

Critérios alternativos para a desconvolução não-supervisionada

Renan D. B. Brotto*, Kenji Nose-Filho**, João M. T. Romano*

*Faculdade de Engenharia Elétrica e de Computação da UNICAMP, Campinas, Brasil

**Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas da UFABC, Santo André, Brasil

e-mail: rbrotto@decom.fee.unicamp.br, kenji.filho@ufabc.edu.br, romano@decom.fee.unicamp.br

Resumo - Este trabalho aborda a aplicação do critério de Máxima Verossimilhança sobre o problema de desconvolução não-supervisionada. Tal critério pode ser obtido a partir do princípio Infomax, comumente usado no contexto de Separação de Fontes. Adotando a distribuição da Gaussiana generalizada, este critério pode ser simplificado através da norma ℓ_p , cujo valor de p varia de acordo com a forma da distribuição. Tais normas, por sua vez, despontam como uma alternativa em potencial para a tarefa de desconvolução preditiva e esta potencialidade é explorada aqui através da dualidade entre as normas ℓ_1 e ℓ_∞ e as distribuições dos sinais a serem recuperados.

Palavras-chave: normas ℓ_p , desconvolução preditiva, estimador de Máxima Verossimilhança, dualidade.

Introdução

O problema de desconvolução linear é de grande relevância para a área de processamento de sinais, tanto por seus aspectos teóricos quanto práticos [1]. Neste problema, busca-se ajustar um filtro $w[n]$ de modo a compensar as distorções ocasionadas por um sistema linear e invariante no tempo, com resposta ao impulso $h[n]$, sobre o sinal de entrada $s[n]$, como ilustrado pelo diagrama de blocos da Figura 1.

Em sua vertente não-supervisionada, o projeto do filtro é conduzido sem o acesso às amostras de $s[n]$ e sem o conhecimento explícito do sistema $h[n]$. No entanto, para conduzir o processo de otimização do filtro de desconvolução, são necessárias algumas premissas sobre o sinal de entrada e/ou o sistema $h[n]$. Neste trabalho, é enfatizada a hipótese de que $s[n]$ é um sinal estacionário, composto por amostras de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) de uma distribuição não Gaussiana.

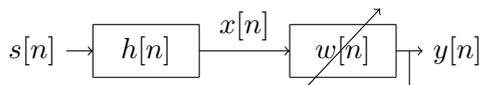


Figura 1 – Diagrama de blocos de um sistema de desconvolução linear e não-supervisionado.

Uma possível abordagem para tal problema é através da desconvolução preditiva, apresentada ini-

cialmente em [2], dentro do contexto de desconvolução sísmica, e que emprega técnicas de predição linear para compensar as distorções observadas. O objetivo, neste caso, é ajustar os coeficientes $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_L]$ do chamado filtro de erro de predição (FEP), que pode ser escrito como [1]:

$$W(z) = 1 - w_1 z^{-1} - w_2 z^{-2} - \dots - w_L z^{-L}, \quad (1)$$

de modo a minimizar uma dada função de custo estabelecida sobre o erro de predição $y[n]$, dado por:

$$y[n] = x[n] - \sum_{i=1}^L w_i x[n-i], \quad (2)$$

com $x[n] = \sum_{k=0}^{P-1} h[k] s[n-k]$.

Classicamente, utiliza-se o Erro Quadrático Médio de predição (MSE, do inglês, *Mean Squared Error*), conduzindo à seguinte função de custo:

$$J(\mathbf{w}) = \mathbb{E}[y^2[n]], \quad (3)$$

o que resulta no problema de otimização:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{arg\,min}} J(\mathbf{w}). \quad (4)$$

Antes de continuarmos, cabe aqui a definição da norma ℓ_p de um vetor \mathbf{y} no espaço \mathbb{R}^M :

$$\|\mathbf{y}\|_p = \left(\sum_{i=1}^M |y_i|^p \right)^{\frac{1}{p}}, \quad (5)$$

onde y_i denota a i -ésima componente do vetor \mathbf{y} . Comparando as equações (3) e (5) vemos que o problema de otimização apresentado na desconvolução preditiva é equivalente à minimização do quadrado da norma ℓ_2 do erro de predição. Contudo, a abordagem clássica ($p = 2$) apresenta a restrição de ser eficaz apenas no tratamento de canais de fase mínima, o que motiva o uso de outros valores para p [3].

