SEPARAÇÃO DE FONTES SONORAS POR FATORAÇÃO DUPLAMENTE DECONVOLUTIVA DE MATRIZES NÃO-NEGATIVAS COM USO DE RESTRIÇÕES

Renan Mariano Almeida

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Luiz Wagner Pereira Biscainho

Rio de Janeiro
Setembro de 2014

SEPARAÇÃO DE FONTES SONORAS POR FATORAÇÃO DUPLAMENTE DECONVOLUTIVA DE MATRIZES NÃO-NEGATIVAS COM USO DE RESTRIÇÕES

Renan Mariano Almeida

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

Prof. Wallace Alves Martins, D.Sc.

Prof. Tadeu Nagashima Ferreira, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
SETEMBRO DE 2014

Almeida, Renan Mariano

Separação de Fontes Sonoras por Fatoração Duplamente Deconvolutiva de Matrizes Não-Negativas com Uso de Restrições/Renan Mariano Almeida. – Rio de Janeiro: UFRJ/COPPE, 2014.

XI, 89 p.: il.; 29, 7cm.

Orientador: Luiz Wagner Pereira Biscainho

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2014.

Referências Bibliográficas: p. 69 – 72.

1. processamento digital de áudio. 2. separação de fontes sonoras. 3. separação de matrizes não-negativas. I. Biscainho, Luiz Wagner Pereira. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Dedico este trabalho a Deus, aos meus pais e aos meus amigos. Sem a luz divina a me iluminar e sem o apoio de pessoas próximas de mim e que me amam, eu jamais conseguiria nada nesta vida.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

SEPARAÇÃO DE FONTES SONORAS POR FATORAÇÃO DUPLAMENTE DECONVOLUTIVA DE MATRIZES NÃO-NEGATIVAS COM USO DE RESTRIÇÕES

Renan Mariano Almeida

Setembro/2014

Orientador: Luiz Wagner Pereira Biscainho

Programa: Engenharia Elétrica

Neste trabalho aborda-se o problema da separação de fontes sonoras. Primeiramente, apresentam-se em ordem cronológica as principais soluções encontradas na literatura científica para solucioná-lo. Em particular, detalha-se o método conhecido como NMF (do inglês, *Non-Negative Matrix Factorization*), seguido dos algoritmos dele derivados, tais como NMF2D (do inglês, *Non-Negative Matrix Factor 2-D Deconvolution*) e SNMF2D (do inglês, *Sparse Non-Negative Matrix Factor 2-D Deconvolution*), que são o estado-da-arte dos métodos para separação de sinais musicais. Essas modificações tornam a NMF capaz de separar toda a informação proveniente de um dado instrumento musical; em particular a SNMF2D inclui um critério de restrição que condiciona o método a realizar uma fatoração esparsa no tempo. A partir daí, o trabalho concentra-se em incluir e testar novos critérios de restrição (e combinações deles) na NMF2D, dando origem a um algoritmo genérico chamado de CNMF2D (do inglês, *Constrained Non-Negative Matrix Factor 2-D Deconvolution*). São realizados testes de diversas versões da CNMF2D sobre misturas sintéticas e naturais, cujos resultados podem sugerir boas escolhas conforme os tipos de sinal presentes na mistura.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

SOUND SOURCE SEPARATION BY NON-NEGATIVE MATRIX FACTORIZATION 2-D DECONVOLUTION USING CONSTRAINTS

Renan Mariano Almeida

September/2014

Advisor: Luiz Wagner Pereira Biscainho

Department: Electrical Engineering

This work deals with sound source separation. Firstly, the main proposals found in the literature to solve this problem are reviewed in cronological order. In particular the NMF (Non-Negative Matrix Factorization) is detailed, followed by derived algorithms such as the NMF2D (Non-Negative Matrix Factor 2-D Deconvolution) and the SNMF2D (Sparse Non-Negative Matrix Factor 2-D Deconvolution), which are the state-of-the-art in separation of musical signals. These modifications enable the NMF to seggregate all information associated with a given musical instrument; in particular the SNMF2D includes a sparsity restriction that conditions the method to perform a temporally sparse factorization. Proceeding from this background, the work focuses on the inclusion and assessment of new restriction criteria (and respective combinations) into the NMF2D, thus producing a generic algorithm called CNMF2D (Constrained Non-Negative Matrix Factor 2-D Deconvolution). Several versions of the CNMF2D are tested over synthetic as well as natural mixtures; the corresponding results can suggest best choices according to the characteristics of the mixed signals.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

1.1 Motivação

O ser humano está exposto a uma série de estímulos sonoros simultâneos o tempo inteiro. Para o cérebro, é fácil distinguir a fonte sonora (emissora) de cada um desses sinais. Por exemplo, a voz de uma pessoa e o canto de um pássaro emitidos ao mesmo tempo em um ambiente, apesar de chegarem juntos às orelhas, são devidamente entendidos pelo cérebro como duas fontes sonoras distintas, permitindo a identificação de cada emissor.

Um problema mais clássico é o da festa de coquetel (do inglês, *Cocktail Party*) [1], em que, mesmo com várias pessoas falando ao mesmo tempo, é possível para o ser humano se ater ao que está sendo dito por determinada pessoa, desconsiderando as outras vozes.

Isso nos parece óbvio, já que o processamento cerebral da audição (assim como o dos outros sentidos) se dá de forma automática e perceptivamente instantânea. Além do mais, o ser humano está acostumado com ele desde quando começa a viver, o que significa que, ao longo da vida, o cérebro foi (e continua sendo) treinado para distinguir os mais variados tipos de fontes sonoras.

Entretanto, deve-se ter em mente que um computador é incapaz de realizar qualquer tipo de processamento por si só. Dessa forma, para aplicações computacionais, faz-se necessário o desenvolvimento de métodos (algoritmos) que realizem a tarefa de separação de fontes sonoras contidas em um sinal. Este sinal que contém (é a soma de) as contribuições de todas as fontes sonoras de interesse será chamado de **sinal de mistura**.

1.2 Aplicações

A separação de fontes encontra diversas aplicações dentro de muitas áreas do conhecimento, como, por exemplo, na área de biomédica, em que sinais cerebrais precisam ser devidamente separados e identificados, já que são misturas de estímulos que podem ser provenientes de qualquer parte do corpo [2].

Também são encontradas aplicações em telecomunicações, como por exemplo, na separação entre um sinal de interesse e interferências de outros sinais em um método de Acesso Múltiplo por Divisão de Código (do inglês, *Code-Division Multiple Acess*, CDMA), utilizado em telefonia celular e rastreamento via satélite (GPS).

Dentro da área de processamento de sinais, a separação de fontes também se aplica na redução de ruído em imagens, áudio e fala. Especificamente no processamento de sinais de áudio, as principais aplicações são [3]: (1) codificação de áudio, em que separar as fontes sonoras contidas em um sinal de mistura para depois codificar cada uma delas separadamente pode melhorar a eficiência e a flexibilidade geral do processo; (2) reconhecimento (classificação) de áudio, onde a separação de fontes é inerente ao sistema, já que primeiramente é necessário obter-se as fontes individuais para depois associá-las aos padrões contidos no banco de dados da aplicação; (3) manipulação de áudio, em que ter cada uma das fontes separadas permite que elas possam ser subtraídas da mistura, individualmente modificadas, misturadas de formas diferentes etc.

1.3 Delimitação

Os objetos de estudo deste trabalho são as misturas de sinais contendo áudio. Considera-se sinal de áudio todo aquele que é audível pelo ser humano, ou seja, situado em uma faixa de frequência entre 20 Hz e 20 kHz. Os sinais musicais (provenientes de instrumentos musicais e voz cantada) e os sinais de fala podem ser considerados sinais de áudio. Como os sinais de fala costumam ser tratados e classificados separadamente devido a suas características e aplicações próprias, esta dissertação atém-se principalmente aos sinais musicais.

1.4 Metodologia estudada

O método de separação de fontes sonoras explorado neste trabalho é a Fatoração de Matrizes Não-Negativas (do inglês, *Non-Negative Matrix Factorization*, NMF), descrita pela primeira vez em [4] para o processamento de imagens. Essa técnica

tem por objetivo fatorar a matriz de espectrograma de um sinal de mistura em duas outras matrizes: uma contendo os vetores de padrões espectrais que constituem o sinal, e a outra contendo a informação de ativação temporal desses padrões.

Em [5] [6], é utilizado o termo *Non-Negative Matrix Approximation* (NNMA) para se referir a esse método. Porém, o termo mais consagrado na literatura científica é NMF. Nesse método, a matriz que é fatorada (o espectrograma) é, por definição, não-negativa, e as matrizes resultantes da fatoração também são não-negativas. Portanto, a tradução para português mais adequada para NMF seria Fatoração de Matrizes Não-Negativas em Matrizes Não-Negativas [7].

1.5 Objetivo

Este trabalho tem por objetivo revisar teoricamente diversas versões da literatura da NMF e propor novas contribuições para o método. São mostrados experimentos de separação de fontes sonoras que permitem verificar as vantagens e desvantagens de cada variação do método, bem como o impacto das contribuições propostas.

1.6 Sistema de separação por NMF

De forma genérica, o sistema de separação de fontes sonoras por NMF pode ser entendido como sendo constituído por cinco etapas, em que cada etapa corresponde a um bloco de processamento, como mostrado na Figura 1.1.

O primeiro bloco corresponde à etapa de Análise do sinal de mistura. Nesta, o sinal é transformado do domínio do tempo para o domínio tempo-frequencial, obtendo-se o seu espectrograma e evidenciando-se, assim, a evolução de suas características espectrais ao longo do tempo. A informação de fase contida no sinal de mistura é descartada nessa etapa.

