A Novel Entropy-based Equalization Performance Measure and Relations to L_p-Norm Deconvolution

Kenji Nose-Filho, Denis G. Fantinato, Romis Attux, Aline Neves and J.M.T. Romano

***Abstract*— A crucial performance measure in the context of the problem of deconvolution is the level of residual intersymbol interference (ISI), a metric that is classically understood and formulated in terms of an L_2-norm perspective. In order to enhance the scope of ISI quantification, we propose a novel entropy-based performance measure, which is called Entropy-based Intersymbol Interference (HISI). Interestingly, this metric is related to an information-theoretic relationship between source distribution and L_p-norms when the error entropy is used as a basis for optimal filtering. The new metric is analytically investigated and illustrated with some simulations.**

***Keywords*— Deconvolution, intersymbol interference, performance measure, L_p-norms, information-theoretic learning, entropy.**

I. Introduction - Deconvolution and Performance Metrics

The inverse problem known as deconvolution plays a central role in modern signal processing theory. This is, at least in part, a consequence of the occurrence of this problem in research fields as diverse as astronomy, biomedicine, speech, seismic and communications [1].

The origins of statistical deconvolution methods can be traced to the seminal contributions of Norbert Wiener and Andrey Kolmogorov, formalized in the first half of the last century [1]. From these contributions, it became possible to define in clear terms, in the context of stationary information signals, optimal deconvolution conditions in terms of the mean-squared error (MSE) or, in other words, from an L_2-norm perspective.

A number of important theoretical results associated with the use of the L_2-norm were obtained over the past decades [2] - mainly under the aegis of the classical hypothesis of Gaussianity - and these results naturally led to performance measures that intrinsically rely upon quadratic entities, such as the already mentioned MSE, the *signal-to-noise ratio* (SNR) and the degree of *intersymbol interference* (ISI).

However, in the presence of sparse and uniformly-distributed error signals, it was shown that the use of L_p-norms distinct from the L_2 can lead to certain improvements for estimation and deconvolution problems [3], [4], [5]. Indeed, these promising results have contributed to encourage the proposal of deconvolution criteria based on other L_p-norms.

Notwithstanding, when it comes to evaluate the performance of L_p systems, the existing metrics, like MSE, SNR and ISI, can be misleading due to their tacit assumption of L_2-norm optimality.

Having these facts in mind, we propose, in this work, a novel deconvolution performance measure that is able to encompass the peculiarities of the L_p-norms without privileging any of the compared L_p-terms. The derivation will be based on points of contact [5] between L_p-norms and a metric belonging to the framework of information theoretical learning (ITL) [6]. Focusing on seismic deconvolution and channel equalization problems, we also present some properties of the new metric, as well as elements of comparison with quadratic ISI measures.

The organization of the paper is as follows. A brief background on the main obtained results by employing the Lp-norms is presented in Section II. In Section III, we present the novel performance measure, and, in the following, in Section IV, some properties concerning the new metric are explored. In Section V, we conduct some simulation for the problem of channel deconvolution and compare the new metric to the conventional ISI. Finally, the conclusions and further research interests are summarized in Section VI.

II. Fundamentals of L_p-Norm Deconvolution

The problem of deconvolution can be stated as follows. Given an observed signal $x(n)$ and a reference signal $s(n-d)$ - where d is a given delay -, as illustrated in Fig. 1, it is desired to find a filter with parameter vector $\mathbf{w}$ that minimizes the L_p-norm of the error signal, given by

$$\|e(n)\|_p = \left(\sum_n |s(n-d) - y(n)|^p\right)^{1/p} \qquad (1)$$

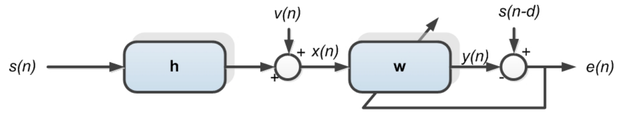


Fig. 1. Deconvolution scheme.

