

# Sobre Filtragem $L_p$

Kenji Nose-Filho, Leonardo T. Duarte, Romis Attux, Everton Z. Nadalin, Rafael Ferrari, João M. T. Romano

**Resumo**—Neste trabalho, investigamos o uso de diferentes normas para a quantificação do erro em problemas de estimação e filtragem, particularmente em desconvolução. Consideramos, em nossa análise, diferentes distribuições de probabilidade dos sinais processados e suas relações com a norma de otimização. A análise é feita inicialmente no caso clássico de estimação de um parâmetro, para em seguida, abordar o ponto de vista de filtragem e dos critérios baseados na entropia do erro. Os resultados obtidos mostram-se promissores no sentido de se obter a norma mais adequada para ser utilizada na otimização de um filtro de desconvolução, tendo em vista as características do sinal que queremos recuperar. Tais resultados são discutidos em duas aplicações de grande relevância: a de comunicações e a de processamento sísmico.

**Palavras-Chave**—Desconvolução, Estimação, Filtragem, Information-Theoretic-Learning, Normas  $L_p$ .

**Abstract**—In this paper, we investigate the use of different norms for the quantification of the error in estimation and filtering problems, particularly in deconvolution. We consider in our analysis different probability distributions of the processed signals and their relationships with the optimization norm. Initially, the analysis is made for the classical estimation problem of one parameter and then, addresses the aspects of filtering and of the error entropy-based criteria. The results are promising in the sense of finding the most suitable norm for the optimization of a deconvolution filter, given the nature of the signals that we want to recover. These results are discussed in the context of two applications of great importance: communications and seismic processing.

**Keywords**—Deconvolution, Estimation, Filtering, Information-Theoretic-Learning,  $L_p$  Norms.

## I. INTRODUÇÃO

Em problemas de filtragem ótima, busca-se ajustar um sistema que, a partir de um dado sinal de entrada, seja capaz de aproximar um dado sinal de referência. Classicamente, este problema é tratado com uma abordagem probabilística, de modo que o processo de ajuste é feito através da minimização do erro quadrático médio entre a saída do filtro e o sinal desejado. Em um contexto determinístico, tal abordagem pode ser interpretada com a minimização da norma  $L_2$  do erro.

Atualmente, há um grande interesse da comunidade de processamento de sinais pela utilização de outras normas para quantificar o erro. Embora resultados importantes tenham sido

Kenji Nose Filho, Romis Attux, Everton Z. Nadalin, Rafael Ferrari e João M. T. Romano, Faculdade de Engenharia Elétrica e Computação, Universidade de Campinas, Campinas-SP, E-mails: knfilho@dmo.fee.unicamp.br, attux@dca.fee.unicamp.br, nadalin@dca.fee.unicamp.br, rferrari@decom.fee.unicamp.br, romano@dmo.fee.unicamp.br.

Leonardo T. Duarte, Faculdade de Ciências Aplicadas, Universidade de Campinas, Limeira-SP, E-mail: leonardo.duarte@fca.unicamp.br.

Este trabalho foi parcialmente financiando pelo CNPq (142714/2011-9).

obtidos mais recentemente, sobretudo no que diz respeito à utilização da norma  $L_1$ , a busca por alternativas à norma  $L_2$  data de décadas [1]. Por exemplo, em 1964, Rice e White analisaram a relação que existe entre a minimização de diferentes normas do erro, no problema de estimação, para erros com diferentes distribuições de probabilidade. Os resultados obtidos relacionam a norma  $L_1$  com as distribuições de Cauchy e de Laplace, a norma  $L_2$  com a triangular e a gaussiana e a  $L_\infty$  com a uniforme, no sentido destas normas fornecerem os estimadores de menor variância [2].

Em trabalhos como o de Debeye e Van Riel de 1990 é ainda possível visualizar a relação entre as normas e a função densidade de probabilidade (pdf, do inglês *probability density function*) da gaussiana generalizada, sendo possível chegar na correspondência entre a  $L_1$  e a Laplaciana, a  $L_2$  e a gaussiana e a  $L_\infty$  e a uniforme [3].

