Blind Deconvolution of Discrete-Valued Signals

Ta-Hsin Li

Department of Statistics
Texas A&M University
College Station, Texas 77843

Abstract — *This paper shows that when the input signal to a linear system is discrete-valued the blind deconvolution problem of simultaneously estimating the system and recovering the input can be solved more efficiently by taking into account the discreteness of the input signal. Two situations are considered. One deals with noiseless data by an inverse-filtering procedure which minimizes a cost function that measures the discreteness of the output of an inverse filter. For noisy data observed from FIR systems, the Gibbs sampling approach is employed to simulate the posteriors of the unknowns under the assumption that the input signal is a Markov chain. It is shown that in the noiseless case the method leads to a highly efficient estimator for parametric systems so that the estimation error decays exponentially as the sample size grows. The Gibbs sampling approach also provides rather precise results for noisy data, even if the initial and transition probabilities of the input signal and the variance of the noise are completely unknown.*

1. Introduction

Blind deconvolution in general deals with the simultaneous estimation of a linear system $\{s_j\}$ and reconstruction of its random input $\{x_t\}$ on the basis of the data $\{y_t\}$ obtained from the convolution

$$y_t = \sum_{j=-\infty}^{\infty} s_j x_{t-j}. \tag{1}$$

Partial information about the statistical properties of $\{x_t\}$ is usually required in order to obtain a sensible solution. It is evident that how well the knowledge of $\{x_t\}$ can be incorporated into the solution plays an important role in this problem.

The current paper is concerned with a special problem of blind deconvolution in which the input signal takes discrete values from a known alphabet — a typical situation encountered frequently in digital communications [8]. Based on the *inverse filtering* approach, a cost function is employed to measure the closeness of the filtered data to a discrete-valued sequence and minimized to obtain an estimate for the unknown system. In the parametric case where the system is characterized by a finite dimensional parameter (e.g., ARMA models) the method is proved to yield highly efficient estimates so that the estimation error may decay *exponentially* as the sample size grows. When the data $\{y_t\}$ are contaminated by Gaussian white noise and the system $\{s_j\}$ has finite length (FIR), the current paper shows that the *Gibbs sampling* procedure can be used to deal with the estimation of $\{s_j\}$ and $\{x_t\}$ under the assumption that $\{x_t\}$ is a Markov chain [2]. This method presents an avenue to incorporate *colored* input signals into the blind deconvolution problem. All these results provide yet another piece of evidence that a digital signal is capable of resisting distortion and contamination if its discreteness can be judiciously utilized in the restoration procedure.

2. An Inverse Filtering Procedure

When $\{x_t\}$ is an i.i.d. sequence of zero-mean random variables and $\{s_j\}$ is a minimum-phase ARMA(p, q) system so that

$$\sum_{j=0}^{p} a_j^* y_{t-j} = \sum_{j=0}^{q} b_j^* x_{t-j} \tag{2}$$

with $a_0^* = b_0^* = 1$, the classical least-squares method [1] provides a solution to the problem by seeking the coefficients $\{a_1, \ldots, a_p, b_1, \ldots, b_q\}$ to minimize the sample variance of the linear prediction error $\{u_t\}$ given by

$$u_t = \sum_{j=0}^{p} a_j y_{t-j} - \sum_{j=1}^{q} b_j u_{t-j}. \tag{3}$$

Since the method approximates the maximum likelihood estimation by ignoring the end-point effect, it is not surprising that minimizing the sample variance of $\{u_t\}$ leads to asymptotically efficient estimates [1] for the ARMA system (2). An alternative to least squares is the method of moments. Although computationally appealing, it does not however provide efficient estimates except for the pure AR systems. The variance of the estimates in both methods is usually proportional to the reciprocal of the sample size, i.e., $O(1/n)$.

