrix factorization in [11]. In particular, we use a variation on the lifting technique pioneered by [7] and followed in [11], recasting Problem 2 as a double minimization in a larger space. Here and in the following sections bold capitals, e.g. $\mathbf{M}$, will denote matrices (tensors actually) with three indices. The ambient space in which the lifted problem objects live is $\mathcal{H}_3 := \mathbb{R}_+^{(N+1)\times(N+1)\times m}$, and specifically on $\mathcal{Y}$, and $\mathcal{W}$, two subsets of $\mathcal{H}_3$ defined below in terms of the given data (Y, U),

$$\mathcal{Y} = \left\{ \mathbf{Y} \in \mathcal{H}_3 : \mathbf{Y}_{i\bullet j} = Y_{ij} \right\},$$
$$\mathcal{W} = \left\{ \mathbf{W} \in \mathcal{H}_3 : \mathbf{W}_{ilj} = h_l U_{i-l,j}, \text{ for some } h \in \mathcal{H} \right\}.$$

Remark 11: As a consequence of Convention 1, all $\mathbf{W} \in \mathcal{W}$ have elements $\mathbf{W}_{ilj} = h_l U_{i-l,j} = 0$ for $i < l$.

Remark 12: For any $\mathbf{W} \in \mathcal{H}_3$ let $W \in \mathbb{R}_+^{(N+1)\times m}$ be its marginal, with elements $W_{ij} := \mathbf{W}_{i\bullet j}$. Note that

$$\mathbf{W} \in \mathcal{W} \Longrightarrow W_{ij} = \sum_l h_l U_{i-l,j}. \tag{III.1}$$

It follows that, if $\mathbf{Y} \in \mathcal{Y} \cap \mathcal{W}$, the data (Y, U) can be described with a perfect model $Y = T(h)U$, since equation (III.1) and the definition of $\mathcal{Y}$, imply that $Y_{ij} = \sum_l h_l U_{i-l,j}$.

We consider below two divergence minimization problems in the ambient space $\mathcal{H}_3$.

Problem 13: Given $\mathbf{W} \in \mathcal{H}_3$, minimize the divergence $\mathcal{I}(\mathbf{Y}||\mathbf{W})$ over $\mathbf{Y} \in \mathcal{Y}$.

Problem 14: Given $\mathbf{Y} \in \mathcal{H}_3$, minimize the divergence $\mathcal{I}(\mathbf{Y}||\mathbf{W})$ over $\mathbf{W} \in \mathcal{W}$.

Both problems have explicit solutions. Problem 13, the first, has already been solved in [11]. For ease of reference, we adapt the result below.

Proposition 15: The solution of Problem 13, denoted $\mathbf{Y}^*$ *or* $\mathbf{Y}^*(\mathbf{W})$*, satisfies*

$$\mathbf{Y}^*_{ilj} = \frac{Y_{ij}}{W_{ij}} \mathbf{W}_{ilj},$$

moreover

$$\mathcal{I}(\mathbf{Y}^*(\mathbf{W})||\mathbf{W}) = \mathcal{I}(Y||W), \tag{III.2}$$

which, if $\mathbf{W} \in \mathcal{W}$*, reads*

$$\mathcal{I}(\mathbf{Y}^*(\mathbf{W})||\mathbf{W}) = \mathcal{I}(Y||T(h)U). \tag{III.3}$$

The solution of Problem 14, the second, is detailed in the next proposition. Here and elsewhere in the paper we use the notation

$$\alpha_k = \sum_{l=0}^{k} U_{l\bullet}, \qquad k = 0, \ldots, N.$$

Proposition 16: Assume that $U_{0\bullet} > 0$*. The solution of Problem 14, denoted* $\mathbf{W}^*$ *or* $\mathbf{W}^*(\mathbf{Y})$*, satisfies*

$$\mathbf{W}^*_{ilj} = h^*_l U_{i-l,j}, \quad \textit{where} \quad h^*_l = \frac{\mathbf{Y}_{\bullet l\bullet}}{\alpha_{N-l}},$$

moreover, if $\mathbf{Y} \in \mathcal{Y}$*, the vector* $h^* \in \mathcal{S} := \{h \in \mathcal{H} : \sum_{k=0}^{N} h_k \alpha_{N-k} = \sum_{ij} Y_{ij}\}$.

Proof: Since $\mathbf{W} \in \mathcal{W}$, we in fact optimize over $h \in \mathcal{H}$. Trivial manipulations of the objective function reduce the problem to the explicit minimization of

$$-\sum_{l=0}^{N} \mathbf{Y}_{\bullet l\bullet} \log h_l + \sum_{l=0}^{N} h_l \alpha_{N-l},$$

which is attained at h^*. Finally, if $\mathbf{Y} \in \mathcal{Y}$, checking that $h^* \in \mathcal{S}$ is immediate,

$$\sum_{k=0}^{N} h^*_k \alpha_{N-k} = \mathbf{Y}_{\bullet\bullet\bullet} = \sum_{ij} Y_{ij}.$$

□

Now we can make the connection between the original minimization Problem 2 and the two partial minimization Problems 13 and 14.

Proposition 17: It holds that

$$\min_{\mathbf{Y}\in\mathcal{Y}} \min_{\mathbf{W}\in\mathcal{W}} \mathcal{I}(\mathbf{Y}||\mathbf{W}) = \min_{h\in\mathcal{H}} \mathcal{I}(Y||T(h)U),$$

moreover, if h^* *is the minimizer on the right and* $\mathbf{W}^*$ *its correspondent in* $\mathcal{W}$*,*

$$\mathcal{I}(\mathbf{Y}^*(\mathbf{W}^*)||\mathbf{W}^*) = \mathcal{I}(Y||T(h^*)U).$$

Proof: Fix $\mathbf{Y} \in \mathcal{Y}$ and $\mathbf{W} \in \mathcal{W}$, and let $\mathbf{Y}^* = \mathbf{Y}^*(\mathbf{W})$ be the solution of Problem 13 with $\mathbf{W}$ as input. From equation (III.3), one has

$$\begin{aligned} \mathcal{I}(\mathbf{Y}|\mathbf{W}) &\geq \mathcal{I}(\mathbf{Y}^*(\mathbf{W})||\mathbf{W}) \\ &= \mathcal{I}(Y||T(h)U) \\ &\geq \inf_{h\in\mathcal{H}} \mathcal{I}(Y||T(h)U). \end{aligned}$$

It follows that

$$\min_{\mathbf{Y}\in\mathcal{Y}} \min_{\mathbf{W}\in\mathcal{W}} \mathcal{I}(\mathbf{Y}||\mathbf{W}) \geq \inf_{h\in\mathcal{H}} \mathcal{I}(Y||T(h)U).$$

