Estudo da NMF auxiliada por partitura aplicada à separação de sons de instrumentos musicais

André O. Françani e Magno T. M. Silva

***Resumo*— Recentemente, a fatoração de matrizes não negativas auxiliada por partitura (SI-NMF – *score-informed non-negative matrix factorization*) tem sido utilizada para separação de sons de instrumentos musicais. Considerando essa aplicação, apresenta-se neste artigo uma comparação entre a SI-NMF e dois algoritmos de separação de fontes baseados na análise de componentes independentes (ICA – *independent component analysis*): o infomax (*information-maximization*) e o JADE (*joint approximate diagonalization of eigenmatrices*).**

***Palavras-Chave*— Separação cega de fontes, análise de componentes independentes, fatoração de matrizes não negativas, processamento de sinais musicais.**

***Abstract*— Recently, score-informed non-negative matrix factorization (SI-NMF) has been used for separation of sounds of musical instruments. Considering this application, we compare in this paper the SI-NMF method with two blind source separation algorithms based on independent component analysis (ICA): infomax (*information-maximization*) and JADE (*joint approximate diagonalization of eigenmatrices*).**

***Keywords*— Blind source separation, independent component analysis, non-negative matrix factorization, music signal processing.**

I. INTRODUÇÃO

A Fig. 1 ilustra o problema de separação cega de fontes, cujo objetivo é obter o vetor $\mathbf{y} = \mathbf{S}\mathbf{x}$ contendo as estimativas das N_f fontes, agrupadas no vetor $\mathbf{s} = [s_1, \cdots, s_{N_f}]^T$. Para isso, deve-se ajustar a matriz de separação $\mathbf{S}$ utilizando o vetor observável $\mathbf{x} = [x_1, \cdots, x_{N_m}]^T = \mathbf{M}\mathbf{s}$, que contém amostras de N_m misturas geradas pela matriz $\mathbf{M}$ [1], [2].

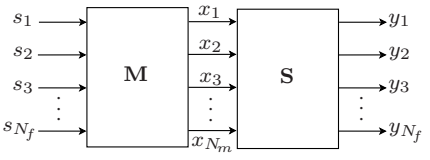

Fig. 1. Diagrama simplificado de separação de fontes.

Um dos métodos mais conhecidos para separação de fontes é o da análise de componentes independentes (ICA), que tem como hipótese a independência das fontes [1], [2]. Diferentes critérios de independência levam a diferentes algoritmos. O *infomax* maximiza a entropia conjunta das fontes utilizando o método de ascensão do gradiente estocástico e uma adequada função não linear $g(\mathbf{y})$ [1], [2]. Em contrapartida, o JADE não necessita da função $g(\mathbf{y})$, pois faz o uso de estatísticas de ordem superior e busca a independência das fontes pela diagonalização conjunta de matrizes de cumulantes [1], [2]. Para que esses algoritmos consigam recuperar as fontes, é necessário que $N_m \geq N_f$, ou seja, o número de sensores deve ser maior ou igual ao número de fontes [2].

André O. Françani e Magno T. M. Silva, , Escola Politécnica, Universidade de São Paulo, São Paulo, SP, Brasil, e-mails: {andre.francani, magno.silva}@usp.br. Este trabalho foi financiado pela FAPESP - (2015/25992-1) e CNPq (304275/2014-0).

Considerando a mistura de sons de instrumentos musicais, foi proposto em 2012 um método de separação baseado na fatoração de matrizes não negativas auxiliada por partitura (SI-NMF – *score-informed non-negative matrix factorization*) [3], [4]. Este artigo tem por objetivo comparar a SI-NMF, descrita na seção seguinte, com o infomax e o JADE para separação de sons de instrumentos musicais. Descrições dos algoritmos de ICA podem ser encontradas, por exemplo, em [1], [2].

II. A TÉCNICA SI-NMF

Dada uma matriz $\mathbf{V} \in \mathbb{R}_+^{M\times N}$ com elementos não negativos, a NMF busca encontrar a decomposição $\mathbf{V} \approx \hat{\mathbf{V}} = \mathbf{W}\mathbf{H}$, em que $\mathbf{W} \in \mathbb{R}_+^{M\times K}$ e $\mathbf{H} \in \mathbb{R}_+^{K\times N}$, ambas de elementos não negativos, que melhor aproxima $\mathbf{W}\mathbf{H}$ a $\mathbf{V}$. O algoritmo se desenvolve a partir da minimização de uma medida de distância, como por exemplo, a divergência de Kullback-Leibler definida como

$$D_{\text{KL}} = \sum_{m=1}^{M}\sum_{n=1}^{N}\left(\mathbf{V}_{m,n}\ln\frac{\mathbf{V}_{m,n}}{\hat{\mathbf{V}}_{m,n}} - \mathbf{V}_{m,n} + \hat{\mathbf{V}}_{m,n}\right), \quad (1)$$