Traditionally, the filter of interest is a finite impulse response (FIR) filter, and preference for the L_2-norm in (1) comes from the historical fact that the resulting cost function is convex and there is a closed-form solution for its global minimum [1]. However, it is already known, as outlined in Section I, that

Kenji Nose-Filho, Denis G. Fantinato, Romis Attux and J.M.T. Romano, School of Electrical and Computer Engineering, University of Campinas, Campinas-SP, Brazil, E-mails: knfilho@dmo.fee.unicamp.br, {denisgf,attux}@dca.fee.unicamp.br, romano@dmo.fee.unicamp.br. Aline Neves, Engineering, Modeling and Applied Social Science Center, Federal University of ABC, Santo André - SP, Brazil, E-mail: aline.neves@ufabc.edu.br This work was partially supported by CAPES and CNPq.

to consider alternative norms can be relevant in this kind of estimation problem, especially when a priori knowledge about the involved signals, such as the distribution of the additive noise and the distribution of the input signal [3], [4], [5], is available.

In view of these facts, analogously to what has been done in [5], the impact of adopting alternative norms will be evaluated for two cases: seismic deconvolution and channel equalization. These two problems are related by the characteristics of the involved signals, as represented in Tab. I. For the seismic deconvolution case, the earth reflectivity, in general, is assumed to be white and sparse. On the other side, the transmitted message in channel equalization is assumed to be white, and the message is composed of symbols belonging to a finite alphabet.

TABLE I
SIGNALS INVOLVED IN THE SEISMIC DECONVOLUTION AND THE CHANNEL EQUALIZATION PROBLEMS.

Signal	Seismic deconvolution	Channel equalization
$s(n)$	earth's reflectivity	transmitted message
$h(n)$	impulse response of the seismic wavelet	impulse response of the communication channel
$x(n)$	seismic trace	received message

The main results obtained for these two problems can be summarized by illustrating the cases where the norms L_1, L_2 and L_∞ are employed. To do so, two different source models will be considered: a sparse signal and a signal with discrete uniform distribution, both in a noiseless scenario and for a generic channel with transfer function $H(z) = 1 + h_1 z^{-1}$. We make use of a two-tap filter $W(z) = w_0 + w_1 z^{-1}$ or, in vector notation, $\mathbf{w} = [w_0\ w_1]^T$ for performing deconvolution.

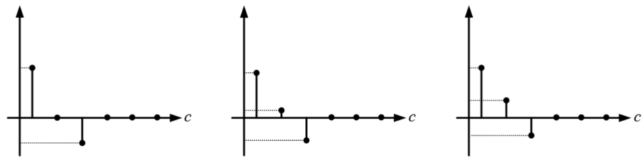
In this context, we adopt a signal $s(n)$ given by a single impulse with amplitude equal to one as a representative sparse signal and, as for the signal with discrete uniform distribution, it is assumed to be a $\{$+1/-1$\}$ independent and identically distributed (i.i.d.) Bernoulli random variable. For both cases, we present the optimum filter according to each criterion in Section II, calculated as in [5].

For the sparse signal case, the best filter is that capable of producing the sparsest signal. In order to identify the suitable norm, we illustrate the combined channel-equalizer impulse response $\mathbf{c} = \mathbf{h} * \mathbf{w}$, where $*$ denotes convolution, for the sparse signal in Fig. 2. We can see that the L_1-norm is the one with less resulting non-zero taps for $\mathbf{c}$, which culminates in a sparser signal at filter output. For the second case, the uniformly-distributed signal, the best filter is that capable of producing the signal with less states. Now, to select the best filter, we present the probability density function (PDF) of the output signal for the uniformly-distributed signal in Fig. 3. It is clear that, for this case, the L_∞-norm provides the most desirable solution.

Interestingly, the best solutions in each case are identical with respect to the value of the filter coefficients, i.e. $\mathbf{w} = [1\ -h_1]^T$, which is obtained with the L_1-norm for the sparse signal and with the L_∞-norm for the uniformly-distributed signal. It is important to remark that, for both cases, the definition of "best" here can be seen from the conceptual point of view of the error entropy [5]. In that sense, the output with less uncertainty / entropy is the one most similar to the source. This will be crucial in the development of a performance measure when dealing with L_p-norms, as will be seen in the next section.

