Blind Estimation of Row Relative Degree Via Constrained Mutual Information Minimization

Jani Even* and Kenji Sugimoto

Graduate School of Information Science,
Nara Institute of Science and Technology,
Nara, 630-0192, Japan
even@is.naist.jp, kenji@is.naist.jp

Abstract. This paper studies a method for blind (input signals being unknown) estimation of the row relative degrees of a system non invertible at infinity. The proposed method uses a blind signal deconvolution scheme: A system, called demixer, is applied to the observed signals and is updated in order to minimize the mutual information. A key point is that the demixer is constrained to be biproper whereas the system is not invertible at infinity, consequently deconvolution is not achievable. But, the row relative degrees can be obtained in two steps: i) minimizing the mutual information at the output of the demixer. ii) using second order statistics of the obtained outputs. Although convergence has not yet been proved, extensive numerical simulation shows the effectiveness of this method.

1 Introduction

Blind signal separation has recently attracted much attention, and many efficient statistical methods have appeared in the last decade. In blind signal separation, the focus is on the recovery of unknown signals using only observed mixtures of these signals. During recovery, the dynamical system corresponding to the mixture is also partially identified (see book [1] for details). This approach is hence promising in control engineering too, because we often encounter the case where some of the input signals are unavailable due to noise, saturation, or failure; see, for example, [4].

Since control systems are often strictly proper, i.e. non invertible at infinity, many theoretical developments assume that the degrees of the rational transfer matrix representing the system are partially known. Some of the most important parameters are the row relative degrees (see Sect.2.2 and [2, 5]). However few methods enable the determination of these parameters, thus trial and error scheme is often used in practical applications.

This paper studies a method that enables the blind estimation of the row relative degrees of a system. However, the blind treatment has a cost in term of

* This work is supported in part by Scientific Research Grant-In-Aid from the Japan Society for the Promotion of Science under "JSPS Postdoctoral Fellowship for Foreign Researchers" program.

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 352–359, 2006.

indeterminacy: Only the difference of relative degree among the rows is obtained. The blind estimation is achieved in two steps:

i) Minimizing the mutual information with some structural constraints on the demixer: It does not result in a blind deconvolution because of the constraints,
ii) exploiting the second order statistics of the output signals obtained after i).

Although convergence has not yet been proved, extensive numerical simulation shows the effectiveness of this method.

The proposed method provides a good insight in the system's structure because the approximate inverse of the system is factorized in two terms, one of those corresponding to the row relative degrees information. Traditional blind signal separation methods do not perform such a factorization because they focus on the recovery of sources. Since the use of blind deconvolution techniques based on mutual information is not widespread yet in control community, this paper is also an attempt to show their potential in this field.

Some notations below will be used in this paper: For a matrix A: $A^{(i,:)}$ denotes the i^{th} row of A and $A^{(i,j)}$ is the element of A in the i^{th} row and the j^{th} column.

$\mathcal{O}_{m\times p}$ is a null matrix of size $m \times p$ and $\mathcal{I}_m$ is the identity matrix of size m. $\delta_{i,j}$ is the Kronecker's delta equal to 1 if $i = j$ and 0 if $i \neq j$. All transfer matrices are assumed to be square of size $m \times m$ with $m > 1$.

2 Preliminaries

2.1 Blind Deconvolution

Throughout the paper we treat discrete-time signals. The goal of blind deconvolution is to recover the unknown input signals, called source signals, $\mathbf{s}(t) = [s_1(t), \ldots, s_m(t)]^T$ applied to an unknown transfer matrix $H(z)$, called "mixer", when only the observed signals $\mathbf{v}(t) = [v_1(t), \ldots, v_m(t)]^T = H(z)\mathbf{s}(t)$ are available. Throughout the paper, $H(z)$ is assumed to be stable rational, proper, and of minimal phase with full normal rank.

A transfer matrix $W(z)$, called "demixer", is applied to the observations in order to obtain the estimates $\mathbf{y}(t) = [y_1(t), \ldots, y_m(t)]^T$ of the sources as illustrated in Fig. 1. A common hypothesis used in blind signal deconvolution is to assume that each source $s_i(t)$ is an independent identically distributed (*i.i.d.*) process and that all sources are mutually statistically independent. It is also necessary to assume that at most one of the sources has a Gaussian distribution. With these hypotheses, blind signal deconvolution can be achieved by adapting the demixer $W(z)$ in order to obtain signals $\mathbf{y}(t)$ whose components are mutually statistically independent [3].