Conversely, fix $h \in \mathcal{H}$ and let $\mathbf{W}$ be the corresponding element in $\mathcal{W}$, i.e. with $\mathbf{W}_{ilj} = h_l U_{i-l,j}$ then, again from equation (III.3),

$$\begin{aligned} \mathcal{I}(Y|T(h)U) &= \mathcal{I}(\mathbf{Y}^*(\mathbf{W})||\mathbf{W}) \\ &\geq \min_{\mathbf{Y}\in\mathcal{Y}} \min_{\mathbf{W}\in\mathcal{W}} \mathcal{I}(\mathbf{Y}||\mathbf{W}), \end{aligned}$$

which yields

$$\inf_{h\in\mathcal{H}} \mathcal{I}(Y||T(h)U) \geq \min_{\mathbf{Y}\in\mathcal{Y}} \min_{\mathbf{W}\in\mathcal{W}} \mathcal{I}(\mathbf{Y}||\mathbf{W}).$$

Next we check the value of the minimum. Proposition 7 guarantees the existence of a minimizer of the right hand side, call it $h^* \in \mathcal{H}$, and let $\mathbf{W}^*$ be the corresponding element of $\mathcal{W}$. Then, using (III.3) once more, one gets $\mathcal{I}(Y||T(h^*)U) = \mathcal{I}(\mathbf{Y}^*(\mathbf{W}^*)||\mathbf{W}^*)$, which shows that $(\mathbf{Y}^*(\mathbf{W}^*), \mathbf{W}^*)$ is a minimizing pair. □

The solutions of the two partial minimization problems share the essential *Pythagorean property* (see [4] and [6]) which, in the present context, is derived below.

Lemma 18: In Problem 13, with $\mathbf{W}$ *fixed, for all* $\mathbf{Y} \in \mathcal{Y}$,

$$\mathcal{I}(\mathbf{Y}||\mathbf{W}) = \mathcal{I}(\mathbf{Y}||\mathbf{Y}^*(\mathbf{W})) + \mathcal{I}(\mathbf{Y}^*(\mathbf{W})||\mathbf{W}). \quad \text{(III.4)}$$

In Problem 14, with $\mathbf{Y}$ *fixed, for all* $\mathbf{W} \in \mathcal{W}$,

$$\mathcal{I}(\mathbf{Y}||\mathbf{W}) = \mathcal{I}(\mathbf{Y}||\mathbf{W}^*(\mathbf{Y})) + \mathcal{I}(\mathbf{W}^*(\mathbf{Y})||\mathbf{W}). \quad \text{(III.5)}$$

Proof: Equation (III.4) follows by a straightforward computation. We proceed to the proof of equation (III.5). We first compute

$$\begin{aligned}\mathcal{I}(\mathbf{Y}||\mathbf{W}) &- \mathcal{I}(\mathbf{Y}||\mathbf{W}^*(\mathbf{Y})) \\ &= \sum_{ilj} \mathbf{Y}_{ilj} \log \frac{\mathbf{W}^*_{ilj}}{\mathbf{W}_{ilj}} + \sum_{ilj} \mathbf{W}_{ilj} - \sum_{ilj} \mathbf{W}^*_{ilj} \\ &= \sum_{l} \mathbf{Y}_{\cdot l \cdot} \log \frac{h^*_l}{h_l} + \sum_{ilj} \left(\mathbf{W}_{ilj} - \mathbf{W}^*_{ilj}\right). \quad \text{(III.6)}\end{aligned}$$

Next we compute

$$\begin{aligned}\mathcal{I}(\mathbf{W}^*(\mathbf{Y})||\mathbf{W}) &= \sum_{ilj} \mathbf{W}^*_{ilj} \log \frac{\mathbf{W}^*_{ilj}}{\mathbf{W}_{ilj}} + \sum_{ilj} \mathbf{W}_{ilj} - \sum_{ilj} \mathbf{W}^*_{ilj} \\ &= \sum_{il} \frac{Y_{\cdot l \cdot}}{\alpha_{N-l}} U_{i-l,\cdot} \log \frac{h^*_l}{h_l} + \sum_{ilj} \left(\mathbf{W}_{ilj} - \mathbf{W}^*_{ilj}\right) \\ &= \sum_{l} Y_{\cdot l \cdot} \log \frac{h^*_l}{h_l} + \sum_{ilj} \left(\mathbf{W}_{ilj} - \mathbf{W}^*_{ilj}\right), \quad \text{(III.7)}\end{aligned}$$

which coincides with (III.6). □

IV. Algorithm

We propose here an iterative algorithm for the solution of the minimization Problem 2. The algorithm is of the classic alternating minimization type, and is derived using the results of the previous section. Abstractly, one starts at an initial $\mathbf{W}^0 \in \mathcal{W}$, and implements the alternating minimization scheme

$$\ldots \mathbf{W}^t \xrightarrow{1} \mathbf{Y}^t \xrightarrow{2} \mathbf{W}^{t+1} \xrightarrow{1} \mathbf{Y}^{t+1} \ldots,$$

where the superscript t denotes the value at the t-th iteration. The arrow $\xrightarrow{1}$ denotes an instance of the first partial minimization, Problem 13, the matrix at the tail of the arrow is the given input, and the matrix at the head is the optimal solution. The symbols $\mathbf{W}^t \xrightarrow{1} \mathbf{Y}^t$ mean that $\mathbf{Y}^t = \mathbf{Y}^*(\mathbf{W}^t)$. The meaning of $\xrightarrow{2}$ is analogous, and represents an instance of the second partial minimization, Problem 14. The symbols $\mathbf{Y}^t \xrightarrow{2} \mathbf{W}^{t+1}$ mean that $\mathbf{W}^{t+1} = \mathbf{W}^*(\mathbf{Y}^t)$. The hope is that the alternating minimizations produce a sequence of iterates $(\mathbf{W}^t, \mathbf{Y}^t)$ converging to the pair $(\mathbf{W}^*, \mathbf{Y}^*(\mathbf{W}^*))$ of Proposition 17, thus solving Problem 2. This is indeed the case, as proved in Section V. Here we concentrate on producing a computational version of the algorithm sketched above in abstract terms.

Note that, at each iteration, $\mathbf{W}^t$ is completely specified by the fixed data U and by the vector $h^t \in \mathcal{H}$. Computationally it is more efficient to work only with the vectors $h^t \in \mathcal{H}$, one therefore has to shunt the $\mathbf{Y}^t$ steps of the alternating minimization, and move directly from $\mathbf{W}^t$ to $\mathbf{W}^{t+1}$. This leads to the following scheme. For given $h^t \in \mathcal{H}$, define the corresponding $\mathbf{W}^t_{ilj} = h^t_l U_{i-l,j}$ and use it as input in the first partial minimization. The solution, computed according to Proposition 15, is

$$\mathbf{Y}^t_{ilj} = Y_{ij} \frac{h^t_l U_{i-l,j}}{\sum_{p=0}^{i} h^t_p U_{i-p,j}}. \quad \text{(IV.1)}$$

Use now $\mathbf{Y}^t_{ilj}$ as input in the second partial minimization. The solution, computed according to Proposition 16, is

$$h^{t+1}_k = \frac{\mathbf{Y}^t_{\cdot k \cdot}}{\sum_{l=0}^{N-k} U_{l\cdot}}, \quad \text{(IV.2)}$$

with

$$\mathbf{Y}^t_{\cdot k \cdot} = \sum_{i=k}^{N} \sum_{j=1}^{m} \frac{Y_{ij} U_{i-k,j}}{\sum_p h^t_p U_{i-p,j}} h^t_k. \quad \text{(IV.3)}$$

To shunt the $\mathbf{Y}^t$ step it is enough to combine equations (IV.1), (IV.2), and (IV.3) to obtain the following iterative algorithm, solely in terms of h^t vectors and original data (U, Y).