No presente trabalho, investigamos diferentes normas de otimização no contexto do problema de filtragem, mais especificamente em desconvolução. Em relação à literatura existente, a abordagem aqui proposta consiste em se colocar a seguinte questão: - Qual a norma mais adequada para ser utilizada na otimização de um filtro de desconvolução, tendo em vista as características do sinal que queremos recuperar? De fato, um objetivo mais amplo de nosso estudo seria definir a melhor norma de otimização para um dado problema de filtragem. Neste trabalho, procura-se dar uma contribuição nesse sentido considerando, numa primeira instância, a aplicação em comunicações e em processamento sísmico.

O trabalho está dividido nas seguintes seções: de modo a introduzir o problema, recapitulamos, na seção II, resultados referentes à minimização de diferentes normas para a estimação de um valor representativo de um conjunto de observações. Na seção III, estendemos a análise para nosso problema de interesse; a contribuição principal está em revelar a melhor norma a ser empregada em cada aplicação, à luz de suas características intrínsecas: no caso de comunicações, busca-se, em última análise, uma situação de “*pdf matching*” entre os sinais transmitido e recebido; em sísmica, busca-se uma solução que maximize a esparsidade do sinal recuperado. Na seção IV, faz-se uma conexão entre os resultados obtidos anteriormente e os critérios baseados na entropia do erro, oriundos da área de aprendizado de máquina baseado na teoria da informação (ITL, do inglês *information-theoretic learning*), recentemente formalizada em [4]. Finalmente, na seção V, tecemos algumas considerações finais a título de conclusão do trabalho.

## II. ESTIMAÇÃO UTILIZANDO DIFERENTES NORMAS

Nesta seção, revisitamos alguns resultados clássicos do problema de estimação de um parâmetro que seja representativo de um conjunto de observações a partir das

normas  $L_1$ ,  $L_2$  e  $L_\infty$ . Tal estudo segue, em linhas gerais, os trabalhos de Rice e White de 1964 [2] e de Claerbout e Muir de 1973 [5].

A norma  $L_p$  de um vetor  $\mathbf{x}$  é dada por,

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^N |x_i|^p}, 1 \leq p < \infty, \quad (1)$$

sendo que,

$$\|\mathbf{x}\|_\infty = \lim_{p \rightarrow \infty} \sqrt[p]{\sum_{i=1}^N |x_i|^p} = \max\{|x_i|, i = 1, \dots, N\}. \quad (2)$$

Vamos considerar inicialmente um problema que consiste na busca de um valor  $\hat{y}$  que melhor descreve um vetor de  $N$  observações de uma variável aleatória  $\mathbf{y} = [y_1, \dots, y_N]$ , capaz de minimizar a norma  $L_p$  do erro, dada por  $\|\hat{y} - \mathbf{y}\|_p$ .

A primeira análise será feita na minimização da norma  $L_1$ , ou seja, do somatório dos valores absolutos do erro,

$$\min_{\hat{y}} J_{L_1} = \min_{\hat{y}} \sum_{i=1}^N |\hat{y} - y_i|. \quad (3)$$

Derivando-se esta expressão em função de  $\hat{y}$  e igualando-se o resultado a zero, obtêm-se,

$$\frac{dJ_{L_1}}{d\hat{y}} = \sum_{i=1}^N \text{sgn}(\hat{y} - y_i) = 0, \quad (4)$$

onde  $\text{sgn}(\cdot)$  é a função sinal. Através desta expressão, é possível ver claramente que a solução  $L_1$  é a mediana dos elementos do vetor  $\mathbf{y}$ , i.e.,

$$\hat{y}_{L_1} = \text{mediana}\{\mathbf{y}\}. \quad (5)$$

A minimização da norma  $L_2$  busca a minimização do somatório do erro quadrático,

$$\min_{\hat{y}} J_{L_2} = \min_{\hat{y}} \sum_{i=1}^N |\hat{y} - y_i|^2. \quad (6)$$

Derivando o lado direito de (6) em função de  $\hat{y}$  e igualando-se o resultado a zero, obtêm-se,

$$\frac{dJ_{L_2}}{d\hat{y}} = \sum_{i=1}^N 2(\hat{y} - y_i) = 0. \quad (7)$$

Assim, a solução  $L_2$  é dada pela média dos elementos do vetor  $\mathbf{y}$ , ou seja,

$$\hat{y}_{L_2} = \frac{1}{N} \sum_{i=1}^N y_i. \quad (8)$$

Por fim, temos a minimização da norma  $L_\infty$ , que corresponde a minimizar o valor máximo do erro,

$$\min_{\hat{y}} J_{L_\infty} = \min_{\hat{y}} \max\{|\hat{y} - y_i|, i = 1, \dots, N\}. \quad (9)$$

A solução  $L_\infty$  é dada pelo ponto médio dos elementos do vetor  $\mathbf{y}$ ,

$$\hat{y}_{L_\infty} = \frac{\max\{\mathbf{y}\} + \min\{\mathbf{y}\}}{2}, \quad (10)$$

a qual resulta em valores iguais para o erro nas suas extremidades.

