Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

An information theoretic approach to hybrid deconvolution problems

Fabio Fagnani * **Sophie M. Fosson** **

* *Dipartimento di Matematica, Politecnico di Torino, Torino, Italy (e-mail: fabio.fagnani@polito.it).*
** *Scuola Normale Superiore, Pisa, Italy (e-mail: s.fosson@sns.it).*

Abstract: Recovering the input of a system from a noisy lecture of the output is both a typical inverse ill-posed problem and a transmission paradigm. If the input-output relation is given by a convolution integral, we are concerned with the well-known deconvolution problem, which occurs in several scientific frameworks. In this paper, we develop an original information theoretic analysis and we design an encoding-decoding scheme for deconvolution. We propose different decoding algorithms to identify the input and we show both theoretic and simulations' results.

Keywords: Recursive identification; bit-MAP decoding.

1. INTRODUCTION

The deconvolution problem is ubiquitous in many scientific and technological areas such as seismology, astrophysics, image processing and medical applications (Banham et al. [1997], Byrne et al. [1982], Jain [1989], Starck et al. [2002]). Its most general formulation is as follows. We consider a time horizon T (possibly infinite), a convolution kernel $\mathcal{K}(t)$ and the input/output system

$$x(t) = \int_0^t \mathcal{K}(t-s)u(s)ds \tag{1}$$

(we implicitly assume that $\mathcal{K}$ and u are s.t. the above integral makes sense). The problem is to estimate u from some noisy version of x.

This is an instance of inverse problem. To see why the problem is difficult we focus on a special case which will be the one treated in this paper: the case when $\mathcal{K} = 1$. In this context, (1) can be written as

$$\dot{x}(t) = u(t)\,, \quad x(0) = 0\,. \tag{2}$$

Since the operation of differentiation is not robust with respect to noise perturbation, the reconstruction of u from x cannot be simply done by differentiation.

To be more specific in this paper we assume that the available output signal is a noisy sampled version of $x(t)$:

$$y_k = x_k + n_k$$

where $\tau > 0$ is some constant sampling time, $x_k = x(\tau k)$ and n_k are noises which we model as independent Gaussian variables of 0 mean and variance σ^2. We will denote by $\mathbf{y}$ the vector of all available measures:

$$\mathbf{y} = (y_1, \dots y_K) \in \mathbb{R}^K$$

where $K = T/\tau$ is assumed to be an integer.

A deconvolution algorithm consists in a function $\Gamma : \mathbb{R}^K \to \mathbb{R}^{[0,T]}$: $\hat{u} = \Gamma(\mathbf{y})$ is the reconstructed input. In general $\hat{u}$ will not coincide with the true input u. What in general we request to a deconvolution algorithm is a bound on the error $u - \hat{u}$ and some consistency property: when the variance of the noise and the sampling time go to 0, the error should converge (in some sense) to 0.

Another important issue is causality: we say that a deconvolution algorithm is causal (with delay $k_0\tau$) if there exists a sequence of functions $\Gamma_k : \mathbb{R}^{k+k_0} \to \mathbb{R}^{[(k-1)\tau,k\tau[}$, where $k = 1, 2, ...,$ such that

$$\hat{u}(t)|_{t\in[(k-1)\tau,k\tau[} = \Gamma_k(y_1, \dots y_{k+k_0})\,.$$

Such an algorithm uses only past information (along with a possible bounded future information) to estimate the unknown function in the current time interval. This is essential in many applications that require delays to remain bounded. Most of the available deconvolution resolution methods (e.g. Tikhonov [1963], Tikhonov et al. [1977]) are not causal: the estimation $\hat{u}$ at any time depends on the whole sequence $\mathbf{y}$. Causal algorithms have been studied in Fagnani et al. [2002, 2003], where bounds on the error have been obtained for the case of bounded noises and regularity assumptions on the input signals u.

In this paper we focus on a different case: we assume u to be a piecewise constant signal with values restricted to a fixed known finite alphabet. This turns out to be a quite important case in the context of hybrid systems where continuous-time systems are driven by discrete digital signals. Specifically, we assume that there is a finite alphabet $\mathcal{U} \subset \mathbb{R}$ and we consider signals of type

$$u(t) = \sum_{k=0}^{K-1} u_k \mathbb{1}_{[k\tau,(k+1)\tau[}(t) \qquad u_k \in \mathcal{U}\,. \tag{3}$$

$u(t)$, with $t \in [0, T[$ is then completely determined by the sequence of samples $u_0, u_1, ..., u_{K-1}$. For simplicity we assume the sampling time τ to be the same than in the output and to have an exact synchronization in the sampling instants. The output signal is now identified by samples $x_1, x_2, ..., x_K \in \mathcal{X}$, where $\mathcal{X} \subset \mathbb{R}$ is a suitable alphabet (recall that we have fixed $x_0 = 0$). Of course, in principle, one could still use the deconvolution algorithms in Fagnani et al. [2002, 2003] or Tikhonov [1963], Tikhonov et al. [1977], however, there would be no way to use inside the algorithm the a priori information on the quantization of u. Instead we now show that, in this case, our deconvolution problem can completely be recasted into a discrete decoding problem. Notice indeed that the input/output system is simply described by

$$\begin{cases} x_0 = 0 \\ x_{k+1} = x_k + \tau u_k, \quad k = 0, ..., K-1 \end{cases} \tag{4}$$

The vector $\mathbf{x} = (x_1, \dots x_K)$ can thus be seen as a coded version of $\mathbf{u} = (u_0, \dots u_{K-1})$: we can write $\mathbf{x} = \mathcal{E}(\mathbf{u})$. Afterwards, $\mathbf{x}$ is transformed as it was transmitted through a classical Additive White Gaussian Noise, or AWGN, channel: the received output being given by $y_k = x_k + n_k$.