Algorithm 19: 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

$$h^{t+1}_k = I_k(h^t) := \frac{h^t_k}{\sum_{l=0}^{N-k} U_{l\cdot}} \sum_{j=1}^{m} \sum_{i=k}^{N} \frac{Y_{ij} U_{i-k,j}}{\sum_p h^t_p U_{i-p,j}}. \quad \text{(IV.4)}$$

If the data satisfy $U_{0\cdot} > 0$, as is the case under Condition 5, any $h^0 > 0$ componentwise is sufficient for $F(h^0) < \infty$.

For further reference it is convenient to introduce the functions G_k defined implicitly as (see equation (IV.4))

$$I_k(h^t) := h^t_k G_k(h^t). \quad \text{(IV.5)}$$

Remark 20: Note that, under the assumption $U_{0\cdot} > 0$, the functions $G_k(h)$ are continuous at all points h such that $Y \ll T(h)U$.

Properties of Algorithm 19:

1) The h^{t+1}_k are convex combinations of the Y_{ij} for $k \geq 1$, with weights depending in the data U_{ij} and the previous vector h^t.
2) The algorithm decreases the divergence $\mathcal{I}(Y||T(h^t)U)$ at each step. Indeed, by construction and Propositions 15 and 16, we have

$$\begin{aligned}\mathcal{I}(Y|T(h^{t+1})U) &= \mathcal{I}(\mathbf{Y}^{t+1}||\mathbf{W}^{t+1}) \\ &\leq \mathcal{I}(\mathbf{Y}^{t}||\mathbf{W}^{t+1}) \\ &\leq \mathcal{I}(\mathbf{Y}^{t}||\mathbf{W}^{t}) = \mathcal{I}(Y|T(h^t)U). \quad \text{(IV.6)}\end{aligned}$$

Proposition 22 will quantify the decrease.

3) If for some t the vector h^t is a perfect model, i.e. $Y = T(h^t)U$, then

$$\sum_{j=1}^{m}\sum_{i=k}^{N}\frac{Y_{ij}U_{i-k,j}}{\sum_p h_p^t U_{i-p,j}} = \sum_{j=1}^{m}\sum_{i=k}^{N}\frac{\sum_p h_p^t U_{i-p,j}U_{i-k,j}}{\sum_p h_p^t U_{i-p,j}} = \sum_{l=0}^{N-k} U_{l\bullet},$$

hence, from equation (IV.4), $h^{t+1} = h^t$, i.e. perfect models are fixed points of the algorithm.

4) If for some t the gradient $\nabla F(h^t) = 0$, i.e. h^t is a stationary point of $F(h)$, then, using equation (II.10) to rewrite the recursion (IV.4),

$$h_k^{t+1} = h_k^t\left(1 - \frac{\nabla F(h^t)_k}{\sum_{l=0}^{N-k} U_{l\bullet}}\right) = h_k^t, \qquad \text{(IV.7)}$$

i.e. stationary points of $F(h)$ are fixed points of the algorithm. Moreover, we recognize a stability property of the recursion. If h^t is such that F is increasing (decreasing) in the k-th coordinate of h^t, then $h_k^{t+1} < h_k^t$ ($h_k^{t+1} > h_k^t$).

5) The vectors h^t belong to the simplex $\mathcal{S}$, as it follows from Proposition 16.

6) Assume the condition of Lemma 6. If a starting value $h_k^0 > 0$, then $h_k^t > 0$ for all $t > 0$.

7) We omit the details of the following trivial consistency check. If $N = 0$, the solution of Problem 2 is $h^* = h_0^* = \frac{Y_{0\bullet}}{U_{0\bullet}}$, the algorithm produces $h^1 = h^*$, and stays there.

Remark 21: Algorithm 19 has multiplicative update rules for the h_k^t and as stated above, see 6), all iterates remain positive. In principle the algorithm risks to get trapped if some component h_k^t is (nearly) zero. But this can only happen if the iterates are close to a point where the divergence is minimized. In fact the (strict) convexity of the objective function excludes the existence of local minima. See further Theorem 25 below, which guarantees that the algorithm converges to the minimizing h, and hence will not get trapped elsewhere.

This is in contrast with algorithms for nonnegative matrix factorization, where given a matrix V one attempts to find a best possible factorisation $V \approx WH$, for instance in the sense of minimal divergence, or minimal least squares. In this case the objective function is not convex, there are local minima, and undesired traps may occur. For further discussion on this issue see [13].

We are now in the position to prove Proposition 7.

Proof of Proposition 7: The impulse response $h = \mathbf{1}$, i.e. $h_k \equiv 1$, gives $Y = T(\mathbf{1})U$ strictly positive, hence Condition 5 is satisfied and by Lemma 6 the I-divergence $F(\mathbf{1}) = \mathcal{I}(Y||T(\mathbf{1})U)$ is finite. Take then $h^0 = \mathbf{1}$ as a starting value of Algorithm 19, which at the first step produces h^1 with $F(h^1) \leq F(h^0)$ according to Equation (IV.6). Moreover, since h^1 is (partly) computed according to the second minimization problem, we have in view of Proposition 16 that $h^1 \in \mathcal{S}$, a compact set. Hence we can confine our search for a minimum of F to $\mathcal{S}$.

The functions $d_{ij} : x \to Y_{ij}\log\frac{Y_{ij}}{x} - Y_{ij} + x$ (for $x \geq 0$) have a minimum at $x = Y_{ij}$, also if $Y_{ij} = 0$. Choose a sufficiently small positive $\varepsilon < \min\{Y_{ij} : Y_{ij} > 0\}$. Then a minimizer of F has to belong to $\mathcal{F} = \{h \in \mathcal{H} : (T(h)U)_{ij} \geq \varepsilon,$ for all i, j such that $Y_{ij} > 0\}$, and thus finding a minimizer of F can be confined to the compact set $\mathcal{S} \cap \mathcal{F}$. We next show that this set is nonempty, for a judiciously chosen $\varepsilon > 0$. Let $\lambda > 0$ and consider $\lambda\mathbf{1}$. Since $U_{0\bullet} > 0$, we can choose λ such that $\lambda\sum_{k=0}^{N} = S$, hence for this λ we have $\lambda\mathbf{1} \in \mathcal{S}$. Redefine, if necessary, $\varepsilon > 0$ such that also $\varepsilon < \min_j(T(\lambda\mathbf{1})U)_{0j}$, then $\lambda\mathbf{1} \in \mathcal{F}$, showing that $\mathcal{S} \cap \mathcal{F}$ is non-void. Since F is continuous on this set, a minimizer indeed exists and, by the strict convexity of F, it is unique. □

Next we quantify the update gain of Algorithm 19 at each step.