Deste modo, podemos relacionar os valores ótimos de  $\hat{y}$  com as estatísticas de tendência central dos elementos do vetor  $\mathbf{y}$ .

A Tabela I ilustra o desempenho dos três estimadores para diferentes variáveis aleatórias geradoras do vetor  $\mathbf{y}$  por meio do valor assintótico da variância do estimador em função do número de amostras  $N$ .

Tabela I. Valores assintóticos da variância do estimador em função de  $N$ , extraído de [2].

Distribuição das variáveis aleatórias geradoras do vetor $\mathbf{y}$ .	$L_1$	$L_2$	$L_\infty$
Uniforme	$\frac{1}{4N}$	$\frac{1}{12N}$	$\frac{1}{2N^2}$
Triangular	*	$\frac{1}{6N}$	$\frac{4N}{4 - \pi}$
Normal	$\frac{\pi}{2N}$	$\frac{1}{N}$	$\frac{4N}{\pi^2}$
Laplace	$\frac{1}{2N}$	$\frac{2}{N}$	$\frac{12 \log(N)}{\pi^2}$
Cauchy	$\frac{\pi^2}{4N}$	$\infty$	$\infty$

\* Não há uma expressão fechada.

Conforme a Tabela I é possível visualizar de maneira clara a relação que existe entre a norma utilizada e a distribuição da variável aleatória envolvida, com respeito ao desempenho obtido em termos da variância do estimador. Para distribuições subgaussianas, como a uniforme, o estimador de menor variância é o que minimiza a norma  $L_\infty$ . Por outro lado, para distribuições supergaussianas, como a de Laplace, e distribuições com caudas longas, como a de Cauchy, o estimador de menor variância é o que minimiza a norma  $L_1$ . Como caso intermediário, surge a densidade gaussiana, no qual o estimador de menor variância é o que minimiza a norma  $L_2$ .

### III. FILTRAGEM UTILIZANDO DIFERENTES NORMAS

Feita esta breve discussão sobre a influência da escolha da norma em um problema clássico de estimação, passemos ao problema de filtragem, que pode ser modelado pelo sistema da Fig. 1. Neste caso, há uma entrada  $x_n$  que passa por um filtro linear e invariante no tempo de resposta ao impulso finita – FIR (do inglês *finite impulse response*), gerando um processo média móvel – MA (do inglês *moving average*)  $y_n$ . A esta saída é adicionado um sinal  $v_n$  que modela a perturbação e deseja-se estimar a entrada através do sinal  $y_n + v_n$ . Para isto, projeta-se um filtro de desconvolução FIR que produz à sua saída um sinal  $\hat{x}_n$ . Os parâmetros do filtro são calculados visando a minimização ou maximização de um critério baseado no erro, dado por  $x_n - \hat{x}_n$ . Em particular, estudaremos os filtros capazes de minimizar as normas  $L_1$ ,  $L_2$  e  $L_\infty$  do vetor de erro.

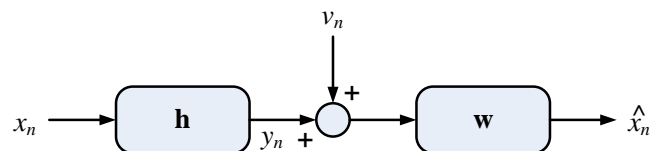


Fig. 1. Modelo do sistema do filtro de convolução e o filtro de desconvolução.

Neste estudo, vamos considerar um caso simples em que  $\mathbf{h} = [1 \ h_1]$ , com  $|h_1| < 1$ ,  $\mathbf{w} = [1 \ w_1]$ ,  $v_n = 0$ , de modo a possibilitar o tratamento analítico do problema.

Posteriormente, será possível estender as análises considerando uma perturbação  $v_n$  com distribuição gaussiana.