O segundo bloco consiste na Fatoração propriamente dita, onde é utilizado como entrada o espectrograma do sinal de mistura obtido na etapa anterior. Os fatores obtidos serão utilizados na reconstrução dos espectrogramas de cada fonte.

O terceiro bloco corresponde à etapa de Processamento, já sobre os espectrogramas de cada fonte, cuja finalidade é realçar as características mais evidentes de cada um deles através de uma filtragem.

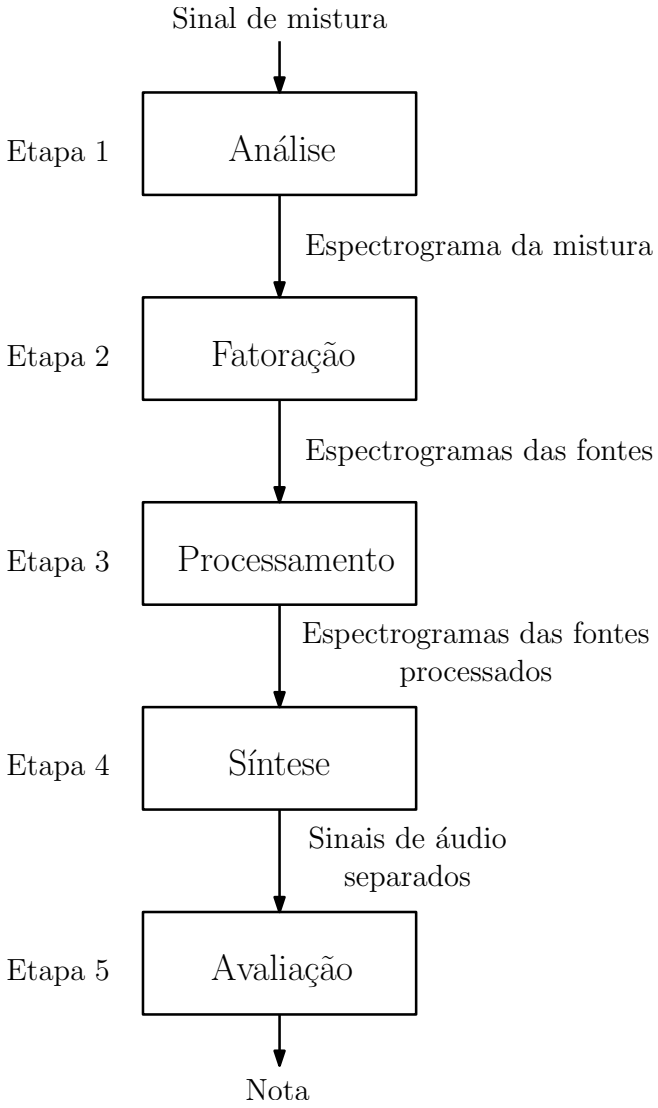

Figura 1.1: Diagrama em blocos de um sistema completo de separação de fontes sonoras por NMF.

O quarto bloco refere-se à etapa de Síntese (estimação) de fase para cada espectrograma (lembrando que a informação de fase é perdida na segunda etapa e, portanto, não é conhecida a fase para os espectrogramas obtidos na terceita etapa), necessária para que se obtenham as fontes separadas no domínio do tempo.

Por fim, o quinto bloco refere-se à etapa de Avaliação dos sinais obtidos no domínio do tempo, realizada através da comparação entre eles e seus respectivos sinais de referência (antes da mistura), quando houver, ou através de critérios subjetivos.

1.7 Descrição dos capítulos

Com base no sistema descrito na seção anterior, esta dissertação está subdividida em 7 capítulos, sendo que este capítulo apresentou a introdução ao trabalho.

No Capítulo 2 é descrita a primeira etapa do sistema de separação, a Análise tempo-frequencial. Além disso, apresenta-se também uma breve discussão sobre os principais métodos de separação encontrados na literatura científica, com exceção da NMF.

No capítulo 3 é discutida em detalhes a etapa de Fatoração por NMF, de forma a conduzir o leitor desde sua forma básica ao estado-da-arte do método.

No Capítulo 4 são apresentadas algumas contribuições para a família de métodos baseados na NMF.

No Capítulo 5 descrevem-se, de forma sucinta, as etapas de Processamento, Síntese e Avaliação da separação.

No Capítulo 6 são descritos os experimentos realizados e apresentam-se os resultados obtidos a partir de tais experimentos, bem como as conclusões que podem ser obtidas de cada um.

Por fim, no Capítulo 7 é feita a conclusão geral do trabalho e a apresentação da perspectiva para trabalhos futuros.

Capítulo 2

Análise da mistura e separação de fontes sonoras

2.1 Análise da mistura

No processamento digital de sinais, há duas formas básicas de representação de um sinal: no domínio do tempo e no domínio da frequência. Enquanto a representação no domínio do tempo permite a observação do comportamento do sinal ao longo do tempo, a representação no domínio da frequência permite a verificação da intensidade e da fase com que cada componente frequencial contribui na sua composição.

Combinando-se as representações no domínio do tempo e da frequência, obtém-se a representação em tempo-frequência. Essa análise é de suma importância na separação de fontes, já que estas podem estar dissociadas em qualquer dos dois domínios. A Análise da mistura é a primeira etapa de um sistema completo de separação. A ferramenta responsável pela obtenção do chamado espectrograma do sinal de mistura a partir de sua representação no domínio do tempo é a versão discreta da **Transformada de Fourier de Curta-Duração** (do inglês, *Short-Time Fourier Transform*, STFT) [8].

Na STFT, o sinal discreto $x(l)$, de tamanho L, representado no domínio do tempo, é dividido em trechos através de sua multiplicação por uma função $w(l)$ chamada de **janela**. Sobre cada trecho, é aplicada a Transformada de Fourier, obtendo-se assim a característica de frequência do sinal para cada intervalo de tempo. A equação da STFT é dada por

$$X_{k,m} = \sum_{l=0}^{L-1} x(l)w(l - mS)e^{\frac{-j2\pi kl}{L}}, \tag{2.1}$$

onde mS é o deslocamento da janela w de forma a abranger todos os trechos de $x(l)$, sendo S o passo de análise e m o índice do segmento considerado. O produto $x(l)w(l-mS)$ para cada valor de m isola um segmento diferente de $x(l)$, modificado pela janela w. A variável k é o índice da frequência discretizada.

Para todos os valores de k e m, considera-se a STFT do sinal $x(l)$ como $\mathbf{X}_e$, uma matriz de elementos $x_{k,m}$. A partir dessa forma matricial, pode-se obter o valor absoluto ou o quadrado do valor absoluto de $\mathbf{X}_e$, caracterizando o espectrograma de magnitude $|\mathbf{X}_e|$ e o espectrograma de potência $|\mathbf{X}_e|^2$, respectivamente. Na presente aplicação, o espectrograma de magnitude é o mais utilizado [3], e, neste trabalho, utiliza-se a seguinte notação:

$$\mathbf{X} = |\mathbf{X}_e|\,, \tag{2.2}$$

ou seja, chama-se de $\mathbf{X}$ o espectrograma.

É importante ressaltar alguns aspectos da etapa de análise tempo-frequencial da mistura. A janela utilizada para segmentar o sinal de mistura e o seu tamanho devem ser escolhidos com cuidado. Uma janela de duração pequena (com poucos pontos) fornece alta resolução no tempo, já que isso significa elevado número de janelas sendo utilizadas para abranger todo o sinal, sendo cada uma delas responsável por um pequeno trecho. Assim, quanto maior for o número de janelas utilizadas na análise, maior é o número de pontos no eixo do espectrograma. Em contrapartida, a janela de duração pequena fornece baixa resolução na frequência. Isso acontece porque na Transformada de Fourier, quanto maior a duração do sinal temporal dado como entrada, maior é a resolução frequencial obtida. Logo, se os quadros forem muito pequenos, a resolução do espectrograma para o eixo das frequências será baixo. A duração da janela é, portanto, um parâmetro importante desta etapa.

Deve-se ter em mente que o sinal sempre sofrerá alterações quando multiplicado por uma janela. Somente uma janela cuja representação no domínio da frequência fosse um impulso não distorceria o espectro do sinal; mas esta corresponderia a uma janela plana de tamanho infinito no domínio do tempo, o que, em outras palavras, seria o mesmo que não realizar nenhum tipo de segmentação.

Sabe-se que a janela retangular, apesar de parecer a opção mais intuitiva e possuir o lobo principal mais estreito dentre todos os tipos de janela, possui lobos secundários muito elevados quando representada no domínio da frequência. Isso significa distanciamento da característica impulsiva desejada, comprometendo as características espectrais do sinal de mistura. Um tipo de janela mais adequado deve possuir as

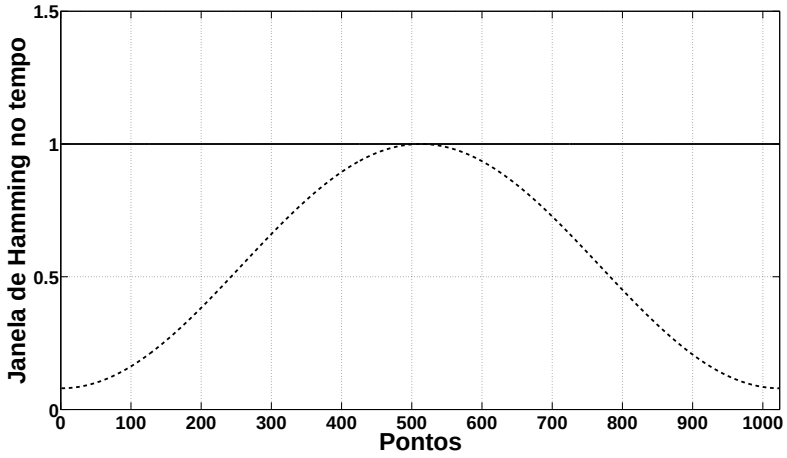

Figura 2.1: Em linha contínua, janela retangular; em linha tracejada, janela de Hamming. Ambas representadas no domínio do tempo com 1024 pontos.

bordas suaves, de forma que os lobos secundários de sua representação na frequência sejam mais atenuados, ao custo de o lobo principal ser mais largo. Portanto, o tipo de janela é também um parâmetro importante da análise tempo-frequencial. Uma janela muito utilizada na análise por STFT é a janela de Hamming. As Figuras 2.1 e 2.2 ilustram essa discussão.