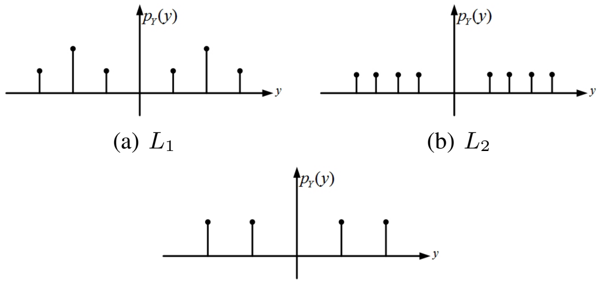
TABLE II
FILTERS FOR DIFFERENT SIGNALS, OPTIMIZED FOR DIFFERENT NORMS.

Norm	Sparse	Uniform
L_1	$[1\ -h_1]$	$\left[1\ \frac{-h_1}{1+\lvert h_1\rvert}\right]$
L_2	$\left[\frac{1+h_1^2}{1+h_1^2+h_1^4}\ \frac{-h_1}{1+h_1^2+h_1^4}\right]$	$\left[\frac{1+h_1^2}{1+h_1^2+h_1^4}\ \frac{-h_1}{1+h_1^2+h_1^4}\right]$
L_∞	$\left[\frac{1+\lvert h_1\rvert}{1+\lvert h_1\rvert+h_1^2}\ \frac{-h_1}{1+\lvert h_1\rvert+h_1^2}\right]$	$[1\ -h_1]$



(a) L_1 (b) L_2 (c) L_∞

Fig. 2. Combined response for the sparse source.



(a) L_1 (b) L_2 (c) L_∞

Fig. 3. PDF of the output of the filters for the uniform signal.

III. PROPOSAL - THE ENTROPY BASED INTERSYMBOL INTERFERENCE

In Section I, we have mentioned that the possibility of dealing with criteria based on norms alternative to the L_2 can lead to relevant performance improvements if the source signal distribution is known. However, the classical performance evaluation metrics, like SNR, MSE and ISI, are strongly related to the L_2 framework, since they have been proposed under the assumption of the quadratic optimality. For example, the usual performance evaluation measure in deconvolution is the ISI, which makes use of a complete knowledge of the channel:

$$ISI_{dB} = 10 \log_{10} \frac{\left(\sum_{i=0}^{M} |c_i|^2\right) - \max_i |c_i|^2}{\max_i |c_i|^2}, \qquad (2)$$

where $\mathbf{c}$ is assumed causal and M is the maximum length of $\mathbf{c}$. Basically, this measure is the ratio between the interference energy and the signal energy, which gravitates around the L_2-norm.

Hence, the use of different norms for error quantification is not trivial in view of the difficulty of defining "unbiased" evaluation metrics. In light of this, the present work gives a step in that direction by seeking a measure able to properly quantify the characteristics of a given problem. Interestingly, the use of ideas and concepts belonging to the field of information theoretical learning (ITL) [6] can contribute to this task as they deal with statistical information in a more extensive manner, being not restricted to the L_2-norm. More specifically, Shannon's entropy measure deserves special attention within this framework, since it is also able to establish points of contact with different norms - as pointed out in [5] - when applied to the error signal.

The combination of the above mentioned ideas culminates in the proposal of the following equalization performance measure, which will be called Entropy-based Intersymbol Interference (HISI):

$$\begin{aligned} HISI_{dB} &= 10\log_{10} H(\alpha|\mathbf{c}|) \\ &= 10\log_{10}\left(-\sum_{i=0}^{M}\alpha|c_i|log_2\left(\alpha|c_i|\right)\right), \end{aligned} \quad (3)$$

where $H(\cdot)$ denotes Shannon's entropy and $\alpha = 1/\sum_i |c_i|$ is a unit-norm correction term that allows the combined response to be treated as probability distribution. A possible interpretation for this novel evaluation metric is that it numerically translates the uncertainty associated with the combined channel-equalizer impulse response. A desired result in channel deconvolution is to keep the HISI as low as possible in order to reduce the presence of interference components. In the following, we will discuss some additional implications concerning this measure.