#### A. Caso de sinais com distribuição uniforme discreta.

Este estudo é interessante, principalmente do ponto de vista do problema de comunicações, onde os sinais de entrada possuem uma distribuição uniforme discreta, i.e., símbolos equiprováveis. No caso de equalização de canais de comunicações,  $x_n$  é a mensagem transmitida,  $\mathbf{h}$  é a resposta impulsiva do canal,  $v_n$  é uma perturbação ocasionada pelo ambiente,  $y_n + v_n$  é o sinal recebido,  $\mathbf{w}$  é o equalizador, sendo que um dispositivo de decisão é também incluído para estimar a mensagem transmitida [6].

Iremos considerar ainda  $\mathbf{x}$  como um vetor de variáveis aleatórias de Bernoulli independentes e identicamente distribuídas - i.i.d., assumindo valores  $-1$  e  $+1$ , de modo que sua pdf é dada por,

$$f_x(x) = \frac{1}{2}\delta(1-x) + \frac{1}{2}\delta(1+x). \quad (11)$$

No primeiro caso, iremos analisar a minimização da norma  $L_1$  do vetor de erro,

$$J_{L_1} = \|\mathbf{x} - \hat{\mathbf{x}}\|_1, \quad (12)$$

dada por,

$$\min_{w_1} \sum_{i=1}^N |(h_1 + w_1)x_{i-1} + w_1 h_1 x_{i-2}|. \quad (13)$$

Como o sinal de entrada assume valor  $+1$  ou  $-1$ , o termo  $|(h_1 + w_1)x_{i-1} + w_1 h_1 x_{i-2}|$  em (13) poderá assumir, com a mesma probabilidade, os valores dados por,

$$\begin{cases} |(h_1 + w_1) + w_1 h_1|, & \text{para } x_{i-1} = x_{i-2}, \\ |(h_1 + w_1) - w_1 h_1|, & \text{para } x_{i-1} = -x_{i-2}. \end{cases} \quad (14)$$

É possível visualizar duas soluções, uma que minimiza um dos termos, e outra que minimiza o outro termo. Estas soluções são dadas pelos valores de  $w_1$  que produzem os valores de  $J_{L_1}$  proporcionais a,

$$\begin{cases} \left| \frac{h_1^2}{1+h_1} \right|, & \text{para } w_1 = \frac{-h_1}{1+h_1}, \\ \left| \frac{h_1^2}{1-h_1} \right|, & \text{para } w_1 = \frac{-h_1}{1-h_1}, \end{cases} \quad (15)$$

como  $|h_1| < 1$ , a solução é dada por  $w_1 = \frac{-h_1}{1+h_1}$ .

No segundo caso, iremos analisar a minimização da norma  $L_2$ ,

$$J_{L_2} = \|\mathbf{x} - \hat{\mathbf{x}}\|_2, \quad (16)$$

dada pela equação,

$$\min_{w_1} \sum_{i=1}^N |(h_1 + w_1)x_{i-1} + w_1 h_1 x_{i-2}|^2. \quad (17)$$

A solução neste caso é a de Wiener, dada por  $w_1 = \frac{-h_1}{1+h_1^2}$ .

No terceiro e último caso, iremos analisar a minimização da norma  $L_\infty$ ,

$$J_{L_\infty} = \|\mathbf{x} - \hat{\mathbf{x}}\|_\infty, \quad (18)$$

dada pela equação,

$$\min_{w_1} \max \{|(h_1 + w_1)x_{i-1} + w_1 h_1 x_{i-2}|, i = 1, \dots, N\}. \quad (19)$$

Novamente, como o sinal de entrada assume valor  $+1$  ou  $-1$ , o termo  $|(h_1 + w_1)x_{i-1} + w_1 h_1 x_{i-2}|$  em (19), poderá assumir, com a mesma probabilidade, os valores dados por (14).

A solução é dada pelo valor de  $w_1$  que consegue minimizar a distância entre os valores extremos do erro que neste caso simples se torna,  $|(h_1 + w_1) + w_1 h_1| = |(h_1 + w_1) - w_1 h_1|$ . Duas possíveis soluções produzem os valores de  $J_{L_\infty}$  dados por,

$$\begin{cases} |h_1|, & \text{para } w_1 = 0, \\ |h_1^2|, & \text{para } w_1 = -h_1, \end{cases} \quad (20)$$

como  $|h_1| < 1$ , a solução é dada por  $w_1 = -h_1$ .