O objetivo deste trabalho é estudar características do sinal $s[n]$ que permitam determinar o valor de p mais adequado para tornar mais eficiente o procedimento de desconvolução preditiva. Em particular investigamos algumas relações de dualidade entre as normas ℓ_p para algumas funções densidade de probabilidade (pdf, do inglês *probability density function*) do sinal $s[n]$. Tal

estudo complementa os resultados observados no contexto supervisionado em [4].

Nossa abordagem aqui consiste em partir de critérios comumente empregados no problema de Separação Cega de Fontes (BSS, do inglês *Blind Source Separation*), adequá-los ao problema de desconvolução para, em seguida, assumindo uma distribuição Gaussiana generalizada para $s[n]$, estabelecer as relações de interesse.

Na seção de **Fundamentação Teórica** é apresentada a relação entre o critério Infomax [1], [5] e o estimador de Máxima Verossimilhança [6] para o problema de Separação de Fontes. Em seguida, ambos os critérios são particularizados para o problema de desconvolução não-supervisionada. Finalmente, o critério de Máxima Verossimilhança é aplicado sobre as distribuições Gaussianas generalizadas, estabelecendo uma relação entre a forma da distribuição e a escolha do parâmetro p a ser empregado na norma ℓ_p .

A seção de **Resultados** mostra a dualidade entre as normas ℓ_1 e ℓ_∞ , tanto para a desconvolução supervisionada quanto para o contexto preditivo. Por fim, a seção de **Conclusões** apresenta as discussões feitas e algumas perspectivas futuras de trabalho.

Fundamentação Teórica

No problema de Separação Cega de Fontes de misturas lineares, uma das formas de se obter a matriz de separação é através da chamada Análise por Componentes Independentes (ICA, do inglês *Independent Component Analysis*). Em ICA, parte-se da premissa de que o número de misturas é igual ao número de fontes e que as fontes são estatisticamente independentes e provenientes de uma distribuição não Gaussiana. Sob estas condições, a recuperação das fontes pode ser dada por uma matriz capaz de prover sinais estatisticamente independentes [1].

Ao longo do tempo, diversas métricas foram propostas com o objetivo de implementar a ideia de ICA. Dentre elas, destacam-se o critério Infomax e o critério de Máxima Verossimilhança. Com base nestes critérios é possível estabelecer algumas relações interessantes entre o critério de norma ℓ_p e a distribuição do sinal $s[n]$ através da distribuição da Gaussiana generalizada.

Critérios Infomax e de Máxima Verossimilhança

O critério Infomax [5] empregado na Separação de Fontes tem sua origem no campo das redes neurais ar-

tificiais [1]. Este critério aplica funções não-lineares nas saídas do sistema separador \mathbf{W} , um sistema com múltiplas entradas e múltiplas saídas, originando os sinais $z_i[n]$, $i = 1, 2, \dots, N$, onde N denota o número de canais combinados:

$$z_i[n] = f_i(y_i[n]). \quad (6)$$

As funções não-lineares $f_i(\cdot)$, $i = 1, \dots, N$ têm por finalidade limitar a entropia das saídas do sistema separador e devem ser monotonicamente crescentes, com $f_i(-\infty) = 0$ e $f_i(\infty) = 1$, $\forall i$.

Assim sendo, os seguintes vetores são definidos:

$$\begin{aligned} \mathbf{s} &= [s_1[n] \ s_2[n] \ \dots \ s_N[n]]^T, \\ \mathbf{x} &= [x_1[n] \ x_2[n] \ \dots \ x_N[n]]^T, \\ \mathbf{z} &= [z_1[n] \ z_2[n] \ \dots \ z_N[n]]^T. \end{aligned}$$

O critério Infomax busca maximizar a informação mútua [7] entre os sinais \mathbf{x} e \mathbf{z} :

$$\max_{\mathbf{W}} I(\mathbf{x}; \mathbf{z}), \quad (7)$$

Na ausência de ruído, a maximização da informação mútua é equivalente à maximização da entropia conjunta das saídas do sistema separador, a qual é dada por [1]:

$$H(\mathbf{z}) = H(\mathbf{x}) + \mathbb{E} \left[\sum_{i=1}^N \log(f'_i(\mathbf{w}_i \mathbf{x})) \right] + \log(|\det(\mathbf{W})|). \quad (8)$$