IV. HISI Applied to L_p-Norm Deconvolution

In order to analyze a representative set of potentially practical situations, the HISI measure will be studied for two cases: a case in which perfect deconvolution (i.e. a zero-forcing (ZF) condition) is attainable and a case in which this is not possible.

A. *Zero-Forcing Condition*

Ideally, it is desired that the equalizer be able to perfectly invert the convolutional effects posed by the unknown system. In such case, the combined channel-equalizer impulse response is given by $c_i = \delta(i-\tau)$, being $\delta(\cdot)$ the Kronecker delta function and τ a given delay. This condition is known as ZF condition. In such case, the value for the measured HISI tends towards $-\infty$ - a direct consequence of the fact that entropy is null when there is no uncertainty. The same singularity occurs for the conventional (quadratic) ISI measure. It is important to remark that both measures do not depend on a scale factor imposed to the filter coefficients, as demonstrated in the Appendix. With this in view, it is possible to state that, when the ZF condition can be attained, the optimal solution leads to the combined channel-equalizer impulse response given by $c_i = \delta(i-\tau)$, which corresponds to the minimum ISI and HISI values.

To illustrate this, we consider the scenario for which the ZF condition is attainable, constituted by an all-pole channel $H(z) = 1/(1+0.6z^{-1})$ without additive noise and an inverse FIR filter with two coefficients. In this case, the filter H(z) can be perfectly inverted by the filter $\mathbf{w} = [1\ 0.6]^T$, which provides the minimum ISI and HISI.

Therefore, it is also expected that the minimization of any L_p-norm provide the same optimal solution. To verify this, we present, in Fig. 4, the contour plots for the L_1, L_2 and L_∞ norms. The source is considered, at first, a sparse signal and, later, a signal with discrete uniform distribution. The samples of the first signal $s(n)$ are given by i.i.d. Bernoulli Gaussian random variables with probability of zero equal to 0.95, and the samples of the second signal are given by i.i.d. Bernoulli {+1/-1} random variables.

In this case, Fig. 4 shows that all minima for the chosen norms are exactly the same, i.e. $\mathbf{w} = [1\ 0.6]^T$, indicating that the L_p-norm is a consistent criteria for cases with an attainable ZF conditions.

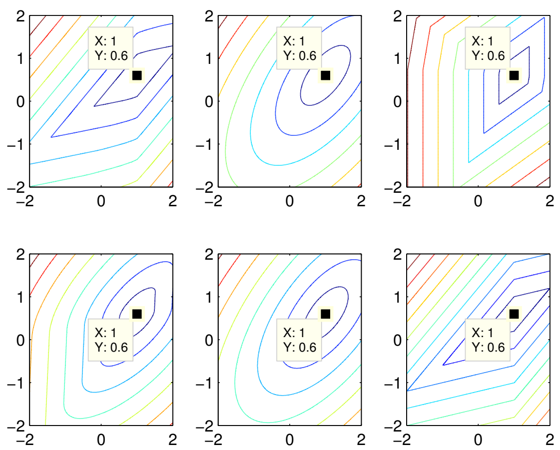


Fig. 4. Contour plots of the L_p norms for the sparse signal (top) and the uniform signal (bottom). L_1 (left), L_2 (middle), L_∞ (right).

B. *Non-Attainable Zero-Forcing Condition*

In cases for which the ZF condition cannot be attained, the combined channel-equalizer impulse response c is composed of more than a single spike, and the differences between ISI and HISI become more pronounced. A direct way to visualize their behavior can be obtained if we observe the values of these metrics by evaluating all possible solutions for the equalizer. For a minimum-phase channel with impulse response $H(z) = 1 + 0.6z^{-1}$, we adopt a two-tap equalizer of the type $\mathbf{w} = [1\ w_1]^T$ (as the performance metrics are invariable to a scalar factor in the filter coefficients), in which we vary w_1 from -3 to 3. The normalized values for the ISI and HISI are presented in Fig. 5. It is clear that the points of minima differ such that the optimum filter coefficients that minimize the ISI and HISI are $\mathbf{w}_{ISI} = [1\ -0.4412]^T$ and $\mathbf{w}_{HISI} = [1\ -0.6]^T$, respectively.