Even under the above conditions the blind identification of the transfer matrix has indeterminacies: We can detect neither a permutation of the outputs, the delay, nor the scale of each output. This is formulated by the relation:

$$W(z)H(z) = P\ \Lambda(z), \tag{1}$$

where P is a permutation matrix and $\Lambda(z)$ is a diagonal transfer matrix with entries of the form $\alpha_i\ z^{-\lambda_i}$.

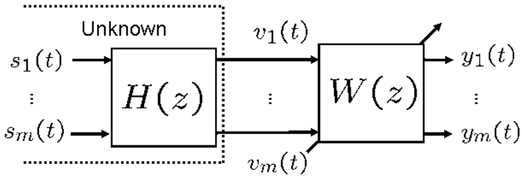

Fig. 1. Blind signal deconvolution scheme

2.2 Row Relative Degree

A transfer matrix $H(z)$ is said to be biproper, i.e. invertible at infinity, if the first matrix that appears in its power series expansion $H(z) = H_0 + H_1 z^{-1} + \dots$ is invertible. A transfer matrix $H(z)$ of size $m \times m$ is said to be non invertible at infinity if $\lim_{|z| \to \infty} H(z) = M$ is not an invertible matrix.

The relative degree of a rational polynomial fraction $N(z)/D(z)$ is $\deg D(z) - \deg N(z)$. The matrix of relative degrees of a polynomial transfer matrix $H(z) = \{H_{ij}(z)\}_{i,j \in [1,m]}$ is $\{d_{ij}\}_{i,j \in [1,m]}$ with d_{ij} relative degree of $H_{ij}(z)$. The i^{th} row relative degree of $H(z)$ is $d_i = \min_j d_{ij}$.

In the remainder of the paper, we assume that $H(z)$ is non invertible at infinity. We further assume that $H(z)$ is such that

$$\operatorname{diag}\left(z^{d_1} \dots z^{d_m}\right) H(z) \tag{2}$$

is invertible at infinity. Namely, shifting each output signal by a number of samples equal to the row relative degree results in a biproper transfer matrix.

Considering the two transfer matrices $H(z)$ and $\widetilde{H}(z) = H(z)\Lambda(z)$, where $\Lambda(z) = \operatorname{diag}(z^{-\alpha_j})_{j \in [1,m]}$, their row relative degrees are $d_i = \min_j d_{ij}$ and $\tilde{d}_i = \min_j (d_{ij} + \alpha_j)$, respectively, and are different. However, the row relative degree differences $r_{pq} = d_p - d_q$ and $\tilde{r}_{pq} = \tilde{d}_p - \tilde{d}_q$ are same if either one of the conditions below is fulfilled

i) $\alpha_j = \alpha$ for all j: $\tilde{r}_{pq} = \min_j (d_{pj}) + \alpha - \min_j (d_{qj}) - \alpha = r_{pq}$,
ii) $d_{ij} = d_i$ for all j: $\tilde{r}_{pq} = d_p + \min_j \left(\alpha_{\sigma_p(j)}\right) - d_q - \min_j \left(\alpha_{\sigma_q(j)}\right) = r_{pq}$.

In the case ii), the indeterminacies of blind deconvolution in Eq.(1) do not prevent to estimate the row relative degree differences (permutation of column has no effect on the row relative degrees).

3 Main Results

3.1 Adaptation of $W(z)$

The proposed method exploits a classical blind deconvolution scheme. However structures of the mixer and demixer are incompatible: The demixer is constrained to be biproper and thus cannot be the inverse of the mixer which is non invertible at infinity.