Proposition 22: It holds that

$$\mathcal{I}(Y||T(h^t)U) - \mathcal{I}(Y||T(h^{t+1})U) = \mathcal{I}(\mathbf{Y}^t||\mathbf{Y}^{t+1}) + \mathcal{I}(\mathbf{W}^{t+1}||\mathbf{W}^t).$$

Proof: Recall that $\mathbf{W}^{t+1}$ is the result of the second minimization problem with $\mathbf{Y}^t$ given. Invoking Equation (III.5), we have

$$\mathcal{I}(\mathbf{Y}^t||\mathbf{W}^t) = \mathcal{I}(\mathbf{Y}^t||\mathbf{W}^{t+1}) + \mathcal{I}(\mathbf{W}^{t+1}||\mathbf{W}^t). \qquad \text{(IV.8)}$$

On the other hand, $\mathbf{Y}^{t+1}$ is the result of the first minimization problem with $\mathbf{W}^{t+1}$ given. Hence Equation (III.4) yields

$$\mathcal{I}(\mathbf{Y}^t||\mathbf{W}^{t+1}) = \mathcal{I}(\mathbf{Y}^t||\mathbf{Y}^{t+1}) + \mathcal{I}(\mathbf{Y}^{t+1}||\mathbf{W}^{t+1}). \qquad \text{(IV.9)}$$

Substitution of (IV.9) into (IV.8) yields

$$\mathcal{I}(\mathbf{Y}^t||\mathbf{W}^t) = \mathcal{I}(\mathbf{Y}^t||\mathbf{Y}^{t+1}) + \mathcal{I}(\mathbf{Y}^{t+1}||\mathbf{W}^{t+1}) + \mathcal{I}(\mathbf{W}^{t+1}||\mathbf{W}^t).$$

To finish the proof apply (III.3) to both $\mathcal{I}(\mathbf{Y}^t||\mathbf{W}^t)$ and $\mathcal{I}(\mathbf{Y}^{t+1}||\mathbf{W}^{t+1})$. □

Notice that the update gain is the sum of two non-negative contributions, one from the first minimization and one from the second. The latter term can be given in an alternative expression, which will be useful later (see proof of Lemma 23). We have

$$\begin{aligned}\mathcal{I}(\mathbf{W}^{t+1}||\mathbf{W}^t) &= \sum_{l=0}^{N} U_{l\bullet}\sum_{k=0}^{N-l}(h_k^{t+1}\log\frac{h_k^{t+1}}{h_k^t} - h_k^{t+1} + h_k^t)\\ &= \sum_{k=0}^{N}(\sum_{l=0}^{N-k} U_{l\bullet})\mathcal{I}(h_k^{t+1}||h_k^t)\\ &= \sum_{k=0}^{N}\alpha_{N-k}\mathcal{I}(h_k^{t+1}||h_k^t).\end{aligned}$$

Recall that each h^t belongs to $\mathcal{S}$, since $\sum_{k=0}^{N} h_k^t\alpha_{N-k} = \sum_{ij}^{N} Y_{ij} =: S$. Let

$$p_k^t := \alpha_{N-k}h_k^t/S, \qquad k = 0, 1 \ldots N,$$

then $p^t := (p_0^t, \dots, p_N^t)$ is a probability vector and

$$\begin{aligned} S\mathcal{I}(p^{t+1}||p^t) &= S\sum_k p_k^{t+1}\log\frac{p_k^{t+1}}{p_k^t} \\ &= \sum_k \alpha_{N-k}h_k^{t+1}\log\frac{h_k^{t+1}}{h_k^t} \\ &= \sum_k (\alpha_{N-k}\mathcal{I}(h_k^{t+1}||h_k^t) + p_k^{t+1} - p_k^t) \\ &= \sum_k \alpha_{N-k}\mathcal{I}(h_k^{t+1}||h_k^t). \end{aligned}$$

It follows that

$$\mathcal{I}(\mathbf{W}^{t+1}||\mathbf{W}^t) = S\mathcal{I}(p^{t+1}||p^t). \tag{IV.10}$$

V. Asymptotics

We turn to the asymptotic behaviour of Algorithm 19. The main result of the section is Theorem 25. The preparatory lemmas, much in the spirit of [3], [20], and [23], are typical of this class of problems. See also [18] for a recent example. Our proofs, contrary to the cited references, rely heavily on the optimality results for the partial minimizations (the Pythagoras rules of Lemma 18). As a consequence proofs are short and transparent.

First we use the Pythagoras rules for the updates $\mathbf{Y}^t$ and $\mathbf{W}^{t+1}$. Since $\mathbf{Y}^t = \mathbf{Y}^*(\mathbf{W}^t)$ and $\mathbf{W}^{t+1} = \mathbf{W}^*(\mathbf{Y}^t)$, from Lemma 18 we get the following identities, valid for any $\mathbf{Y} \in \mathcal{Y}$ and $\mathbf{W} \in \mathcal{W}$,

$$\mathcal{I}(\mathbf{Y}||\mathbf{W}^t) = \mathcal{I}(\mathbf{Y}||\mathbf{Y}^t) + \mathcal{I}(\mathbf{Y}^t||\mathbf{W}^t) \tag{V.11}$$

$$\mathcal{I}(\mathbf{Y}^t||\mathbf{W}) = \mathcal{I}(\mathbf{Y}^t||\mathbf{W}^{t+1}) + \mathcal{I}(\mathbf{W}^{t+1}||\mathbf{W}). \tag{V.12}$$

Moreover, from Proposition 15 we also have

$$\mathcal{I}(\mathbf{Y}^t||\mathbf{W}^t) = \mathcal{I}(Y||T(h^t)U). \tag{V.13}$$

Suppose that h^∞ is a fixed point of Algorithm 19, with corresponding $\mathbf{W}^\infty \in \mathcal{W}$ and let $\mathbf{Y}^\infty = \mathbf{Y}^*(\mathbf{W}^\infty)$. Then we also have

$$\mathcal{I}(\mathbf{Y}^\infty||\mathbf{W}^\infty) = \mathcal{I}(Y||T(h^\infty)U). \tag{V.14}$$

For simplicity throughout this section we assume, without loss of generality, that $S = \sum_{ij} Y_{ij} = 1$, see Remark 3. Then we have that $p_k^t = \alpha_{N-k}h_k^t$. The update equation (IV.2) is equivalent to

$$p_k^{t+1} = \mathbf{Y}_{\bullet k}^t. \tag{V.15}$$

In correspondence to the fixed point h^∞, let us define p^∞ as $p_k^\infty = \alpha_{N-k}h_k^\infty$, then

$$p_k^\infty = \mathbf{Y}_{\bullet k}^\infty. \tag{V.16}$$

Since p^t and p^∞ are probability vectors, by the lumping property of the I-divergence, see [6, Lemma 4.1], it holds that

$$\mathcal{I}(p^\infty||p^{t+1}) \le \mathcal{I}(\mathbf{Y}^\infty||\mathbf{Y}^t). \tag{V.17}$$

We will also need the following

Lemma 23: Limit points of the sequence (h^t) are fixed points of Algorithm 19.