Como a entropia conjunta das misturas $H(\mathbf{x})$ não depende da matriz de separação \mathbf{W} , o critério Infomax resulta no seguinte problema de otimização:

$$\max_{\mathbf{W}} \mathbb{E} \left[\sum_{i=1}^N \log(f'_i(\mathbf{w}_i \mathbf{x})) \right] + \log(|\det(\mathbf{W})|) \quad (9)$$

onde $f'_i(\cdot)$ denota a derivada de $f_i(\cdot)$ e \mathbf{w}_i denota a i -ésima linha de \mathbf{W} .

Diante das restrições que as funções $f_i(\cdot)$ devem obedecer, uma escolha natural para tais funções são as funções de densidade de probabilidade acumulada das fontes, o que conduz ao critério de Máxima Verossimilhança apresentado em [8] e explicitado a seguir:

$$\max_{\mathbf{W}} \mathbb{E} \left[\sum_{i=1}^N \log(p_{s_i}(\mathbf{w}_i \mathbf{x})) \right] + \log(|\det(\mathbf{W})|) \quad (10)$$

Ambos os critérios podem ser adaptados para o problema de desconvolução não-supervisionada, bastando trocar a diversidade espacial das fontes pela diversidade temporal das amostras dos sinais. Deste modo, a notação matemática adotada passa a ser dada por:

$$\begin{aligned}\mathbf{s} &= [s[n] \ s[n-1] \ \dots \ s[n-N+1]]^T, \\ \mathbf{x} &= [x[n] \ x[n-1] \ \dots \ x[n-N+1]]^T, \\ \mathbf{z} &= [z[n] \ z[n-1] \ \dots \ z[n-N+1]]^T,\end{aligned}$$

e \mathbf{W} passa a ser dada pela matriz de convolução associada ao filtro $w[n]$.

Como a matriz de convolução é uma matriz triangular inferior, com diagonal unitária, temos que $\det(\mathbf{W}) = 1$. Pelo fato de estarmos tratando um problema monocanal com sinais estacionários, temos que $f_1(\cdot) = f_2(\cdot) = \dots = f_N(\cdot) = f(\cdot)$ e $p_{s_1}(\cdot) = p_{s_2}(\cdot) = \dots = p_{s_N}(\cdot) = p_s(\cdot)$, de modo que os critérios (9) e (10) passam a ser escritos da seguinte forma:

$$\max_{\mathbf{W}} J_{Infomax}(\mathbf{W}) = \max_{\mathbf{W}} \mathbb{E} \left[\sum_{i=1}^N \log(f'(\mathbf{w}_i \mathbf{x})) \right], \quad (11)$$

$$\max_{\mathbf{W}} J_{ML} = \max_{\mathbf{W}} \mathbb{E} \left[\sum_{i=1}^N \log(p_s(\mathbf{w}_i \mathbf{x})) \right]. \quad (12)$$

Ao reescrever a função de Máxima Verossimilhança em termos da divergência de Kullback-Leibler, como em [1], este critério pode ser visto como um critério de casamento de pdfs. Através do Teorema de Benveniste, Goursat e Rouget, a desconvolução de um sinal composto por amostras de variáveis aleatórias i.i.d. de uma distribuição não Gaussiana pode ser obtida através do casamento da pdf do sinal de entrada $s[n]$ com o sinal de saída do filtro de desconvolução $y[n]$. Desta forma, o critério de Máxima Verossimilhança poderia ser utilizado na desconvolução preditiva de sinais com estas características.