To generalize the idea of least squares, it is crucial to observe that $\{u_t\}$ in (3) is the output of an *inverse filter* corresponding to the ARMA system in (2). Therefore, the least-squares method calls for the *minimization of variance* of the output sequence obtained by filtering the data $\{y_t\}$ with an inverse filter. For an arbitrary parametric system with $s_j = s_j(\theta^*)$ in (1), one may consider the output sequence

$$u_t(\theta) = \sum_{j=-\infty}^{\infty} s_j^{-1}(\theta)\, y_{t-j} \tag{4}$$

where $\{s_j^{-1}(\theta)\}$ is the inverse of $\{s_j(\theta)\}$. Since the variance alone is no longer sufficient for the discrimination of nonminimum-phase systems, higher-order moments of $\{u_t\}$ have to be involved in the selection of optimal filters [3], [4], [5], [9], [10]. Minimization of $E(|u_t|^k - r_k)^2$ with $r_k = E(|x_t|^{2k})/E(|x_t|^k)$ and $k > 1$, for example, was suggested in [5], whereas maximization of $|c_k(u_t)|/(c_2(u_t))^{k/2}$ with $c_k(u_t)$ being the k-th order cumulant of u_t for $k > 2$ was discussed in [3], [4]. The *stationarity* of $\{x_t\}$ is a crucial requirement in all these procedures, and many of them further require some moments of $\{x_t\}$ to be available. The estimation accuracy of these procedures is usually $O(1/n)$ [4].

This accuracy limit, however, can be significantly improved when the discreteness of the input signal is taken into account. In fact, for an m-ary signal whose alphabet is $\mathcal{A} = \{a_i, i = 1, \ldots, m\}$, a highly efficient estimator can be obtained by minimizing

$$\hat{J}_n(\theta) = \frac{1}{2n+1} \sum_{t=-n}^{n} \prod_{i=1}^{m} |\hat{u}_t(\theta) - a_i|^2, \tag{5}$$

where $\{\hat{u}_t(\theta)\}$ results from the inverse filtering

$$\hat{u}_t(\theta) = \sum_{j=-2n}^{2n} s_{t-j}^{-1}(\theta)\, y_j \tag{6}$$

using only the observed data $\{y_t, t = -2n, \ldots, 2n\}$ (assuming $y_t = 0$ for all $|t| > 2n$). This criterion measures the *closeness of* $\{\hat{u}_t(\theta)\}$ *from being an* $\mathcal{A}$*-valued discrete sequence*. It can be shown [6], [7] that the minimizer of $\hat{J}_n(\theta)$, denoted by $\hat{\theta}_n$, is a consistent estimator for the true parameter θ^* and, more importantly, that the estimation error $\|\hat{\theta}_n - \theta^*\|$ is bounded by the tail behavior of the true inverse system so that

$$\|\hat{\theta}_n - \theta^*\| \le c \sum_{|j|\ge n} |s_j^{-1}(\theta^*)| \tag{7}$$

where $c > 0$ is a constant. For ARMA systems, this implies that the error of $\hat{\theta}_n$ decays as an *exponential function* rather than the square-root reciprocal of the sample size n. In other words, minimization of $\hat{J}_n(\theta)$ would produce "super-efficient" estimates for the blind deconvolution problem. If the system is autoregressive with finite order, the super-efficiency yields

$$\lim_{n\to\infty} \Pr(\hat{\theta}_n = \theta^*) = 1.$$

In other words, the minimizer of $\hat{J}_n(\theta)$ would be *equal* to the true values of the parameter with probability tending to unity as n increases. It is also important to point out that all these results can be obtained without requiring the x_t to have the same distribution as long as they are independent [6], [7]. Therefore the super-efficiency applies even to *nonstationary* signals.