Outro aspecto da análise tempo-frequencial do sinal que merece atenção é decorrente do uso de uma janela de bordas suaves. Enquanto esse tipo de janela reduz os efeitos negativos da segmentação no sinal em termos de frequência, ele produz um efeito negativo no domínio do tempo: os trechos do sinal que estão sob os efeitos das bordas das janelas ficam atenuados. Isso é resolvido sobrepondo-se as janelas de forma que, se um trecho do sinal estiver mal contemplado por uma, ele estará melhor contemplado pelas janelas adjacentes, conforme mostra a Figura 2.3. Logo, o número de pontos (ou passo) de sobreposição é outro parâmetro importante desta análise.

2.2 Separação de fontes sonoras

2.2.1 O conceito de fonte

O termo "fonte" não é preciso quando se lida com sinais musicais [7]. De forma intuitiva, pode-se entender como fonte um instrumento musical. Assim, uma mistura contendo os sons de um violão e de um piano, por exemplo, conteria duas fontes. Porém, um violão possui 6 cordas; logo, cada uma das cordas também poderia

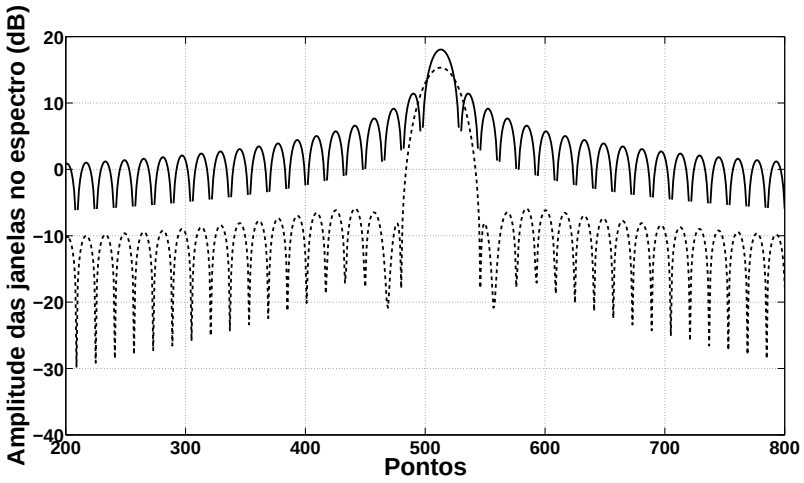

Figura 2.2: Espectro da janela retangular (em linha contínua) e da janela de Hamming (em linha tracejada). Observa-se que os lobos secundários da janela de Hamming são mais atenuados que os da janela retangular. Foram utilizados 1024 pontos para esta representação, porém estão mostrados somente os valores situados entre os pontos 200 e 800.

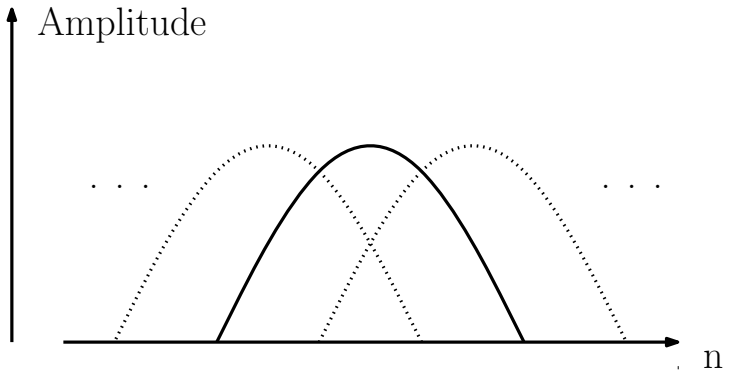

Figura 2.3: As janelas em pontilhado devem sobrepor-se à janela central de forma a compensar as atenuações causadas por ela na amplitude do sinal.

ser considerada como uma fonte distinta. No caso do piano, a partir da batida do martelo (quando se pressiona uma tecla), os sinais correspondentes a diferentes notas musicais também podem ser considerados como provenientes de fontes diferentes. Além disso, em sua maioria, as teclas do piano percutem mais de uma corda. Cada uma dessas cordas poderia ser considerada uma fonte diferente.

Outro caso é o da bateria, que, por ser composta por bumbos, pratos etc, pode ser vista como um instrumento subdividido em outros instrumentos, cada qual sendo uma fonte. Já um naipe de violinos tocando em uníssono tanto pode ser considerado como um conjunto de fontes (cada qual representada por um violino), como uma fonte única. Sendo assim, a definição de fonte depende da aplicação do sistema e, portanto, será explicitado ao longo do trabalho o que está sendo considerado como fonte.

2.2.2 Separação não-supervisionada

A separação de fontes sonoras pode ser realizada através de diversos métodos. São chamados de métodos **supervisionados** aqueles que utilizam informação prévia das fontes como parâmetro de entrada do algoritmo. Dessa forma, o algoritmo é "treinado" adequadamente para encontrar determinados padrões dentro das misturas.

Entretanto, na maioria dos problemas práticos, tem-se disponível pouca informação das fontes individuais. Portanto, novos métodos de separação tiveram que ser desenvolvidas para tentar contornar esse problema, fazendo uso de determinados princípios teóricos [9]. Esses são os chamados métodos **não-supervisionados**. Os algoritmos que serão apresentados ainda nesta seção são técnicas **não-supervisionadas** de separação, pois realizam a separação de fontes sonoras sem que informações prévias das fontes individuais que compõem as misturas sejam utilizadas para alimentar o algoritmo.

2.2.3 Modelo geral

Os métodos não-supervisionados de separação de fontes, em geral, possuem base em um mesmo modelo matemático que descreve o sinal de mistura a ser processado. Nesse modelo, o vetor-coluna de magnitude $\mathbf{x}_t$ do espectrograma $\mathbf{X}$, com $t = 1, 2, ..., T$, é descrito como sendo aproximadamente uma combinação linear de M vetores (ou funções) de base frequencial $\mathbf{b}_m$ [9] [10], ou seja,

$$\mathbf{x}_t \approx \sum_{m=1}^{M} g_{m,t}\mathbf{b}_m, \tag{2.3}$$

em que $g_{m,t}$ é o ganho da m-ésima função de base no quadro t.

Cabe aqui introduzir o conceito de "componente" na separação de fontes. De maneira geral, para os métodos não-supervisionados, é chamada de componente cada parte em que o sinal de mistura é decomposto pelo método de separação. De maneira específica, com base no modelo descrito na equação (2.3), componente é o produto entre um vetor de base $\mathbf{b}_m$ e seus respectivos ganhos $g_{m,t}$, para $t = 0, 1, ..., T$. Logo, em cada quadro t, $\mathbf{x}_t$ é constituído por até M componentes (caso nenhum dos ganhos seja nulo), cada qual referente a uma função de base $\mathbf{b}_m$.

Uma fonte pode corresponder a apenas uma componente, como também pode corresponder a um grupo de componentes. Conforme já foi dito na Subseção 2.2.1, o conceito de fonte vai depender da aplicação.

Também é importante ressaltar que, apesar de o sinal de mistura no domínio do tempo ser modelado simplesmente como uma soma de diversos sinais, isso não significa que a soma entre as magnitudes dos sinais originais correponde exatamente à magnitude do sinal de mistura [3] [9], pois a informação de fase é descartada na STFT. Entretanto, como já dito, o uso do espectrograma de magnitude produz bons resultados.

2.2.4 Principais métodos

Os métodos não-supervisionados de separação de fontes sonoras descritos a seguir estão entre os mais conhecidos da literatura científica e se baseiam no modelo descrito na equação (2.3). Omitiu-se aqui a NMF, que, por sua importância para o trabalho, será abordada separadamente no próximo capítulo.

A **Análise de Componentes Principais** (do inglês *Principal Component Analysis*, PCA) é um método linear de separação que realiza a extração das características de uma mistura com base na descorrelação entre elas [11]. Na PCA, são devolvidas as componentes em ordem decrescente de energia (chamadas de componentes principais), todas descorrelacionadas entre si. Devido à possibilidade de haver descarte das componentes de menor energia, a PCA pode ser vista apenas como uma etapa prévia de redução de dimensão do problema original, e não como método de separação propriamente.