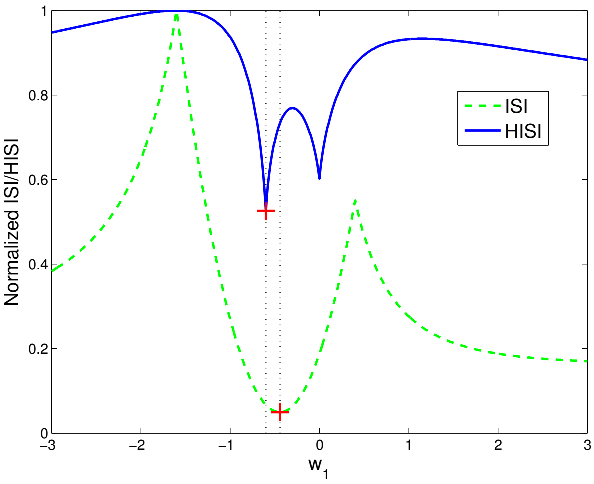


Fig. 5. Normalized ISI/HISI for a varying filter.

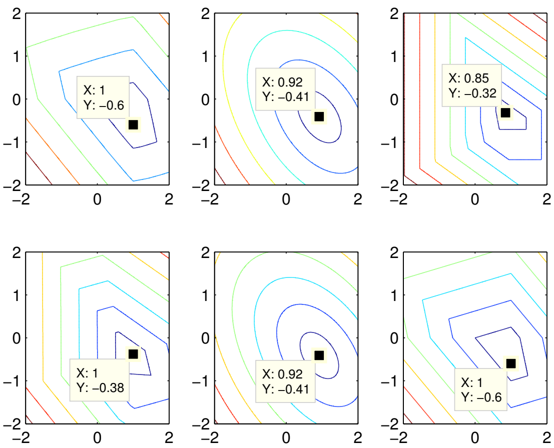


Fig. 6. Contour plots of the L_p norms for the sparse signal (top) and the uniform signal (bottom). L_1 (left), L_2 (middle), L_∞ (right).

The resulting combined channel-equalizer impulse response for $\mathbf{w}_{ISI}$ and $\mathbf{w}_{HISI}$ is exactly as in Figs. 2(b) and 2(a), respectively. While the first solution reflects the optimality of Wiener filters, the second one emphasizes the combined response c that has the least number of impulses so that the uncertainty or entropy of c can be minimized.

In this case, the optimal solutions for the filter associated with each norm are all different. If we consider the same scenario employed in the ZF case - except for the channel, which is now an FIR minimum-phase filter $H(z) = 1 + 0.6z^{-1}$, we can see that the minima for L_1, L_2 and L_∞ in the contour plots in Fig. 6 are those of Tab. II and the desired solutions between them are that exactly equal to $\mathbf{w}_{HISI} = [1 \ -0.6]^T$ i.e. the optimal filter that minimizes the HISI.

It is possible to state that, given a second-order channel of the form $H(z) = 1 + h_1 z^{-1}$ and a two-tap inverse FIR model, the HISI will always provide an optimal solution for the coefficients of the type $\mathbf{w}_{HISI} = [1 \ -h_1]^T$, which is in consonance with the best solutions pointed out for the L_p-norm minimization presented in Section II.

In summary, the correspondence between the optimal solutions for the inverse filter according to the minimization of HISI and the minimization of the L_∞-norm for uniformly distributed signals and the L_1-norm for sparse signals, for both attainable and non-attainable ZF conditions, lead us to conclude that the use of the L_∞- and the L_1-norm, considering the specific scenario, reaches the core of the concept associated with the minimization of channel uncertainty. Moreover, these connections indicate that the HISI measure can serve as a good equalization performance metric when the optimization criterion is based on generic L_p-norms.