To demonstrate these results, let us consider a simple nonminimum-phase MA(2) system [6]

$$y_t = -1.5x_t + 3.5x_{t-1} - x_{t-2}$$

where $\{x_t\}$ is a *binary* sequence with $\Pr(x_t = 0) = p_t$ and $\Pr(x_t = 1) = 1 - p_t$. For the general MA(2) model $y_t = b_0x_t + b_1x_{t-1} + b_2x_{t-2}$, we assume $b_0 + b_1 + b_2 = 1$ and reparametrize the resulting two-parameter system with the zeros of the polynomial $b_0z^2 + b_1z + b_2$ denoted by $\theta = (z_1, z_2)$. Therefore, in this example, $\theta^* = (z_1^*, z_2^*) = (1/3, 2)$. To compare with other methods which make no use of the discreteness of the input signal, we consider the well-known procedure of maximizing the *standardized skewness* [4], [3], [9]

$$\hat{S}_n(\theta) = \frac{|\hat{c}_3(\hat{u}_t)|}{(\hat{c}_2(\hat{u}_t))^{3/2}}$$

where $\hat{u}_t = \hat{u}_t(\theta)$ is the output of the inverse filter in (6) and $\hat{c}_k(\hat{u}_t)$ the k-th order sample cumulant of $\hat{u}_t$. Two cases are considered: In Case 1 the input signal $\{x_t\}$ is stationary with $p_t = 0.6$ for all t, while in Case 2 it is nonstationary with $p_t = \Phi(\sin(3t\pi/128))$ where $\Phi(\cdot)$ is the distribution function of the standard normal random variable. In both cases a random sample of size $n = 1000$ is used in the computation of $\hat{J}_n(\theta)$ and $\hat{S}_n(\theta)$, and the contour plots of these criteria are presented in Figures 1–4.

As we can see from Figs. 1 and 3, the (binariness) criterion $\hat{J}_n(\theta)$ has a very sharp valley near the true value θ^* (indicated by +) in *both* stationary and nonstationary cases. This implies that minimizing $\hat{J}_n(\theta)$ will produce very precise estimates for both stationary and nonstationary input signals. On the other hand,

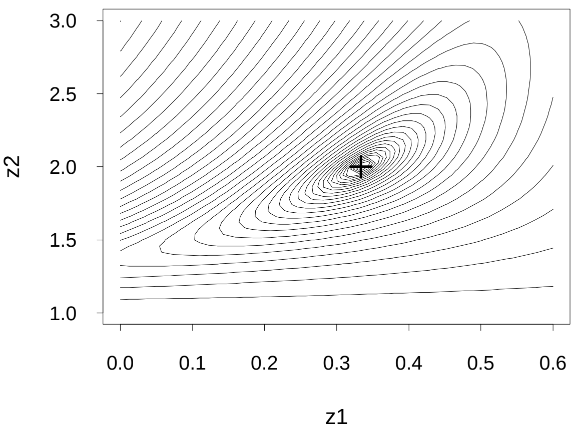

Fig. 1. Contour of $\hat{J}_n(\theta)$: Stationary case.

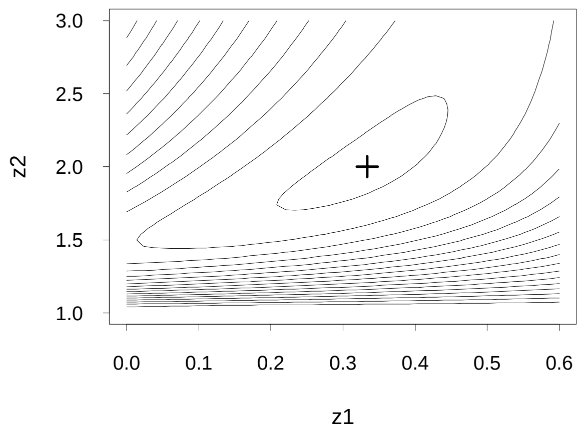

Fig. 2. Contour of $\hat{S}_n(\theta)$: Stationary case.

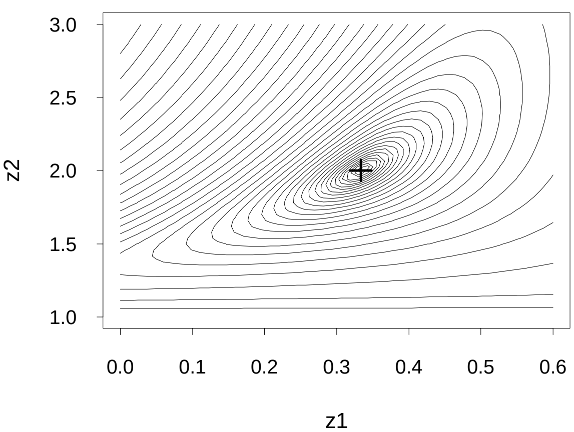

Fig. 3. Contour of $\hat{J}_n(\theta)$: Nonstationary case.

