Adaptive Blind Deconvolution of Linear Channels Using Renyi's Entropy with Parzen Window Estimation

Deniz Erdogmus, *Member, IEEE*, Kenneth E. Hild, *Member, IEEE*, Jose C. Principe, *Fellow, IEEE*, Marcelino Lazaro, *Member, IEEE*, and Ignacio Santamaria, *Member, IEEE*

***Abstract*—Blind deconvolution of linear channels is a fundamental signal processing problem that has immediate extensions to multiple-channel applications. In this paper, we investigate the suitability of a class of Parzen-window-based entropy estimates, namely Renyi's entropy, as a criterion for blind deconvolution of linear channels. Comparisons between maximum and minimum entropy approaches, as well as the effect of entropy order, equalizer length, sample size, and measurement noise on performance, will be investigated through Monte Carlo simulations. The results indicate that this nonparametric entropy estimation approach outperforms the standard Bell–Sejnowski and normalized kurtosis algorithms in blind deconvolution. In addition, the solutions using Shannon's entropy were not optimal either for super- or sub-Gaussian source densities.**

***Index Terms*—Blind deconvolution, Parzen windowing, Renyi's entropy.**

I. INTRODUCTION

BLIND signal processing has become an important contemporary research topic due to the wide range of engineering applications that could benefit from such techniques. Online principal components analysis (PCA) is probably the first blind technique based on the second-order statistics of the data [1]. An immediate extension of PCA is independent component analysis (ICA) [2], [3], which seeks independence between the separated components. An important class of blind adaptation problems that predated ICA is single linear channel deconvolution (or equalization) [4]–[7]. In a classic paper, Donoho presented an overview of the initial approaches taken to solve the blind deconvolution (BD) problem [6]. These methods were broadly identified as *minimum entropy deconvolution* algorithms, in spite of the fact that none of the approaches presented utilized an explicit estimate of entropy from the samples. Instead, the algorithms concentrated on higher order normalized cumulants of the signals, which mimicked the behavior of entropy when operated on by a linear filter. Bussgang methods are an alternative classical way of achieving BD through the use of nonlinearities [8].

A generalized class of blind signal processing problems that encompasses both ICA and single-channel linear BD is linear blind source separation (BSS). This general framework deals with separating source signals that are filtered and mixed with each other with an unknown mixing process. Since BSS might involve convolutive mixtures, it is also sometimes referred to as multichannel blind deconvolution [9]–[11], for which closed-form solutions can be obtained using frequency-domain methods [12]–[14]. The blind signal processing research on multi-input multi-output channels finds applications in many practical problems, multiple access communication systems being the major target application [15]–[17].

The popular contemporary approach of tackling BD relies on the higher order cumulants or spectra of the signals to achieve the separation/deconvolution task [8], [18]–[20]. The main exception is the Bell–Sejnowski algorithm, which uses the maximization of Shannon's entropy of the deconvolved outputs following a nonlinearity that is tuned to the cumulative density function (cdf) of the source signal [21]. The utility of entropy-based criteria is well known and studied in the BD literature [4], [6], [8], [19], [21]–[25]. The main remaining problem is to obtain robust and data-efficient estimates of this quantity. This drawback is usually the reason for resorting to higher order cumulants when approaching the BD problem. Although the nonparametric estimation of Shannon's entropy is a well-studied area for which there exists abundant literature (see [26] for a review), many of these methods are not suitable for adaptive filtering. Since gradient-based methods form the backbone of adaptation methods, a continuous and differentiable entropy estimator is useful for this purpose. This is why Parzen density estimation [27] appears to be suitable for the task of entropy estimation in the context of adaptive filtering. Since Parzen windowing is a consistent estimator, by simply plugging in the corresponding density estimate in the entropy definition, it is possible to directly obtain a *plug-in entropy estimator*, as termed by Beirlant *et al.* [26]. In fact, such an estimator was employed by Viola in the context of mutual-information manipulation of images [28].

The combination of the Parzen density estimator (using Gaussian kernels) with Renyi's quadratic entropy has been extensively studied by Principe and co-workers in the context of blind source separation and information theoretic feature

Manuscript received June 19, 2002; revised August 5, 2003. This work was supported in part by the National Science Foundation under Grant ECS-0300340. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Athina Petropulu.

D. Erdogmus, K. E. Hild, and J. C. Principe are with the Electrical and Computer Engineering Department, University of Florida, Gainesville, FL 32611 USA (e-mail: deniz@cnel.ufl.edu; principe@cnel.ufl.edu; k.hild@ieee.org).

M. Lazaro is with the Signals and Communications Theory Department, University of Carlos III, Madrid, Spain (e-mail: marce@ieee.org).

I. Santamaria is with the Communications Engineering Department, University of Cantabria, Santander, Spain (e-mail: nacho@gtas.dicom.unican.es).

Digital Object Identifier 10.1109/TSP.2004.827202

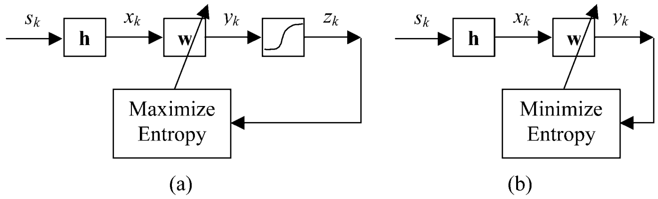


Fig. 1. Schematic diagram of (a) maximum (b) minimum entropy deconvolution.

extraction.[1] We are specifically interested in this entropy estimator because recent studies showed that in instantaneous BSS and feature extraction, it resulted in improved performance and data efficiency [31]–[33]. These good results motivated the investigation of the performance of this nonparametric Renyi's entropy estimator in adaptive BD.

Blind deconvolution of a linear channel can be achieved both by maximizing and minimizing the output entropy of an adaptive equalizer. The schematic diagrams of these approaches are shown in Fig. 1. Here, $\mathbf{h}$ denotes the unknown channel impulse response, and $\mathbf{w}$ denotes the finite impulse response (FIR) equalizer impulse response. Besides the channel, the actual source waveform s_k is also unknown, except for its statistics. The task is to determine the inverse of the channel $\mathbf{h}$ by adapting the equalizer coefficients $\mathbf{w}$. This paper investigates the performance of the Parzen-window-based nonparametric estimator for Renyi's entropy in solving the BD problem. Since Renyi's entropy includes Shannon's as a special case, the analysis will also provide a comparison of these two types of entropy when estimated using the same plug-in methodology.[2]

Theoretical results that support the application of Renyi's entropy to BD, as well as simulations results that indicate the effect of entropy order, adaptive filter length, sample size, and measurement noise, will be presented in Sections II–V.

