

Aplicação da Técnica de Fatoração de Matrizes Não-Negativas à Separação de Fontes Sonoras em Misturas contendo Elementos Harmônicos e Percussivos

Wellington Fonseca, Zélia Peixoto, Flávia Magalhães, Marcelo Santos

Programa de Pós-Graduação em Engenharia Elétrica
Pontifícia Universidade Católica de Minas Gerais
Belo Horizonte, Minas Gerais – Brasil

wellington.fonseca@sga.pucminas.br,
{assiszmp, flaviamagfreitas}@pucminas.br,
marcelo.borba.ti@hotmail.com

Abstract. *This paper deals with the separation of audio signals of which there is not a priori information. More specifically, the aim is to separate each instrument in a mixture which includes harmonic and percussive elements. Among the usual techniques, the method of non-negative matrix factorization (NMF) was chosen, once the quality of separation provided by this method is not conditional upon the number of available observations. It was analyzed in Matlab[®] environment, using the divergences of Kullback-Leibler and Itakura-Saito. The results showed no significant differences between the methods, stimulating new researches that include constraints on NMF method, such as the sparsity and smoothness, among others.*

Resumo. *Este trabalho trata da separação de sinais de áudio dos quais não se tem informação a priori, visando separar os instrumentos em uma mistura composta por elementos harmônicos e percussivos. Dentre as técnicas usuais, foi escolhido o método de Fatoração de Matrizes Não-Negativas, uma vez que a qualidade da separação dessa técnica não está condicionada ao número de observações disponíveis. São apresentados os resultados da implementação em ambiente Matlab[®] utilizando as divergências de Kullback-Leibler e Itakura-Saito. Os resultados não indicaram diferenças significativas entre os métodos, incentivando novas investigações sobre a incorporação de restrições ao método NMF, tais como a esparsidade e smoothness, dentre outros.*

1. Introdução

Com os recentes avanços da tecnologia digital as quantidades de dados gerados e a serem tratados pelos sistemas de processamento, em geral, vêm tornando obsoletas as ferramentas clássicas de análise [Chen et al. 2011] [Berry et al. 2006]. Tarefas envolvendo a localização de informações e/ou revelações de características intrínsecas dos sinais demandam, cada vez mais, de técnicas mais eficientes.

Nesse contexto, a técnica de Fatoração de Matrizes Não-Negativas (NMF – *Nonnegative Matrix Factorization*) vem se destacando dentre as demais técnicas disponíveis. A NMF é, basicamente, uma técnica que realiza a decomposição de uma matriz aproximando-a a uma matriz não-negativa composta por um produto de matrizes de menor posto, também não-negativas [Guan et al. 2012].

Dentre as áreas nas quais as técnicas NMF podem ser aplicadas, conforme Krömer et al. (2010), destacam-se a mineração de dados, análise de texto e detecção de

intrusos em redes de computadores; separação de sinais acústicos e simulação de multicanais, 5.1 ou 7.1 por exemplo; eliminação de ruídos diversos em equipamentos biomédicos e redução da dimensão de modelos matemáticos [Krömer et al. 2010] [Lee e Seung 2001].

Contudo, problemas apresentados pela NMF, como por exemplo, a dificuldade de agrupamento de padrões espectrais em uma fonte alvo [Kitamura et al. 2013], têm incentivado modificações da técnica original para um melhor desempenho quanto à separação de sinais de áudio. Além disso, para alguns instrumentos musicais cujas notas apresentam, além de ataque percussivo, sustentação harmônica, como guitarra, baixo e piano, dentre outros, a NMF apresenta dificuldades na representação desta nota como componentes, uma vez que seus espectros variam no tempo [Tygel, 2009].

O aprimoramento de técnicas de separação de sinais de áudio podem possibilitar melhorias quanto à remoção de ruídos de *crossstalk* em gravações com mais de um microfone, aumento da qualidade no reconhecimento de voz, análises de cenários musicais, transcrição automática de partituras, restauração e remasterização de obras antigas [Mahmoud et al. 2009] [Kitamura et al. 2013].