It is on the basis of these measures that we have to estimate the 'information signal' $\mathbf{u}$. Notice that the real time t is completely out of the problem at this point and everything can be considered at the discrete sampling clock time. In the coding theory language, a decoder is exactly a function $\mathcal{D} : \mathbb{R}^K \to \mathcal{U}^K$ which allows to construct an estimation of the input signal: $\hat{\mathbf{u}} = \mathcal{D}(\mathbf{y})$. Even in this context we can talk about causal algorithm if there exists a sequence of functions $\mathcal{D}_k : \mathbb{R}^{k+k_0} \to \mathcal{U}$ such that

$$\hat{u}_{k-1} = \mathcal{D}_k(y_1, \dots y_{k+k_0}) \, .$$

In this paper we cast the problem in a purely probabilistic setting assuming a uniform statistics on the input sequence. In Section 2, we present a natural decoding procedure which minimizes the mean square value of the input error and which, in some cases, corresponds to the popular bit-MAP (Maximum A Posteriori) decoding rule in coding theory. The BCJR algorithm gives a practical implementation of this scheme and is recalled in Section 3. We also discuss causal variations of BCJR, while in Section 4 we introduce another causal sub-optimal low complexity algorithm. In Section 5, we propose some numerical simulations which prove the goodness of our simple algorithm. Finally, in Section 6 a theoretic analysis of this algorithm is carried on using techniques of Markov chains in random environments.

2. DECODING PERFORMANCE: THE MEAN SQUARE COST

To complete the statistical description of our problem we need to fix an a priori probability distribution on the information signals $\mathbf{u}$. Notice that, from now on, we will use the following notation: r.v. is short for random variable; P indicates the probability on discrete r.v.'s, while f is the probability density of continuous r.v.'s and also of hybrid events, i.e., events involving both continuous and discrete random variables; $\mathbb{E}$ is the mean. For the sake of this paper we will assume that each u_k is the realization of a uniformly distributed r.v. U_k and that the U_k's are independent among themselves. We put $\mathbf{U} = (U_0, \dots, U_{K-1})$. Corresponding to the coding rule (4) we obtain a r.v. $\mathbf{X} = (X_1, \dots, X_K)$, a received r.v. $\mathbf{Y} = (Y_1, \dots, Y_K)$ ($Y_k = X_k + n_k$), and, finally, a decoded r.v. $\widehat{\mathbf{U}} = (\widehat{U}_0, \dots, \widehat{U}_{K-1}) = \mathcal{D}(\mathbf{Y})$.

A fundamental issue in the deconvolution problem is the choice of the norm respect to which errors are estimated. For our purpose, we shall consider the mean square cost:

$$\overline{d}(\mathbf{U}, \widehat{\mathbf{U}}) = \mathbb{E}\left[||\mathbf{U} - \widehat{\mathbf{U}}||^2\right] = \sum_{k=0}^{K-1} \mathbb{E}\left[(U_k - \widehat{U}_k)^2\right]$$

The best estimate $\widehat{\mathbf{u}}(\mathbf{y})$ is that one that minimizes $\overline{d}(\mathbf{U}, \widehat{\mathbf{U}})$. In order to find this value consider that, for any $v \in \mathcal{U}$,

$$\mathbb{E}\left[(U_k - v)^2\right] = \mathbb{E}\left[\mathbb{E}\left[(U_k - v)^2 | \mathbf{Y} = \mathbf{y}\right]\right]$$
$$= \int_{\mathbb{R}^K} \sum_{u \in \mathcal{U}} (u - v)^2 P(U_k = u | \mathbf{Y} = \mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}$$

hence

$$\widehat{u}_k(\mathbf{y}) = \operatorname*{argmin}_{v \in \mathcal{U}} \sum_{u \in \mathcal{U}} (u - v)^2 P(U_k = u | \mathbf{Y} = \mathbf{y}) \tag{5}$$

This is a finite combinatorial problem based on the probabilities $P(U_k = u | \mathbf{Y} = \mathbf{y})$ which can be computed by means of a marginalization procedure and a Bayesian inversion:

$$P(U_k = u | \mathbf{Y} = \mathbf{y}) = \sum_{\substack{\mathbf{u} \in \mathcal{U}^K: \\ u_k = u}} \frac{f(\mathbf{Y} = \mathbf{y} | \mathbf{U} = \mathbf{u}) P(\mathbf{U} = \mathbf{u})}{f(\mathbf{Y} = \mathbf{y})}$$

In the special case when $|\mathcal{U}| = 2$ the solution is particularly simple. Indeed if $\mathcal{U} = \{u_0, u_1\}$, (5) simply becomes

$$\widehat{u}_k(\mathbf{y}) = \operatorname*{argmax}_{u \in \mathcal{U}} P(U_k = u | \mathbf{Y} = \mathbf{y}) \tag{6}$$