Uma forma interessante de visualizar os resultados obtidos é através da pdf do sinal de saída. Deste modo obtemos, para cada norma utilizada na otimização do filtro de desconvolução, uma pdf diferente do sinal de saída. Tais pdf's são apresentadas na Fig. 2.

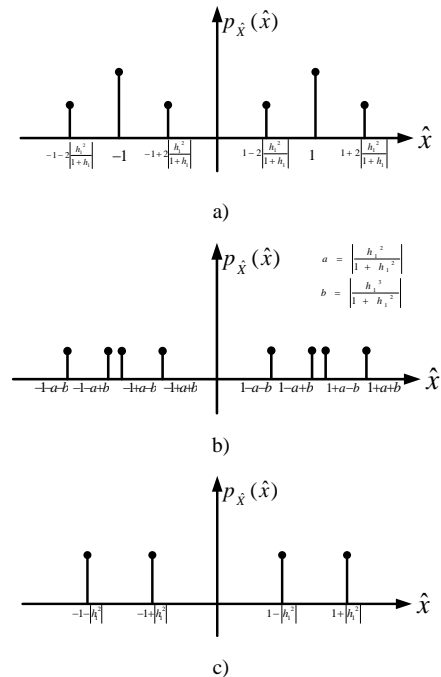


Fig. 2. Pdf do sinal de saída do filtro de desconvolução, para os casos a)  $L_1$ , b)  $L_2$  e c)  $L_\infty$ .

Através da Fig. 2, e sabendo-se que,

$$|h_1^2| \leq \left| \frac{h_1^2}{1+h_1^2} \right| + \left| \frac{h_1^3}{1+h_1^2} \right| \leq 2 \left| \frac{h_1^2}{1+h_1^2} \right|. \quad (21)$$

É possível afirmar que, do ponto de vista do teorema de Benveniste-Goursat-Rouget<sup>1</sup> (BGR), o filtro ótimo  $L_\infty$  seria o melhor dos três, uma vez que é o que melhor realiza um critério de *pdf matching*, neste caso, baseado na menor distorção do diagrama de olho.

#### B. Caso de sinais esparsos.

O estudo de desconvolução considerando sinais esparsos é de interesse em diversas aplicações, como por exemplo, em problemas de processamento sísmico. Neste caso,  $x_n$  modela a refletividade da terra, que pode ser vista como um sinal esparsos,  $\mathbf{h}$  representa a chamada wavelet sísmica, oriunda da emissão da onda acústica,  $v_n$  é uma perturbação ocasionada pelo ambiente,  $y_n + v_n$  é o traço sísmico,  $\mathbf{w}$  é o filtro de desconvolução e  $\hat{x}_n$  é a refletividade estimada [7].

<sup>1</sup> Segundo o teorema BGR, se as pdfs do sinal transmitido (não gaussiano) e do sinal de saída do equalizador são iguais, então o canal está perfeitamente equalizado [5].

Para este estudo, iremos considerar uma entrada impulsiva de modo que os vetores  $\mathbf{x}$ ,  $\mathbf{y}$  e  $\hat{\mathbf{x}}$  são dados por,

$$\mathbf{x} = [1 \ 0 \ 0 \ 0 \dots], \quad (22)$$

$$\mathbf{y} = [1 \ h_1 \ 0 \ 0 \dots], \quad (23)$$

$$\hat{\mathbf{x}} = [1 \ h_1 + w_1 w_1 h_1 \ 0 \dots]. \quad (24)$$

No primeiro caso, iremos analisar a minimização da norma  $L_1$  do vetor de erro,

$$J_{L_1} = \|\mathbf{x} - \hat{\mathbf{x}}\|_1, \quad (25)$$

dada pela equação,

$$\min_{w_1} |h_1 + w_1| + |w_1 h_1|. \quad (26)$$

Derivando a expressão acima em função de  $w_1$  e igualando a zero, obtemos,

$$\text{sgn}|h_1 + w_1| + h_1 \text{sgn}|w_1 h_1| = 0. \quad (27)$$

Como  $|h_1| \neq 1$ , as possíveis soluções são dadas nas descontinuidades da expressão acima. Estas soluções são dadas pelos valores de  $w_1$  que produzem os valores de  $J_{L_1}$  dados por,

$$\begin{cases} |h_1|, & \text{para } w_1 = 0, \\ |h_1^2|, & \text{para } w_1 = -h_1, \end{cases} \quad (28)$$

como  $|h_1| < 1$ , a solução é dada por  $w_1 = -h_1$ .