A literatura técnico-científica aborda também a utilização de técnicas supervisionadas da NMF nas quais, em linhas gerais, cria-se uma base de treinamento com informações da fonte de interesse, quando se deseja incrementar a qualidade da separação [Kitamura et al. 2013], embora nem sempre ocorra a disponibilidade do sinal para treinamento. Além disso, a literatura aborda modificações nas funções de custo, ou penalizações, visando atender a determinada restrição, como por exemplo, extensão da NMF para sinais estéreo [Sawada et al. 2012].

Estudos de cunho mais teóricos abordando, dentre outros, diferenças de resultados, em termos de qualidade e velocidade de convergência, entre métodos diferentes de otimização, problemas da técnica quanto a convergência e formas de melhoria dos mesmos, descrição dos algoritmos e descrição das áreas mais comuns de aplicação podem, também, ser encontrados [Wang e Zhang 2013] [Guan et al. 2012] [Lin 2007] [Berry et al. 2006].

O presente trabalho tem por objetivo estudar a aplicação da NMF à separação de misturas contendo fontes sonoras percussivas e harmônicas, caracterizando um conjunto musical padrão. Os resultados apresentados referem-se a sinais monoaurais, sendo a extensão a sinais estéreo ou outras formas multi-canais realizadas pelo uso de tensores, técnica de Fatoração de Tensores Não-Negativos (NTF - *Nonnegative Tensor Factorization*) ou modificações adicionais da técnica NMF para a adequação a esse tipo de sinal, como em [Wang e Zhang 2013] [Sawada et al. 2012].

2. Fundamentos Matemáticos

Fatoração de matrizes é um tópico largamente estudado em áreas como processamento de sinais e álgebra linear. A ideia de que um elemento, tão simples quanto possível, possa descrever fenômenos físicos de alto nível de complexidade, tornando-os mais simples e de fácil entendimento, é um dos motivadores do avanço deste tipo de técnica. Fatores como a enorme quantidade de dados envolvidos na observação de um problema, o inter-relacionamento entre variáveis em fenômenos físicos complexos e modelos matemáticos de problemas físicos restritos a números não negativos são motivadores da NMF, descrito formalmente como [Chchocki et al. 2009] [Berry et al. 2006] [Lee and Seung 2001] [Wang and Zhang 2013]:

$$V \approx \check{V} = W \times H \quad (1)$$

onde $V \in \mathbb{R}_+^{MN}$ representa a matriz que contém o espectrograma da mistura a ser separada, obtido por meio da Transformada de Fourier de Curta Duração (STFT – *Short-Time Fourier Transform*), $W \in \mathbb{R}_+^{MR}$ é a chamada matriz de base e representa o padrão espectral e $H \in \mathbb{R}_+^{RN}$, a matriz de ativação, representa os coeficientes da combinação linear que gera cada fonte especificamente. O coeficiente R expressa o posto das matrizes envolvidas na aproximação e representa o número de fontes contidas na mistura avaliada. Como se trata de um índice matricial, $R \in \mathbb{Z}_+^*$ e, para que de fato ocorra uma redução dimensional, $R < \min(M, N)$ ou $M \times R + R \times N < M \times N$ [Lin 2007].

$$D_{EUC}(V||\check{V}) = \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^N (v_{mn} - \check{v}_{mn})^2 \quad (2)$$

$$D_{KL}(V||\check{V}) = \sum_{m=1}^M \sum_{n=1}^N \left(v_{mn} \ln \frac{v_{mn}}{\check{v}_{mn}} - v_{mn} + \check{v}_{mn} \right) \quad (3)$$

$$D_{IS}(V||\check{V}) = \sum_{m=1}^M \sum_{n=1}^N \left(\frac{v_{mn}}{\check{v}_{mn}} - \ln \frac{v_{mn}}{\check{v}_{mn}} - 1 \right) \quad (4)$$

Basicamente, a NMF é uma técnica iterativa, na qual as regras de atualização das entradas das matrizes W e H são derivadas de funções de custo ou funções objetivo, frequentemente, a Distância Euclidiana, Divergência de Kullback-Leibler e Divergência de Itakura-Saito, sendo essa última, a mais utilizada em aplicações de áudio [Sawada et al. 2012] [Lee and Seung 2001] [Kitamura et al. 2013], todas derivadas da Divergência de Bregman [Chchocki et al. 2009]. As Equações (2), (3), (4) referem-se à Distância Euclidiana, e as Divergências de Kullback-Leibler e Itakura-Saito, respectivamente.