V. Simulation Results

In order to test the effectiveness of the HISI as a performance measure in the context of adaptive channel equalization, we consider the two scenarios already exposed in this work, those characterized by sparse and uniformly-distributed sources. Furthermore, we will present elements of comparison with the conventional ISI metric. In order to perform parameter optimization under different norms, it will be necessary to employ adaptive algorithms, such as the sign-error LMS (Least-Mean-Squared), the LMS, and the LMF (Least-Mean-Fourth), for the minimization of the L_1, L_2 and L_4 norms, respectively, in which the last is considered as an approximation for the L_∞-norm. More details about these algorithms can be found in [7].

For the first scenario, the source signal $s(n)$ is generated by an i.i.d. Bernoulli Gaussian random variable with probability of zero equal to 0.95 and imposed to the minimum-phase channel $H(z) = 1 + 0.6z^{-1}$ (without additive noise). We employ a two-tap filter initialized at zero. The chosen step-sizes were μ_{L_1}=μ_{LMS}=0.001 and μ_{L_4}=$2e{-}5$, so that the variance of their respective filter coefficients were similar after convergence. The performance measures in terms of ISI and HISI when training these filters are shown in Fig. 7, as an average of 50 experiments.

With respect to ISI, the sign-error LMS (L_1-norm) carries a higher interference level, while LMS (L_2-norm) seems to be the optimum case, followed by the LMF (L_4-norm). Now, when considering the performance in respect to HISI, we see that the sign-error LMS algorithm is the only one that attains lower levels of HISI, which is in consonance with the desired solution for sparse sources. The other methods converge to positive values of HISI. It is important to remark that the HISI $= 0dB$ case is equivalent to a two-tap combined response representative of a uniform probability distribution (i.e. same amplitude), while, for ISI $= 0dB$, it corresponds to equal energy (or quadratic) levels for signal and interference. Also, it is clear that there is no direct correspondence between the values for ISI and HISI, since the first one presents larger variations while the second remains closer to zero. For this reason, it is not strictly necessary to express the HISI in dB, but we shall do so for a better comparison with ISI.

For the second scenario, we keep the same configuration of the sparse case, except for the source signal $s(n)$, which now is composed of {+1/-1} i.i.d. samples. The step-size for LMF is μ_{L_4}=0.002, while the others remain unchanged. Fig.

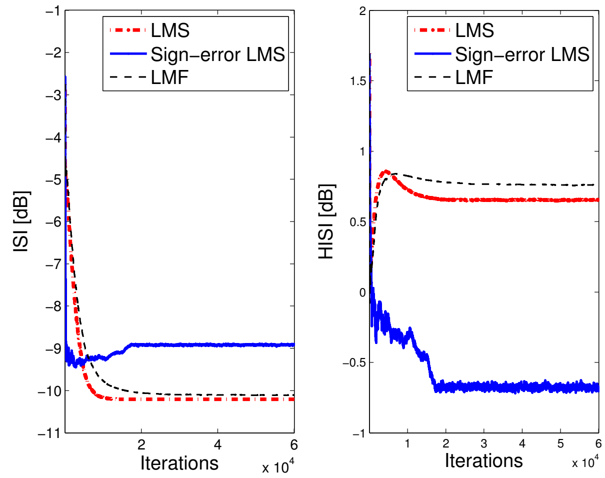


Fig. 7. ISI and HISI performance for a sparse source.

8 shows the resulting ISI and HISI performances for this case.

We see that, again, the LMS algorithm converges to the lowest ISI level and, now, the LMF presents the highest level of residual ISI. In terms of HISI, on the other hand, the LMF is the algorithm that is able to attain the lowest levels of interference, agreeing with the desired solution for the uniformly distributed source case. However, the residual HISI interference after convergence is larger than that achieved with the sign-error LMS in the sparse case, which can be explained by the fact that the L_4-norm is only an approximation of the L_∞-norm. The HISI, as a performance measure, was, in the analyzed cases, capable of successfully extracting the filter solution associated to the most suitable norm - given the source signal distribution - and, consequently, to reduce the uncertainty with respect to the combined response.