No segundo caso, analisamos a minimização da norma  $L_2$ ,

$$J_{L_2} = \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2, \quad (29)$$

dada pela equação,

$$\min_{w_1} |h_1 + w_1|^2 + |w_1 h_1|^2. \quad (30)$$

A solução neste caso é a de Wiener, dada por  $w_1 = \frac{-h_1}{1+h_1^2}$ .

No terceiro e último caso, iremos analisar a minimização da norma  $L_\infty$ ,

$$J_{L_\infty} = \|\mathbf{x} - \hat{\mathbf{x}}\|_\infty, \quad (31)$$

dada por,

$$\min_{w_1} \max(|h_1 + w_1|, |w_1 h_1|). \quad (32)$$

A solução é dada pelo valor de  $w_1$  que consegue minimizar a distância entre os valores extremos do erro, que neste caso simples se torna  $|h_1 + w_1| = |w_1 h_1|$ . Duas possíveis soluções são dadas pelos valores de  $w_1$  que produzem os valores de  $J_{L_\infty}$  dados por,

$$\begin{cases} \left| \frac{h_1^2}{1+h_1} \right|, & \text{para } w_1 = \frac{-h_1}{1+h_1}, \\ \left| \frac{h_1^2}{1-h_1} \right|, & \text{para } w_1 = \frac{-h_1}{1-h_1}, \end{cases} \quad (33)$$

como  $|h_1| < 1$ , a solução é dada por  $w_1 = \frac{-h_1}{1+h_1}$ .

Neste caso, podemos visualizar o resultado observando os elementos do vetor de saída  $\hat{\mathbf{x}}$ . Se fizermos os cálculos, o filtro que produz a saída mais esparsa, no sentido de possuir mais elementos nulos, é a do filtro  $L_1$ .

### C. Relações entre o uso de diferentes normas e sinais com diferentes distribuições.

Através dos resultados obtidos, elaboramos a Tabela II, que ilustra claramente a dualidade que existe entre sinais de distribuição uniforme discreta e sinais esparsos e a minimização das normas  $L_1$  e  $L_\infty$ . Nesta tabela, destaca-se o

filtro ótimo  $[1 - h_1]$ , capaz de produzir a saída uniforme de menor dispersão, para o caso de uma entrada uniforme, obtido através da minimização da norma  $L_\infty$ , e a saída mais esparsa, no caso de uma entrada esparsa, obtido com a minimização da norma  $L_1$ , tal como visto nas seções III. A e III. B, respectivamente.

Tabela II. Valores do parâmetro ótimo segundo cada critério e levando em consideração a distribuição do sinal de entrada.

Distribuição do vetor $\mathbf{x}$	$L_1$	$L_2$	$L_\infty$
Esparsa (entrada impulsiva)	$-h_1$	$\frac{-h_1}{1+h_1^2}$	$\frac{-h_1}{1+ h_1 }$
Uniforme (bernoulli, iid e média nula)	$\frac{-h_1}{1+ h_1 }$	$\frac{-h_1}{1+h_1^2}$	$-h_1$

Se considerarmos agora o efeito de uma perturbação, modelada por um ruído aditivo gaussiano, esperamos obter que, à medida que a SNR diminui, a pdf do sinal de entrada do filtro de desconvolução se torna mais parecida com a pdf de uma gaussiana, fazendo com que a norma  $L_2$  se torne a mais adequada.

Com base neste resultado, elaboramos o diagrama da Fig. 3, no qual se ilustra basicamente que o melhor filtro de desconvolução para um sinal de entrada uniforme com alta SNR é o filtro obtido com a minimização da norma  $L_\infty$ , enquanto que para um sinal de entrada esparsa, com alta SNR, é o filtro obtido com a minimização da norma  $L_1$ . A medida que a SNR diminui, os filtros ótimos tendem ao filtro que minimiza a norma  $L_2$ .

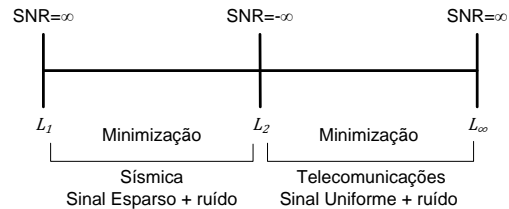


Fig. 3. Relação entre as normas, os casos de interesse e a SNR do sinal de entrada do filtro de desconvolução em dB.