Proof: Since the divergence $\mathcal{I}(Y|T(h^t)U)$ is decreasing in t, it has a limit. Hence we obtain from Proposition 22 that $\mathcal{I}(\mathbf{W}^{t+1}||\mathbf{W}^t) \to 0$. From (IV.10) it follows that $\mathcal{I}(p^{t+1}||p^t) \to 0$. Suppose that h^∞ is a limit point of (h^t), then p^∞ is a limit point of (p^t). Let $\tilde{h}$ be the iteration of the algorithm if h^t is replaced with h^∞ and $\tilde{p}$ be its counterpart, so $\tilde{h} = I(h^\infty)$. By continuity of $I(\cdot)$, which follows from the continuity of the G_k, we then get $\mathcal{I}(\tilde{p}||p^\infty) = 0$ and hence $\tilde{p} = p^\infty$, which entails $\tilde{h} = h^\infty$, so h^∞ is a fixed point of the algorithm. □

We are now ready to prove

Lemma 24: Let h^∞ be a limit point of Algorithm 19, then $\mathcal{I}(p^\infty||p^t)$ is decreasing in t.

Proof: From (V.17) and (V.11) with $\mathbf{Y} = \mathbf{Y}^\infty$ we have

$$\begin{aligned} \mathcal{I}(p^\infty||p^{t+1}) &\le \mathcal{I}(\mathbf{Y}^\infty||\mathbf{Y}^t) \\ &= \mathcal{I}(\mathbf{Y}^\infty||\mathbf{W}^t) - \mathcal{I}(\mathbf{Y}^t||\mathbf{W}^t). \end{aligned}$$

Applying the second Pythagorean rule (III.5) to the first term in the right hand side, with $\mathbf{Y} = \mathbf{Y}^\infty$ and hence $\mathbf{W}^* = \mathbf{W}^\infty$, we get

$$\mathcal{I}(Y^\infty||\mathbf{W}^t) = \mathcal{I}(\mathbf{Y}^\infty||\mathbf{W}^\infty) + \mathcal{I}(\mathbf{W}^\infty||\mathbf{W}^t).$$

By Lemma 23 a limit point of the sequence (h^t) is also a fixed point of the algorithm. Hence we have $\mathbf{Y}^\infty = \mathbf{Y}^*(\mathbf{W}^\infty)$ and we deduce from Proposition 15 that $\mathcal{I}(\mathbf{Y}^\infty||\mathbf{W}^\infty) = \mathcal{I}(Y||T(h^\infty)U)$. A direct computation, similar to that leading to (IV.10), yields $\mathcal{I}(\mathbf{W}^\infty||\mathbf{W}^t) = \mathcal{I}(p^\infty||p^t)$. By also using (V.13), we finally obtain

$$\begin{aligned} \mathcal{I}(p^\infty||p^{t+1}) &\le \mathcal{I}(p^\infty||p^t) - \mathcal{I}(Y||T(h^t)U) \\ &\quad + \mathcal{I}(Y||T(h^\infty)U) \\ &\le \mathcal{I}(p^\infty||p^t), \end{aligned}$$

since Proposition 22 implies that $\mathcal{I}(Y||T(h^t)U)$ is decreasing in t and hence $\mathcal{I}(Y||T(h^\infty)U) \le \mathcal{I}(Y||T(h^t)U)$. □

The main result on the asymptotic behavior of Algorithm 19 is given in the next theorem.

Theorem 25: The sequence of iterates h^t converges to a limit h^∞ which minimizes $h \to \mathcal{I}(Y||T(h)U)$.

Proof: Since all h^t belong to the simplex, see property 5 in the list above, which is compact, the sequence (h^t) has a convergent subsequence, $h^{t_n} \to h^\infty$, for some h^∞. For the corresponding sequence (p^t) it holds that $p^{t_n} \to p^\infty$. By continuity of the I-divergence in the second argument, $\mathcal{I}(p^\infty||p^{t_n}) = \sum_{k:p_k^\infty>0} p_k^\infty \log\frac{p_k^\infty}{p_k^{t_n}}$, we then have $\mathcal{I}(p^\infty||p^{t_n}) \to 0$. The monotonicity result of Lemma 24 then yields $\mathcal{I}(p^\infty||p^t) \to 0$, which implies $p^t \to p^\infty$, equivalently $h^t \to h^\infty$. Recall from Lemma 23 that the limit h^∞ is a fixed point of the algorithm. Hence we have from (IV.7)

$$h_k^\infty = h_k^\infty\left(1 - \frac{\nabla F(h^\infty)_k}{\sum_{l=0}^{N-k} U_{l\bullet}}\right).$$

If $h_k^\infty > 0$, then $\nabla F(h^\infty)_k = 0$. We now consider the case where some $h_k^\infty = 0$. Consider (IV.4) and (IV.5), and write it as the product

$$h_k^{t+1} = h_k^t G_k(h^t).$$

It follows that $h_k^{t+1} = h_k^0 \prod_{j=0}^t G_k(h^j)$. Since we have convergence of the h_k^t, we must have $G_k(h^\infty) \leq 1$, otherwise the product would explode. Indeed, suppose $G_k(h^\infty) > 1$, hence $G_k(h^\infty) > 1 + \varepsilon$ for some $\varepsilon > 0$. Continuity of $G_k(\cdot)$ at h^∞, which holds since $F(h^\infty) < \infty$, yields $\lim_{t\to\infty} G_k(h^t) \geq 1 + \varepsilon$, hence eventually $G_k(h^t) > 1 + \varepsilon/2$, which contradicts that the h^t convergence. We conclude $\nabla F(h^\infty)_k \geq 0$. Altogether, we obtain that for the limit h^∞ the Kuhn-Tucker conditions (II.8), (II.9) for F are satisfied. Since these conditions are also sufficient in view of the convexity of F, [24, Th. 2.9], h^∞ minimizes F. □

Although Theorem 25 establishes convergence of the algorithm, it does not give any information on the rate of convergence. In fact, it is possibly a hard grind to get results in this direction. The following example shows that even in a simple case, depending on the exact circumstances, different rates may occur.