This is the so-called bit-MAP decoding, an optimal decoding procedure whose decision rule is based on the maximization of the a posteriori probability on each bit. In this framework, we introduce the Bit Error Rate (also denoted by BER or $P_b(e)$), generically defined as the mean probability of error on the information symbols:

$$P_b(e) = \frac{1}{K} \sum_{k=0}^{K-1} P(\widehat{U}_k \neq U_k) \tag{7}$$

If $\mathcal{U} = \{0, 1\}$, $P_b(e) = \frac{1}{K}\mathbb{E}\left[||\mathbf{U} - \widehat{\mathbf{U}}||^2\right] = \frac{1}{K}\overline{d}(\mathbf{U}, \widehat{\mathbf{U}})$. Hence, *minimize* $\overline{d}(\mathbf{U}, \widehat{\mathbf{U}})$ *corresponds to bit-MAP decoding and is equivalent to minimize* $P_b(e)$. From now on we will assume that $\mathcal{U} = \{0, 1\}$.

Up to this point we have not enforced any causality on the decoding procedure. If we impose instead that $\widehat{U}_{k-1}$ ($k = 1, \dots, K$) can only depend on the observations $\mathbf{Y}_1^{k+k_0} = (Y_1, \dots, Y_{k+k_0})$, the decoding procedure minimizing the cost functional $\overline{d}(\mathbf{U}, \widehat{\mathbf{U}})$, becomes

$$\widehat{u}_{k-1}(\mathbf{y}) = \operatorname*{argmax}_{u \in \mathcal{U}} P(U_{k-1} = u | \mathbf{Y}_1^{k+k_0} = \mathbf{y}_1^{k+k_0}) \tag{8}$$

If we pick $k_0 = 0$ then we have a purely causal decoding procedure.

By definition 7, the BER is a parameter that evaluates the *mean* performance of the transmission model and it does not consider its behavior for *each* possibile sent sequence. As our system turns out to be sensible to the input (roughly speaking, some sequences are decoded with more reliability than others, both with bit-MAP and other decoding methods), we are interested in studying its behavior in function of the input. This is why we introduce also the Conditional Bit Error Rate, CBER for short, which is a function of $\mathbf{U}$:

$$P_b(e|\mathbf{U}) = \frac{1}{K}\sum_{k=0}^{K-1} P(\widehat{U}_k \neq U_k|\mathbf{U}) \qquad (9)$$

3. RECOVERING THE INPUT SIGNAL: THE BCJR ALGORITHM

In practice, the bit-MAP decoding can be performed with the well-known BCJR algorithm, named for the authors of paper Bahl et al. [1974]. This algorithm computes the probabilities of states and transitions of a Markov source, given the observed channel outputs; in other words, it provides the so-called APP (a posteriori probabilities) on states and transitions, therefore on coded and information symbols.

Let us briefly remind the BCJR procedure in our context. Given the output r.v. $\mathbf{Y} = (Y_1, \dots, Y_K)$ (or its realization $\mathbf{y} = (y_1, \dots, y_K)$), we indicate by $\mathbf{Y}_a^b$ (or $\mathbf{y}_a^b$) its components from time a to time b. For simplicity, from now on we suppose $\tau = 1$; given $\mathcal{U} = \{0, 1\}$, then $\mathcal{X} = \{0, 1, \dots K\}$. For $i, j = 0, 1, 2, ..., K$, we define the following probability density functions:

$$\alpha_k(i) = f(X_k = i, \mathbf{Y}_1^k = \mathbf{y}_1^k) \qquad k = 1, ..., K$$
$$\beta_k(i) = f(\mathbf{Y}_{k+1}^K = \mathbf{y}_{k+1}^K | X_k = i) \qquad k = 0, ..., K-1$$
$$\Gamma_k(i, j) = f(X_k = j, Y_k = y_k | X_{k-1} = i) \qquad k = 1, ..., K$$

For any $k = 1, ..., K$, the relevant APP are given by:

$$\lambda_k(i) = f(X_k = i, \mathbf{Y} = \mathbf{y})$$
$$\sigma_k(i, j) = f(X_k = j, X_{k-1} = i, \mathbf{Y} = \mathbf{y})$$

divided for $f(\mathbf{Y} = \mathbf{y})$. Given

$$\alpha_0(i) = P(X_0 = i) = \delta_0(i) = \begin{cases} 1 & \text{if } i = 0 \\ 0 & \text{otherwise.} \end{cases}$$
$$\beta_K(i) = P(\emptyset) = 1 \text{ for any } i = 0, ..., K$$

then, for $k = 1, ..., K$,

$$\lambda_k(i) = \alpha_k(i)\beta_k(i)$$
$$\sigma_k(i, j) = \alpha_{k-1}(i)\Gamma_k(i, j)\beta_k(j) \qquad (10)$$

where sequences α_k and β_k can be respectively computed with a forward and a backward recursions:

$$\alpha_k = \alpha_{k-1}\Gamma_k \qquad \beta_k = \Gamma_{k+1}\beta_{k+1}$$

Notice that, in our case, $\alpha_k(i)$, $i = 0, 1, ..., K$, is null for any $i > k$, because, starting from state $X_0 = 0$, X_k cannot exceed the value k. In fact, at each step, the state X_k can only remain constant (this occurs if the transmitted information bit is 0) or increase of one (if the transmitted information bit is 1). For the same reason, matrices Γ_k and σ_k are non-null only on diagonal and superdiagonal. In particular, recalling that the transition between X_k and Y_k is modeled by an AWGN channel, the corresponding probability density is

$$f(Y_k = y_k | X_k = x_k) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_k - j)^2}{2\sigma^2}\right)$$