II. Parzen-Window-Based Nonparametric Renyi's Entropy Estimator

A simple and useful plug-in estimator for Renyi's entropy can be obtained by substituting the Parzen window probability density function (pdf) estimate in the entropy definition. Suppose that we are given the independent and identically distributed (iid) samples $\{y_1, \dots, y_L\}$ of a random variable Y with the pdf $f_Y(\cdot)$. The Parzen window estimate of this pdf, using the kernel function $\kappa_\sigma(\cdot)$, is given by

$$\hat{f}_Y(y) = \frac{1}{L}\sum_{i=1}^{L}\kappa_\sigma(y - y_i) \tag{1}$$

where σ is the kernel size, which controls the tradeoff between estimation bias and estimation variance [27]. Typically, the kernel function is selected to be a zero-mean, symmetric, and differentiable pdf. Renyi's entropy of order α for the random variable Y, on the other hand, is defined as [39]

$$\begin{aligned} H_\alpha(Y) &= \frac{1}{1-\alpha}\log\int_{-\infty}^{\infty} f_Y^\alpha(y)dy \\ &= \frac{1}{1-\alpha}\log E_Y[f_Y^{\alpha-1}(Y)]. \end{aligned} \tag{2}$$

This is a parametric family of entropy measures that include Shannon's entropy definition as a limiting special case for $\alpha \to 1$, which is observed easily by applying L'Hopital's rule (for other characteristic properties of Renyi's entropy see Appendix A). Shannon's entropy for Y is defined as

$$H_S(Y) = -\int_{-\infty}^{\infty} f_Y(y)\log f_Y(y)dy = -E_Y[\log f_Y(Y)]. \tag{3}$$

When estimating these entropy measures from a finite number of samples, where the source pdf is unknown, nonparametric density estimates can be used. Specifically, approximating the expectation operator with the sample mean and substituting the Parzen window estimate in (2) and (3) for the pdf, nonparametric Renyi's entropy estimator becomes[3]

$$\hat{H}_\alpha(Y) = \frac{1}{1-\alpha}\log\left[\frac{1}{L^\alpha}\sum_{j=1}^{L}\left(\sum_{i=1}^{L}\kappa_\sigma(y_j - y_i)\right)^{\alpha-1}\right]. \tag{4}$$

Similarly, the Parzen window estimator for Shannon's entropy, which is also the limit of (4) as $\alpha \to 1$, is

$$\hat{H}_S(Y) = -\frac{1}{L}\sum_{j=1}^{L}\log\left(\frac{1}{L}\sum_{i=1}^{L}\kappa_\sigma(y_j - y_i)\right). \tag{5}$$

In adaptation scenarios where maximization or minimization of Renyi's entropy is required, the gradient of (4) or (5) with respect to the weights of the adaptive system that generated the samples of Y can be employed in a steepest ascent or descent algorithm. These gradients, which are evaluated over a window of L samples, are given by (6), shown at the bottom of the next page.

In offline adaptation (batch training), the complete training set can be used in calculating every gradient update, whereas in online adaptation, the gradients in (6) can be used with a sliding window of samples over the signal. The computational complexity (6) increases as $O(L^2)$, where L is the number of samples in the window. Incidentally, a simpler update rule known as the stochastic gradient approach, can be applied to the entropy definition to obtain a stochastic information gradient (SIG) [40], which has $O(L)$ complexity. These stochastic gradient algorithms, however, will increase the *misadjustment* of the weights in an online setting.

[1]It should be noted that the idea of combining Gaussian kernels with Renyi's quadratic entropy could be traced back to the early work of Frideman and Tukey in projection pursuit [29], as well as in nonlinear dynamics [30].

[2]Various other entropy definitions are available. These are generally structurally very similar to Renyi's definition or are generalizations of Renyi's entropy [34]–[38]. In this paper, we specifically focus on Renyi's entropy, as it is the first generalization stemming from Shannon's definition. However, future work will be conducted on investigating the possible benefits that might be obtained from utilizing alternative entropy definitions that are generalizations of Renyi's definition in blind deconvolution and other blind signal processing settings.

[3]Since this estimator is based on the substitution of the consistent Parzen window density estimate in Renyi's entropy, the resulting nonparametric entropy estimate is a *plug-in* type. Therefore, all mathematical properties of such estimators pointed out in [13], and the references therein, apply.

III. Maximum Entropy Deconvolution Using Renyi's Entropy

While the vast majority of the BD algorithms proposed utilize cumulants and related statistical quantities, information theoretic approaches have been recently popular in the literature [21], [41], [42]. Bell and Sejnowski described how Shannon's entropy could be used in conjunction with a nonlinearity tuned to the cdf of the source signal in order to solve the ICA and BD problems. The schematic diagram of this BD approach is shown in Fig. 1(a).

It is well known that when the pdf of the range limited Z [in Fig. 1(a)] is forced to uniform by maximizing its entropy, the pdf of Y matches the pdf that is the derivative of the nonlinearity. In the BD context, the nonlinearity is selected to be the cdf of the source signal S. According to the Benveniste–Goursat theorem, under some restrictions, the pdf of Y can match that of S if and only if the impulse response $\mathbf{h}^*\mathbf{w}$ is a δ-function [5]. Therefore, under the assumptions that the equalizer $\mathbf{w}$ is of sufficient length to invert the channel $\mathbf{h}$ and that the nonlinearity matches the cdf of the source signal, it is possible to determine the optimal equalizer weights by maximizing the entropy at the output of the nonlinearity. In order to estimate the entropy of Z, we can use either (4) or (5); thus, the gradients in (6) are directly applicable to this situation.

IV. Minimum Entropy Deconvolution Using Renyi's Entropy

The schematic diagram of the minimum entropy deconvolution approach is shown in Fig. 1(b). Donoho discusses the principle behind this architecture in detail [6]. Bercher and Vignat recently proposed an estimator for Shannon's entropy and, inspired by Donoho, suggested its minimization for BD, although they did not treat this problem in detail [22]. Here, our aim is to demonstrate the applicability of Renyi's entropy in this setup. Consider the following theorem on the entropy of linear combinations of random variables [23], [24].

Theorem 1: Let $S_i \ \ (i=1,\ldots,n)$ be independent identically distributed random variables with pdf $p_S(\cdot)$. Let $H\alpha(\cdot)$ denote the order-α Renyi's entropy for a continuous random variable. If $a_i \ \ (i=1,\ldots,n)$ are n nonzero real coefficients in $Y=a_1S_1+\cdots+a_nS_n$, then for $\alpha>1$, we have $H_\alpha(Y)\geq H_\alpha(S)+\log|a_1\cdots a_n|^{1/n}$, where we have (6), shown at the bottom of the page, where equality of the two entropies occurs if and only if $a_i=\delta_{ij}$, where δ denotes the Kronecker-delta function. The inequality is reversed for $\alpha<1$.

Proof: A simple proof is provided in Appendix B. See [24] for an alternative (more complicated) proof.[4]

Notice that the BD problem is structurally identical to the linear combination of iid random variables described in Theorem 1. The coefficients a_i are replaced by the impulse response coefficients of the overall filter $\mathbf{h}^*\mathbf{w}$. This theorem, therefore, shows that the minimization of Renyi's entropy at the output of the equalizer can achieve BD. An important issue in designing cost functions for minimum entropy BD is that of scale invariance. Since the differential entropy of a continuous random variable depends on its standard deviation, the global minimum of a minimum entropy BD criterion occurs for zero equalizer weights (corresponding to a zero equalizer output and $-\infty$ differential entropy). This unwanted situation could be prevented by modifying the cost function with the equalizer output variance. It is easy to show that the following modified cost function is scale-invariant and that the zero-weight situation is not a minimum.

$$J(Y)=H_\alpha(Y)-\frac{1}{2}\log[\mathrm{Var}(Y)]. \tag{7}$$

In practice, the output variance needs to be estimated from the data. This can be effectively achieved making use of the fact that

$$\mathrm{Var}(Y)=\mathbf{w}^T\mathbf{R_X}\mathbf{w} \tag{8}$$

where $\mathbf{X}$ is random vector representing the input vector of the equalizer, and $\mathbf{R_X}=E[\mathbf{x}\mathbf{x}^T]$ is the corresponding covariance matrix (assuming that the input is zero mean). If the source is wide sense stationary and the channel is time-invariant, this input-covariance-based output variance estimate can provide a very accurate correction term. Alternative to the output variance correction term, Bercher and Vignat suggest the normalization of the maximum-tap-weight to unity after every update [24]. At the true solution, these two methods become identical.