A independência entre variáveis aleatórias é uma restrição mais forte do que a descorrelação entre elas, e em várias aplicações deseja-se obter fontes independentes entre si. Porém, a PCA não garante o atendimento dessa propriedade. Para isso, é

necessária a utilização de uma técnica mais poderosa. A **Análise de Componentes Independentes** (do inglês, *Independent Component Analysis*, ICA) é um método estatístico e não-linear de separação que, a partir de um conjunto de misturas, entrega as fontes (maximamente) independentes entre si. A ICA é uma técnica amplamente conhecida na literatura científica, sendo utilizada como método em diversas aplicações para a separação de fontes. Um estudo abrangente da ICA pode ser encontrado em [12].

O modelo básico da ICA possui duas grandes limitações. Uma delas é que a ICA só pode ser aplicada quando o número de misturas disponíveis é maior que ou igual ao número de fontes em questão. E isso é muito pouco prático na maioria das situações, em que só se dispõe de um único sinal de mistura (apenas um canal, caracterizando o que se chama de sinal *monaural*, ou simplesmente *mono*). A outra limitação é que a ICA lida apenas com misturas *instantâneas*, sendo que grande parte das misturas são naturalmente *convolutivas*, ou seja, são produzidas sob efeito de atrasos de propagação e reverberação do ambiente.

Com o objetivo de contornar a limitação quanto ao número mínimo de misturas do modelo básico da ICA, em [13] propõe-se a utilização da **Análise de Subespaços Independentes** (do inglês *Independent Subspace Analysis*, ISA). A ISA pode ser vista como uma generalização da ICA. Nela, é aplicada a STFT sobre o sinal de mistura, resultando no seu espectrograma. A ISA faz uso do espectrograma de um único sinal de mistura para encontrar as fontes que o compõem. Dentro da ISA, a PCA e a ICA são passos do algoritmo. Após a redução da dimensão do espectrograma da mistura por PCA, os quadros (ou colunas) do espectrograma reduzido são passados como parâmetros de entrada para a ICA, que os entende como se fossem misturas diferentes. A ICA então devolve as bases frequenciais, que devem ser agrupadas por algum tipo de processo de *clusterização* para formar os subespaços independentes. Cada subespaço, por sua vez, corresponde a uma fonte.

A limitação da ISA se refere ao número de componentes requeridas para identificação das fontes que devem ser retidas pela PCA. O número adequado varia de acordo com os sinais que compõem a mistura e com a amplitude relativa entre esses sinais (lembrando que a PCA se baseia na energia das componentes do espectrograma, retendo as maiores), não sendo em geral possível determiná-lo de forma automática.

Outra técnica de separação de fontes sonoras é conhecida como **Codificação Esparsa** (do inglês, *Sparse Coding*, SC) [14], que se assenta sobre o conceito de esparsidade. Um sinal de mistura é esparso no domínio do tempo quando apenas

uma das fontes que o compõem emite (está ativa) em um determinado instante; em outras palavras, para sinais musicais, é quando uma nota proveniente de um único instrumento ou apenas um instrumento está tocando em determinado trecho da mistura. A esparsidade no domínio da frequência se dá quando apenas alguns canais de frequência correspondem às frequências contidas na mistura. Na Codificação Esparsa, assume-se a esparsidade do sinal de mistura, o que é razoável para sinais de áudio. Isso faz com que a maioria dos elementos da matriz de ganhos do modelo sejam nulos. A estimativa das fontes que compõem a mistura é feita através da maximização das suas distribuições a posteriori [9]. A esparsidade pode ser um critério adicional ao método da NMF, conforme será visto mais adiante.

Capítulo 3

Fatoração de Matrizes Não-Negativas

3.1 Discussão inicial

Ao final do capítulo anterior, foram apresentados de maneira resumida alguns dos principais métodos não-supervisionados de separação de fontes sonoras: Análise de Componentes Principais, Análise de Componentes Independentes, Análise de Subespaços Independentes e Codificação Esparsa. Agora, será discutido em detalhes o método estudado nesta dissertação, a Fatoração de Matrizes Não-Negativas.

O modelo geral de representação do sinal de mistura apresentado na equação (2.1) pode ser escrito na forma matricial

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{BG}, \tag{3.1}$$

em que $\mathbf{X}$ é o espectrograma de magnitude do sinal de mistura, $\hat{\mathbf{X}}$ é a estimativa de $\mathbf{X}$, $\mathbf{B}$ é a matriz de base frequencial e $\mathbf{G}$ é a matriz de ganhos. A dimensão de $\mathbf{X}$ é NxT, em que N é o número de canais (ou raias) de frequência e T é o número de quadros de tempo. A dimensão de $\mathbf{B}$ é NxM, em que M é o número de componentes (cada qual representada por um vetor da base frequencial). E a dimensão de $\mathbf{G}$ é MxT.

O número de componentes M deve ser menor que N e T, o que faz com que o produto $\mathbf{BG}$ seja uma versão comprimida do espectrograma $\mathbf{X}$ [15], já que poucas componentes são utilizadas para representar as frequências contidas no sinal de mistura. Na realidade, a NMF procura representar em poucos vetores de base as frequências mais proeminentes do sinal. Portanto, o produto $\mathbf{BG}$ é uma aproximação de $\mathbf{X}$, conforme representado na equação (3.1).

As colunas de $\mathbf{B}$ contêm os vetores $\mathbf{b}_m$ (para $m = 1, 2, ..., M$) da base frequencial. Cada linha m em $\mathbf{G}$ contém uma sequência temporal de ganhos, $g_{m,t}$ (para $t = 1, 2, ..., T$), que multiplicam o vetor $\mathbf{b}_m$. Assim, para cada m, descreve-se a contribuição de sua respectiva componente de frequência, com padrão espectral representado por $\mathbf{b}_m$, no sinal de mistura para todos os intervalos de tempo.

A coluna m de $\mathbf{B}$ (ou seja, o vetor $\mathbf{b}_m$) associada à linha m de $\mathbf{G}$ (ou seja, sua linha correspondente em $\mathbf{G}$) é modelada como o espectrograma da componente m, $\mathbf{X}_m$,

$$\mathbf{X}_m \approx \hat{\mathbf{X}}_m = \mathbf{b}_m \begin{bmatrix} g_{m,1} & g_{m,2} & \cdots & g_{m,T} \end{bmatrix}, \tag{3.2}$$

em que $\hat{\mathbf{X}}_m$ é a estimativa de $\mathbf{X}_m$. A soma de todas as componentes gera o espectrograma $\hat{\mathbf{X}} = \mathbf{BG}$, ou seja,

$$\mathbf{X} \approx \hat{\mathbf{X}} = \sum_{m=1}^{M} \mathbf{X}_m. \tag{3.3}$$

Analisando de um outro ponto de vista, cada coluna t de $\mathbf{G}$ representa os ganhos de todas as componentes m para o quadro t. Assim, tem-se a descrição de todas as componentes de frequência do sinal de mistura durante o quadro t.

O objetivo da NMF é encontrar as matrizes $\mathbf{B}$ e $\mathbf{G}$ a partir do espectrograma do sinal de mistura dado, $\mathbf{X}$. Como o espectrograma é não-negativo por definição, é natural restringir os valores de $\mathbf{B}$ de forma que eles sejam não-negativos [9], já que $\mathbf{B}$ é uma matriz que contém dados de frequência. Além disso, faz sentido restringir os valores de $\mathbf{G}$ de forma que eles também sejam não-negativos[1], pois dessa forma as componentes serão puramente aditivas: um "todo" sendo descrito como a soma de várias "partes" [7]. Assim, preserva-se a natureza não-negativa da representação.

As restrições de não-negatividade impostas para as matrizes $\mathbf{B}$ e $\mathbf{G}$, além de sempre conferirem sentido físico para os valores obtidos (vetores da base frequencial e ganhos com valores negativos, mesmo que matematicamente corretos dentro do modelo, afastam-se do sentido físico do problema), são suficientes, sob certas condições, para a separação de fontes [16]. A ICA, além de não garantir valores não-negativos para suas matrizes, baseia-se no conceito de independência estatística entre as fontes, o que não é necessário na NMF.

[1] Implicitamente, evita-se uma indesejável informação de fase

3.2 Medida de distorção

Uma medida de distorção precisa ser utilizada para medir o quão próximo o espectrograma estimado $\mathbf{X} \approx \mathbf{BG}$ está do espectrograma real $\mathbf{X}$. As medidas de distorção mais utilizadas e que foram propostas em [15] são o quadrado da distância Euclidiana,

$$D_{\text{euc}} = \frac{1}{2}\left\|\mathbf{X} - \hat{\mathbf{X}}\right\|_F^2, \tag{3.4}$$

e a divergência de Kullback-Leibler generalizada,

$$D_{\text{kl}} = \left|\mathbf{X} \odot \ln \frac{\mathbf{X}}{\hat{\mathbf{X}}} - \mathbf{X} + \hat{\mathbf{X}}\right|, \tag{3.5}$$

em que o operador $\|.\|$ é a função Norma e $\|.\|_F$ especificamente indica a norma de Frobenius[2], o operador $|.|$[3] indica soma de todos os elementos, e o operador $\odot$ em D_{kl} indica multiplicação ponto-a-ponto entre matrizes (também chamado de produto de Hadamard). A divisão entre matrizes em D_{kl} também é ponto-a-ponto.

A medida D_{kl} não representa uma distância, pois não se trata de uma função simétrica, ou seja, a divergência de $\mathbf{X}$ para $\hat{\mathbf{X}}$ não é igual à divergência de $\hat{\mathbf{X}}$ para $\mathbf{X}$. Na realidade, D_{kl} não poderia nem ser chamada de divergência, devido ao fato de que seus argumentos não representam distribuições de probabilidade. Entretanto, essa é a nomenclatura utilizada na literatura científica sobre NMF.