Moreover,

$$P(X_k = j | X_{k-1} = i) = \begin{cases} 1/2 & \text{if } j = i, i+1 \\ 0 & \text{otherwise} \end{cases}$$

therefore we have

$$\Gamma_k(i, j) = f(Y_k = y_k | X_k = j) P(X_k = j | X_{k-1} = i)$$
$$= \frac{1}{2\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_k - j)^2}{2\sigma^2}\right) \qquad \text{for } j = i, i+1$$

	Number of computations	Storage locations	Delay to decode the kth bit
BCJR	$O(K^2)$	$6K^2{+}K$ bits	$(K-k)\tau$
caus. BCJR	$O(K^2)$	$6K^2{+}K$ bits	$k_0\tau$ (bounded)
BBB	$O(K)$	1+ K bits	0

Table 1. The complexities of the algorithms

Finally, to estimate the transmitted bit at time k we compare the a posteriori probabilities on the corresponding transition: if it is more likely that the state remains the same, the decoding outputs a 0, otherwise a 1. The bit-MAP decoding rule at time k is then the following one:

$$\widehat{u}_{k-1} = \begin{cases} 0 & \text{if } \displaystyle\sum_{i=0}^{k-1} \sigma_k(i, i+1) \le \sum_{i=0}^{k-1} \sigma_k(i, i) \\ 1 & \text{otherwise} \end{cases} \qquad (11)$$

3.1 Causal BCJR algorithm

Analogous causal versions of the BCJR algorithm can be used to compute the decoding rule (8) with delay k_0.

For $k = 1, ..., K - k_0$, the APP on the transitions becomes

$$\sigma_k(i, j) = f(X_k = j, X_{k-1} = i, \mathbf{Y}_1^{k+k_0} = \mathbf{y}_1^{k+k_0}) = \alpha_{k-1}(i)\Gamma_k(i, j)\widetilde{\beta}_k(j) \qquad (12)$$

where the function α_k and Γ_k are defined as above, while $\widetilde{\beta}_k(j) = f(\mathbf{Y}_{k+1}^{k+k_0} = \mathbf{y}_{k+1}^{k+k_0} | X_k = j)$, while for $k > K - k_0$ we come back to the classical formulation (10).

In the purely causal case, namely the special case when $k_0 = 0$, the functions $\widetilde{\beta}_k$ do not show up at all, and we have just the forward recursion of the functions α_k.

4. RECOVERING THE INPUT SIGNAL: THE ITERATIVE BIT-BY-BIT ALGORITHM

Remaining in the context of causal decoding, we propose a recursive algorithm which is still less complex than the causal BCJR. This algorithm decodes the sequence bit by bit and requires, at each step, the estimation and storage of a state value $\hat{x}_k$ that resumes all past information. The decoder is then a function $\mathcal{D}_k : \mathbb{R}^2 \to \mathcal{U}$ such that $\hat{u}_{k-1} = \mathcal{D}_k(y_k, \hat{x}_{k-1})$. The pattern of the algorithm is the following one (d_E indicates the euclidean distance):

(1) Initialize state: $\widehat{x}_0 = 0$;
(2) For $k = 1, ..., K$, given the received symbol $y_k \in \mathbb{R}$,

$$\widehat{u}_{k-1} = \underset{u \in \{0,1\}}{\text{argmax}}\ P(U_{k-1} = u | Y_k = y_k, \hat{X}_{k-1} = \widehat{x}_{k-1}) = \begin{cases} 0 & \text{if } d_E(y_k, \widehat{x}_{k-1}) \le d_E(y_k, \widehat{x}_{k-1} + 1) \\ 1 & \text{otherwise} \end{cases}$$
$$\widehat{x}_k = \widehat{x}_{k-1} + \widehat{u}_{k-1}$$

At each step, the decoder estimates the achieved state x_k by $\widehat{x}_k$ and decides on the current bit with the MAP method, that here consists of a simple comparison between two euclidean distances. This straightforward procedure, that we name iterative bit-by-bit algorithm or BBB algorithm, requires a lower number of operations and a smaller storage than the causal BCJR. A detailed comparison between the complexity's characteristics of the three decoding methods considered in this work can be seen in table 1.

5. SIMULATIONS' OUTCOMES

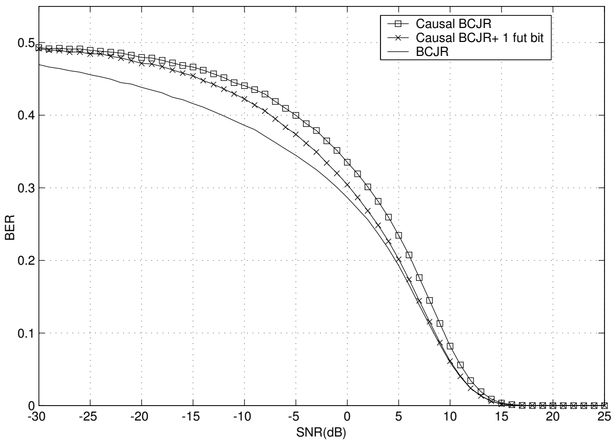

Figure 1. The Bit Error Rates derived from simulations with classical and causal BCJR algorithms

In this section, we present the simulations' outcomes of the different decoding methods above described. The simulations are performed on random sequences of $K = 96$ bits and the outcomes are represented in terms of BER, which corresponds to $T = K$ times the mean square cost. The BER curves are represented in function of the *Signal-to-noise ratio*, here defined as SNR $= \frac{\tau^2}{\sigma^2} = \frac{1}{\sigma^2}$, and are expressed in dB. In figure 1, we show the BER curves derived from simulations with different BCJR decodings. First of all, we notice that all BCJR methods are consistent: the BER converges to zero for high SNR values, i.e., for low noise values. On the other hand, if the noise is very large, the BER tends to $\frac{1}{2}$: the decoding becomes absolutely random.