Lemma 1: Assume that the distribution of the output signal Y is not $\delta(\cdot)$. Provided that the entropy estimate asymptotically converges as L goes to infinity (or roughly, on the average), the entropy estimate in (4) provides an upper bound to the actual entropy of Y, i.e., $\hat{H}_\alpha(Y)\geq H_\alpha(Y)$ for $\alpha>1$ and a lower

[4]The authors of [23] and [24] have independently arrived at the same theorem and utilized it for blind deconvolution, and interestingly, both results were presented at the same conference

$$\frac{\partial\hat{H}_\alpha(Y)}{\partial\mathbf{w}}=\frac{-\sum_{j=1}^{L}\left(\sum_{i=1}^{L}\kappa_\sigma(y_j-y_i)\right)^{\alpha-2}\left(\sum_{i=1}^{L}\kappa'_\sigma(y_j-y_i)\left(\frac{\partial y_j}{\partial\mathbf{w}}-\frac{\partial y_i}{\partial\mathbf{w}}\right)\right)}{\sum_{j=1}^{L}\left(\sum_{i=1}^{L}\kappa_\sigma(y_j-y_i)\right)^{\alpha-1}}$$

$$\frac{\partial\hat{H}_S(Y)}{\partial\mathbf{w}}=-\frac{1}{L}\sum_{j=1}^{L}\frac{\left(\sum_{i=1}^{L}\kappa'_\sigma(y_j-y_i)\left(\frac{\partial y_j}{\partial\mathbf{w}}-\frac{\partial y_i}{\partial\mathbf{w}}\right)\right)}{\left(\sum_{i=1}^{L}\kappa_\sigma(y_j-y_i)\right)} \tag{6}$$

bound $\hat{H}_\alpha(Y) \le H_\alpha(Y)$ for $\alpha < 1$. Equality is possible if and only if the kernel size is set to zero.

Proof: The proof is found in Appendix B.

Lemma 1 allows the substitution of the entropy estimator in place of the actual entropy since the estimated cost function asymptotically (as the number of samples approach ∞) provides an upper bound for the true cost, which needs to be minimized. Another advantage of the Parzen-window estimator for Renyi's entropy is that the kernel size can be reduced asymptotically to zero (e.g., inversely proportional to the number of samples) such that the upper bound provided by the nonparametric estimator asymptotically approaches the true value of the entropy (i.e., the bound becomes tighter).

The kernel function in the Parzen estimator plays another crucial role in Renyi's entropy-based BD. It is well known that the motivation behind using the minimum entropy criterion for BD is the central limit theorem. Due to the addition of many random variables, the signal pdf at the output of the channel approaches a Gaussian (especially for long channel impulse responses). Shannon's entropy, attaining its maximum value for a Gaussian density under the fixed variance constraint, is therefore the ideal Gaussanity measure for this situation. Minimizing Shannon's entropy, which is the special case of Renyi's entropy for $\alpha = 1$, while keeping the output variance constant, guarantees the maximization of non-Gaussianity at the equalizer output. However, when using Renyi's entropy orders other than 1, an inconvenience arises: These definitions of entropy do not have their maximum for a Gaussian density under the same circumstances.

For example, consider the generalized Gaussian family of densities, which are given by $C_1 e^{-C_2|x|^\beta}$. This family encompasses the Laplacian ($\beta = 1$), Gaussian ($\beta = 2$), and uniform ($\beta \to \infty$) densities as special cases. When we evaluate Renyi's entropies (for various values of α) for unit-variance pdfs selected from the generalized Gaussian family, we observe that Renyi's entropy has a peak at $\beta > 2$ for $\alpha > 1$ and $\beta < 2$ for $\alpha < 1$. Some of these entropy curves are shown in Fig. 2 to demonstrate this fact.

Observing the plots shown in Fig. 2, we can arrive at a number of conclusions. These curves demonstrate the theoretical result in Lemma 1. In addition, they verify the need to use different entropy orders for the BD of sub- and super-Gaussian sources. Assuming the channel output is very close to Gaussian, BD of a super-Gaussian source ($\beta < 2$) can be achieved by minimizing the Renyi's entropy estimate of order $\alpha > 1$. On the other hand, BD of a sub-Gaussian source ($\beta > 2$) can be achieved by minimizing the Renyi's entropy estimate of order $\alpha < 1$. Based on the curves in Fig. 2(c) and our previous experience with this entropy estimator [32] (which disclosed links with convolution smoothing [43]), we suggest using a reasonably large kernel size to produce a cost surface with fewer local minima. A demonstration of how the cost surface becomes smoother with increasing kernel size is presented for a small sample case in [23].

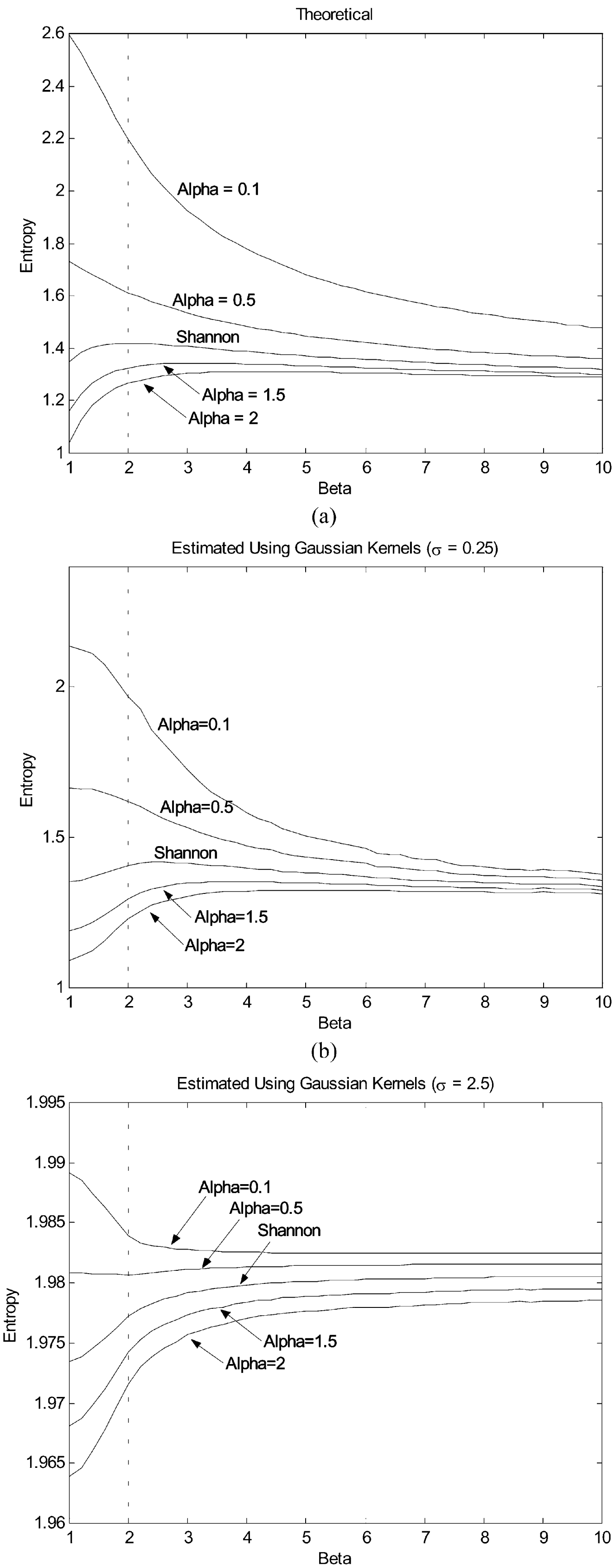


Fig. 2. Renyi's entropy for the generalized Gaussian family. (a) Theoretical. (b) Estimated with small Gaussian kernels. (c) Estimated with large Gaussian kernels.