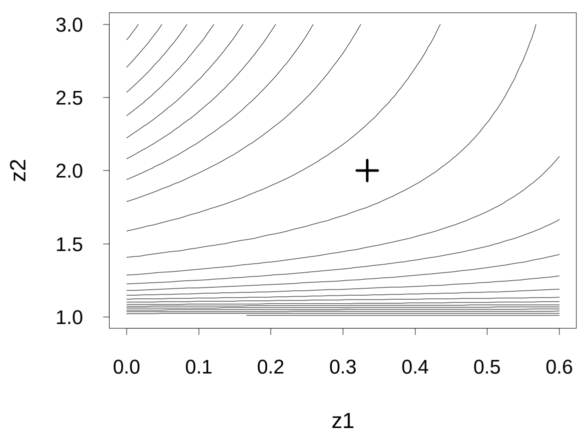

Fig. 4. Contour of $\hat{S}_n(\theta)$: Nonstationary case.

the standardized skewness $\hat{S}_n(\theta)$ has a rather broad peak near θ^* in the stationary case (Fig. 2). Although a solution to the deconvolution problem is provided by maximizing $\hat{S}_n(\theta)$, the *broad* peak in $\hat{S}_n(\theta)$ as shown by Fig. 2 may yield inaccurate estimates for θ^*. To make things even worse, the peak completely disappears in $\hat{S}_n(\theta)$ for the nonstationary signal (Fig. 4). This reveals how crucial the stationarity may be to the successful implementation of procedures like maximization of the standardized skewness. It is evident that the advantage of $\hat{J}_n(\theta)$ comes primarily from its utilization of the discreteness of the input signal.

3. A Gibbs Sampling Procedure

Suppose $\{s_j\}$ in (1) is an FIR system operated in a noisy environment so that $\{y_t\}$ is obtained from

$$y_t = \sum_{j=0}^{q} \phi_j x_{t-j} + \epsilon_t \tag{8}$$

where $\{\epsilon_t\}$ is Gaussian white noise with unknown variance σ^2. For the input signal, we assume that $\{x_t\}$ is a first-order Markov chain with state space $\mathcal{A}$, unknown initial probabilities $\theta_i = \Pr(x_{1-q} = a_i)$, and unknown transition probabilities $\theta_{ij} = \Pr(x_t = a_j | x_{t-1} = a_i)$. The blind deconvolution (or restoration) problem becomes the joint estimation of all the unknown parameters $\boldsymbol{\phi} = [\phi_0, \ldots, \phi_q]^T$, $\boldsymbol{\theta} = \{\theta_i, \theta_{ij}\}$, and σ^2, and the recovery of the unknown input $\boldsymbol{x} = \{x_{1-q}, \ldots, x_n\}$, solely from a finite data set $\boldsymbol{y} = \{y_1, \ldots, y_n\}$. It should be pointed out that most of the previously mentioned methods of blind deconvolution do not directly apply to this situation since the input signal $\{x_t\}$ is *colored* and its moments *unknown*.

To deal with this problem, Chen and this author have recently combined the Bayesian approach with a Gibbs sampling procedure [2]. The gist of method can be summarized as follows. Upon regarding all the unknowns as independent random variables/vectors, a multivariate Gaussian distribution and an inverse chi-squire distribution are used as priors for $\boldsymbol{\phi}$ and σ^2, respectively, so that

$$\boldsymbol{\phi} \sim \mathrm{N}(\boldsymbol{\phi}_0, \Sigma_0) \quad \text{and} \quad \sigma^2 \sim \chi^{-2}(\nu; \lambda)$$