As expected, the classical BCJR gives the best performance, i.e., the BER is the lowest. Afterwards, we can appreciate the gain obtained with a causal BCJR with just one future bit with respect to the causal BCJR: for SNR$>$ 5 dB, the first one is very close to the optimal curve. From these observations, we can then deduce that decodings considering some future bits perform considerably better than the purely causal ones.

Afterwards, in figure 2, we compare the iterative bit-by-bit decoding with the causal BCJR. The corresponding curves are not distant: this means that, in the context of causal methods, the *incomplete* past information resumed by a state value represents the *whole* past sufficiently well. Our iterative bit-by-bit algorithm is then suitable because its low complexity does not incur a considerable performance loss.

6. THEORETIC ANALYSIS OF THE ITERATIVE BIT-BY-BIT ALGORITHM

In the next, we propose an exhaustive theoretic analysis of our bit-by-bit algorithm and we provide a formal setting in which analytically compute the performance in terms of BER. According to definitions 7 and 9 in section 2, the analysis is divided into two parts: first, we consider the mean case and we explicitly compute the BER through

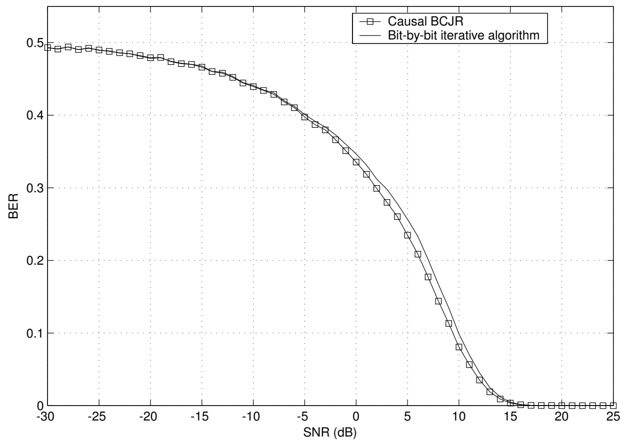

Figure 2. The Bit Error Rates derived from simulations with causal decodings

denumerable Markov chains arguments; afterwards, we calculate the CBER in the stochastic framework of Markov chains in random environments (or MCRE's, see Cogburn [1984, 1986], Nawrotzki [1981, 1982]). In fact, a perfect correspondence between the BBB decoding procedure and a particular instance of MCRE can be identified.

Consider the generic transmission of $K \to \infty$ bits and suppose to decode by bit-by-bit method; the starting point of our analysis is the definition, at any step $k = 1, 2, 3...$, of the r.v. D_k taking values in $\mathbb{Z}$:

$$D_k = \hat{X}_k - X_k \tag{13}$$

Having fixed the initial value $D_0 = 0$, the following recursive relationship holds

$$D_{k+1} = D_k + \hat{U}_k - U_k. \tag{14}$$

6.1 The mean BER

According to 14, $(D_k)_{k=0,1,\dots}$ is a denumerable homogeneous Markov chain on state space $\mathbb{Z}$, with transition probabilities

$$\overline{\mathbf{P}}_{x,y} = P(D_{k+1} = y | D_k = x) = \frac{1}{2}[\mathbf{P}_{x,y}(0) + \mathbf{P}_{x,y}(1)]$$

where

$$\mathbf{P}_{x,y}(u) = P(D_{k+1} = y, U_k = u | D_k = x) \qquad u \in \{0, 1\}$$

In particular, $\mathbf{P}(0)$ and $\mathbf{P}(1)$ have, for any $d \in \mathbb{Z}$, the following non-null entries:

$$\mathbf{P}_{d,d+1}(0) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{d + \frac{1}{2}}{\sqrt{2}\sigma}\right) \qquad \mathbf{P}_{d,d}(0) = 1 - \mathbf{P}_{d,d+1}(0)$$

$$\mathbf{P}_{d,d}(1) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{d - \frac{1}{2}}{\sqrt{2}\sigma}\right) \qquad \mathbf{P}_{d,d-1}(1) = 1 - \mathbf{P}_{d,d}(1)$$

Hence, $\overline{\mathbf{P}}$ is tridiagonal and for any $x, y \in \mathbb{Z}$, $\overline{\mathbf{P}}_{x,y} = \overline{\mathbf{P}}_{-x,-y}$.

We note that all the states of (D_k) *communicate*, i.e., for any $x, y \in \mathbb{Z}$ there exist $n, m \in \mathbb{N}$ s.t. $\overline{\mathbf{P}}^n_{x,y} > 0$ and $\overline{\mathbf{P}}^m_{y,x} > 0$; such a Markov chain is said *irreducible*. Further, let $\tau_j = \min\{n > 0 | D_n = j\}$: state j is said *positive recurrent* if $\mathbb{E}(\tau_j | D_0 = j) < \infty$.