Uma vez que se define a função de custo a ser utilizada, algum algoritmo deve ser utilizado para minimizá-la. As técnicas de otimização são as ferramentas matemáticas que, quando aplicadas às funções de custo, geram o algoritmo de atualização das matrizes de base e ativação da NMF, e incluem gradiente descendente, gradiente projetado, descida coordenada, mínimos quadrados alternados, dentre outros. A escolha do método de otimização é um compromisso básico entre velocidade e convergência. O método mais utilizado na NMF é o método do gradiente descendente. A primeira regra de atualização publicada para a versão original da NMF, na forma de regra multiplicativa, é obtida aplicando-se o método do gradiente descendente à Distância Euclidiana, conforme Equações (6) e (7) [Lee and Seung 2001] [Guan et al. 2012] [Berry et al. 2006]. A Equação (5) descreve o método do gradiente descendente.

$$= x_k - \alpha \times \nabla f(x_k) \quad (5)$$

onde x_{k+1} e x_k representam os estados futuro e atual da variável x , respectivamente, α é o passo ou peso e $\nabla f(x_k)$, a derivada parcial de f em relação à variável x .

As Equações (6) a (11) mostram as regras multiplicativas para a Distância Euclidiana, e as Divergências de Kullback-Leibler e Itakura-Saito, respectivamente, derivadas a partir do método do gradiente descendente.

$$W \leftarrow W \otimes \frac{VH^T}{WHH^T} \quad (6)$$

$$H \leftarrow H \otimes \frac{W^T H}{W^T W H} \quad (7)$$

$$W \leftarrow W \otimes \frac{V}{\frac{WH}{UH^T} H^T} \quad (8)$$

$$H \leftarrow H \otimes \frac{W^T \frac{V}{WH}}{W^T U} \quad (9)$$

$$W \leftarrow W \otimes \frac{\frac{V}{(WH)^2} H^T}{\frac{U}{WH} H^T} \quad (10)$$

$$H \leftarrow H \otimes \frac{W^T \frac{V}{(WH)^2}}{W^T \frac{U}{WH}} \quad (11)$$

onde $U \in \mathbb{R}_+^{MN}$ é uma matriz cujos elementos são todos iguais a 1 e \otimes representa o produto de Hadamard ou elemento a elemento. As divisões também são definidas elemento a elemento, tal que as dimensões envolvidas sejam todas compatíveis [Santos 2015].

3. Metodologia Aplicada

Como já dito na Seção 1, tipicamente uma base de testes é composta por músicas sintetizadas, do tipo MIDI. Entretanto, com o objetivo de testar as técnicas em uma mistura mais condizente com as músicas comerciais (músicas profissionais gravadas em estúdio), gravou-se uma música contendo uma guitarra, um baixo e uma bateria, sendo esta última fonte composta por prato de condução, bumbo e caixa.

Para a gravação dos sinais foi utilizado um microfone para cada fonte, sendo este microfone caracterizado na tabela 1.

Além disso, como é bastante comum o uso de processadores de efeitos nas músicas comerciais, a guitarra foi gravada com o efeito de distorção, o que implica na inserção de mais harmônicos na mistura. Dessa forma, objetiva-se comparar o desempenho das Divergências de Kullback-Leibler e Itakura-Saito nesse contexto. A tabela 2 mostra os parâmetros usados na simulação.

Os algoritmos foram implementados em ambiente *Matlab*[®], utilizando um notebook com processador *Intel*[®] 3230m de 2,6GHz e 8GB de memória RAM.