A aplicação do critério de Máxima Verossimilhança requer o conhecimento *a priori* das distribuições de probabilidade dos sinais de entrada. No entanto, diversas pdfs com aplicações em cenários reais, tais como as distribuições Laplaciana, Gaussiana e uniforme, podem ser unificadas na distribuição de probabilidade Gaussiana generalizada [4], cuja equação é dada a seguir:

$$p_S(s) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-\left(\frac{|s-\mu|}{\alpha}\right)^\beta}, \quad (13)$$

onde o parâmetro α controla a sua dispersão, o parâmetro β define a sua forma e ambos se relacionam através de $\alpha^\beta = \beta \mathbb{E}[|s|^\beta]$. Utilizando (13) em (12) e considerando um sinal de entrada de média nula

($\mu = 0$) e variância unitária ($\alpha = \sqrt{2}$), obtemos:

$$\begin{aligned}J_{ML}(\mathbf{W}) &= \mathbb{E}[\ln(p_S(\mathbf{W}\mathbf{x}))] = \\ &= \mathbb{E} \left[\ln \left(\frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-\left(\frac{|y|}{\alpha}\right)^\beta} \right) \right] = \\ &= \mathbb{E} \left[\ln \left(\frac{\beta}{2\alpha\Gamma(1/\beta)} \right) - \frac{|y|^\beta}{\alpha^\beta} \right].\end{aligned} \quad (14)$$

Consequentemente, o seguinte problema de otimização pode ser formulado:

$$\max_{\mathbf{W}} J_{ML}(\mathbf{W}) = \max_{\mathbf{W}} -|y|^\beta = \min_{\mathbf{W}} |y|^\beta \quad (15)$$

Em termos gerais, este resultado mostra o nosso interesse de utilizar valores de p menores do que 2 para distribuições com valor de β menor do que 2 e valores de p maiores do que 2 para distribuições com valor de β maior do que 2. Em lados opostos, vemos o interesse do uso da norma ℓ_1 para sinais com distribuições esparsas (uma vez que $\beta = 1$ recai na distribuição de Laplace) e o uso da norma ℓ_∞ para sinais com distribuições uniformes (uma vez que $\beta = \infty$ recai na distribuição uniforme).

Resultados

Nesta seção iremos investigar a posição dos zeros do filtro de erro de predição ótimo para dois casos duais: a minimização da norma ℓ_1 do erro de predição para sinais esparsos e a minimização da norma ℓ_∞ do erro de predição para sinais uniformemente distribuídos de um alfabeto finito. Para termos de comparação, também são obtidos os zeros do filtro de desconvolução para os mesmos casos em um contexto supervisionado.

Os sinais esparsos foram gerados segundo uma distribuição de Bernoulli, com $Pr(s[n] = 1) = 0.02$. Para os sinais com distribuição uniforme, foi adotado um alfabeto binário com $s[n] = \pm 1$. Os experimentos foram realizados para os três casos típicos de canais: fase mínima $H_1(z) = 1 + 0.5z^{-1} + 0.3z^{-2}$; fase mista $H_2(z) = 0.5 + 1z^{-1} + 0.3z^{-2}$; fase máxima $H_3(z) = 1 + 1.5z^{-1} + 1.8z^{-2}$.

Para o caso supervisionado, foi utilizado um filtro de desconvolução com dois coeficientes $W(z) = w_0 + w_1z^{-1}$; após a adaptação de w_0 e w_1 , os filtros obtidos foram normalizados no valor do primeiro coeficiente (w_0), apenas para fins de comparação. Já para o caso não-supervisionado, foi utilizado um FEP com apenas um coeficiente, ou seja, $W(z) = 1 - w_1z^{-1}$.

Um ponto bastante relevante do caso supervisionado é quanto ao atraso ótimo para o sinal de referência: atraso zero para o canal de fase mínima, atraso de

uma amostra para o canal de fase mista e atraso de três amostras para o canal de fase máxima.

Os coeficientes dos filtros obtidos com a minimização da norma ℓ_1 foram ajustados através do algoritmo LMS (do inglês, *least mean squares*) sinal do erro [9], enquanto que os coeficientes do filtro obtidos com a minimização da norma ℓ_∞ foram otimizados através da meta-heurística populacional de evolução diferencial [10].

A tabela 1 apresenta os filtros obtidos através das normas ℓ_1 e ℓ_∞ dentro do contexto supervisionado e não-supervisionado.

Tabela 1 – Filtros de desconvolução obtidos no contexto supervisionado e não-supervisionado.