The demixer $W(z)$ is a finite impulse response (FIR) system

$$W(z) = W_0 + W_1 z^{-1} + \ldots + W_l z^{-l}.$$

The matrices W_j are adapted with a batch algorithm based on the on-line method proposed in [1]. The method minimizes the mutual information of the outputs

$$\mathrm{MI}[\mathbf{y}(t)] = -H(\mathbf{y}(t)) + \sum_{i=1}^{m} H(y_i(t)),$$

where $H(X) = -\int P_X(X) \ln P_X(X) dX$ is the entropy. Adaptation rule is derived from the relative gradient of $\mathrm{MI}[\mathbf{y}(t)]$.

Let $W_i(k)$ denote the matrix W_i at iteration k and $\mu(k)$ be the positive adaptation step used at iteration k. The adaptation law is

$$W_i(k+1) = W_i(k) - \mu(k) \Delta W_i(k),$$

$$\text{with} \quad \Delta W_i(k) = \sum_{j=0}^{i} (\delta_{j,0} \mathcal{I}_m - < \psi(\mathbf{y}(t)) \mathbf{y}^T(t-j) >_t) W_{i-j}(k)$$

$$\Delta W_0(k) = (\mathcal{I}_m - < \psi(\mathbf{y}(t)) \mathbf{y}^T(t) >_t) \ W_0(k)$$

where $< . >_t$ denotes time average on the data block. $\psi(.) = [\psi_1(.) \ldots \psi_m(.)]^T$ is a vector containing approximations of the score functions associated with the source signals: $\psi_{\text{real}}(s) = -\partial ln[P_s(s)]/\partial s$ (see [1] for derivation and discussion on approximation of score). In a single iteration k, the whole block of signal $\mathbf{y}(t)$ has to be computed (hence this is a batch algorithm).

At initialization $W(z) = \mathcal{I}_m$. An important property of the adaptation law is that when $W(z)$ is initialized to a biproper filter, it remains biproper during adaptation [1]. Hence blind deconvolution cannot be attained because $H(z)$ is not biproper and $\mathrm{MI}[\mathbf{y}(t)]$ reaches a local minimum by the above adaptation.

For simplicity, we illustrate our discussion with 2×2 transfer matrices. Assume that first and second rows of $H(z)$ have relative degrees $d_1 = 0$ and $d_2 = d > 0$, respectively. In this case, the power series expansion of $H(z)$ is

$$H(z) = \left[\frac{H_0^{(1,:)}}{\mathcal{O}_{1\times 2}} \right] + \ldots + \left[\frac{H_{d-1}^{(1,:)}}{\mathcal{O}_{1\times 2}} \right] z^{-(d-1)} + H_d z^{-d} + H_T z^{-(d+1)} + \ldots$$

Conjecture. *If the demixer $W(z)$ is initialized to identity and adapted with the above adaptation law, then the cascade $G(z) = W(z)H(z)$ (for an even d) converges to*

$$G(z) \approx \left[\frac{G_0^{(1,:)}}{\mathcal{O}_{1\times 2}} \right] + \left[\frac{G_1^{(1,:)}}{\mathcal{O}_{1\times 2}} \right] z^{-1} + \ldots + \left[\frac{G_{r-1}^{(1,:)}}{\mathcal{O}_{1\times 2}} \right] z^{-(r-1)}$$
$$+ \left[\frac{G_r^{(1,:)}}{G_r^{(2,:)}} \right] z^{-r} + \left[\frac{\mathcal{O}_{1\times 2}}{G_{r+1}^{(2,:)}} \right] z^{-r-1} + \ldots + \left[\frac{\mathcal{O}_{1\times 2}}{G_d^{(2,:)}} \right] z^{-d} \quad (3)$$

with the integer $r = d/2$ and the two rows of G_r being orthogonal.

Let us explain a ground of this conjecture. First note that minimizing MI[$\mathbf{y}(t)$] is to force that $G(z) = W(z)H(z)$ has statistically orthogonal outputs that have non Gaussian distributions [3]. Namely, both of the following conditions are to be fulfilled:

i) The two rows $G_j^{(1,:)}$ and $G_j^{(2,:)}$ of each matrix G_j are orthogonal.
ii) The vectors $[G_1^{(i,:)}, \ldots, G_l^{(i,:)}]$ for $i = 1, 2$ have only one non null element.