Tabela 1. Especificações do microfone MX150

Característica	Especificação Técnica
Tipo	Dinâmico

Polaridade	Unidirecional
Faixa de Frequência	56Hz a 14kHz
Sensibilidade	56dB +/- 3dB (0dB = 1V/PA a 1kHz)
Impedância	250 Ohms +/- 30% em 1kHz

Tabela 2. Parâmetros de Simulação

STFT	Tamanho da Janela (Amostras)	1024
	Sobreposição (Amostras)	512
	NFFT	1024
Alg	Número de Iterações	100
	Número de Fontes	5

4. Resultados

Utilizando-se das configurações de simulação descritas na tabela 2, testou-se o resultado obtido pelo algoritmo implementado baseado nas relações sinal/ruído (SNR). Estas métricas consideram que a fonte estimada da mistura (s_j) é igual à fonte original (s_{salvo}) adicionada às interferências (e_{inter}) causadas por outras fontes e erros de quantização ou artefatos (e_{artef}) no processo de aquisição/separação. Inclui-se, ainda, uma parcela e_{noise} referente ao ruído contido na mistura original [Vincent et al. 2006]. A representação matemática da fonte estimada da mistura é apresentada na Equação (12):

$$s_j = s_{salvo} + e_{inter} + e_{artef} + e_{noise} \quad (12)$$

As métricas propostas por [Vincent et al. 2006] são apresentadas a seguir:

- Razão Fonte-Distorção (SDR – *Source to Distortio Ratio*): Quantifica a qualidade da separação de uma forma geral, com base na distorção existente entre a fonte original e a fonte estimada. A Equação (13) apresenta a expressão utilizada para o cálculo da SDR

$$SDR_{dB} = 10 \log \left(\frac{\|s_{salvo}\|^2}{\|e_{inter} + e_{artef} + e_{noise}\|^2} \right) \quad (13)$$

- Razão Fonte-Interferência (SIR – *Source to Interference Ratio*): Mensura a interferência que a fonte sofreu das demais fontes presentes na mistura, dada pela expressão (14),

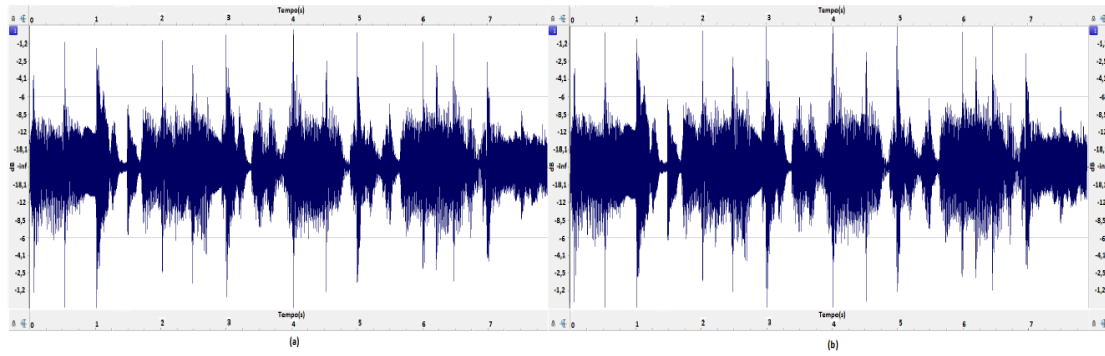


Figura 1. Mistura Considerada: a - Real; b – Estimada

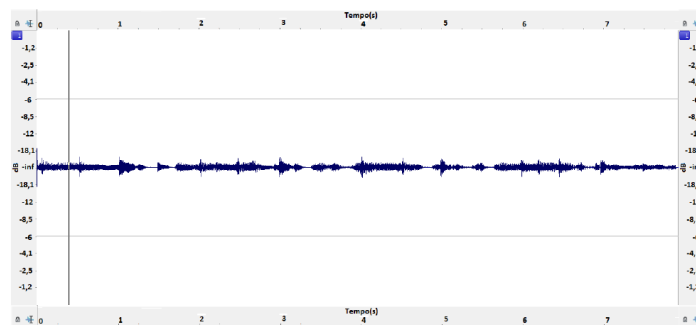


Figura 2. Erro do Algoritmo

$$SIR_{dB} = 10 \log \left(\frac{\|s_{alvo}\|^2}{\|e_{interf}\|^2} \right) \quad (14)$$