V. SIMULATIONS

In this section, we present results from Monte Carlo simulations that demonstrate the performance of the Parzen-window-based Renyi's entropy approaches for BD. Both minimum and maximum entropy approaches are discussed, and the effect of various parameters such as entropy order, filter length, sample size, and measurement signal-to-noise-ratio (SNR) is studied. We will evaluate the performance of three classes of algorithms under the various settings using Monte Carlo simulations:

- maximum entropy with Renyi's entropy using Parzen windowing (RMAX);
- minimum entropy with Renyi's entropy using Parzen windowing (RMIN);
- maximum entropy with Bell-Sejnowski's [21] approach (INFOMAX).

We originally planned to include the normalized kurtosis method, where the cost function is given by $J(Y) = E[y^4]/E[y^2]^2$ [6], as a benchmark for the minimum entropy method; however, results obtained with this approach under identical conditions were much worse than the performance demonstrated by the three approaches listed above. Therefore, we do not include detailed results for the normalized kurtosis measure here.

In this section, the performances of the algorithms are measured in terms of the signal-to-interference-ratio (SIR), which is defined as

$$\mathrm{SIR(dB)} = 10\log_{10}\frac{[\max(|a_i|)]^2}{\sum_i |a_i|^2 - [\max(|a_i|)]^2} \tag{9}$$

where a_i are the tap weights of the overall filter from the source to the equalizer output, i.e., $\boldsymbol{a} = \mathbf{h}^* \mathbf{w}$.[5]

A. Effect of Entropy Order

In this set of Monte Carlo simulations, our aim is to investigate the effect of entropy order on the performance of the minimum and maximum entropy deconvolution algorithms described above. For this purpose, a fixed third-order all-poles channel with poles at 0.1, 0.5, and 0.9 is utilized on one Laplacian and one uniformly distributed source. A four-tap FIR equalizer is trained using 3000 randomly generated samples from both source distributions, which are mixed through the specified channel. The gradients in (6) are used with a sliding window of $L = 100$ samples and a step size of 10^{-5} for minimum entropy and 10^{-6} for maximum entropy methods. The size of the Gaussian kernels in Parzen windowing is set to 10 and 0.1 for minimum and maximum entropy methods, respectively. In the minimum entropy case, kernel size adjustment based on the variance of the samples was employed (i.e., the base kernel size given above was multiplied by the norm of the equalizer tap-weight vector). For ten randomly generated training data sets of both sources, the equalizers were trained, updating once per sample (i.e., 3000 updates), using the entropy orders $\alpha \in \{0.1, 0.2, \ldots, 0.9, 1, 0.5, 2, \ldots, 6\}$, where $\alpha = 1$ corresponds to Shannon's definition. The average of the final SIR values obtained for both approaches are presented in Fig. 3. In all cases, the standard deviation associated with the mean values was less than 1 dB. Interestingly, the performance of the minimum entropy approach does not significantly depend on the entropy order for both Laplacian (super-Gaussian) and uniform (sub-Gaussian) source densities. On the other hand, the performance of the maximum entropy approach is affected by the entropy order. Specifically, if we could generalize from the presented results, the minimum entropy approach is better than the maximum entropy approach, even though the maximum entropy algorithms used additional information in the form of the optimal nonlinearity that is matched to the cdf of the source. Second, if maximum entropy is used, then higher entropy orders should be preferred for sub-Gaussian sources, and lower entropy orders should be preferred for super-Gaussian sources.

[5]Note that in cases where measurement noise is present, SIR does not reflect the distortions caused by the noise.

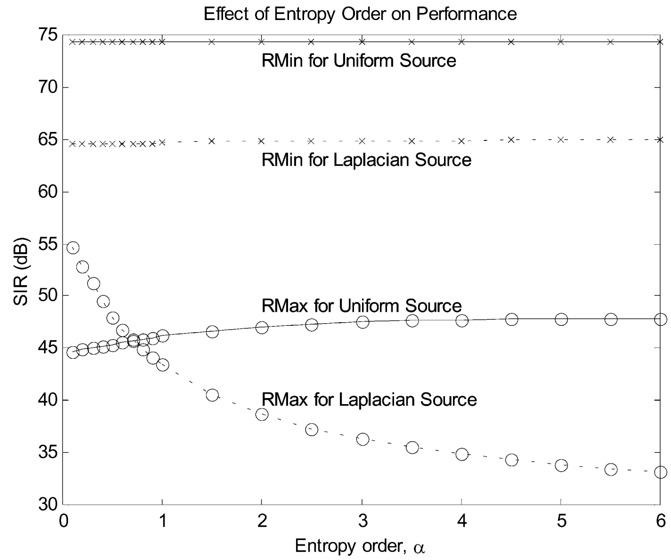


Fig. 3. Average SIR of minimum and maximum Renyi's entropy algorithms versus entropy order for Laplacian and uniform sources. Shannon's entropy corresponds to $\alpha = 1$.

B. Effect of Equalizer Length

In these simulations, we evaluate the effect of the equalizer length on the final performance of the algorithms RMax, RMin, and Infomax. Recall that Bell–Sejnowski's Infomax is basically a maximum entropy approach that optimizes Shannon's entropy after the nonlinearity using an update rule based on the Jacobian of the criterion with respect to the weights [21]. In the RMax and RMin approaches, we will consider two entropy orders that are of interest to us: Shannon's entropy (SMax and SMin), which corresponds to $\alpha = 1$, is appealing because of its historical significance, as well as the additivity properties that it satisfies, whereas quadratic entropy (QMax and QMin), which corresponds to $\alpha = 2$, is appealing because it provides the simplest functional form in the estimator and the gradient updates, thus minimizing computational complexity of algorithms. For these experiments, five different allpole channels are utilized. The poles of these channels were located at $[0.5]$, $[0.1, 0.9]$, $[0.1, 0.5, 0.9]$, $[0.1, 0.37, 0.63, 0.9]$, and $[0.1, 0.3, 0.5, 0.7, 0.9]$. The number of FIR taps required to equalize these channels range from 2 to 6. However, for each of these channels, we have trained five FIR equalizers with lengths changing from 2 to 6. If the equalizer is shorter than required, we expect the performance to be low due to insufficient equalization capability. If the equalizer is longer than necessary, we expect the performance to slowly drop due to the misadjustment of the unnecessary weights (in the stochastic update framework). For each