Na situação ideal, quando o produto $\mathbf{BG}$ é exatamente igual ao espectrograma $\mathbf{X}$, ambas as medidas de distorção resultam iguais a zero. Entretanto, dentre as duas, a divergência de Kullback-Leibler generalizada é a mais sensível para baixas energias, assemelhando-se mais ao sistema auditivo humano [10]. Por isso, essa medida é tida como mais adequada para aplicações da NMF em sinais de áudio.

3.3 Algoritmo básico

Todo algoritmo de NMF se baseia na minimização de uma função-custo para a obtenção das matrizes $\mathbf{B}$ e $\mathbf{G}$. No caso da NMF sem restrições, a função-custo é a própria medida de distorção escolhida. O objetivo do algoritmo é encontrar as matrizes $\mathbf{B}$ e $\mathbf{G}$ não-negativas que, multiplicadas, mais se aproximam do espectrograma $\mathbf{X}$.

[2] A norma de Frobenius de uma matriz $\mathbf{A}$ é escrita como $\|\mathbf{A}\|_F = \left(\sum_i \sum_j |a_{ij}|^2\right)^{\frac{1}{2}}$.

[3] Essa notação pouco usual é adotada em diversos artigos da área.

A minimização é normalmente realizada através do método do gradiente descendente [17] com um passo suficientemente pequeno. Dessa forma, as equações de atualização para $\mathbf{B}$ e $\mathbf{G}$ considerando o quadrado da distância Euclidiana como função-custo são [15]:

$$\mathbf{B} = \mathbf{B} \odot \frac{\mathbf{X}\mathbf{G}^T}{\mathbf{B}\mathbf{G}\mathbf{G}^T}, \tag{3.6}$$

$$\mathbf{G} = \mathbf{G} \odot \frac{\mathbf{B}^T\mathbf{X}}{\mathbf{B}^T\mathbf{B}\mathbf{G}^T}. \tag{3.7}$$

Da mesma forma, para a divergência de Kullback-Leibler generalizada, as equações de atualização são [15]:

$$\mathbf{B} = \mathbf{B} \odot \frac{\frac{\mathbf{X}}{\bar{\mathbf{X}}}\mathbf{G}^T}{\mathbf{1}\mathbf{G}^T}, \tag{3.8}$$

$$\mathbf{G} = \mathbf{G} \odot \frac{\mathbf{B}^T\frac{\mathbf{X}}{\bar{\mathbf{X}}}}{\mathbf{B}^T\mathbf{1}}. \tag{3.9}$$

Vale mencionar novamente que em todas essas equações a operação $\odot$ indica multiplicação ponto-a-ponto, e que a divisão entre matrizes também é feita ponto-a-ponto. A matriz $mathbf1$ tem todos os elementos iguais a 1.

As deduções das equações de atualização da NMF, tanto para o quadrado da distância Euclidiana quanto para a divergência de Kullback-Leibler generalizada, podem ser vistas na Seção A.1 do Apêndice A. O fato de as equações de atualização serem multiplicativas faz com que seja preservada a não-negatividade de $\mathbf{B}$ e $\mathbf{G}$, desde que essas matrizes sejam inicializadas com valores não-negativos.

As matrizes $\mathbf{B}$ e $\mathbf{G}$ são obtidas de forma alternada em cada iteração, pois o problema não é convexo em $\mathbf{B}$ e $\mathbf{G}$ ao mesmo tempo. Isso significa que, após a atualização de $\mathbf{B}$, a função-custo é calculada novamente e o novo valor de $\mathbf{B}$ é utilizado na equação de atualização de $\mathbf{G}$. Isso é feito de forma análoga partindo-se da atualização de $\mathbf{G}$, até que o algoritmo convirja. Essas considerações podem ser vistas em [15], onde é apresentada a prova de convergência desse algoritmo para ambas as funções-custo.

Em [5] é citada a divergência de Bregman, que resulta em equações de atualização gerais para $\mathbf{B}$ e $\mathbf{G}$. Dependendo da escolha de determinados parâmetros descritos nesse artigo e contidos em suas equações, podem-se obter as equações de atualização baseadas no quadrado da distância Euclidiana ou na divergência de Kullback-Leibler

generalizada, sendo estas portanto casos particulares das equações de atualização baseadas na divergência de Bregman.

3.4 *Non-Negative Matrix Factor Deconvolution* (NMFD)

O modelo básico da NMF descrito na seção anterior, apesar de ser uma ferramenta útil na análise espectral do sinal de mistura, lida bem apenas com padrões de frequências que são inteiramente representados dentro de um único quadro de tempo. Entretanto, um padrão de frequência de uma fonte pode evoluir com o tempo. Notas de instrumentos, no geral, possuem padrões de frequência que extrapolam a duração de um quadro, como por exemplo, uma nota de piano, que possui um ataque percussivo (batida na tecla) e uma sustentação aproximadamente harmônica [7].

A NMFD [18] se propõe a representar notas de instrumentos. Para isso, permite que os padrões espectrais ocupem mais de um quadro; com isso, em vez de uma base de vetores, trabalha com uma base de matrizes, cada uma correspondendo a um padrão espectral. O conceito de componente agora se estende para nota musical: uma componente representa uma nota. Assim, o modelo da NMFD é expresso por

$$\mathbf{X} \approx \hat{\mathbf{X}} = \sum_{l=0}^{\tau-1} \mathbf{B}^l \overrightarrow{\mathbf{G}}^{l}, \tag{3.10}$$

em que τ é o número de quadros permitidos para cada nota e o operador $\overrightarrow{(.)}^{l}$ representa deslocamento horizontal dos elementos da matriz. Um exemplo de aplicação desse operador é descrito a seguir: considerando

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}, \tag{3.11}$$

tem-se que

$$\overrightarrow{\mathbf{A}}^{0} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}, \overrightarrow{\mathbf{A}}^{1} = \begin{bmatrix} 0 & a_{1,1} & a_{1,2} \\ 0 & a_{2,1} & a_{2,2} \\ 0 & a_{3,1} & a_{3,2} \end{bmatrix}, \overrightarrow{\mathbf{A}}^{2} = \begin{bmatrix} 0 & 0 & a_{1,1} \\ 0 & 0 & a_{2,1} \\ 0 & 0 & a_{3,1} \end{bmatrix}, \cdots. \tag{3.12}$$

Da mesma forma,

$$\overset{\leftarrow 0}{\mathbf{A}} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}, \overset{\leftarrow 1}{\mathbf{A}} = \begin{bmatrix} a_{1,2} & a_{1,3} & 0 \\ a_{2,2} & a_{2,3} & 0 \\ a_{3,2} & a_{3,3} & 0 \end{bmatrix}, \overset{\leftarrow 2}{\mathbf{A}} = \begin{bmatrix} a_{1,3} & 0 & 0 \\ a_{2,3} & 0 & 0 \\ a_{3,3} & 0 & 0 \end{bmatrix}, \cdots . \quad (3.13)$$

Conforme pode ser observado através dessa exemplificação, e generalizando para matrizes quadradas $\mathbf{A}$ de dimensão KxK, cada elemento $a_{i,j}$ da matriz $\mathbf{A}$, quando é aplicado sobre ela o operador $\overset{\rightarrow l}{(.)}$, com $0 \leq l < K$, assume o valor $a_{i,j-l}$ para $j > l$, e zero para $j \leq l$. Ao se aplicar o operador $\overset{\leftarrow l}{(.)}$, cada elemento de $\mathbf{A}$ assume o valor $a_{i,j+l}$ para $j + l \leq K$, e zero para $j + l > K$.

As matrizes $\mathbf{B}^l$, para $l = 0, 1, ..., \tau - 1$, possuem dimensão NxM. Cada $\mathbf{B}^l$ contém os padrões espectrais das componentes para o quadro l. Uma única matriz $\mathbf{G}$ é formada tomando-se a média elemento a elemento das matrizes $\mathbf{G}^l$, para $l = 0, 1, ..., \tau - 1$, tendo cada qual sido atualizada fazendo-se uso da matriz $\mathbf{B}^l$ associada. As matrizes $\mathbf{G}^l$ (e consequentemente a matriz $\mathbf{G}$) possuem dimensão MxT. Assim como na NMF básica, todas as matrizes são restritas a ter somente elementos não-negativos.

As equações de atualização de $\mathbf{B}^l$, $\mathbf{G}^l$ e $\mathbf{G}$ para a divergência de Kullback-Leibler generalizada como função-custo[4], cujas deduções estão na Seção A.2 do Apêndice A, são [18]:

$$\mathbf{B}^l = \mathbf{B}^l \odot \frac{\frac{\mathbf{X}}{\hat{\mathbf{X}}}(\overset{\rightarrow l}{\mathbf{G}^T})}{\mathbf{1}(\overset{\rightarrow l}{\mathbf{G}^T})} \text{ para } l = 0, 1, ..., \tau - 1, \quad (3.14)$$

$$\mathbf{G}^l = \mathbf{G}^l \odot \frac{\mathbf{B}^{l^T}(\frac{\overset{\leftarrow l}{\mathbf{X}}}{\bar{\mathbf{X}}})}{\mathbf{B}^{l^T}\mathbf{1}} \text{ para } l = 0, 1, ..., \tau - 1, \quad (3.15)$$

$$\mathbf{G} = \frac{\sum_{l=0}^{\tau-1} \mathbf{G}^l}{\tau}. \quad (3.16)$$

[4]Daqui por diante (para este método e para os métodos apresentados posteriormente), só serão mostradas as equações de atualização referentes à divergência de Kullback-Leibler generalizada, visto que esta medida de distorção é considerada mais adequada para o tratamento de sinais musicais.