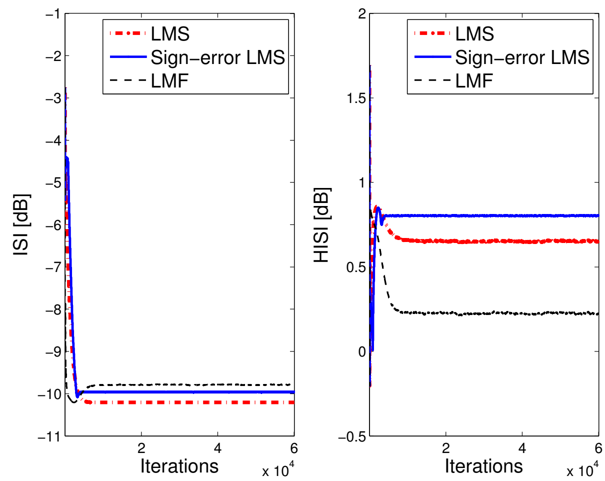


Fig. 8. ISI and HISI performance for a uniformly distributed source.

VI. Conclusions

In this work, we have proposed a novel entropy-based equalization performance measure, the HISI, which is suitable to describe the problem of deconvolution when the use of general L_p-norms is considered. We have shown that the new measure is a promising solution from the standpoint of successful problem characterization, by indicating that the minimization of HISI coincides with the optimal solution of the L_∞-norm for uniformly distributed signals and with the solution of L_1-norm for sparse signals.

The HISI measure is associated, in essence, to the minimization of the combined response channel-equalizer uncertainty; thus, we can conclude that the use of the L_∞- and the L_1-norm, considering specific scenarios, can sufficiently extract from the received signal the mentioned uncertainty. As extensions of this initial effort, we highlight the possibility of establishing points of contact between error entropy criterion and HISI, given the connections between Shannon's entropy and the norms [5]. Also, we intend to provide a blind estimation for HISI, as well as a blind algorithm based on this metric.

Acknowledgements

The authors thank CAPES and CNPq for the financial support.

References

[1] J.M.T. Romano, R. Attux, C.C. Cavalcante, R. Suyama, *Unsupervised Signal Processing: Channel Equalization and Source Separation*. CRC Press, 2010.

[2] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 2001

[3] J.R. Rice, J.S. White, "Norms for smoothing and estimation," *SIAM Review*, v. 6, pp. 243-255, 1964.

[4] H. W. J. Debeye, P. Van Riel, "Lp-Norm Deconvolution," *Geophysical Prospecting*, 1990.

[5] K. Nose-Filho, L.T. Duarte, R.R.F. Attux, E. Nadalin, R. Ferrari, J.M.T. Romano, "Sobre Filtragem L_p," *XXX Simpósio Brasileiro de Telecomunicações*, 2012.

[6] J. C. Principe, *Information Theoretic Learning: Renyi's entropy and kernel perspectives*. Springer, 2010.

[7] A. H. Sayed, *Fundamentals of adaptive filtering*. Wiley, 2003.

Appendix

We show bellow that the performance measures ISI and HISI are invariant of a scale factor, say β, on the filter coefficients $\mathbf{w}$. It is possible to state that $\mathbf{h} * (\beta\mathbf{w}) = \beta\mathbf{c}$. Based on this, we analyze the ISI metric,

$$ISI(\beta\mathbf{c}) = \frac{\left(\sum_{i=0}^{M} |\beta c_i|^2\right) - \max_i |\beta c_i|^2}{\max_i |\beta c_i|^2} = \frac{|\beta|^2}{|\beta|^2} \frac{\left(\sum_{i=0}^{M} |c_i|^2\right) - \max_i |c_i|^2}{\max_i |c_i|^2} = ISI(\mathbf{c}),$$

where we have suppressed the term $10\log_{10}$ for the sake of simplicity. In respect to the HISI, we should first determine the unit-norm correction. Originally, $\alpha = 1/\sum_i |c_i|$, but with a scale factor, it is possible to express:

$$\frac{1}{\sum_i |\beta c_i|} = \frac{1}{|\beta| \sum_i |c_i|} = \frac{\alpha}{|\beta|}$$

Substituting in the HISI formula, Eq. (3), we obtain:

$$HISI(\beta\mathbf{c}) = H(\frac{\alpha}{|\beta|}|\beta\mathbf{c}|) = H(\alpha|\mathbf{c}|) = HISI(\mathbf{c}).$$