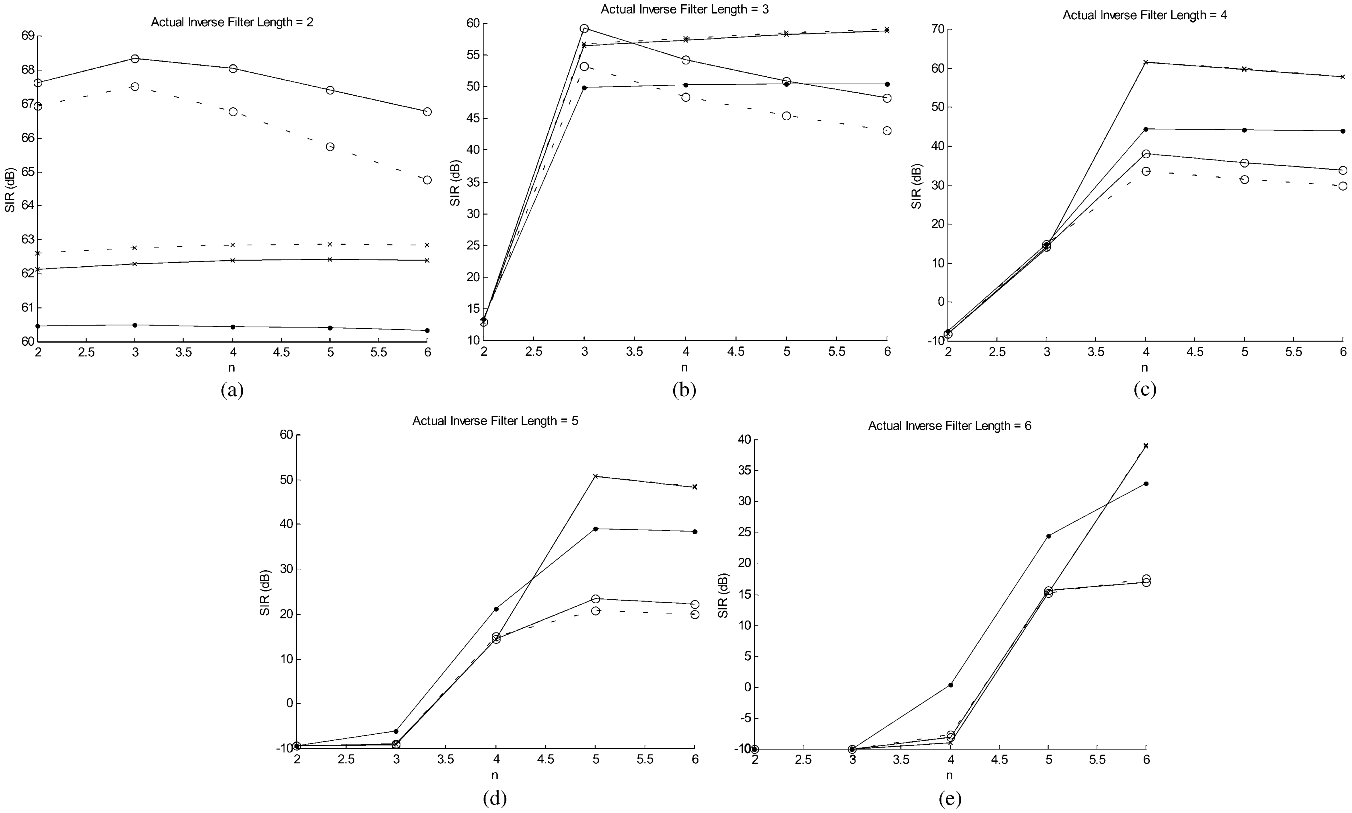


Fig. 4. Average SIR of QMax (dotted o), QMin (dotted x), SMax (solid o), SMin (solid x), and Infomax (solid •) versus equalizer length for various required/sufficient (r/s) equalizer lengths. (a) Two-tap FIR (r/s). (b) Three-tap FIR (r/s). (c) Four-tap FIR (r/s). (d) Five-tap FIR (r/s). (e) Six-tap FIR (r/s).

channel-equalizer combination, we have performed ten Monte Carlo runs. For each run, 5000 samples were generated by a Laplacian source, which were then passed through the designated channel. In addition, each equalizer is trained with the five algorithms (Infomax, SMin, SMax, QMin, and QMax) using updates computed over a sliding window of $L = 100$ samples. The kernel sizes and the step sizes for RMin and RMax algorithms are set to the same values as in the previous set of simulations. The step size for Infomax was set to 10^{-6}, and its nonlinearity was set to the optimal cdf, as before. For this set of experiments, although we do not present detailed results associated with the standard normalized kurtosis algorithm, it is important to note that the performance of this approach never exceeded an SIR of 10 dB when using semi-stochastic updates computed using the 100-sample sliding windows. However, the use of batch training with normalized kurtosis (i.e., every update uses all training samples), using an equalizer of correct length and 10 000 samples, resulted in an SIR around 25 dB on average.

The results in Fig. 4 indicate that when the equalizer is sufficiently long [except for the case in Fig. 4(a), where a two-tap equalizer is required], minimum entropy methods using Parzen window estimates outperform Infomax, which outperforms the maximum entropy methods that use Parzen windows. It is interesting to note that when the equalizer length is slightly smaller than required, Infomax achieves better deconvolution than all Parzen-window-based Renyi's entropy methods [especially observed in Fig. 4(d) and (e)], which indicates some form of robustness to insufficient model order. This robustness is likely to be a consequence of the close relationship between Infomax and maximum-likelihood solutions [3]. An additional observation is that the maximum entropy methods SMax and QMax perform particularly well for the short-equalizer situations in Fig. 4(a) and (b), in contrast to their relatively poor performances in training long-equalizers.

C. *Effect of Sample Size and Measurement SNR*

In the previous case studies, the effect of sample/window size and measurement noise on the performance of the algorithm is not addressed. In this section, we analyze the effect of these parameters on the performance of Renyi's entropy-based algorithms in a single set of Monte Carlo simulations. The Qmin, QMax, SMin, and SMax approaches are considered for the deconvolution of a unit-power Laplacian source from measurements through a unit-gain channel with two complex-conjugate poles located at $0.2 \pm j0.1$ via a three-tap FIR adaptive filter. Since the effect of sample size on the performance is sought, the experiments here are performed on smaller training sets using batch-updates for 5000 iterations (we observed that in general, this was sufficient to achieve convergence with all algorithms). For Qmin and SMin, the step size was 10^{-2} and the kernel size was 2, and for QMax and SMax, the step size was 10^{-3} and the kernel size was 0.1. By eliminating stochastic update misadjustment, batch training adds the advantage that the SIR loss (from the perfect solution) is attributed only to the finite number of samples and measurement noise. For sample sizes of $L = 150$, 350, 500, and 1000 and measurement noise levels of SNR= 10 and 30 dB, we performed ten Monte Carlo simulations, whose

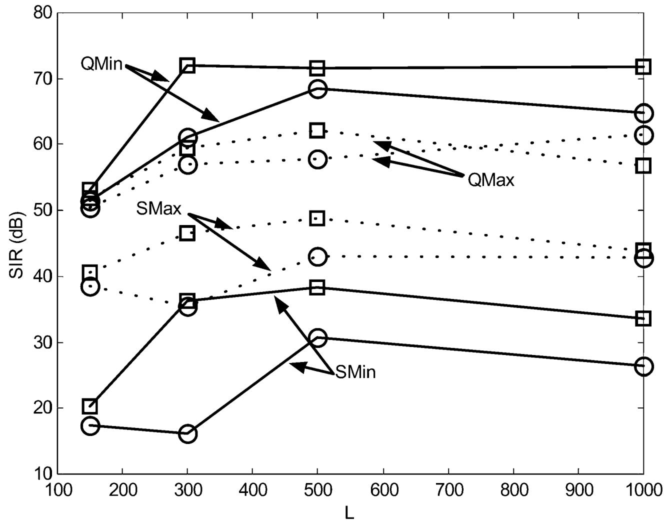


Fig. 5. Average SIR of Qmin, QMax, SMax, and SMin versus training set sample size for measurement SNR levels of 10 dB (o) and 30 dB ($\Box$).