Lemma 1. (D_k) is positive recurrent (that is, all its states are so).

Proof It suffices to apply the following criterium proposed in Stroock [2005]: if there exists a function $\mathbf{g} \in \mathbb{R}^{+\mathbb{Z}}$ so that $\mathbf{g}_x \geq (\overline{\mathbf{P}}\mathbf{g})_x + \varepsilon$ for any $x \in \mathbb{Z} \setminus \{y\}$ and for some $\varepsilon > 0$, then y is a positive recurrent state.
In our case, it is easy to prove that $y = 0$ is a positive recurrent state considering $\mathbf{g}_x = |x|$. Moreover, given that the chain is irreducible, if one state is positive recurrent, all the states are so. ■

We remind that a probability vector $\Phi \in \mathbb{R}^{+\mathbb{Z}}$ such that $\Phi^T\overline{\mathbf{P}} = \Phi^T$ is called *invariant* for the denumerable Markov chain. We have the following result:

Proposition 2. These statements hold:

(1) (D_k) admits a unique invariant probability vector Φ;
(2) Φ is defined by the recursive relation

$$\Phi_d = \Phi_0 \prod_{i=1}^{|d|} \frac{\overline{\mathbf{P}}_{i-1,i}}{\overline{\mathbf{P}}_{i,i-1}} \tag{15}$$

where $\Phi_0 = \left[1 + 2\sum_{d=1}^{\infty}\prod_{i=1}^{|d|}\overline{\mathbf{P}}_{i-1,i}/\overline{\mathbf{P}}_{i,i-1}\right]^{-1}$.

Proof (1) It follows from a well-known result (see, for instance, Stroock [2005]): any irreducible and positive recurrent Markov chain admits a unique invariant probability measure.
(2) By $(\Phi^T\overline{\mathbf{P}})_d = \Phi_d^T$, for any $d \in \mathbb{Z}$, it follows that

$$\Phi_{d-1}\overline{\mathbf{P}}_{d-1,d} - \Phi_d\overline{\mathbf{P}}_{d,d-1} = c \quad (c \text{ constant}) \tag{16}$$

In particular, as $\Phi_d = \Phi_{-d}$ for any $d \in \mathbb{Z}$ (this is due to the uniqueness of the invariant and to the symmetry of $\overline{\mathbf{P}}$), it suffices to substitute values $d = 0$ and $d = 1$ in 16 to conclude that $c = 0$; hence, relation 15 holds. ■

Since $\overline{\mathbf{P}}_{i-1,i}/\overline{\mathbf{P}}_{i,i-1} < 1$ for $i \geq 1$, Φ_d has its maximum in $d = 0$ and it is a symmetric monotone function exponentially decreasing towards zero.
The invariant Φ describes the long-time behavior of our system and enables us to compute the BER for $K \to \infty$ and initial state $D_0 = 0$, as stated by the following

Theorem 3.

$$\lim_{K\to\infty} P_b(e) = \sum_{d\in\mathbb{Z}} \overline{A}_d\Phi_d \tag{17}$$

Proof Define $\overline{A}_d = P[\hat{U}_k \neq U_k | D_k = d] = \overline{\mathbf{P}}_{d,d+1} + \overline{\mathbf{P}}_{d,d-1}$. Hence,

$$P_b(e) = \frac{1}{K}\sum_{k=0}^{K-1}\sum_{d\in\mathbb{Z}}\overline{A}_d P(D_k = d) = \frac{1}{K}\sum_{k=0}^{K-1}\sum_{d\in\mathbb{Z}}\overline{A}_d\overline{\mathbf{P}}_{0,d}^k$$

The thesis is now a direct consequence of the standard result: $1/K\sum_{k=0}^{K-1}\overline{\mathbf{P}}_{x,d}^k \to \Phi_d$ for $K \to \infty$, $\forall x \in \mathbb{Z}$ (Stroock [2005]). ■

6.2 The conditional BER

Note that, from 14, we can also interpret (D_k) as a Markov chain in a random environment (Cogburn [1984, 1986]), thinking of the input U_k as the *random environment* at time k. This is the right way to look at (D_k) if we want to understand its behavior with respect to typical instances of the input $\mathbf{U} = (U_0, U_1, \dots)$.
For any $x, y \in \mathbb{Z}$, we have

$$P(D_{k+1} = y | D_k = x, D_{k-1}, \dots, D_0; \mathbf{U}) = \mathbf{P}_{x,y}(U_k)$$

As Cogburn states in Cogburn [1984], a MCRE can be modeled as a Markov process and this leads to the possibility to apply standard ergodic theorems of Markov processes to describe its behavior. In the next, we are going to illustrate this key idea in our context.