(i.e., $\nu\lambda/\sigma^2 \sim \chi^2(\nu)$). Dirichlet distributions are employed as priors for the θ's, namely

$$\begin{aligned}(\theta_1, \ldots, \theta_m) &\sim \mathrm{D}(\alpha_1, \ldots, \alpha_m) \quad \text{and} \\ (\theta_{i1}, \ldots, \theta_{im}) &\sim \mathrm{D}(\alpha_{i1}, \ldots, \alpha_{im}),\end{aligned}$$

so that $p(\theta_1, \ldots, \theta_m) \propto \prod \theta_i^{\alpha_i}$ with $\sum \theta_i = 1$ and $p(\theta_{i1}, \ldots, \theta_{im}) \propto \prod \theta_j^{\alpha_{ij}}$ with $\sum_j \theta_{ij} = 1$. Selection of the parameters in these priors reflects the *a priori* information about the unknowns. For instance, small values of ν and λ or large variances in Σ_0 correspond to less informative priors suitable for the situations where information about $\boldsymbol{\phi}$ and σ^2 is limited. Jeffray's non-informative Dirichlet prior for $(\theta_1, \ldots, \theta_m)$ corresponds to $\alpha_i = -1/2$ while in general $\alpha_i > -1$.

According to the Bayesian approach, one is interested in seeking the conditional expectation $E(x_t|\boldsymbol{y})$ or the mode of the conditional probability $p(x_t|\boldsymbol{y})$, for instance, as estimates of x_t. The difficulty is that any direct computation of these estimates seems impossible because of the complexity of the problem (more unknowns than observations). Alternatively, one may employ the Monte Carlo method with a Gibbs sampler. The idea of Gibbs sampling is to construct a Markov chain by recursively generating random samples from the *conditional* posterior distribution of an individual or a subset of the unknowns given the data $\boldsymbol{y}$ and the rest unknowns. This procedure continues until the sampling Markov chain converges in distribution. In this case, the random samples generated by the Gibbs sampler can be regarded as ergodic samples from the *joint* posterior distribution $p(\boldsymbol{x}, \boldsymbol{\phi}, \sigma^2, \boldsymbol{\theta}|\boldsymbol{y})$, so the simple average of the x_t components and the maximum relative frequency of $x_t = a_i$ obtained from these samples, for example, will approximate the conditional expectation (MMSE estimator) $E(x_t|\boldsymbol{y})$ and the MAP estimator $\mathrm{mode}\{p(x_t|\boldsymbol{y})\}$, respectively.

It is not too difficult to derive for the Gibbs sampler the conditional posterior distributions of the unknowns in our problem. As a matter of fact, it can be shown [2] that the conditional posterior distribution of $\boldsymbol{\phi}$ given $\boldsymbol{y}$ and the rest unknonws is Gaussian with mean vector $\boldsymbol{\phi}_1$ and covariance matrix Σ_1, i.e.,

$$p(\boldsymbol{\phi} \,|\, \mathrm{rest}, \boldsymbol{y}) \sim \mathrm{N}(\boldsymbol{\phi}_1, \Sigma_1)$$

where

$$\begin{aligned}\Sigma_1^{-1} &= \sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^T/\sigma^2 + \Sigma_0^{-1} \quad \text{and} \\ \boldsymbol{\phi}_1 &= \Sigma_1 \left(\sum_{t=1}^{n} \boldsymbol{x}_t y_t/\sigma^2 + \Sigma_0^{-1}\boldsymbol{\phi}_0 \right)\end{aligned}$$

with $\boldsymbol{x}_t = [x_t, \cdots, x_{t-q}]^T$. Similarly, it can be shown [2] that $p(\sigma^2|\,\mathrm{rest}, \boldsymbol{y}) \sim \chi^{-2}(\mu + n; \nu\lambda + s^2)$,

$$\begin{aligned}&p(\theta_1, \ldots, \theta_m|\,\mathrm{rest}, \boldsymbol{y}) \\ &\quad \sim \mathrm{D}(\alpha_1 + \delta_1, \ldots, \alpha_m + \delta_m), \quad \text{and} \\ &p(\theta_{i1}, \ldots, \theta_{im}|\,\mathrm{rest}, \boldsymbol{y}) \\ &\quad \sim \mathrm{D}(\alpha_{i1} + n_{i1}, \ldots, \alpha_{im} + n_{im}),\end{aligned}$$