average SIR results are presented in Fig. 5.[6] Recall that SIR does not include noise corruption on the deconvolved signals.[7]

We observe that all algorithms show improvement in performance as measurement SNR increases from 10 to 30 dB. The average performance also tends to increase as the sample size increases. According to Fig. 5, minimum Shannon's entropy is the algorithm that is most affected by a small sample size and measurement noise. Although the maximum Shannon's entropy approach performed better than the latter, both minimum and maximum Renyi's quadratic entropy algorithms outperformed their Shannon counterparts with a large margin in terms of their data efficiency and robustness to measurement noise. Finally, notice that since the total number of samples seen by the deconvolution algorithms in these experiments (≤ 1000) is smaller than the total number of samples seen in the previous case studies (3000 and 5000 for entropy order and equalizer length examples), the average performances of the algorithms decreased (even in the 30-dB SNR case, which is virtually noiseless). With the gradient in (6), updates using more than $L = 500$ samples become computationally cumbersome due to the $O(L^2)$ complexity. These results indicate that in practice, training of an equalizer with these criteria for large data sets should be performed using a small sliding window (e.g., $L = 100$). This approach will take full advantage of the available training data with a reasonable computational requirements. Results indicate that the performance gain from the utilization of additional data outweighs the additional misadjustment caused by the semi-batch updates from a sliding window.

[6]The curves for Infomax are not presented to avoid crowding the figure further; however, it suffices to say that the performance of this algorithm was similar to that of SMax.

[7]Nevertheless, if SIR + SNR at the equalizer output is the main consideration and if various accurate $(\mathrm{SIR} > 20\,\mathrm{dB})$ algorithms are considered, the SNR of the equalizer outputs for these different solutions will be similar, since the weight vectors are close to the optimal zero-forcing equalizer. Therefore, it is reasonable to assume that the SIR levels reflect the relative overall quality of the deconvolved output including both convolutive interference and measurement noise.

VI. Conclusions

Blind deconvolution is a basic problem in signal processing, which has applications extending to numerous areas of communications, speech processing, and image processing, to list a few. Although contemporary research focuses on multi-input multi-output blind deconvolution scenarios (mostly encouraged by the increased interest in blind source separation and multiple-access communication applications), most solutions under investigation are still inspired by the early algorithms and criteria originally proposed for the single-input single-output (SISO) blind deconvolution scenario. Therefore, the SISO case still attracts much attention from many researchers.

In this paper, we have investigated the performance of a data efficient, robust, and computationally simple estimator for Renyi's entropy in solving the SISO blind linear channel deconvolution problem. In previous work, we have shown that algorithms designed using this Parzen-window-based Renyi's entropy estimator were able to solve the ICA (or instantaneous-mixture BSS) problem, having superior performance to their counterparts, both in terms of convergence speed, which is useful in tracking nonstationary mixtures, and data efficiency, which is important in situations where collecting additional data for more accuracy is expensive. Motivated by those results, in this paper, we have investigated the performance of the Parzen window entropy estimator in the blind deconvolution context.

Using the Renyi family of entropy measures, we have studied the two well-known approaches to blind deconvolution: maximum entropy after a suitably selected nonlinearity and minimum entropy. As the benchmark for these approaches, we have considered the classical normalized kurtosis and Infomax algorithms. In Monte Carlo simulations designed to evaluate the effect of the entropy order on performance, we have determined that for minimum entropy algorithms, there was insignificant dependency, whereas for maximum entropy algorithms, the performance greatly depended on the selected entropy measure. Interestingly, Shannon's entropy did not turn out to be optimal for either super-Gaussian or sub-Gaussian sources. We have also investigated the performance degradation due to insufficient or excessive equalizer length. The former case results in incomplete equalization, whereas the latter results in unnecessary misadjustment in the equalizer weights. Experiments performed over various equalizer lengths demonstrated that the minimum entropy approach is superior to maximum entropy, as well as the Infomax and normalized kurtosis. Finally, results from Monte Carlo simulations designed to analyze the effect of sample size and measurement noise on the performance of these algorithms demonstrated that, as expected, as the number of samples seen by the training algorithm and the measurement signal-to-noise ratio increased, the performance of the algorithms (in determining the zero-forcing equalizer for the unknown channels) increased.

Appendix A

In this Appendix, we provide a brief discussion of characteristic properties that the Renyi parametric family of entropies exhibit.

- $H_\alpha(\cdot)$ is a decreasing function of the entropy order α.
- $\lim_{\alpha\to\infty} H_\alpha(Y) = \log(\max_y f_Y(y))$.
- $\lim_{\alpha\to 0} H_\alpha(Y) = \log(\text{vol support} \quad f_Y(y)\})$, where vol support$(\cdot)\}$ of a pdf denotes the volume of the support of the density.

These last two properties might be useful in understanding the behavior of RMax versus entropy order in Fig. 3. Generalizing from the presented results, for super-Gaussian sources (e.g., Laplacian), as kurtosis increases, the main support of the density becomes the discriminating factor. Hence, for smaller α, better performance is obtained. On the sub-Gaussian side (e.g., uniform), the maximum value of the source becomes the determining factor, and a larger α provides better performance.

Shannon's entropy satisfies a condition called subset independence [44], which is stated as follows for continuous random variables:

- Let sets $\{P_i\}$ $i = 1,\ldots,n$ be a partitioning of the support of a random variable Y, i.e., the support D of Y is given by the following union: $D = \bigcup_{i=1}^n P_i$. If $\{p_i\}$ are the probabilities of their respective partitions, i.e., $p_i = \text{Prob}\{Y \in P_i\} = \int_{y\in P_i} f_Y(y)dy$ and Y_i are random variables defined on each P_i with pdf $f(y)/p_i$, where $y \in P_i$ and zero everywhere else, then $H_S(Y) = \sum_{i=1}^n p_i H_S(Y_i)$.

Although Renyi's entropy (for $\alpha \neq 1$) does not satisfy this arithmetic mean property, it does satisfy an exponential mean [39]

- $f(H_\alpha(Y)) = \sum_{i=1}^n p_i f(H_\alpha(Y_i))$, where $f(x) = 2^{(\alpha-1)/x}$.

This exponential weighting of the individual partitions of the support might be related to the observed data efficiency of Renyi's entropy estimators (compared to Shannon). Specifically, depending on the selected entropy order, the estimator can be forced to emphasize the lobes or tails of the distribution under consideration. For example, it can be deduced from this property that as $\alpha \to 0$ (for $\alpha < 1$), Renyi's entropy increasingly relies on the low-density regions (partitions), whereas as for $\alpha \to \infty$ (for $\alpha > 1$), it depends more on the high-density regions. A discussion of how the emphasis on each training sample is distributed in the gradients of the Parzen window-based nonparametric Renyi estimator is presented in [32].