Deste modo, do ponto de vista de filtragem e estimação, a escolha de um critério de otimização deveria ser determinada a partir da distribuição dos sinais envolvidos. No entanto, esta não é a regra geral, como por exemplo, na teoria clássica de filtragem ótima (Wiener) que tem por base o uso da norma  $L_2$  do erro.

### IV. CRITÉRIOS BASEADOS NA ENTROPIA DO ERRO

A área de aprendizado de máquina baseado na teoria da informação tem se afirmado, ao longo dos últimos anos, como um dos campos de pesquisa mais promissores em processamento de sinais [4]. Um motivo para isto é a perspectiva de, por meio do uso de critérios alternativos ao critério quadrático de Wiener, empregar efetivamente estatísticas de ordem superior do sinal de erro [8].

A partir da análise apresentada na seção anterior, propomos agora uma interpretação alternativa para a aplicabilidade de ITL, interpretação esta que, até onde sabemos, não foi relatada na literatura. Para isto, consideremos um critério específico de ITL, o critério de mínima entropia diferencial do erro, definido como,

$$h(e) = -E\{\ln(f_E(e))\}. \quad (34)$$

A ideia básica, nesse caso, é que, minimizando essa entropia, ter-se-á um estimador preciso no sentido de que, intuitivamente, haverá mínima incerteza quanto ao valor de interesse.

Consideremos o caso da pdf da gaussiana generalizada [9], de média nula, dada por,

$$f_E(e) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} \exp\left(-\left(\frac{|e|}{\alpha}\right)^\beta\right), \quad \beta > 0, \alpha > 0, \quad (35)$$

sendo  $\Gamma(\cdot)$ , a função gamma.

Neste caso mais geral, temos que  $h(e)$  é igual a,

$$h(e) = -\ln\left(\frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)}\right) + E\left\{\left(\frac{|e|}{\alpha}\right)^\beta\right\}, \quad (36)$$

sendo que  $\alpha$  pode ser reescrito em função de  $\beta$  e  $|e|$  [10],

$$\alpha^\beta = \beta E\{|e|^\beta\}. \quad (37)$$

Substituindo (37) em (36), obtemos a seguinte expressão,

$$h(e) = \frac{1}{\beta}[1 + \ln(\beta)] - \ln\left(\frac{\beta}{2\Gamma\left(\frac{1}{\beta}\right)}\right) + \ln\left(\beta \sqrt[\beta]{E\{|e|^\beta\}}\right) \quad (38)$$

Considerando  $\beta = 1$ , temos a pdf da laplaciana de média nula [10] e  $h(e)$  é dado por,

$$h(e) = 1 - \ln(0,5) + \ln(E\{|e|\}), \quad (39)$$

sendo proporcional a,

$$h(e) \propto \ln(E\{|e|\}) \propto E\{|e|\} \propto \|e\|_1. \quad (40)$$

Isto significa que minimizar a entropia diferencial do erro será equivalente, do ponto de vista de otimização, a minimizar a sua norma  $L_1$ . Portanto, do ponto de vista de ITL, por meio do critério de entropia do erro, estaríamos implicitamente selecionando a norma mais adequada para o caso em questão. Se tomarmos a densidade gaussiana, veremos que considerações idênticas podem ser feitas com respeito à norma  $L_2$ .

Para o caso de uma densidade uniforme, é, à primeira vista, mais difícil verificar uma conexão. No entanto, sabemos que, quando  $\beta \rightarrow \infty$ , a pdf da gaussiana generalizada converge para a pdf da distribuição uniforme  $[-\alpha, \alpha]$  [10]. Deste modo, o critério de entropia diferencial do erro de uma variável aleatória, com uma distribuição uniforme, poderia ser visto como o caso limite de  $\beta \rightarrow \infty$  na Eq. (38). Assim, temos que  $h(e)$  é igual a,

$$h(e) = \lim_{\beta \rightarrow \infty} \frac{1}{\beta}[1 + \ln(\beta)] - \ln\left(\frac{\beta}{2\Gamma\left(\frac{1}{\beta}\right)}\right) + \ln\left(\beta \sqrt[\beta]{E\{|e|^\beta\}}\right), \quad (41)$$

sendo que,

$$\lim_{\beta \rightarrow \infty} \frac{1}{\beta}[1 + \ln(\beta)] - \ln\left(\frac{\beta}{2\Gamma\left(\frac{1}{\beta}\right)}\right) = -\ln(0,5), \quad (42)$$

logo, proporcional a,

$$h(e) \propto \lim_{\beta \rightarrow \infty} \ln\left(\beta \sqrt[\beta]{E\{|e|^\beta\}}\right) \propto \lim_{\beta \rightarrow \infty} \beta \sqrt[\beta]{E\{|e|^\beta\}} \propto \|e\|_\infty. \quad (43)$$