Our MCRE naturally evolves in the state space $\Omega = \mathbb{Z} \times \{0,1\}^{\mathbb{N}}$. Let $\mathcal{A}$ and $\mathcal{B}$ be the discrete σ-fields respectively on $\mathbb{Z}$ and $\{0,1\}$ and $\mathcal{B}_0^{+\infty} = \prod_0^{+\infty}\mathcal{B}$; then $\mathcal{F} = \mathcal{A}\times\mathcal{B}_0^{+\infty}$ is a σ-field on Ω. We also provide Ω of the measure $\mu = \kappa\times\pi$, where κ is the counting measure on $\mathbb{Z}$ and π is the usual uniform Bernoulli measure on $\{0,1\}^{\mathbb{N}}$.
Now, let $x, y \in \mathbb{Z}$, $\mathbf{U} = (U_0, U_1, \dots) \in \{0,1\}^{\mathbb{N}}$ and $B \in \mathcal{B}_0^{+\infty}$; in the next, will denote by ω or by $(x, \mathbf{U})$ a generic element in Ω. Considering the shift operator T on $\{0,1\}^{\mathbb{N}}$, that is, $T\mathbf{U} = (U_1, U_2, \dots)$, we define

$$P((x,\mathbf{U}); \{y\}\times B) = \mathbf{P}_{x,y}(U_0)\mathbb{1}_B(T\mathbf{U}) \tag{18}$$

which turns out to be a transition probability in the Foguel [1969] sense, hence we have well defined a Markov process in $(\Omega, \mathcal{F}, \mu)$ modeling our MCRE. We name it *extended* Markov process, EMP for short.

A measure ψ on $(\Omega, \mathcal{F})$ is *invariant* for the EMP if

$$\int_{\omega\in\Omega} P(\omega, F)\psi(d\omega) = \psi(F) \tag{19}$$

for any $F \in \mathcal{F}$. By explicit computation through Φ, we prove that an invariant probability exists for our EMP:

Theorem 4. Let ϕ be a probability measure on $(\mathbb{Z}, \mathcal{A})$ defined by $\phi(\{d\}) = \Phi_d$ for any integer d. Then, $\psi = \phi\times\pi$ is an invariant probability measure for the EMP.

Proof It suffices to prove that $\psi = \phi\times\pi$ verifies 19 for any $F = \{y\}\times B$, $y \in \mathbb{Z}$, $B \in \mathcal{B}_0^{+\infty}$. Let $\omega = (x, \mathbf{U})$, then

$$\int_\Omega P(\omega, F)\psi(d\omega) = \sum_{x\in\mathbb{Z}}\overline{\mathbf{P}}_{x,y}\Phi_x\pi(B) = \Phi_y\pi(B) = \psi(F)$$

where we have exploited the independence of the U_k's to split the integral on $\mathbf{U}$ into a sum on U_0 and an integral on $(U_1, U_2 ...) = T\mathbf{U}$. ■

Remark 1. By definition, $\psi \simeq \mu$ (that is $\psi << \mu$ and $\mu << \psi$). Further, as a consequence of corollary 3.4 in Cogburn [1984], ψ is the unique invariant probability measure absolutely continuous with respect to μ.

The evolution of our EMP is strictly linked to the behavior of the CBER; in particular, in the asymptotic case, the CBER can be computed by means of the ergodic theorem of Markov processes:

Theorem 5.

$$\lim_{K\to\infty} P_b(e|\mathbf{U}) = \sum_{d\in\mathbb{Z}}\overline{A}_d\Phi_d \quad \text{for } \pi\text{-a.e. } \mathbf{U} \tag{20}$$

Proof Defining $A_d(U_k) = P[\hat{U}_k \neq U_k | D_k = d, U_k] = \mathbf{P}_{d,d+1}(U_k) + \mathbf{P}_{d,d-1}(U_k)$, for any $K \in \mathbb{N}$ and given $D_0 = 0$, the CBER can be expressed as follows:

$$P_b(e|\mathbf{U}) = \frac{1}{K}\sum_{k=0}^{K-1}\sum_{d\in\mathbb{Z}} A_d(U_k)\mathbf{P}_{0,d}(U_0, U_1, \dots U_{k-1}) \tag{21}$$

where $\mathbf{P}(U_0, U_1, \dots, U_{k-1}) = \mathbf{P}(U_0)\mathbf{P}(U_1)\cdots\mathbf{P}(U_{k-1})$.
A set $F \in \mathcal{F}$ is said *invariant* if $P(w, F) = 1$ for μ-a.e.

$\omega \in F$ and $P(w, F) = 0$ for μ-a.e. $\omega \notin F$. We denote by $\mathcal{F}_i$ the σ-field of the invariant sets an we say that $\mathcal{F}_i$ is *trivial* if for every $F \in \mathcal{F}_i$, $\mu(F) = 0$ or $\mu(\Omega \setminus F) = 0$.

A condition to verify triviality is the following (Cogburn [1986]): if for each $x, y \in \mathbb{Z}$ and a.e. $\mathbf{U}$ there exist an $n = n(x, y, \mathbf{U})$ and a $z = z(x, y, \mathbf{U}, n) \in X$ such that $P_{x,z}(U_0, \ldots, U_n)P_{y,z}(U_0, \ldots U_n) > 0$, then $\mathcal{F}_i$ is trivial. In our context, take any couple of starting states x and y with distance $|x - y| = d$: after $n > d$ steps, we have a non-null probability of having joined a common state z, then triviality holds.