Conditions i) and ii) cannot be satisfied completely because the demixer is biproper, but the adaptation tries to attain them. In particular, the first matrix W_0 is solution of:

$$W_0 \left[\frac{H_0^{(1,:)}}{\mathcal{O}_{1\times m}} \right] = \left[\frac{W_0^{(1,1)} H_0^{(1,:)}}{W_0^{(2,1)} H_0^{(1,:)}} \right] = G_0.$$

At initialization $W_0^{(1,1)} = 1$ and $W_0^{(2,1)} = 0$, consequently i) implies that during adaptation the second row of G_0 remains null and the first row is proportional to $H_0^{(1,:)}$.

The second row of G_d is initialized to $G_d^{(2,:)} = W_0^{(2,2)} H_d^{(2,:)}$. During adaptation, W_0 is constrained to be invertible and $W_0^{(2,1)}$ is null consequently $W_0^{(2,2)}$ cannot be null. Therefore the second row of G_d remains non null during adaptation but i) implies that the first row $G_d^{(1,:)}$ converges to zero.

Ideally all the other matrices G_j should be set to zero during adaptation in order to fulfill "as well as possible" the condition ii). For $j > d$ it is possible to do so. But due to the constraints imposed by the non invertibility of $H(z)$ and biproperness of $W(z)$, all the coefficients cannot be simultaneously set to zero for $j \in [1, d-1]$. However, because of the structure of $H(z)$, most of the G_j have one null row. Consequently i) is fulfilled: All G_j have orthogonal rows. But ii) is not achieved and the algorithm obtains Eq.(3). Extensive numerical simulation shows that the repartition of these non null coefficients is balanced between the two rows. (*Note: If d is odd with $r = (d-1)/2$ then second row of G_j for $j \in [0, r]$ and first row for $j \in [r+1, d]$ are null.*)

3.2 Row Relative Degree Difference Estimation

After minimizing the mutual information MI[$\mathbf{y}(t)$], the row relative degree difference between the rows of $H(z)$ are determined by using the off-diagonal terms of the covariance $\Gamma(\mathbf{y}, \tau)$. Considering the 2×2 case, the off-diagonal term is:

$$\mathcal{C}_{12}(\mathbf{y}, \tau) = \mathcal{E}\{y_1(t) y_2(t+\tau)^T\}. \quad (4)$$

By hypothesis the source signals are statistically independent and have unit variance, as a result their covariance is: $\mathcal{E}\{s_1(p) s_2(q)^T\} = \delta_{p,q}$. Since the transfer from sources to output signals is of the form Eq.(3), thus Eq. (4) gives:

$$\begin{aligned} \mathcal{C}_{12}(\mathbf{y}, \tau) = G_0^{(1,:)} G_d^{(2,:)T} \delta_{\tau,d} + \left[G_0^{(1,:)} G_{d-1}^{(2,:)T} + G_1^{(1,:)} G_d^{(2,:)T} \right] \delta_{\tau,d-1} + \ldots \\ + \left[G_{r-1}^{(1,:)} G_r^{(2,:)T} + G_r^{(1,:)} G_{r+1}^{(2,:)T} \right] \delta_{\tau,1} \end{aligned} \quad (5)$$

The covariance is null if τ is not in $[1, d]$. Therefore after minimizing $\text{MI}[\mathbf{y}(t)]$, the row relative degree difference d can be estimated by inspecting the covariance of the output signals: d is equal to the largest delay τ for which $\mathcal{C}_{12}(\mathbf{y}, \tau)$ is not null. (*Note: In the case of a negative relative degree difference d, the covariance is null if τ is not in $[d, -1]$.*)

3.3 Proposed Method

In practice, a threshold β, function of estimation variance, is chosen and the largest delay τ_0 such that $\mathcal{C}_{12}(\mathbf{y}, \tau) > \beta$ is the row relative degree difference estimation $\widehat{r_{12}} = \tau_0$. But finding such a threshold β is not an easy task.