As matrizes $\mathbf{B}^l$ e $\mathbf{G}^l$ devem ser atualizadas de forma alternada devido ao problema da não-convexidade existente quando se consideram as duas matrizes juntas. Em cada iteração, $\mathbf{B}^l$ ou $\mathbf{G}^l$ é atualizada para todos os valores de l e, a partir do resultado obtido, é calculado o espectrograma estimado $\hat{\mathbf{X}}$.

Observa-se através da equação (3.10) que cada coluna de $\hat{\mathbf{X}}$, ou seja, cada vetor $\hat{\mathbf{x}}_t$ (para $t = 1, 2, ..., T$), é representado como uma mistura convolutiva entre as matrizes $\mathbf{B}^l$ e $\mathbf{G}$ [19]. Assim, em outras palavras, a NMFD pode ser vista como uma modificação da NMF básica que contempla misturas convolutivas, onde existe dependência entre quadros sucessivos do espectrograma [20].

3.5 *Non-Negative Matrix Factor 2-D Deconvolution* (NMF2D)

A NMFD consegue representar notas de um instrumento musical ao permitir o uso de mais de um quadro para descrever os padrões epectrais das componentes. Todavia, quando se deseja representar em uma componente o instrumento como um todo, a NFMD não é suficiente. Portanto, a fim de permitir a separação entre instrumentos, foi desenvolvida a NMF2D [21], que é uma extensão da NMFD.

Além de permitir o uso de mais de um quadro para representar uma nota musical, a NMF2D permite que o padrão espectral contido nesses quadros seja deslocado para cima ou para baixo no eixo das frequências, modelando assim todo o conjunto de notas que podem ser emitidas por um instrumento com um mesmo padrão espectral de emissão. Em outras palavras, na NMF2D, um instrumento musical é modelado como um fonte emissora de timbre único para várias notas, como se existisse uma "assinatura" acústica do instrumento presente em todas as notas.

O modelo da NMF2D é expresso por

$$\mathbf{X} \approx \hat{\mathbf{X}} = \sum_{l=0}^{\tau-1} \sum_{p=0}^{\phi-1} \mathbf{B}^l \overset{\downarrow p}{\underset{\rightarrow l}{\mathbf{G}^p}}, \tag{3.17}$$

em que τ e o operador $\overset{\rightarrow l}{(.)}$ possuem o mesmo significado visto para o caso da NMFD, ou seja, número de quadros permitidos para representar as notas e operador de deslocamento horizontal dos elementos da matriz, respectivamente. O parâmetro ϕ representa o número de deslocamentos no eixo das frequências e o operador $\overset{\downarrow p}{(.)}$ representa deslocamento vertical dos elementos da matriz. O uso desse operador

pode ser exemplificado da seguinte forma: considerando novamente

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}, \tag{3.18}$$

tem-se que

$$\overset{\downarrow 0}{\mathbf{A}} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}, \overset{\downarrow 1}{\mathbf{A}} = \begin{bmatrix} 0 & 0 & 0 \\ a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}, \overset{\downarrow 2}{\mathbf{A}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ a_{1,1} & a_{1,2} & a_{1,3} \end{bmatrix}, \cdots . \tag{3.19}$$

Da mesma forma,

$$\overset{\uparrow 0}{\mathbf{A}} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}, \overset{\uparrow 1}{\mathbf{A}} = \begin{bmatrix} a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \\ 0 & 0 & 0 \end{bmatrix}, \overset{\uparrow 2}{\mathbf{A}} = \begin{bmatrix} a_{3,1} & a_{3,2} & a_{3,3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \cdots . \tag{3.20}$$

Conforme pode ser observado através dessa exemplificação, e generalizando para matrizes quadradas $\mathbf{A}$ de dimensão KxK, cada elemento $a_{i,j}$ da matriz $\mathbf{A}$, quando é aplicado sobre ela o operador $\overset{\downarrow p}{(.)}$, com $p \leq K$, assume o valor $a_{i-p,j}$ para $i > p$, e zero para $i \leq p$. Ao se aplicar o operador $\overset{\uparrow p}{(.)}$, cada elemento de $\mathbf{A}$ assume o valor $a_{i+p,j}$ para $i + p \leq K$, e zero para $i + p > K$.

Na NMF2D, as matrizes $\mathbf{B}^l$, para $l = 0, 1, ..., \tau - 1$, têm a mesma estrutura que na NMFD, ou seja, possuem dimensão NxM, e cada uma contém os padrões espectrais de todos os M instrumentos para o respectivo quadro l. Já as matrizes $\mathbf{G}^p$ possuem dimensão MxT, e cada uma contém a descrição temporal de ocorrências da nota p de cada instrumento.

As equações de atualização de $\mathbf{B}^l$ e $\mathbf{G}^p$ para a divergência de Kullback-Leibler generalizada como função-custo, cujas deduções estão na Seção A.3 do Apêndice A,

são [21]:

$$\mathbf{B}^l = \mathbf{B}^l \odot \frac{\sum_{p=0}^{\phi-1} \left(\frac{\mathbf{X}}{\hat{\mathbf{X}}}\right)^{\uparrow p} . \overrightarrow{\mathbf{G}^p}^{l\,T}}{\sum_{p=0}^{\phi-1} \mathbf{1} . \overrightarrow{\mathbf{G}^p}^{l\,T}} \text{ para } l = 0, 1, ..., \tau, \text{ e} \tag{3.21}$$

$$\mathbf{G}^p = \mathbf{G}^p \odot \frac{\sum_{l=0}^{\tau-1} \mathbf{B}^{l\,\downarrow p\,T} . \left(\frac{\mathbf{X}}{\hat{\mathbf{X}}}\right)^{\leftarrow l}}{\sum_{l=0}^{\tau-1} \mathbf{B}^{l\,\downarrow p\,T} . \mathbf{1}} \text{ para } p = 0, 1, ..., \phi. \tag{3.22}$$

As matrizes $\mathbf{B}^l$ e $\mathbf{G}^p$ também devem ser atualizadas de forma alternada. Em cada iteração, $\mathbf{B}^l$ e $\mathbf{G}^p$ são atualizadas para todos os valores de l e p respectivamente, uma de cada vez, e, a partir do resultado obtido, o espectrograma estimado $\hat{\mathbf{X}}$ é calculado, até que o algoritmo atinja a convergência.

Conforme indica a equação (3.17), a NMF2D representa o vetor $\hat{\mathbf{x}}_t$ (para $t = 1, 2, ..., T$) de $\hat{\mathbf{X}}$ como uma mistura duplamente convolutiva, havendo tanto dependência entre quadros sucessivos no tempo, quanto dependência entre canais sucessivos na frequência; daí vem o nome do modelo.

No caso da NMF2D, o bloco de Análise de sinais de um sistema completo de separação por NMF, descrito no Capítulo 2, é realizado de forma diferente. O deslocamento de padrões espectrais no eixo das frequências deve estar de acordo com a escala empregada para as notas musicais, ou seja, o deslocamento deve estar na escala de igual temperamento, que divide uma oitava[5] em 12 intervalos (semitons) cujas alturas em Hz são organizadas em proporções geométricas de razão $2^{\frac{1}{12}}$. Isso quer dizer que a razão entre as frequências de duas notas sucessivas perfeitamente aplicadas—separadas de um semitom—deve ser sempre considerada igual a $2^{\frac{1}{12}}$ [6].

Entretanto, o espaçamento das frequências no espectrograma resultante da STFT é linear e não está de acordo com a escala musical. Assim, faz-se necessário o uso do espectrograma logarítmico do sinal de mistura, de forma que cada deslocamento unitário no eixo das frequências resulte em um intervalo constante entre notas musicais.

[5] Uma oitava é o intervalo entre uma nota musical e outra com a metade ou o dobro de sua frequência.

[6] Por razões perceptivas, a afinação de instrumentos com extensão muito ampla tende a se afastar dessa lei de formação.

Um espectrograma de escala logarítmica pode ser obtido de duas formas: através do uso da Transformada de Q Constante (*Constant-Q Transform*, CQT) [22] [23] em vez da STFT na segmentação do sinal de mistura, ou através de algum tipo de mapeamento da escala linear para a escala logarítmica sobre o espectrograma linear obtido pela STFT.

Nos testes realizados neste trabalho, foi utilizado o mapeamento logarítmico [24] [25], por apresentar resultados em geral ligeiramente melhores do que os resultados obtidos através da CQT, como pode ser verificado em [7]. O método tem por objetivo agrupar as raias (canais) de frequência para que se tenha o mesmo de número de raias por oitava, ou seja, enquanto o número de raias na representação linear é variável, na representação logaritmica ele é fixo, adequando a representação espectral do sinal de mistura à escala de igual temperamento. Isso é feito através de uma matriz de tranformação $\mathbf{C}$, de modo que

$$\hat{\mathbf{X}}^{\log} = \mathbf{C}\hat{\mathbf{X}}, \tag{3.23}$$

em que $\hat{\mathbf{X}}^{\log}$ é a estimativa do espectrograma linear $\mathbf{X}$ na escala logarítmica. Como a dimensão de $\hat{\mathbf{X}}$ é NxT, $\mathbf{C}$ deve possuir dimensão $N^{\log}$xN (sendo $N^{\log}$ o número de canais de frequência na escala logarítmica) de forma que $\hat{\mathbf{X}}^{\log}$ tenha dimensão $N^{\log}$xT. Assim, tem-se o mapeamento de N para $N^{\log}$; através da matriz $\mathbf{C}$, as oitavas de $\hat{\mathbf{X}}$ são mapeadas nas oitavas de $\hat{\mathbf{X}}^{\log}$.