where $s^2 = \sum_{t=1}^{n}(y_t - \sum_{j=0}^{q} \phi_j x_{t-j})^2$,

$$n_{ij} = \#\{(x_t, x_{t-1}) = (a_i, a_j)\},$$

and $\delta_i = 1$ if $x_{1-q} = a_i$ and $\delta_i = 0$ if $x_{1-q} \neq a_i$. For any fixed $t' \in \{1-q, \ldots, n\}$, the conditional posterior distribution of $x_{t'}$ can be expressed as

$$\Pr(x_{t'} = a_i|\,\mathrm{rest}, \boldsymbol{y}) \propto p(\boldsymbol{x}'|\boldsymbol{\theta}) \, \exp(-s'^2/(2\sigma^2))$$

where $\boldsymbol{x}' = \{x'_{1-q}, \ldots, x'_n\}$ with $x'_{t'} = a_i$ and $x'_t = x_t$ for $t \neq t'$, and $s'^2 = \sum_{t=1}^{n}(y_t - \sum_{j=0}^{q} \phi_j x'_{t-j})^2$. Note that under the Markovian assumption of $\{x_t\}$ we have $p(\boldsymbol{x}|\boldsymbol{\theta}) = (\prod \theta_i^{\delta_i})(\prod \theta_{ij}^{n_{ij}})$.

As an example of the Gibbs sampling procedure, let us consider the MA(3) system

$$\begin{aligned}y_t = &-0.1833x_t + 0.9162x_{t-1} \\ &+ 0.4812x_{t-2} - 0.1987x_{t-3} + \epsilon_t\end{aligned}$$

where $\{x_t\}$ is a four-level Markov chain with $\mathcal{A} = \{-3, -1, 1, 3\}$, $\theta_i = 1/4$, and

$$[\theta_{ij}] = \begin{bmatrix} .4 & .2 & .2 & .2 \\ .2 & .4 & .2 & .2 \\ .1 & .3 & .4 & .2 \\ .1 & .2 & .3 & .4 \end{bmatrix}.$$

A realization of $\{x_t\}$ with $n = 100$ is shown in Fig. 5(a) and the corresponding $\{y_t\}$ shown in Fig. 5(b). The sample variance of $\{\epsilon_t\}$ is adjusted so that the signal-to-noise ratio in $\{y_t\}$ equals 15 dB. The parameters in the prior distributions are chosen as follows: $\boldsymbol{\phi}_0 = \mathbf{0}$, $\Sigma_0 = 1000$ I, $\nu = 2$, $\lambda = 0.3$, and $\alpha_i = \alpha_{ij} = 1$. Fig. 1(c) shows the i.i.d. uniform initial guess for $\{x_t\}$ in the Gibbs sampler, and Figs. 5(d) and 5(e) present the conditional mean and mode of x_t given $\boldsymbol{y}$, i.e., $E(x_t|\boldsymbol{y})$ and mode$\{p(x_t|\boldsymbol{y})\}$, respectively, calculated from the last 500 samples of the total 1000 iterations of Gibbs sampling. The constraints $\phi_1 \geq 0.4$ and $\phi_1 \geq |\phi_i| + 0.2$ for $i \neq 1$ are used to remove the sign and shift ambiguities in the solution. Estimates of $\boldsymbol{\phi}$ and θ_{ij} are given in the form of $E(\cdot|\boldsymbol{y}) \pm \sqrt{V(\cdot|\boldsymbol{y})}$ by, respectively,

$$(-0.1539, 0.8942, 0.4692, -0.1781) \pm (0.0191, 0.0195, 0.0202, 0.0211)$$

and

$$\begin{bmatrix} .45 & .15 & .22 & .18 \\ .21 & .36 & .30 & .13 \\ .05 & .34 & .42 & .19 \\ .17 & .12 & .39 & .32 \end{bmatrix} \pm \begin{bmatrix} .11 & .08 & .09 & .08 \\ .07 & .08 & .08 & .06 \\ .03 & .07 & .08 & .06 \\ .08 & .06 & .10 & .09 \end{bmatrix}.$$