- Razão Fontes-Artefatos (SAR – *Source to Artifacts Ratio*): Essa razão visa medir a quantidade de imperfeições geradas nas fases de aquisição/processamento e pode ser avaliada por meio da Equação (15).

$$SAR_{dB} = 10 \log \left(\frac{\|s_{alvo} + e_{interf} + e_{noise}\|^2}{\|e_{artef}\|^2} \right) \quad (15)$$

A fig. 1-a mostra a forma de onda da mistura considerada, enquanto a fig. 1-b mostra a forma de onda da mistura estimada. A fig. 2 mostra a diferença de amplitude entre a mistura considerada e a mistura estimada, via divergência de Itakura-Saito, evidenciando a pequena diferença entre as duas.

As figuras 3 e 4-a mostram os espectrogramas da mistura considerada e da guitarra presente na mistura, respectivamente, enquanto a fig. 4-b mostra o espectrograma da guitarra separada via algoritmo implementado utilizando a divergência de Itakura-Saito.

Os resultados obtidos para as Divergências de Kullback-Leibler e Itakura-Saito são apresentados na tabela 3. Nota-se que, de maneira geral, não é possível afirmar qual das divergências apresentou melhor resultado para a mistura considerada, uma vez que houve alternância de acordo com o instrumento considerado. A divergência de Kullback-Leibler apresentou melhores resultados considerando guitarra, prato de condução e bumbo enquanto que a divergência de Itakura-Saito separou melhor considerando baixo e caixa, fontes de menor potência espectral.

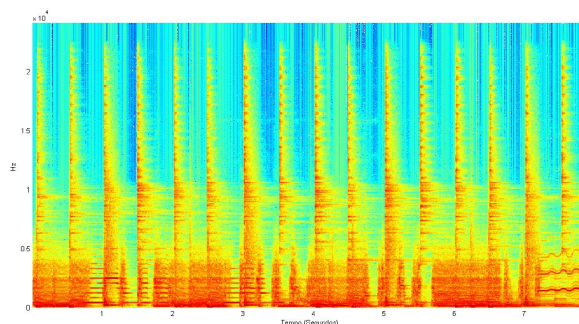


Figura 3. Espectrograma da Mistura

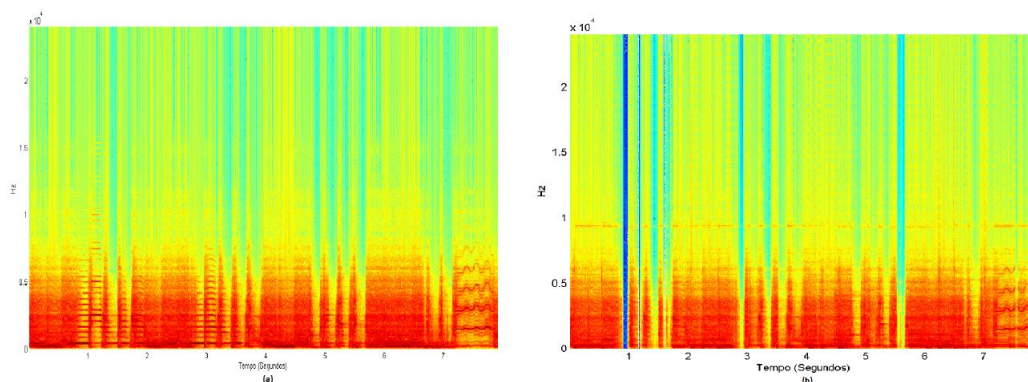


Figura 4. Espectrograma da Guitarra: a - Real; b – Estimado

Tabela 3. Resultado da Separação via NMF

Fonte	SDR(dB)		SIR(dB)		SAR(dB)	
	KL	IS	KL	IS	KL	IS
Guitarra	5,918	4,999	19,994	18,393	6,134	5,265
Baixo	-19,186	-6,543	-18,043	-4,421	5,280	3,346
Condução	5,944	-0,007	12,898	5,995	7,139	2,142
Bumbo	-0,007	-7,030	1,419	-6,092	7,878	7,131
Caixa	-8,839	-5,013	-7,543	-1,084	4,970	0,825