em que $\mathbf{V}_{m,n}$ representa o elemento (m,n) da matriz $\mathbf{V}$. Para garantir que os elementos de $\mathbf{W}$ e $\mathbf{H}$ sejam sempre não negativos, considera-se o método do gradiente descendente com um passo de adaptação específico na minimização de D_{KL}, o que leva às seguintes regras de atualização multiplicativas [4]

$$\mathbf{W} \leftarrow \mathbf{W} \odot \frac{\left(\frac{\mathbf{V}}{\hat{\mathbf{V}}}\right)\mathbf{H}^T}{\mathbf{J}\mathbf{H}^T + \varepsilon\mathbf{L}_w} \quad \text{e} \quad \mathbf{H} \leftarrow \mathbf{H} \odot \frac{\mathbf{W}^T\left(\frac{\mathbf{V}}{\hat{\mathbf{V}}}\right)}{\mathbf{W}^T\mathbf{J} + \varepsilon\mathbf{L}_h}, \quad (2)$$

em que $\odot$ denota a multiplicação elemento a elemento de matrizes, as divisões que aparecem em (2) também são feitas elemento a elemento, ε é uma constante utilizada para evitar a divisão por zero, $\mathbf{J}$, $\mathbf{L}_w$ e $\mathbf{L}_h$ são matrizes com elementos iguais a um e possuem dimensões $M \times N$, $M \times K$ e $K \times N$, respectivamente. Em geral, $\mathbf{W}$ e $\mathbf{H}$ são inicializadas de maneira aleatória com elementos não negativos.

Quando se aplica a NMF à separação de sinais musicais, é comum utilizar restrições nas inicializações de $\mathbf{W}$ e $\mathbf{H}$, provenientes da partitura. A NMF que considera essas restrições é denotada aqui por SI-NMF (*score-informed*-NMF). A partitura é representada pelo arquivo MIDI (*musical instrument digital interface*) da música, que, por sua vez, contém as frequências fundamentais das notas e os instantes de tempo em que elas são executadas. Na SI-NMF, a matriz $\mathbf{V}$ é inicializada com o espectrograma de magnitude da música (mistura), calculado com a transformada de Fourier de curto prazo (STFT – *short-time fourier transform*). A matriz $\mathbf{W}$ é inicializada de maneira harmônica, isto é, o elemento (m,k) dessa matriz é

inicializado como $\mathbf{W}_{m,k} = \sum_{n=1}^{N_h} \varphi\left(m - nf_0^k\right)$, em que $\varphi(\cdot)$ corresponde ao espectro de magnitude da janela de análise, f_0^k é a frequência fundamental da coluna k e N_h o número de harmônicos da frequência fundamental. A matriz $\mathbf{H}$ é inicializada com informações temporais, ou seja, se atribui valor um no intervalo de tempo em que a nota é tocada e zero quando ela não é tocada [4]. O esquema geral da SI-NMF é mostrado na Figura 2. Observa-se que é necessário conhecer o arquivo MIDI dos instrumentos da mistura, pois suas informações são usadas para inicializar as matrizes $\mathbf{W}_{i,0}$ e $\mathbf{H}_{i,0}$ e gerar os sinais sintetizados de cada instrumento i separadamente. O espectrograma de magnitude do som sintetizado do instrumento i é usado como matriz $\mathbf{V}_{i,\text{sint}}$ na fase de aprendizagem. Utilizando essas inicializações, a NMF é executada, obtendo-se as matrizes $\mathbf{W}_{i,1}$ e $\mathbf{H}_{i,1}$ de cada instrumento. Essas matrizes são então concatenadas e utilizadas na fase de separação. Nessa fase, a matriz $\mathbf{V}$ corresponde ao espectro de magnitude da mistura. Executando-se novamente a NMF, as matrizes $\mathbf{W}$ e $\mathbf{H}$ são obtidas. Em seguida, calcula-se a máscara de Wiener de cada instrumento i dada por $\mathbf{M}_i = \mathbf{W}_i\mathbf{H}_i/(\mathbf{WH})$ em que a divisão é feita elemento a elemento e as matrizes $\mathbf{W}_i$ e $\mathbf{H}_i$ são extraídas das matrizes $\mathbf{W}$ e $\mathbf{H}$ a partir de informações do arquivo MIDI. Calcula-se $\mathbf{Y}_i = \mathbf{M}_i\mathbf{X}$ para cada instrumento, em que $\mathbf{X}$ é o espectrograma complexo da mistura. Por fim, calculando-se a STFT inversa de $\mathbf{Y}_i$, consegue-se obter os sinais y_i de cada instrumento no tempo [4].

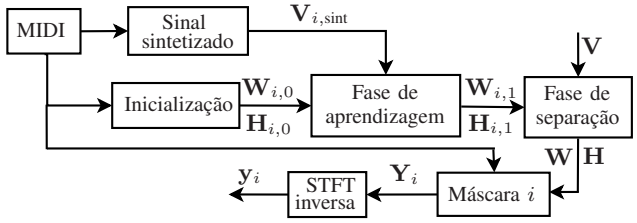

Fig. 2. SI-NMF para recuperação do som do instrumento i.