Portanto, minimizar a entropia diferencial do erro de uma variável aleatória com uma distribuição uniforme será equivalente a minimizar a sua norma  $L_\infty$ .

## V. CONCLUSÃO

As análises e os resultados apresentados neste trabalho nos permitem tirar conclusões que, mesmo que preliminares, se mostram bastante promissoras, tendo em vista o interesse em melhor adequar a norma de otimização ao problema em questão e às características dos sinais processados.

No problema de estimação, a melhor norma  $L_p$  é aquela que fornece o estimador de menor variância, ficando clara a relação entre a conveniência de se usar  $p$  menor ou maior do que dois para os casos de variáveis aleatórias supergaussianas ou subgaussianas, respectivamente.

No tocante ao problema de filtragem, mostramos como o critério mais adequado se relaciona com as características da aplicação. Assim, no caso de desconvolução de canal em comunicações, devemos minimizar a norma que melhor leve ao “*pdf matching*”, satisfazendo assim o teorema BGR. Já em desconvolução sísmica, a melhor norma é aquela cuja minimização leve à recuperação do sinal mais esparso.

Finalmente, ao associarmos os resultados obtidos com a teoria de ITL, percebemos que o critério mais adequado é aquele que minimiza a entropia do erro. Ao proceder esta minimização, confirmamos de uma forma alternativa os resultados anteriores.

Assim, no sentido de contribuir com a teoria de filtragem usando normas  $L_p$ , a abordagem aqui desenvolvida estabelece uma interessante conexão tripartite envolvendo: i-as distribuições do sinal tratado (subgaussiana, supergaussiana ou mesmo gaussiana); ii-a natureza do problema de desconvolução (recuperar uma situação de “*pdf matching*”, ou de esparsidade, ou mesmo intermediária se há ruído gaussiano aditivo); e iii-os critérios oriundos de aprendizado de máquina baseado na teoria da informação.

## AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq pelo apoio financeiro.

## REFERÊNCIAS

- [1] R. Gonin e A. H. Money, *Nonlinear Lp-Norm Estimation*. Taylor, 1989.
- [2] J. R. Rice e J. S. White, “Norms for smoothing and estimation”, *SIAM Review*, v. 6, pp. 243-255, Julho 1964.
- [3] H. W. J. Debye e P. Van Riel, “ $L_p$ -norm deconvolution”, *Geophysical Prospecting*, v. 38, pp. 381-403, Maio 1990.
- [4] J. C. Principe, *Information Theoretic Learning: Renyi's entropy and kernel perspectives*. Springer, 2010.
- [5] J. F. Claerbout e F. Muir, “Robust modeling with erratic data”, *Geophysics*, v. 38, n. 5, pp. 826-844, Outubro 1973.
- [6] A. O. Neves, R. R. F. Attux, R. Suyama, M. D. Miranda, J. M. T. Romano, “Sobre critérios para equalização não-supervisionada”, *Revista Controle & Automação*, v. 17, n. 3, Julho, Agosto e Setembro 2006.
- [7] O. Yilmaz, *Seismic Data Analysis: Processing, Inversion and Interpretation of Seismic Data*. SEG, v. 1, 2001.
- [8] D. Erdogmus e J. C. Principe, “From linear adaptive filtering to nonlinear information processing”, *IEEE Signal Processing Magazine*, v. 23, pp. 14-33.
- [9] S. Nadarajah, “A generalized normal distribution”, *Journal of Applied Statistics*, v. 32, n. 7, pp. 685-694, Setembro 2005.
- [10] H. C. Wu, “Blind source separation using information measures in the time and frequency domains”, Ph.D. dissertation, Dept. Elec. and Comp. Eng., Florida Univ., Gainesville, FL, 1999.