Example 26: Here we continue the toy Example 9. The update equation (IV.4) for h_1^t becomes

$$h_1^{t+1} = h_1^t \frac{y_1}{h_0^t u_1 + h_1^t u_0}.$$

Assume again the second case, $y_1u_0 - y_0u_1 < 0$, and $y_1 > 0$ to avoid a trivial recursion. Choose $\varepsilon \in (0, \frac{y_0u_1 - y_1u_0}{u_0+u_1})$. We know from Theorem 25 that $h_0^t \to \frac{y_0+y_1}{u_0+u_1}$ and $h_1^t \to 0$. Hence $h_0^t u_1 + h_1^t u_0 \to \frac{y_0+y_1}{u_0+u_1}u_1$, and thus for some $t_0 > 0$ and $t \leq t_0$ one has $h_0^t u_1 + h_1^t u_0 > \frac{y_0+y_1}{u_0+u_1}u_1 - \varepsilon$ and therefore

$$h_1^{t+1} \leq h_1^t \frac{y_1(u_0+u_1)}{(y_0+y_1)u_1 - \varepsilon(u_0+u_1)} =: h_k^t g_\varepsilon.$$

Hence we have, at least asymptotically, convergence of $h_1^t \to 0$ at an exponential rate, since $g_\varepsilon < 1$ by the choice of ε. Note that, in the notation of the proof of Theorem 25, we have $G_1(h^\infty) = \frac{y_1(u_0+u_1)}{(y_0+y_1)u_1} = g_0 < 1$.

The convergence of the h_0^t could possibly be slower than exponential, since $G_0(h^\infty) = 1$. This will be investigated now. The update equation for h_0^1 reads

$$h_0^{t+1} = \frac{y_0}{u_0+u_1} + h_0^t \frac{y_1u_1}{(u_0+u_1)(h_0^t u_1 + h_1^t u_0)}.$$

Let $v_0^t := h_0^t - h_0^\infty = h_0^t - \frac{y_0+y_1}{u_0+u_1}$. Tedious computations lead to the recursion for v_0^t,

$$v_0^{t+1} = -\frac{y_1u_0}{(u_0+u_1)(h_0^t u_1 + h_1^t u_0)} h_1^t.$$

Since the factor in front of h_1^t stabilizes around its limit value $-\frac{y_1u_0}{u_1(y_0+y_1)}$ and h_1^t converges exponentially fast to zero, the latter property is shared by v_0^t.

Next we investigate the case where an exact solution exists, $y_1u_0 - y_0u_1 \geq 0$. Let $v_k^t = h_k^t - h_k^\infty$ and $y_1^t = h_0^t u_1 + h_1^t u_0$. Putting the v_k^t in a vector $V^t = (v_0^t, v_1^t)^\top$, one arrives after more tedious computations at the recursion

$$\begin{aligned} V^{t+1} &= \frac{u_1}{y_1^t} \begin{pmatrix} \frac{u_0}{u_0+u_1} \\ -1 \end{pmatrix} \begin{pmatrix} h_1^\infty & -h_0^\infty \end{pmatrix} V_t \\ &\approx \frac{u_1}{y_1} \begin{pmatrix} \frac{u_0}{u_0+u_1} \\ -1 \end{pmatrix} \begin{pmatrix} h_1^\infty & -h_0^\infty \end{pmatrix} V_t =: AV^t. \end{aligned}$$

Clearly the matrix A in front of V_t at the right hand side is singular. Its eigenvalues are 0 and $\frac{u_1(y_0+y_1)}{(u_0+u_1)y_1}$, where the latter one is smaller than 1 if we assume the strict inequality $y_1u_0 - y_0u_1 > 0$. Hence, also here one has exponential stability.

What is left is the case $y_1u_0 - y_0u_1 = 0$. Now the matrix A has an eigenvalue equal to 1. We investigate the exact equation for V^t in this case,

$$V^{t+1} = \frac{u_1}{y_1^t} \begin{pmatrix} 0 & -\frac{y_0}{u_0+u_1} \\ 0 & \frac{y_0}{u_0} \end{pmatrix} V^t.$$

It follows that for $t \geq 1$

$$v_0^t = -\frac{u_0}{u_0+u_1} v_1^t,$$

and hence $y_1^t = y_1 + \frac{u_0^2}{u_0+u_1} v_1^t$. This leads to the recursion

$$v_1^{t+1} = \frac{y_1}{y_1 + wv_1^t} v_1^t,$$

with $w = \frac{u_0^2}{u_0+u_1}$. This recursion has the solution

$$v_1^t = \frac{v_1^0 y_1}{wv_1^0 t + y_1}.$$

We conclude that now v_1^t and hence also v_0^t tend to zero at rate $1/t$ instead of exponentially.

VI. Statistics

In the previous sections we focused on the minimization of $\mathcal{I}(Y||T(h)U)$, 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. Recall that $Y, U \in \mathbb{R}^{(N+1)\times m}$. We will give limit results for the optimizing $h = h^m$, when $m \to \infty$ and the pair of columns (Y^i, U^i) of Y, U $(i = 1, \ldots, m)$ form an i.i.d. sample. For each fixed m, Algorithm 19 can be used to find h^m, which now becomes a random vector as well.

Write $\mathcal{I}(Y||T(h)U) = \sum_{i=1}^m \mathcal{I}(Y^i||T(h)U^i)$, with the Y^i and U^i the columns of the matrices Y and U respectively. We assume that the pairs (Y^i, U^i) are i.i.d. In what follows, we let, *contrary to the previously employed notation*, (Y, U) be a random vector that has the same distribution as each of the (Y^i, U^i). Moreover we assume for the entries Y_j of Y and $(T(h)U)_j$ of $T(h)U$ the 'true' relationship

$$Y_j = (T(h^*)U)_j \delta_j, \tag{VI.1}$$

where h^* is an interior point and the $\delta_j \geq 0$ are assumed to be independent of U. In the present context it is more appropriate to have a multiplicative disturbance δ_j, than an additive one as in e.g. least squares estimation.

The displayed relationship can be summarized as

$$Y = \Delta T(h^*)U,$$

where Δ is diagonal with entries δ_j, and U and Δ independent. We impose $\mathbb{E}\,\Delta = I$, the identity matrix, so $\mathbb{E}\,\delta_j = 1$.

Lemma 27: Assume the model (VI.1), $\mathbb{E}\,U_j < \infty$, $\mathbb{E}\,\delta_j = 1$, *and* $\mathbb{E}\,\delta_j \log \delta_j < \infty$. *Then it holds that*

$$\mathbb{E}\,\mathcal{I}(Y||T(h)U) = \mathbb{E}\,\mathcal{I}(T(h^*)||T(h)U) + \sum_j (\mathbb{E}\,(T(h^*)U)_j \mathbb{E}\,(\delta_j \log \delta_j).$$

Proof: Let us first compute $\mathbb{E}\,\mathcal{I}(Y_j||(T(h)U)_j)$. We get, using the definition of divergence and Equation (VI.1),

$$\begin{aligned}\mathbb{E}\,\mathcal{I}(Y_j||(T(h)U)_j) &= \mathbb{E}\,\{Y_j \log \frac{Y_j}{(T(h)U)_j} - Y_j + (T(h)U)_j\} \\ &= \mathbb{E}\,\{(T(h^*)U)_j\delta_j \log \frac{(T(h^*)U)_j\delta_j}{(T(h)U)_j} \\ &\quad - (T(h^*)U)_j\delta_j + (T(h)U)_j\} \\ &= \mathbb{E}\,\{(T(h^*)U)_j\delta_j (\log \frac{(T(h^*)U)_j}{(T(h)U)_j} + \log \delta_j) \\ &\quad - (T(h^*)U)_j\delta_j + (T(h)U)_j\}.\end{aligned}$$