It is evident by comparing Figs. 5(d) and 5(e) with Fig. 5(a) that the MAP estimator mode$\{p(x_t|\boldsymbol{y})\}$ completely recovers the input signal $\{x_t\}$ from the noisy data while the recovery by the MMSE estimator $E(x_t|\boldsymbol{y})$ is almost complete except for the last point, even though the sample size is relatively small. The estimates for the system parameters and the transition probabilities are reasonably accurate given that n is merely 100. This demonstrates again the impact of the discreteness of input signals on the improvement of blind deconvolution solutions.

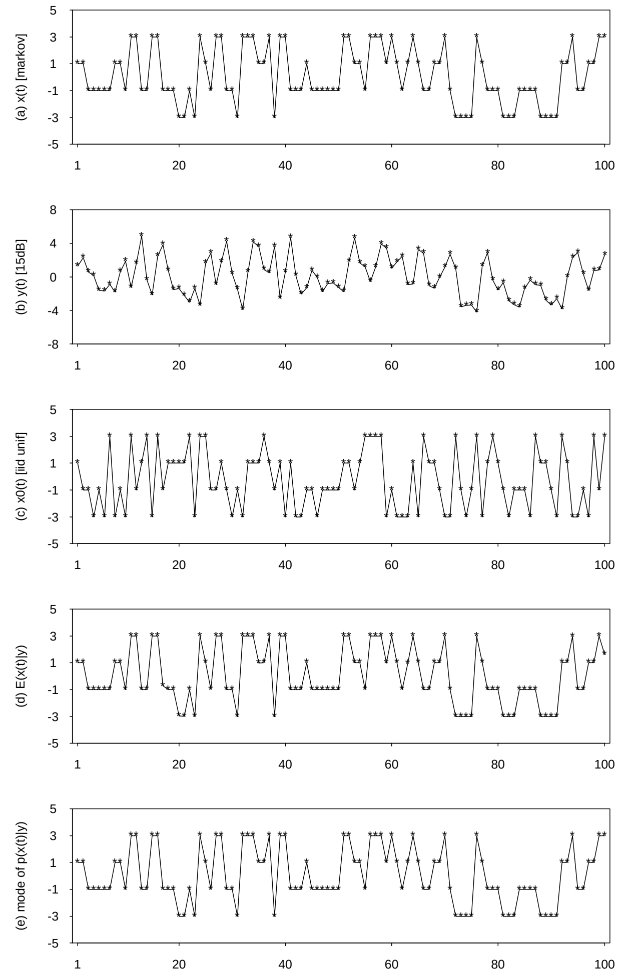

Fig. 5. Deconvolution by Gibbs sampling.

References

[1] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, 2nd Ed., New York: Springer, 1991.

[2] R. Chen and T.H. Li, "Blind restoration of linearly degraded discrete signals by Gibbs sampler," Tech. Rep. 197, Dept. of Statist., Texas A&M University, College Station, 1993.

[3] Q. Cheng, "Maximum standardized cumulant deconvolution of non-Gaussian processes", *Ann. Statist.*, vol. 18, pp. 1774–1783,1990.

[4] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*, D. Findley, Ed., New York: Academic, 1981.

[5] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.

[6] T.H. Li, "Blind identification and deconvolution of linear systems driven by binary random sequences," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 26–38, Jan. 1992.

[7] T.H. Li, "Blind deconvolution of linear systems with nonstationary multilevel inputs," *Proc. IEEE Signal Process. Workshop on Higher-Order Statist.*, S. Lake Tahoe, CA, pp. 160–163, June 1993.

[8] J.G. Proakis, *Digital Communications*, 2nd Edn, New York: McGraw-Hill, 1989.

[9] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 312–321, Mar. 1990.

[10] J.K. Tugnait, "Inverse filter criteria for estimation of linear parametric models using higher order statistics," *Proc. ICASSP-91*, pp. 3101–3104.