However, a nice property appears when a relatively hight threshold β is chosen in order to avoid selecting delay out of $[1, d]$. The estimated row relative degree difference is $\widehat{r_{12}} = r_{12} - \epsilon$ with $0 \le \epsilon < r_{12}$ an integer representing the error. Then consider the shifted observations:

$$\begin{bmatrix} v_1(t + \widehat{r_{12}}) \\ v_2(t) \end{bmatrix} = \begin{bmatrix} z^{d_1 - d_2 - \epsilon} & 0 \\ 0 & 1 \end{bmatrix} H(z)s(t) = \begin{bmatrix} z^{d_1 - \epsilon} & 0 \\ 0 & z^{d_2} \end{bmatrix} H(z)s(t - d_2)$$

- When $\epsilon = 0$, $\text{diag}(z^{d_1} \; z^{d_2})H(z)$ is biproper. Consequently, using the same adaptation rule to adapt $W(z)$ after shifting the observation results in a blind deconvolution. Thus $\mathcal{C}_{12}(\mathbf{y}, \tau)$ is null for all τ because output signals are statistically independent.
- When $\epsilon \neq 0$ the row relative degree difference of the system whose outputs are the shifted observations is $\epsilon \in [1, r_{12} - 1]$. Thus using the same adaptation rule to adapt $W(z)$ after shifting the observation leads again to a cascade of the form Eq.(3) and $\mathcal{C}_{12}(\mathbf{y}, \tau)$ still presents non null values.

Thus iterating the same procedure ensures that $\epsilon \to 0$. In order to exploit this property, the proposed method is iterative:

1. initialization $\widehat{r_{12}} := 0$,
2. adapt the biproper demixer to minimize the mutual information,
3. compute $\mathcal{C}_{12}(\mathbf{y}, \tau)$ for delay in $[-\tau_{\max}, \tau_{\max}]$,
4. if $\mathcal{C}_{12}(\mathbf{y}, \tau) < \beta$ for all $\tau \in [-\tau_{\max}, \tau_{\max}]$ then stop iteration,
5. otherwise: Select τ_0 the largest delay such that $\mathcal{C}_{12}(\mathbf{y}, \tau) > \beta$, update the estimation $\widehat{r_{12}} := \widehat{r_{12}} + \tau_0$, shift the first observation $v_1(t) := v_1(t + \tau_0)$ and go to step 2.,

4 Numerical Simulation

Consider the system $H(z) = \begin{bmatrix} \frac{\frac{1}{z+0.7}}{} & \frac{0.9}{z-0.6} \\ \frac{-0.5}{(z-0.3)(z+0.7)(z+0.4)} & \frac{1}{(z+0.2)(z-0.4)(z+0.6)} \end{bmatrix}$, whose impulse response is given in Fig.2-(a). The row relative degrees are $d_1 = 1$ and $d_2 = 3$. Unknown source signals are *i.i.d.* processes uniformly distributed with zero mean and unit variance. The number of samples used in this example was $T = 10000$. 100 experiments were performed. The FIR filter has $l = 20$ coefficients.

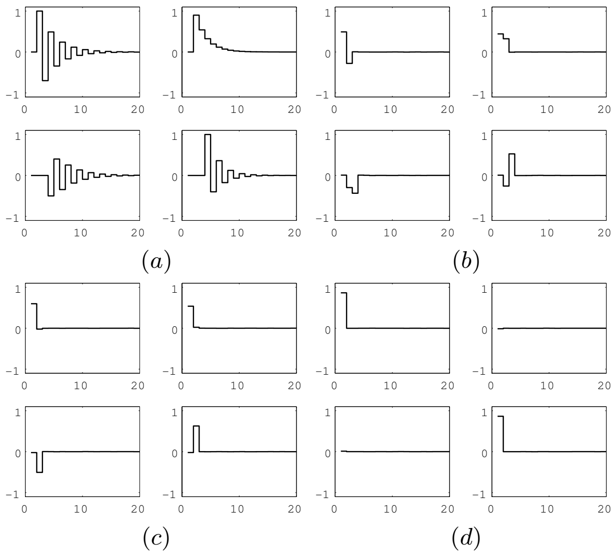

Fig. 2. Impulse response of $H(z)$ (a) and mean (variances are all less than $2.5e^{-4}$) of the impulse response of $W(z)H(z)$ after: First, second and third iterations respectively in (b), (c) and (d) (the subplot ij is the transfer from input j to output i)