Using the independence of U and δ_j, we proceed by rewriting the last expression as

$$\mathbb{E}\,(T(h^*)U)_j \log \frac{(T(h^*)U)_j}{(T(h)U)_j}\mathbb{E}\,\delta_j + \mathbb{E}\,(T(h^*)U)_j\mathbb{E}\,(\delta_j \log \delta_j) - \mathbb{E}\,(T(h^*)U)_j\mathbb{E}\,\delta_j + \mathbb{E}\,(T(h)U)_j.$$

Recalling $\mathbb{E}\,\delta_j = 1$, we obtain that this equals

$$\mathbb{E}\,(T(h^*)U)_j \log \frac{(T(h^*)U)_j}{(T(h)U)_j} + \mathbb{E}\,(T(h^*)U)_j\mathbb{E}\,(\delta_j \log \delta_j) - \mathbb{E}\,(T(h^*)U)_j + \mathbb{E}\,(T(h)U)_j$$

Rearranging terms, we obtain for this, using again the definition of divergence,

$$\mathbb{E}\,\mathcal{I}((T(h^*)U)_j||(T(h)U)_j) + \mathbb{E}\,(T(h^*)U)_j\mathbb{E}\,(\delta_j \log \delta_j).$$

Summation over j yields the result. □

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

Proposition 28: Let $\mathbb{P}(U_0 > 0) > 0$ *and* $\mathbb{E}\,U_j^2 < \infty$ *for all* j. *The limit criterion* $h \mapsto \mathbb{E}\,\mathcal{I}(Y||T(h)U)$ *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: The proof of strict convexity is similar to the proof of Lemma 6. We show that the Hessian $H(h)$ 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

$$H(h)_{kl} = \mathbb{E} \sum_j \frac{(T(h^*)U)_j}{(T(h)U)_j^2} U_{j-k}U_{j-l}.$$

Clearly, $H(h)$ is finite in a neighborhood of h^*. Hence

$$x^\top H(h)x = \mathbb{E} \sum_j \frac{(T(h^*)U)_j}{(T(h)U)_j^2} (U * x)_j^2.$$

Hence the expression inside the expectation can only be zero if $U * x = 0$ a.s. Using $\mathbb{P}(U_0 > 0) > 0$, we argue as in the proof of Lemma 6 to deduce that $x = 0$ iff $x^\top H(h)x = 0$. Clearly, the limit criterion has a minimum equal to zero at $h = h^*$. Conversely, $\mathbb{E}\,\mathcal{I}(T(h^*)U||T(h)U) = 0$ iff $\mathcal{I}(T(h^*)U||T(h)U) = 0$ a.s., which happens iff $T(h^*)U = T(h)U$ a.s. Writing this equality elementwise, $(T(h^*)U)_j = (T(h)U)_j = 0$, we obtain $h = h^*$ under the condition that $\mathbb{P}(U_0 > 0) > 0$. We conclude that $h = h^*$ is the unique minimizer if $\mathbb{P}(U_0 > 0) > 0$. □

Proposition 29: Let $\mathbb{P}(U_0 > 0) > 0$ *and* $\mathbb{E}\,U_j^2 < \infty$ *for all* j, *moreover assume that* h^* *is an interior point. The estimators* $\hat{h}^m$, *defined as the minimizers of the objective function* $\sum_{i=1}^m \mathcal{I}(Y^i||T(h)U^i)$ *are consistent. Moreover, this sequence is asymptotically normal, for some positive definite* $\Sigma \in \mathbb{R}^{(N+1)\times(N+1)}$ *we have* $\sqrt{m}(\hat{h}^m - h^*) \xrightarrow{d} N(0, \Sigma)$.

Proof: The limit criterion $h \mapsto \mathbb{E}\,\mathcal{I}(Y||T(h)U)$ is strictly convex, therefore from [22, Problem 5.27] we conclude that the conditions of [22, Th. 5.7] are satisfied and consistency follows. To show that the estimators $\hat{h}^m$ are asymptotically normal with covariance function as given in [22, Th. 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 28 taking $h = h^*$. □

VII. Numerical Experiments

In this section we provide the results of three numerical experiments that illustrate the behaviour of Algorithm 19. The first two examples investigate whether the algorithm is capable of retrieving the *true* parameter vector h^*, when the output data are actually generated by h^*. The two examples refer to perfect and noisy observations respectively. In the third example 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.

We have observed experimentally that the iterative algorithm converges very fast, which led us to cut to 50 the number of iterations in all examples. For the sake of graph readability in the examples reproduced here the length $N + 1$ of the individual time series is limited to 6, leading to FIR impulse responses h of length 6. Each of the three graphs shows the iterates h_k^t $(k = 0, \ldots, 5)$ 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 colors representing the different k's. In Figures 1 and 2 the diamonds at the right end of the graph indicate the true h^* target values. The precise features underlying the experiments are further detailed below.

A. True Convolutional System With Perfect Observations

In this example we set $m = 10$, $N = 5$. The elements of the true vector h^* (the target values of the algorithm) and of the input matrix U have been randomly generated from a uniform distribution on the interval [0.1, 10], and the output computed as $Y = T(h)U$, see (II.2). The algorithm has been initialized at a randomly chosen $h^0 > 0$ and run for 50 iterations. We observe from Figure 1 fast convergence of the iterates h_k^t to their target values, as should be expected from Remark 8 and Theorem 25. The SCILAB implementation

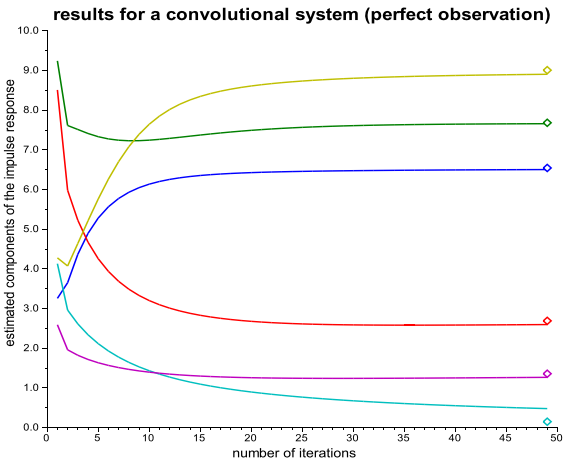

Fig. 1. Perfect observations, $m = 10$, $N = 5$.