Now, consider any non-negative function $g \in L_1(\Omega, \mathcal{F}, \psi)$; defining $Pg(\omega) = \int_\Omega g(\omega')P(\omega, d\omega')$, the ergodic theorem (Foguel [1969]) states that, under the hypothesis that the invariant sets are trivial:

$$\lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} P^k g(\omega) = \mathbb{E}^{(\psi)}(g) \quad \text{for } \psi\text{-a.e. } \omega \qquad (22)$$

where $\mathbb{E}^{(\psi)}$ denotes the mean with respect to the invariant probability ψ. This general result can be applied to our case considering the function $g_d(x, \mathbf{U}) = A_d(U_0)\delta_d(x)$. In fact, $A_d(U_k)\mathbf{P}_{x,d}(U_0, \ldots U_{k-1}) = P^k g_d(x, \mathbf{U})$ and by the ergodic theorem:

$$\lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} P^k g_d(\omega) = \mathbb{E}^{(\psi)}(g_d) = \overline{A}_d \Phi_d \quad \text{for } \psi\text{-a.e. } \omega$$

where $\omega = (x, \mathbf{U})$. Fixing $x = 0$ and recalling that $\psi = \phi \times \pi$, we obtain the thesis. ■

Theorem 5 is stronger than theorem 3 because it states that for *almost all* $\mathbf{U} \in \{0, 1\}^{\mathbb{N}}$ the CBER associated to the BBB decoding tends to the *mean* limit. Hence, we supply a characterization not only of the mean behavior of the algorithm, but also of each possible input occurrence except for a π-negligible set.

Notice also that in theorems 3 and 5 we have considered the initial state fixed to $D_0 = 0$; nevertheless, it follows from their proofs that asymptotic results will be the same choosing any other $D_0 = x \in \mathbb{Z}$.

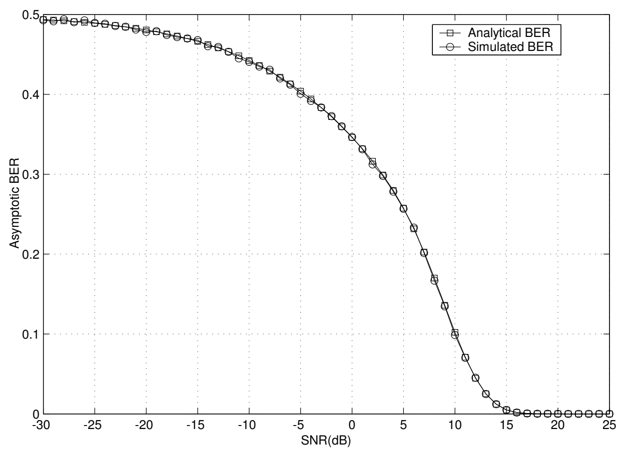

Figure 3. Asymptotic case: analytical vs simulated BER

In figure 3, we can appreciate that the asymptotic BER (or CBER) is very close to the BER obtained by a 96 bits simulation with the BBB method. This suggests that our system rapidly achieves the equilibrium distribution.

7. CONCLUSION

In this work, we have presented a deconvolution problem and proposed an information approach to solve it. We have discussed different consistent decoding methods and in particular we have focused on the good features of the iterative bit-by-bit algorithm, whose evolution turns out to be an instance of MCRE.

Under an information theoretic point of view, our transmission system results sensible to the input, i.e., the BER is different for each transmitted bit sequence. However, we have proved that the asymptotic performance is the same for almost all sequences using the BBB decoding.

Our future work will envisage wider input sources, including, for instance, discrete signals switching at unknown time instants and continuous functions.

REFERENCES

L. Bahl, J. Cocke, F. Jelinek and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans. on Information Theory*, volume IT-20(2), pages 284–287, 1974.

M.R. Banham and A.K. Katsaggelos. Digital image restoration. *IEEE Signal Proc. Magazine*, volume 14(2), pages 24–41, 1997.

C.L. Byrne and R.M. Fitzgerald. Reconstruction from partial information with applications to tomography. *SIAM J. Appl. Math.*, volume 42(4), pages 933–940, 1982.

R. Cogburn. The ergodic theory of Markov chains in random environments. *Zeitschrift Fur Wahrscheinl Und Verwandte Gebiete*, volume 66, pages 109–128, 1984.

R. Cogburn. On products of random stochastic matrices. *Contemporary Mathematics*, volume 50, pages 199–213, 1986.

F. Fagnani and L. Pandolfi. A singular perturbation approach to a recursive deconvolution problem. *SIAM J. Control Optim.*, volume 40, pages 1384–1405, 2002.

F. Fagnani and L. Pandolfi. A recursive algorithm for the approximate solution of Volterra integral equations of the first kind of convolution type. *Inverse problems*, volume 19, pages 23–47, 2003.

S.R. Foguel. The ergodic theory of Markov processes. Van Nostrand, Princeton, 1969.

A.K. Jain. Fundamentals of digital image processing. Prentice-Hall, Englewood Cliffs, 1989.

K. Nawrotzki. Discrete open systems or Markov chains in a random environment I. *EIK*, volume 17(11/12), pages 569–599, 1981.

K. Nawrotzki. Discrete open systems or Markov chains in a random environment II. *EIK*, volume 18(1/2) , pages 83–98, 1982.

J.L. Starck, E. Pantin and F. Murtagh. Deconvolution in astronomy: a review. *Publications of the Astronomical Society of the Pacific*, pages 1051-1069, 2002.

D.W. Stroock. An introduction to Markov processes. *Springer*, 2005.

A.N. Tikhonov. On the solution of ill-posed problems and the method of regularization. *Soviet Math. Dokl.*, volume 4 , pages 1035–1038, 1963.

A.N. Tikhonov and V.Y. Arsenin. Solution of ill-posed problems. Winston and Wiley, Washington, 1977.