No algoritmo, cada raia linear é considerada como ocupando uma banda que vai desde a metade da distância entre sua frequência central e a frequência central representada pela raia imediatamente inferior até a metade da distância entre sua frequência central e a frequência central representada pela raia imediatamente superior. Então, cada raia é avaliada individualmente e, conforme sua frequência central, escolhe-se para qual raia logarítmica ela será mapeada, ou seja, para onde vai a sua energia. Caso a raia linear esteja exatamente entre duas raias logarítmicas, ela é mapeada nas duas raias, sendo a energia dividida entre elas.

Vale mencionar que, em baixas frequências, existem mais raias logarítmicas do que raias lineares. À medida que a frequência aumenta, passam a existir mais raias lineares do que raias logarítmicas.

Um parâmetro importante do mapeamento logarítmico é a resolução que se deseja para o espectrograma na escala logarítmica. Essa resolução é representada pela letra b. Por exemplo, se $b = 12$, tem-se que que o espectrograma terá 1 raia representando 1 intervalo semitom, que é justamente a resolução mínima para conseguir representar

os 12 intervalos de semitom que compõem uma oitava na escala musical. Entretanto, se $b = 24$, tem-se que um intervalo de semitom será representado por 2 raias, o que aumenta a precisão, aumentando o custo computacional. E assim por diante.

O algoritmo utilizado neste trabalho para a construção da matriz $\mathbf{C}$ está descrito no Apêndice A de [7] e se baseia em [24] e [25].

3.6 Adicionando restrições à NMF básica

Algumas restrições podem ser adicionadas ao algoritmo básico da NMF [26] de forma a adequar o modelo a determinados objetivos, melhorando a qualidade da separação. Neste caso, a função-custo ganha novos termos somados às medidas de distorção baseadas no quadrado da distância Euclidiana ou na divergência de Kullback-Leibler generalizada. Esses novos termos podem estar em função tanto da matriz de base frequencial $\mathbf{B}$ quanto da matriz de ganhos $\mathbf{G}$. De forma geral, a função-custo D_{custo} a ser minimizada pode ser representada por

$$D_{\text{custo}} = D_{\text{kl}}(\mathbf{X}, \hat{\mathbf{X}}) + \alpha_1 c_{1,B}(\mathbf{B}) + \alpha_2 c_{2,B}(\mathbf{B}) + \cdots + \beta_1 c_{1,G}(\mathbf{G}) + \beta_2 c_{2,G}(\mathbf{G}) + \cdots , \tag{3.24}$$

em que D_{kl} representa a divergência de Kullback-Leibler generalizada, podendo também ser substituída pelo quadrado da distância Euclidiana. A função $c_{1,B}$ representa o primeiro critério de restrição sobre a matriz $\mathbf{B}$, $c_{2,B}$ representa o segundo critério de restrição sobre a matriz $\mathbf{B}$, e assim sucessivamente para $c_{3,B}, c_{4,B}, ...$ Analogamente, $c_{1,G}, c_{2,G}, ...$ representam os critérios de restrição sobre a matriz $\mathbf{G}$. Os valores $\alpha_1, \alpha_2, ...$ representam os pesos aplicados aos critérios de restrição sobre a matriz $\mathbf{B}$, assim como os valores $\beta_1, \beta_2, ...$ representam os pesos aplicados aos critérios de restrição sobre a matriz $\mathbf{G}$. Esses pesos são utilizados para dar mais ou menos importância aos critérios associados a eles no momento da execução do algoritmo.

Considerando a divergência de Kullback-Leibler generalizada como medida de distorção, as equações de atualização tornam-se [6]

$$\mathbf{B} = \mathbf{B} \odot \frac{\frac{\mathbf{X}}{\hat{\mathbf{X}}}\mathbf{G}^T}{\mathbf{1}\mathbf{G}^T + \alpha_1 \nabla_{\mathbf{B}} c_{1,B} + \alpha_2 \nabla_{\mathbf{B}} c_{2,B} + ...}, \tag{3.25}$$

$$\mathbf{G} = \mathbf{G} \odot \frac{\mathbf{B}^T\frac{\mathbf{X}}{\hat{\mathbf{X}}}}{\mathbf{B}^T\mathbf{1} + \beta_1 \nabla_{\mathbf{G}} c_{1,G} + \beta_2 \nabla_{\mathbf{G}} c_{2,G} + ...}, \tag{3.26}$$

em que $\nabla_{(.)}$ indica a operação de derivada em relação às matrizes sobre as quais os critérios incidem. As deduções das equações (3.25) e (3.26) estão na Seção A.4 do

Apêndice A.

Ao se adicionar restrições à NMF, deve-se tomar cuidado com a convergência, pois elas podem tornar o algoritmo instável, além de produzir matrizes $\mathbf{B}$ e $\mathbf{G}$ com elementos negativos. Entretanto, ao escolher funções com derivadas positivas para representar os critérios desejados, nota-se que a convergência do algoritmo não é comprometida [5].

Na literatura científica, entre os critérios utilizados estão o de esparsidade [10] [27], o de continuidade temporal [10] e o de redução de correlação cruzada [5] [6].

3.7 *Sparse Non-Negative Matrix Factor 2-D Deconvolution* (SNMF2D)

Seguindo a linha de evolução dos algoritmos de NMF, apresenta-se agora a SNMF2D. Esse algoritmo combina a NMF2D com o critério de esparsidade, e foi desenvolvido em [19].

O critério de esparsidade, c_{ep}, incide sobre as matrizes $\mathbf{G}^p$, para $p = 0, 1, ..., \phi - 1$ e é aplicado quando se deseja representar misturas cujas fontes sejam esparsas, isto é, não emitam sempre ao mesmo tempo e o tempo todo. Isso é razoavelmente verdadeiro quando se trata de sinais musicais. Fazendo-se uso deste critério, as matrizes $\mathbf{G}^p$ deverão ter a maioria de seus elementos nulos após a convergência do algoritmo. Definindo $\mathbf{G}$ como uma matriz que representa todas as matrizes $\mathbf{G}^p$,

$$\mathbf{G} \triangleq \left[\begin{array}{ccccccc} \mathbf{G}^0 & \vdots & \mathbf{G}^1 & \vdots & \dots & \vdots & \mathbf{G}^{\phi-1} \end{array} \right], \tag{3.27}$$

é sabido que um conjunto de funções que podem impor a restrição de esparsidade nas matrizes $\mathbf{G}^p$ são as normas L_γ, para $\gamma > 0$, dadas por

$$c_{\text{ep}}(\mathbf{G}) = \|\mathbf{G}\|_\gamma = \left(\sum_{p,m,t} \left| G^p_{m,t} \right|^\gamma \right)^{\frac{1}{\gamma}}, \tag{3.28}$$

em que $G^p_{m,t}$ representa cada elemento de $\mathbf{G}^p$. Assim, tem-se que

$$\nabla_{\mathbf{G}^p}(c_{\text{ep}}) = \frac{\mathbf{G}^{p\bullet(\gamma-1)}}{\|\mathbf{G}\|_\gamma^{\gamma-1}}, \tag{3.29}$$

em que $\mathbf{G}^{p\bullet(\gamma-1)}$ indica que cada elemento de $\mathbf{G}^p$ é elevado a $(\gamma - 1)$.

No modelo da SNMF2D proposto em [19], as matrizes $\mathbf{B}^l$, para $l = 0, 1, ..., \tau - 1$ devem ser devidamente normalizadas em cada iteração de forma a evitar que o termo de esparsidade seja minimizado simplesmente fazendo com que as matrizes $\mathbf{G}^p$ vão a zero enquanto as matrizes $\mathbf{B}^l$ vão ao infinito. Representando-se cada matriz $\mathbf{B}^l$ normalizada por $\tilde{\mathbf{B}}^l$, tem-se que

$$\tilde{B}^l_{n,m} = \frac{B^l_{n,m}}{\|\mathbf{B}_m\|_F} = \frac{B^l_{n,m}}{\sqrt{\sum_{l',n'} \left(B^{l'}_{n',m}\right)^2}}, \tag{3.30}$$

onde $B^l_{n,m}$ representa cada elemento de $\mathbf{B}^l$. Agora, fazendo

$$\mathbf{X} \approx \tilde{\hat{\mathbf{X}}} = \sum_{l,p} \overset{\downarrow p}{\tilde{\mathbf{B}}}{}^l \overset{\rightarrow l}{\mathbf{G}}{}^p, \tag{3.31}$$

tem-se que a função-custo para a divergência de Kullback-Leibler generalizada considerando o critério de esparsidade baseado na norma L_γ é

$$D_{\text{custo}} = \left| \mathbf{X} \odot \ln \frac{\mathbf{X}}{\tilde{\hat{\mathbf{X}}}} - \mathbf{X} + \tilde{\hat{\mathbf{X}}} \right| + \alpha \|\mathbf{G}\|_\gamma, \tag{3.32}$$

em que α é o peso atribuído ao critério de esparsidade. Por fim, com base na normalização apresentada na equação (3.30) e considerando a divergência de Kullback-Leibler como medida de distorção, as equações de atualização para a SNMF2D são [19]