5. Conclusão

A técnica de fatoração de Matrizes Não-Negativas, quando utilizada para a separação cega de fontes (BSS - Blind Source Separation) pode, dependendo da finalidade, ser considerada bem sucedida. Apesar dos resultados obtidos a partir dos critérios objetivos apresentarem, na maioria dos testes, valores próximos à 0dB (relação 1 para 1) ou até mesmo negativos, vale enfatizar que a mistura escolhida para teste trata-se de uma música próxima a músicas comerciais.

Como trabalhos futuros pretende-se explorar a extensão para sinais estéreo, via NTF, além da aplicação de restrições aos métodos NMF/NTF, como por exemplo a esparsidade e smoothness. Assim, trabalha-se com um formato de áudio padrão (estéreo) e com potencial de melhoria de resultados (através das restrições).

Referências

Berry, M. W., Browne, M., Langville, A. N., Puaça, V. P., e Plemmons, R. J. (2006). Algorithms and applications for approximate nonnegative matrix factorization. In Computational Statistics and Data Analysis, pages 155-173.

- Chchocki, A., Zdunek, R., Phan, A. H., and Amari, S.-I. (2009). Nonnegative matrix and tensor factorizations: applications to exploratory multiway data analysis and blindsource separation. John Wiley & Sons, Ltd, Chichester, Sussex do Oeste, Inglaterra, 1 edition.
- Chen, Y., Bao, H., and He, X. (2011). Non-negative local coordinate factorization for image representation. pages 569–574.
- Guan, N., Tao, D., Luo, Z., and Yuan, B. (2012). Nnmf: An optimal gradient method for nonnegative matrix factorization. *Signal Processing, IEEE Transactions on*, 60(6):2882–2898.
- Kitamura, D., Saruwatari, H., Yagi, K., Shikano, K., Takahashi, Y., and Kondo, K. (2013). Robust music signal separation based on supervised nonnegative matrix factorization with prevention of basis sharing. In *Signal Processing and Information Technology (ISSPIT), 2013 IEEE International Symposium on*, pages 000392–000397.
- Krömer, P., Platos, J., and Snasel, V. (2010). Data mining using nmf and generalized matrix inverse. In *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on*, pages 409–414.
- Lee, D. D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In *In NIPS*, pages 556–562. MIT Press.
- Lin, C.-J. (2007). On the convergence of multiplicative update algorithms for nonnegative matrix factorization. *Neural Networks, IEEE Transactions on*, 18(6):1589–1596.
- Mahmoud, A., Ammar, R., Eladawy, M., and Hussien, M. (2009). Improving the performance of the instantaneous blind audio source separation algorithms. In *Signal Processing and Information Technology (ISSPIT), 2009 IEEE International Symposium on*, pages 519–526.
- Santos, M. B. (2015). Aplicação e análise de desempenho da técnica de fatoração de matrizes não-negativas para a separação de fontes acústicas percussivas. Master's thesis, PPGEE/PUCMG, Minas Gerais, Brasil.
- Sawada, H., Kameoka, H., Araki, S., and Ueda, N. (2012). Efficient algorithms for multichannel extensions of itakura-saito nonnegative matrix factorization. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 261–264.
- Tygel, A. F. (2009). Métodos de fatoração de matrizes não-negativas para separação de sinais musicais. Master's thesis, COPPE/UFRJ, Rio de Janeiro, Brasil.
- Vincent, E., Gribonval, R., and Fevotte, C. (2006). Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4):1462–1469.
- Wang, Y.-X. and Zhang, Y.-J. (2013). Nonnegative matrix factorization: A comprehensive review. *Knowledge and Data Engineering, IEEE Transactions on*, 25(6):1336–1353.