Appendix B

A. Proof of Theorem 1

Let S_1 and S_2 be independent, and let $Y = a_1S_1 + a_2S_2$. The pdf of Y is

$$p_Y(y) = \frac{1}{|a_1|}p_{S_1}\left(\frac{y}{a_1}\right) * \frac{1}{|a_2|}p_{S_2}\left(\frac{y}{a_2}\right). \tag{B.1}$$

Recall the definition of Renyi's entropy for Y given in (2). Notice that we can write

$$\begin{aligned} e^{(1-\alpha)H_\alpha(Y)} &= \int_{-\infty}^{\infty} p_Y^\alpha(y)dy \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\frac{1}{|a_1a_2|}p_{S_1}\left(\frac{\tau}{a_1}\right)\right. \\ &\qquad \left.\times\, p_{S_2}\left(\frac{y-\tau}{a_2}\right)d\tau\right]^\alpha dy. \end{aligned} \tag{B.2}$$

Using Jensen's inequality for convex and concave cases, we get the inequalities shown in (B.4), where $V_\alpha(\cdot)$ is called the order-α information potential. For a random variable X with pdf $p_X(\cdot)$, the information potential is given by [32], [45]

$$V_\alpha(X) = \int_{-\infty}^{\infty} p_X^\alpha(x)dx. \tag{B.3}$$

Notice that the information potential is the argument of the *log* in the entropy definition and is named after its resemblance to potentials of physical particles [31].

$$\begin{aligned} e^{(1-\alpha)H_\alpha(Y)} \underset{\alpha<1}{\overset{\alpha>1}{\lesseqgtr}} &\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\frac{1}{|a_1|}p_{S_1}\left(\frac{\tau}{a_1}\right)\right. \\ &\qquad\left.\cdot\left[\frac{1}{|a_2|}p_{S_2}\left(\frac{y-\tau}{a_2}\right)\right]^\alpha d\tau\right]dy \\ &= \int_{-\infty}^{\infty}\frac{1}{|a_1|}p_{S_1}\left(\frac{\tau}{a_1}\right) \\ &\quad\cdot\left[\int_{-\infty}^{\infty}\left[\frac{1}{|a_2|}p_{S_2}\left(\frac{y-\tau}{a_2}\right)\right]^\alpha dy\right]d\tau \\ &= \int_{-\infty}^{\infty}\frac{1}{|a_1|}p_{S_1}\left(\frac{\tau}{a_1}\right)V_\alpha(a_2S_2)d\tau \\ &= V_\alpha(a_2S_2)\cdot\int_{-\infty}^{\infty}\frac{1}{|a_1|}p_{S_1}\left(\frac{\tau}{a_1}\right)d\tau \\ &= V_\alpha(a_2S_2). \end{aligned} \tag{B.4}$$

Reorganizing the terms in the last inequality and using the relationship between entropy and information potential, regardless of the value of α and the direction of the inequality, we arrive at the conclusion $H_\alpha(Y) \geq H_\alpha(S_i) + \log|a_i|$, $i = 1, 2$.

Now, consider the case where Y is the linear combination of n iid random variables. We get n inequalities corresponding to the nonzero coefficients

$$\begin{aligned} H_\alpha(Y) &\geq H_\alpha(S) + \log|a_1| \\ H_\alpha(Y) &\geq H_\alpha(S) + \log|a_2| \\ &\vdots \\ H_\alpha(Y) &\geq H_\alpha(S) + \log|a_n|. \end{aligned} \tag{B.5}$$

Adding these inequalities, we get the desired result. The necessary and sufficient condition for the equality of entropies is obvious from the formulation. If $a_i = \delta_{ij}$, then $Y = S$, and therefore, the entropies are equal. If $a_i \neq \delta_{ij}$, then due to Theorem 1, the entropy of Y is greater than the entropy of S. $\square$

B. Proof of Lemma 1

Recall that the expected value of the Parzen window pdf estimate given in (1) is the convolution of the actual pdf underlying the samples and the kernel function (as N goes to infinity, the pdf estimate converges to this value since Parzen windowing is consistent). We can consider the average pdf as the pdf of a random variable, which is the sum of two independent random variables: one with the same pdf as the data samples and the second with pdf equal to the kernel function. If we define $S = Y + K$, where Y and K correspond to the independent random variables mentioned in the previous sentence, we can apply Theorem 1 to conclude that the entropy of S is larger than either of the entropies of Y and K. $\square$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their critical and constructive comments that helped improve this paper considerably.

REFERENCES

[1] E. Oja, *Subspace Methods of Pattern Recognition*. New York: Wiley, 1983.

[2] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, no. 1, pp. 1–10, 1991.

[3] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[4] R. A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, vol. 16, pp. 21–35, 1978.

[5] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communication," *IEEE Trans. Automat. Contr.*, vol. 25, pp. 385–199, 1980.

[6] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*. New York: Academic, 1981, pp. 565–609.

[7] J. J. Bussgang, "Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals," Mass. Inst. Technol. Res. Lab. Electron., Cambridge, MA, Tech. Rep. 216, 1952.

[8] S. Haykin, Ed., *Blind Deconvolution*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[9] I. Sabala, A. Cichocki, and S.-I. Amari, "Relationships between instantaneous blind source separation and multichannel blind deconvolution," in *Proc. IEEE Int. Joint Conf. Neural Networks*, 1998, pp. 39–44.

[10] J. K. Tugnait, "On blind separation of convolutive mixtures of independent linear signals in unknown additive noise," *IEEE Trans. Signal Processing*, vol. 46, pp. 3117–3123, Nov. 1998.

[11] Y. Inouye and K. Hirano, "Cumulant-based blind identification of linear multi-input–multi-output systems driven by colored inputs," *IEEE Trans. Signal Processing*, vol. 45, pp. 1543–1552, June 1997.

[12] K. I. Diamantaras, A. P. Petropulu, and B. Chen, "Blind two-input–two-output FIR channel identification based on second order statistics," *IEEE Trans. Signal Processing*, vol. 48, pp. 534–542, Feb. 2000.

[13] B. Chen and A. P. Petropulu, "Frequency domain MIMO system identification based on second and higher order statistics," *IEEE Trans. Signal Processing*, vol. 49, pp. 1677–1688, Aug. 2001.

[14] O. Grellier, P. Comon, B. Mourrain, and P. Trebuchet, "Analytical blind channel identification," *IEEE Trans. Signal Processing*, vol. 50, pp. 2196–2207, Sept. 2002.

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

[16] Y. Inouye and T. Sato, "Iterative algorithms based on multistage criteria for multichannel blind deconvolution," *IEEE Trans. Signal Processing*, vol. 47, pp. 1759–1764, June 1999.

[17] C.-Y. Chi and C.-H. Chen, "Cumulant-Based inverse filter criteria for MIMO blind deconvolution: Properties, algorithms, and application to DS/CDMA systems in multipath," *IEEE Trans. Signal Processing*, vol. 49, pp. 1282–1299, July 2001.

[18] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[19] S. Haykin, Ed., *Unsupervised Adaptive Filtering, Volume 2: Blind Deconvolution*. New York: Wiley, 2000.

[20] J. A. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE Signal Processing Mag.*, vol. 13, no. 3, pp. 24–42, 1996.

[21] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[22] J.-F. Bercher and C. Vignat, "Estimating the entropy of a signal with applications," *IEEE Trans. Signal Processing*, vol. 48, pp. 1687–1694, June 2000.