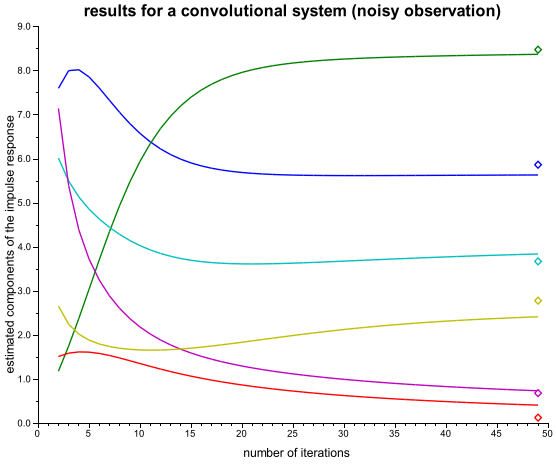

Fig. 2. Noisy observations, $m = 100$, $N = 5$.

of the algorithm on a laptop produced this figure almost instantly.

B. Noisy Observations of True Convolutional System

In the second example we consider data generated by the statistical model of Section VI, i.e. noisy observations of a true convolutional system. We set $N = 5$ and $m = 100$. Recall that here m represents the sample size, and that we want to assess the consistency of the estimates produced by the algorithm. The true h^* and the inputs U have been generated as in the previous example. The disturbances δ, present in every element of Y, see (VI.1), have been generated from a uniform distribution on [0.6, 1.4]. Note that this distribution has mean 1, in agreement with the modelling assumptions of Section VI. We observe from Figure 2 that the estimates are converging towards the true h^*, as predicted by Proposition 29. The SCILAB implementation of the algorithm on a laptop produced Figure 2 within a few seconds.

C. Bold Approximation

In this example there is no true underlying system, i.e. no true impulse response h^*, therefore we only deal with an approximation problem. We set $N = 5$ and $m = 10$.

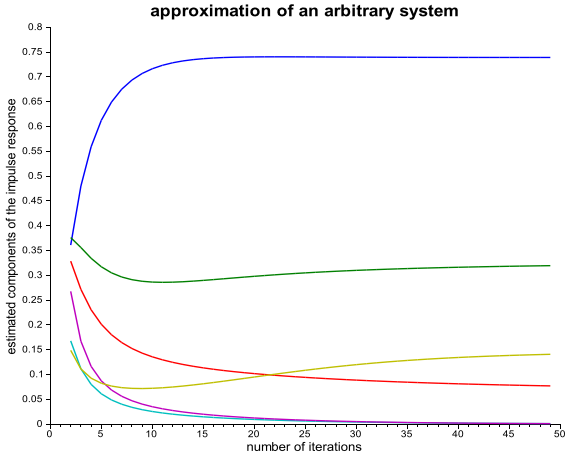

Fig. 3. Arbitrary system, $m = 10$, $N = 5$.

The matrices of the inputs and of the outputs $U, Y \in \mathbb{R}^{6\times 10}$ were randomly generated from the same uniform distribution as in the first example. The aim is to find the vector h which yields the best convolutional approximation to Y. We conclude from Figure 3 that the algorithm quickly stabilises. In agreement with Remark 10 we note that two of the components of h converge to zero.

VIII. Conclusions

We posed the nonparametric approximation problem for scalar nonnegative input/output systems via finite impulse response convolutions, based on repeated 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. Numerical experiments confirm the asymptotic results and exhibit fast convergence to the minimizer of the objective function.

References

[1] B. D. O. Anderson, M. Deistler, L. Farina, and L. Benvenuti, "Nonnegative realization of a linear system with nonnegative impulse response," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 43, no. 2, pp. 134–142, Feb. 1996.

[2] L. Benvenuti and L. Farina, "A tutorial on the positive realization problem," *IEEE Trans. Autom. Control*, vol. 49, no. 5, pp. 651–664, May 2004.

[3] T. M. Cover, "An algorithm for maximizing expected log investment return," *IEEE Trans. Inf. Theory*, vol. 30, no. 2, pp. 369–373, Mar. 1984.

[4] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Ann. Probab.*, vol. 3, no. 1, pp. 146–158, Feb. 1975.

[5] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *Ann. Statist.*, vol. 19, no. 4, pp. 2032–2066, Dec. 1991.

[6] I. Csiszár and P. C. Shields, "Information theory and statistics: A tutorial," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 4, pp. 417–528, 2004.

[7] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statist. Decisions*, supplement issue 1, pp. 205–237, 1984.

[8] D. A. Dewasurendra, P. H. Bauer, and K. Premaratne, "Evidence filtering," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5796–5805, Dec. 2007.
[9] L. Farina and L. Benvenuti, "Positive realizations of linear systems," *Syst. Control Lett.*, vol. 26, no. 1, pp. 1–9, Sep. 1995.
[10] L. Farina and S. Rinaldi, *Positive Linear Systems: Theory and Applications*. New York, NY, USA: Wiley, 2000.
[11] L. Finesso and P. Spreij, "Nonnegative matrix factorization and I-divergence alternating minimization," *Linear Algebra Appl.*, vol. 416, nos. 2–3, pp. 270–287, Jul. 2006.
[12] L. Gurvits, R. Shorten, and O. Mason, "On the stability of switched positive linear systems," *IEEE Trans. Autom. Control*, vol. 52, no. 6, pp. 1099–1103, Jun. 2007.
[13] C.-J. Lin, "On the convergence of multiplicative update algorithms for nonnegative matrix factorization," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1589–1596, Nov. 2007.
[14] Y. Liu and P. H. Bauer, "Frequency domain limitations in the design of nonnegative impulse response filters," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4535–4546, Sep. 2010.
[15] Y. Liu and P. H. Bauer, "Fundamental properties of non-negative impulse response filters," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 6, pp. 1338–1347, Jun. 2010.
[16] B. Nagy and M. Matolcsi, "Minimal positive realizations of transfer functions with nonnegative multiple poles," *IEEE Trans. Autom. Control*, vol. 50, no. 9, pp. 1447–1450, Sep. 2005.
[17] B. Nagy, M. Matolcsi, and M. Szilvási, "Order bound for the realization of a combination of positive filters," *IEEE Trans. Autom. Control*, vol. 52, no. 4, pp. 724–729, Apr. 2007.
[18] J. A. O'Sullivan and J. Benac, "Alternating minimization algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 26, no. 3, pp. 283–297, Mar. 2007.
[19] Z. Shu, J. Lam, H. Gao, B. Du, and L. Wu, "Positive observers and dynamic output-feedback controllers for interval positive linear systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 10, pp. 3209–3222, Nov. 2008.
[20] D. L. Snyder, T. J. Schulz, and J. A. O'Sullivan, "Deblurring subject to nonnegativity constraints," *IEEE Trans. Signal Process.*, vol. 40, no. 5, pp. 1143–1150, May 1992.
[21] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. New York, NY, USA: Wiley, 1977.
[22] A. W. van der Vaart, *Asymptotic Statistics* (Cambridge Series in Statistical and Probabilistic Mathematics), vol. 3. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[23] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Statist. Assoc.*, vol. 80, no. 389, pp. 8–20, 1985.
[24] W. I. Zangwill, *Nonlinear Programming: A Unified Approach*. Upper Saddle River, NJ, USA: Prentice-Hall, 1969.

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.

Peter Spreij obtained his MSc 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), and since 1999 with the Korteweg-de Vries Institute for Mathematics (Universiteit van Amsterdam). 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.