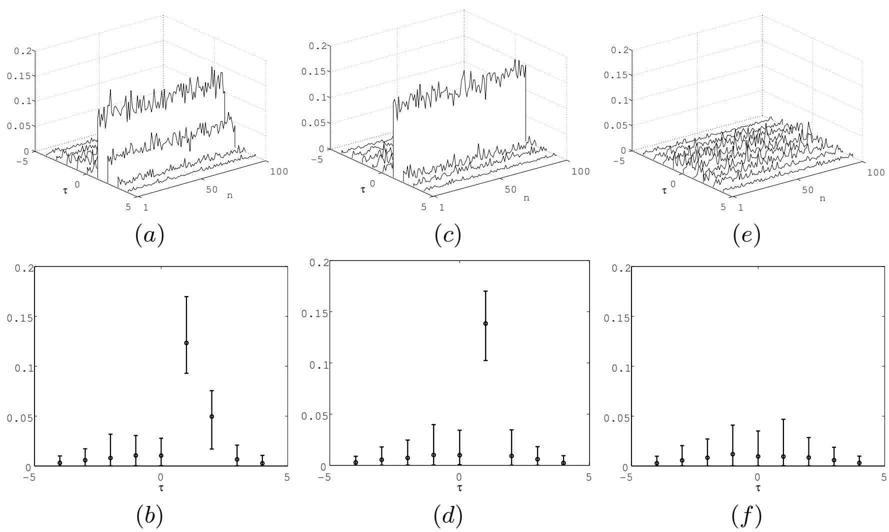

Fig. 3. $\mathcal{C}_{12}(\mathbf{y}, \tau)$ for: First (a)(b), second (c)(d) and third (e)(f) iterations. The upper row shows results for all experiments (index n) and the bottom row shows the mean, the minimum and the maximum of the covariance computed on all experiments.

The evolution of the impulse response of the cascade is presented in Fig.2-(b), (c) and (d). After adapting $W(z)$, the impulse response has the form of Eq.(3) in Fig.2-(b) and (c) but finally after the row relative degree difference was estimated the cascade is equal to identity because the blind deconvolution is achieved, Fig.2-(d).

The evolution of the covariance is depicted in Fig.3: First, second and third iterations (in left, middle and right column respectively). Fig.3-(a), (c) and (e) show $\mathcal{C}_{12}(\mathbf{y}, \tau)$ versus the delay τ for all the 100 experiments. The mean, minimum and maximum value of $\mathcal{C}_{12}(\mathbf{y}, \tau)$ are also plotted in Fig.3-(b), (d) and (f). During first iteration, the largest values are obtained for $\tau = 1, 2$ as expected from Eq.(5). But a threshold β such that $\tau = 2$ is selected for all experiments does not exist, see Fig.3-(b). For second iteration (after shifting first observation by one sample for all experiments: $\widehat{r_{12}} = 1$), the value $\tau = 1$ is selected in all experiments. Thus the true row relative degree difference of two, i.e. $\widehat{r_{12}} + 1 = 2$, is correctly estimated for all experiments. Consequently, after shifting again of one sample the first observation for all experiments and minimize the mutual information, the covariance has only very small values and the algorithm stops.

5 Conclusion

In this paper we show how to blindly estimate the row relative degree difference of a class of transfer matrices non-invertible at infinity by means of an iterative method based on a blind signal deconvolution setting.

References

1. Cichocki, A., Amari, S.: Adaptive Blind Signal and Image Processing : Learning Algorithms and Applications. John Wiley & Sons. (2002)
2. Kailath, T.: Linear systems. Prentice-Hall. (1980)
3. Shalvi, O., Weinstein, E.: New criteria for blind deconvolution of non-minimum phase systems. IEEE Trans. Inf. Theory. **36(2)** (1990) 312–321
4. Sugimoto, K., Suzuki, A., Nitta, M., Adachi, N.: Polynomial Matrix Approach to Independent Component Analysis: (Part II) Application. IFAC World Congress, Prague, Czech Republic (2005)
5. Wolovich, W.A., Falb, P.L.: Invariants and Canonical Forms under Dynamical Compensation. SIAM J. Control and Optimization. **14** (1976) 996–1008