Supervisionado	Não-supervisionado
Canal de fase mínima	
$W_1(z) = 1 - 0.4997z^{-1}$	$W_1(z) = 1 - 0.5z^{-1}$
$W_\infty(z) = 1 - 0.5000z^{-1}$	$W_\infty(z) = 1 - 0.5z^{-1}$
Canal de fase mista	
$W_1(z) = 1 - 0.2993z^{-1}$	$W_1(z) = 1 - 0.3z^{-1}$
$W_\infty(z) = 1 - 0.2997z^{-1}$	$W_\infty(z) = 1 - 0.3z^{-1}$
Canal de fase máxima	
$W_1(z) = 1 - 1.4969z^{-1}$	$W_1(z) = 1 - 1.2z^{-1}$
$W_\infty(z) = 1 - 1.4885z^{-1}$	$W_\infty(z) = 1 - 1.2z^{-1}$

Da tabela apresentada é possível observar três resultados interessantes:

- Os zeros obtidos no contexto supervisionado e não-supervisionado são próximos, sobretudo para os canais de fase mínima e mista. Já para o caso de fase máxima, a diferença observada na posição dos zeros dos filtros se deve, possivelmente, às diferentes estruturas adotadas.
- Os filtros obtidos com a minimização da norma ℓ_1 em um ambiente esparso são os mesmos que os obtidos com a minimização da norma ℓ_∞ em um ambiente com sinais uniformemente distribuídos de um alfabeto finito. Este resultado, em particular, ilustra a relação de dualidade entre as normas ℓ_p com as distribuições do sinal original, sendo que um sinal esparso pode ser visto como o dual de um sinal uniformemente distribuído, levando-se em consideração os limites da distribuição da Gaussiana generalizada.
- As normas ℓ_1 e ℓ_∞ permitem obter filtros de erro de predição com zeros fora do círculo de raio unitário, como elucidado pelo experimento com o canal de fase máxima. Tais normas, portanto, têm a potencialidade de conduzir a desconvolução em canais de fase arbitrária.

Conclusões

Neste trabalho, o critério de Máxima Verossimilhança para a desconvolução não-supervisionada foi obtido como um caso particular do critério Infomax, explorando, deste modo, a relação entre os problemas de desconvolução e Separação Cega de Fontes.

Diante dos resultados apresentados, a partir do estimador de Máxima Verossimilhança, estabelecemos uma relação entre as normas ℓ_p e as distribuições Gaussianas generalizadas. Desta forma, chega-se à escolha mais adequada da norma a ser empregada na desconvolução preditiva.

Finalmente, mostramos que as normas ℓ_p com $p \neq 2$ apresentaram a potencialidade de lidar com canais de fase não mínima, sendo que este fato será melhor investigado em trabalhos futuros.

Agradecimentos

Os autores do trabalho agradecem à CAPES e à FAPESP (2017/13025-2) pelo suporte financeiro.

Referências

- [1] J. M. T. Romano, R. R. de F. Attux, C. C. Cavalcante, R. Suyama *Unsupervised Signal Processing - Channel Equalization and Source Separation*, 1st ed., CRC PRESS, 2011.
- [2] E. A. Robinson *Predictive Decomposition of Time Series with Applications to Seismic Exploration*, PhD. Thesis, Massachusetts Institute of Technology, 1954.
- [3] K. Nose-Filho and J. M. T. Romano “Adaptive sparse linear prediction: a promising tool for blind deconvolution”, in *Proceedings of the 2016 IEEE Workshop on Sensor Array and Multichannel Signal Processing*, 2016.
- [4] K. Nose-Filho, L. T. Duarte, R. Attux, E. Nadalin, R. Ferrari, J. M. T. Romano “Sobre filtragem L_p ”, in *Proceedings of the XXX Simpósio Brasileiro de Telecomunicações*, 2012.
- [5] A. J. Bell, T. J. Sejnowski “An information-maximization approach to blind separation and blind deconvolution”, *Neural Computation*, 1995.
- [6] S. M. Kay *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall; 1st ed, 1993.
- [7] T. M. Cover, J. A. Thomas *Elements of Information Theory*, Wiley-Interscience; 2nd ed., 2006.
- [8] J.-F. Cardoso “Blind Signal Separation: Statistical Principles”, *Proceedings of IEEE*, Vol. 86, No. 10, 1998.
- [9] S. S. Haykin, *Adaptive Filter Theory*, Pearson, 5th ed, 2013.
- [10] K. Price, R. M. Storn, J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, Natural Computing Series, Springer, 2005.