III. Simulações

A avaliação dos métodos de separação foi feita por meio da razão fonte/distorção (*source-to-distortion ratio* – SDR) definida como $\text{SDR} = 10\log_{10}(\|s_{\text{alvo}}\|^2/\|y - s_{\text{alvo}}\|^2)$, em que $\|\cdot\|$ corresponde à norma Euclidiana e s_{alvo} e y representam o sinal a ser estimado e o sinal estimado, respectivamente [2].

A base de dados utilizada na comparação é um arranjo para um quinteto de sopros de madeira, composto originalmente por Ludwig Von Beethoven para um quarteto de cordas (op.18 n.5, III. Andante Cantabile, var. V.) [4]. A mistura (música) considerada se encontra em perfeito sincronismo com o arquivo MIDI correspondente. O espectrograma do sinal foi obtido por meio de uma janela Hanning de 4096 pontos (93ms), frequência de amostragem $f_a = 44{,}1$ kHz e $87{,}5\%$ de sobreposição. A inicialização harmônica da matriz $\mathbf{W}$ teve como parâmetro $N_h = 50$ e a inicialização temporal da matriz $\mathbf{H}$ teve como parâmetros $t_i = t_f = 100$ ms (ver [4] para mais detalhes sobre esses parâmetros). Utilizaram-se também o JADE e o infomax [função $g(\mathbf{y})$ sigmoidal e passo de adaptação $\mu = 0{,}001$].

Os valores de SDR se encontram na Tabela I, cujos melhores resultados estão marcados em negrito. Os valores sublinhados na tabela correspondem ao melhor desempenho utilizando-se uma análise subjetiva realizada com 29 pessoas, que não tinham envolvimento com áudio e/ou música. Nessa análise, os voluntários ouviam o sinal original e depois escolhiam o áudio que julgavam mais próximo do mesmo sem saber o método utilizado na separação. Analisando a Tabela I, é possível concluir que, nem sempre o maior valor de SDR coincide com o melhor resultado da análise subjetiva, já que a SI-NMF apresenta o melhor desempenho para a flauta e a trompa e o algoritmo JADE para o fagote e o oboé. Observando os sinais estimados no tempo, nota-se que há uma distorção quando comparados aos sinais originais. Esse fato será melhor investigado em um trabalho futuro. O formulário utilizado na análise subjetiva, os sinais originais e os separados estão disponíveis em [5]. Cabe observar que há diversos fatores que interferem na estimativa da SI-NMF, tais como a adequada escolha de parâmetros nas operações STFT e STFT inversa, a correta escolha do número de iterações na NMF para que a aproximação seja boa o suficiente e, principalmente, a sincronização entre a partitura e os sinais [4].

TABELA I
VALORES DE SDR OBTIDOS COM OS ALGORITMOS DE SEPARAÇÃO.

Instrumento	SDR (dB) infomax	SDR (dB) JADE	SDR (dB) SI-NMF
fagote	**39,338**	34,252	11,558
clarinete	26,094	**37,240**	13,168
flauta	26,575	**31,619**	15,986
trompa	**35,569**	31,487	10,657
oboé	**37,337**	33,799	5,5981

IV. Conclusões

Deste estudo, observa-se que a SI-NMF depende de vários fatores para estimar as fontes: (1) a disponibilidade do arquivo MIDI, (2) a necessidade de um sintetizador para sintetizar os sons a partir das informações da partitura, (3) a adequada escolha dos parâmetros da STFT e (4) a sincronização entre a partitura e os sinais sintetizados e gravados. O item (4) é essencial para se obter um bom resultado de separação, o que exige técnicas adicionais de sincronismo quando isso não ocorre [3]. Em contrapartida, o *infomax* necessita apenas do conhecimento prévio das distribuições de probabilidade das fontes para a correta escolha da função não linear $g(\mathbf{y})$ e o JADE não necessita desse conhecimento prévio. Pelos resultados de simulação, observa-se que os algoritmos de ICA podem apresentar um desempenho superior à SI-NMF com um custo computacional da mesma ordem de grandeza. Diante disso, cabe questionar a vantagem de se usar a SI-NMF dados o desempenho inferior em alguns casos e a quantidade de informações necessárias para realizar a separação.

Referências

[1] A. Hyvärinen, J. Karhunen, and E. Oja, "*Independent Component Analysis*", John Wiley & Sons, Inc., 2001.

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

[3] E. Vincent, N. Bertin, R. Gribonval, and F. Bimbot, "From blind to guided audio source separation: How models and side information can improve the separation of sound," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 107–115, May 2014.

[4] J. Fritsch, "*High quality musical audio source separation*", M. S. thesis, UPMC/IRCAM/TELECOM Paris-Tech, Paris, França, 2012.

[5] Resultado da simulação com o arranjo para um quinteto de sopros de madeira: www.lps.usp.br/magno/matlab/francani/quinteto.htm. Página web acessada em 12/04/2017.