[23] D. Erdogmus, J. C. Principe, and L. Vielva, "Blind deconvolution with Renyi's minimum entropy," in *Proc. EUSIPCO*, vol. 2, Toulouse, France, 2002, pp. 71–74.

[24] J. F. Bercher and C. Vignat, "A Renyi entropy convolution inequality with application," in *Proc. EUSIPCO*, Tolouse, France, 2002.

[25] I. Santamaria, C. Pantaleon, L. Vielva, and J. C. Principe, "A fast algorithm for adaptive blind equalization using order-α Renyi's entropy," in *Proc. ICASSP*, vol. 3, Orlando, FL, 2002, pp. 2657–2660.

[26] J. Beirlant, E. Dudewicz, L. Gyorfi, and E. Van der Meulen, "Nonparametric entropy estimation," *Int. J. Math. Statist. Sci.*, vol. 6, no. 1, pp. 17–39, 1997.

[27] E. Parzen, "On estimation of a probability density function and mode," in *Time Series Analysis Papers*. San Francisco, CA: Holden-Day, 1967.

[28] P. Viola, N. Schraudolph, and T. Sejnowski, "Empirical entropy manipulation for real-world problems," in *Proc. NIPS*, 1995, pp. 851–857.

[29] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Comput.*, vol. 23, pp. 881–889, 1974.

[30] C. Diks, J. Houwelingen, F. Takens, and J. deGoede, "Detecting differences between delay vector distributions," *Phys. Rev. E*, vol. 53, pp. 2169–2176, 1996.

[31] J. C. Principe, D. Xu, and J. W. Fisher, "Information theoretic learning," in *Unsupervised Adaptive Filtering, vol. 1: Blind Source Separation*, S. Haykin, Ed. New York: Wiley, 2000, vol. 1, pp. 265–319.

[32] D. Erdogmus and J. C. Principe, "Generalized information potential criterion for adaptive system training," *IEEE Trans. Neural Networks*, vol. 13, pp. 1035–1044, Sept. 2002.

[33] K. E. Hild II, D. Erdogmus, and J. C. Principe, "Blind source separation using Renyi's mutual information," *IEEE Signal Processing Lett.*, vol. 8, pp. 174–176, June 2001.

[34] J. D. Aczel and Z. Daroczy, "Uber verallegmeinerte quasilineare mittelwerte die mit gewichtsfunktionnen gebildet sind," *Publications Mathematicae*, vol. 10, pp. 171–190, 1963.

[35] J. N. Kapur, "Generalized entropy of order and type," in *Mathematics Seminar*, vol. 4, 1967, pp. 78–94.

[36] J. Havrda and F. Charvat, "Quantification method of classification process: Concept of structural α-entropy," *Kybernetika*, vol. 3, pp. 30–35, 1967.

[37] B. D. Sharma and I. J. Taneja, "Entropy of type and other generalized additive measures in information theory," *Metrika*, vol. 22, pp. 205–215, 1975.

[38] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *J. Statist. Phys.*, vol. 52, pp. 479–487, 1988.

[39] A. Renyi, *Probability Theory*. New York: Elsevier, 1970.

[40] D. Erdogmus, "Information theoretic learning: Renyi's entropy and its applications to adaptive system training," Ph.D. dissertation, Univ. Florida, Gainesville, FL, 2002.

[41] S.-I. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. First IEEE Signal Process. Workshop Signal Process. Advances Wireless Commun.*, 1997, pp. 101–104.

[42] A. Touzni, I. Fijalkow, M. G. Larimore, and J. R. Treichler, "A globally convergent approach for blind MIMO adaptive deconvolution," *IEEE Trans. Signal Processing*, vol. 49, pp. 1166–1178, June 2001.

[43] R. Y. Rubinstein, *Simulation and the Monte Carlo Method*. New York: Wiley, 1981.

[44] D. Erdogmus, K. E. Hild II, and J. C. Principe, "Blind source separation using Renyi's α-Marginal entropies," Special Issue on BSS and ICA, Neurocomputation, 2002, to be published.

[45] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. 26, pp. 26–37, 1980.

Deniz Erdogmus (M'00) received the B.S. degree in electrical and electronics engineering and mathematics in 1997 and the M.S. degree in electrical and electronics engineering, with emphasis on systems and control, in 1999, all from the Middle East Technical University, Ankara, Turkey. He received the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, in 2002.

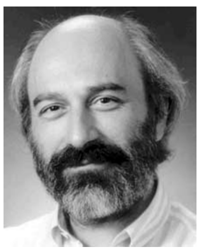
His current research interests include applications of information theory to adaptive systems, signal processing, communications, and control.

Dr. Erdogmus is a member of Tau Beta Pi and Eta Kappa Nu.

Kenneth E. Hild (M'03) received the B.S. degree in electrical engineering in 1992 and the M.S. degree in electrical engineering in 1996, both from the University of Oklahoma (OU), Norman. The emphasis during his graduate work at OU was in the areas of communications, signal processing, and controls. Currently, he is pursuing the Ph.D. degree at the University of Florida, Gainesville, where he is studying information theoretic learning and blind source separation.

From 1993 to 1995, he served in several different assistantships with OU, one of which involved conducting research for Seagate Technologies, Inc., Oklahoma City, OK. From 1995 to 1999, he worked full-time with Seagate, where he served as an advisory development engineer in the Advanced Concepts group.

Mr. Hild is a member of Tau Beta Pi and Eta Kappa Nu.



Jose C. Principe (F'00) is a Distinguished Professor of Electrical and Computer Engineering and Biomedical Engineering at the University of Florida, Gainesville, where he teaches advanced signal processing, machine learning and artificial neural networks (ANN's) modeling. He is BellSouth Professor and the Founder and Director of the University of Florida Computational NeuroEngineering Laboratory (CNEL). His primary area of interest is processing of time-varying signals with adaptive neural models. The CNEL Lab has been studying signal and pattern recognition principles based on information theoretic criteria (entropy and mutual information).

Dr. Principe is a member of the ADCOM of the IEEE Signal Processing Society, Member of the Board of Governors of the International Neural Network Society, and Editor-in-Chief of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING. He is a member of the Advisory Board of the University of Florida Brain Institute and has more than 90 publications in refereed journals, ten book chapters, and 200 conference papers. He has directed 35 Ph.D. dissertations and 45 Master theses. He recently wrote an interactive electronic book entitled *Neural and Adaptive Systems: Fundamentals Through Simulation* (New York: Wiley).

Marcelino Lazaro (M'01) was born in Carriazo (Cantabria), Spain, in 1972. He received the Telecommunication Engineer degree and the Doctor (Ph.D.) degree from the Universidad de Cantabria, in 1996 and 2001 respectively.

From 1996 to 2002, he was with the Departamento de Ingenieria de Comunicaciones, Universidad de Cantabria. In 2003, he joined the Departamento de Teoria de la Senal y Comunicaciones, Universidad Carlos III, Madrid, Spain. His research interest includes digital signal processing and neural networks.

Ignacio Santamaria (M'96) was born in Vitoria, Spain, in 1967. He received the Telecommunication Engineer degree and the Doctor degree from the Universidad Politecnica de Madrid (UPM), Madrid, Spain, in 1991 and 1995, respectively.

In 1992, he joined the Departamento de Ingenieria de Comunicaciones, Universidad de Cantabria, Cantabria, Spain, where he is currently an Associate Professor. His research interests include digital signal processing, nonlinear systems, and neural networks.