$$\mathbf{B}^l = \tilde{\mathbf{B}}^l \odot \frac{\sum_{p=0}^{\phi-1} \left(\frac{\overset{\uparrow p}{\mathbf{X}}}{\hat{\mathbf{X}}}\right) . \overset{\rightarrow l}{\mathbf{G}}{}^{p\,T} + \tilde{\mathbf{B}}^l . \text{diag}\left\{\sum_{l=0}^{\tau-1} \mathbf{1}. \left[\left(\mathbf{1}. \overset{\rightarrow l}{\mathbf{G}}{}^{p\,T}\right) \odot \tilde{\mathbf{B}}^l\right]\right\}}{\sum_{p=0}^{\phi-1} \mathbf{1}. \overset{\rightarrow l}{\mathbf{G}}{}^{p\,T} + \tilde{\mathbf{B}}^l . \text{diag}\left\{\sum_{l=0}^{\tau-1} \mathbf{1}. \left[\left(\left(\frac{\overset{\uparrow p}{\mathbf{X}}}{\hat{\mathbf{X}}}\right) . \overset{\rightarrow l}{\mathbf{G}}{}^{p\,T}\right) \odot \tilde{\mathbf{B}}^l\right]\right\}}, \tag{3.33}$$

$$\mathbf{G}^p = \tilde{\mathbf{G}}^p \odot \frac{\sum_{l=0}^{\tau-1} \overset{\downarrow p}{\mathbf{B}}{}^{l\,T} . \left(\frac{\overset{\leftarrow l}{\mathbf{X}}}{\hat{\mathbf{X}}}\right)}{\sum_{l=0}^{\tau-1} \overset{\downarrow p}{\mathbf{B}}{}^{l\,T} . \mathbf{1} + \alpha \nabla_{\mathbf{G}^p}(\|\mathbf{G}\|_\gamma)}. \tag{3.34}$$

Nessas equações é visto que $\tilde{\mathbf{B}}^l$ e $\tilde{\mathbf{G}}^p$ são atualizados para $\mathbf{B}^l$ e $\mathbf{G}^p$, respectivamente. Isso acontece porque $\mathbf{B}^l$ e $\mathbf{G}^p$ devem ser sempre normalizados no início de cada iteração para que então se tenha os novos $\tilde{\mathbf{B}}^l$ e $\tilde{\mathbf{G}}^p$ a serem utilizados.

Na equação (3.34), nota-se o termo adicionado ao denominador ($\alpha\nabla_{\mathbf{G}^p}(\|\mathbf{G}\|_\gamma)$) referente ao critério de esparsidade. Na equação (3.33), comparada com a equação (3.21) da NMF2D, nota-se um termo extra adicionado ao somatório no numerador,

$$\tilde{\mathbf{B}}^l \text{diag}\left\{\sum_{l=0}^{\tau-1}\mathbf{1}.\left[\left(\mathbf{1}.\overrightarrow{\mathbf{G}^p}^{l\,T}\right)\odot\tilde{\mathbf{B}}^l\right]\right\}, \tag{3.35}$$

e outro no denominador,

$$\tilde{\mathbf{B}}^l \text{diag}\left\{\sum_{l=0}^{\tau-1}\mathbf{1}.\left[\left(\left(\frac{\overset{\uparrow p}{\mathbf{X}}}{\hat{\mathbf{X}}}\right).\overrightarrow{\mathbf{G}^p}^{l\,T}\right)\odot\tilde{\mathbf{B}}^l\right]\right\}, \tag{3.36}$$

em que diag(.) denota uma matriz diagonal cujos elementos do argumento estão contidos na diagonal. Não é discutida em [19] a origem desses termos. Porém, na Seção A.5 do Apêndice A deste trabalho, é mostrado como eles são obtidos e que eles decorrem da utilização da matriz normalizada $\tilde{\mathbf{B}}^l$ nas iterações da SNMF2D.

É importante mencionar que a NMF2D, em geral, não possui solução única. Já a SNMF2D, ao impor matrizes $\mathbf{G}^p$ esparsas (que nesta aplicação são mais interpretáveis fisicamente), evita múltiplas soluções, convergindo na maioria das vezes para a mesma estrutura de matrizes. Portanto, a maior vantagem da SNMF2D sobre a NMF2D é se aproximar mais da unicidade na fatoração.

3.8 Outros aprimoramentos

Além dos aprimoramentos sobre o algoritmo básico da NMF descritos até aqui, muitos outros foram sendo desenvolvidos ao longo dos últimos anos desde a primeira aparição da NMF em [4].

Uma extensão da NMF básica é a *Non-Negative Tensor Factorization (NTF)*, que lida com sinais estéreo, ou seja, sinais que utilizam dois canais [28]. Nesse caso, as matrizes são representadas como tensores, por possuírem mais do que duas dimensões.

Outra modificação da NMF básica é a chamada NMF Complexa [29]. Trata-se do uso da NMF em representações de tempo-frequência complexas dos sinais de mistura, ou seja, representações em que as fases dos sinais não são descartadas. O uso da fase na NMF faz com que o modelo de aditividade dos espectrogramas de cada fonte seja exato, e descarta o bloco de Síntese do sistema completo da NMF,

já que a fase dos sinais é considerada na fatoração. Entretanto, o algoritmo se torna muito mais complicado e lento.

Em [7] são mostrados outros aprimoramentos sobre a NMF2D. Um deles é a *Linear Non-Negative Matrix Factor 2-D Deconvolution (LNMF2D)*, publicada inicialmente em [30]. Conforme já mencionado, a NMF2D utiliza a CQT ou o mapeamento da escala linear para a logarítmica para obter o espectrograma na escala de temperamente igual. Entretanto, a reversibilidade desse espectrograma ao domínio do tempo não é exata. O objetivo da LNMF2D é contornar a necessidade do uso do espectrograma na escala logarítmica, aplicando um operador de deslocamento de acordo com a escala de igual temperamento sobre o espectrograma linear.

Outro problema da NMF2D consiste em considerar que um instrumento musical emite o mesmo padrão espectral para todas as notas. Esta pode ser considerada uma aproximação válida para intervalos pequenos de frequência, porém para intervalos maiores o erro pode ser muito grande. Assim, em [7] é também proposta uma adaptação da base espectral para contemplar a especificidade de cada instrumento em intervalos maiores de frequência.

Ainda em [7], propõe-se um algoritmo que execute a NMF de forma *online*, ou seja, em tempo real. Nesse caso, o sinal de mistura é dividido em blocos de tamanhos iguais e sobre cada bloco é aplicado o sistema completo da NMF mostrado na Figura 1.1, utilizando como estimativa o resultado obtido para o bloco anterior.

Capítulo 4

CNMF2D: Novas restrições sobre a NMF2D

4.1 Discussão inicial

No Capítulo 3, foram apresentados os principais métodos e aprimoramentos feitos sobre o algoritmo básico da NMF, proposto em [15]. O último método discutido foi a SNMF2D, que permite representar instrumentos—em vez de canais de frequências específicos ou notas—como fontes, além de adicionar um critério de esparsidade utilizando normas L_γ, com $\gamma > 0$, conforme pode ser visto na equação (3.28). Este é o estado-da-arte dos algoritmos de NMF para sinais musicais.

O presente capítulo tem por objetivo apresentar uma nova contribuição ao algoritmo da NMF2D. Se a NMF2D pode ser modificada de forma a contemplar um critério de esparsidade (sendo assim chamada de SNMF2D), é intuitivo pensar que critérios de outras naturezas também possam ser agregados a ela, de forma a atingir outros objetivos desejados ao fim do processo de convergência do algoritmo. É sobre esse aspecto que as linhas seguintes deste capítulo versarão.

4.2 *Constrained Non-Negative Matrix Factor 2-D Deconvolution (CNMF2D)*

A ideia aqui é apresentar um algoritmo mais genérico, que adicione outros critérios de restrição ao algoritmo da NMF2D, e não somente o critério de esparsidade com normas L_γ. A normalização vista na equação (3.30) e que é empregada na SNMF2D pode ser mantida para todos os outros critérios. Convencionou-se chamar esse algoritmo genérico de *Constrained Non-Negative Matrix Factor 2-D Deconvolution (CNMF2D)*.

Os critérios da CNMF2D propostos neste trabalho estão divididos em 3 grupos: critérios de esparsidade, sendo cada um referido por c_{ep}; critérios de correlação cruzada, sendo cada um referido por c_{cc}; e critérios de continuidade temporal, sendo cada um referido por c_{ct}. A seguir, serão vistos com detalhes cada um desses grupos e os critérios a eles associados.

4.2.1 Critérios de esparsidade

As normas L_γ não são o único grupo de funções que podem ser utilizadas para inserir o critério de esparsidade na NMF. Em [6] é sugerida também a função $c_{\text{ep}}(\mathbf{G}) = -|\log(1 + \mathbf{G} \odot \mathbf{G})|$ sobre o algoritmo básico da NMF. Pelo fato de essa equação ser função da matriz $\mathbf{G}$, ao adaptá-la para uso na CNMF2D, considera-se o número de deslocamentos no eixo das frequências ϕ (assim como foi feito para o critério com normas L_γ em [19]). Assim, o critério de esparsidade com logaritmo para a CNMF2D pode ser escrito como

$$c_{\text{ep}}(\mathbf{G}) = -\sum_{p=0}^{\phi-1} \left|\log\left(1 + \mathbf{G}^p \odot \mathbf{G}^p\right)\right|. \tag{4.1}$$

Partindo da equação anterior, a derivada $\nabla_{\mathbf{G}^p}(c_{\text{ep}}) = \frac{\partial c_{\text{ep}}(\mathbf{G})}{\partial \mathbf{G}^p}$, cujo cálculo é necessário durante o processo de convergência do algoritmo, conforme visto na Seção 3.6, é

$$\nabla_{\mathbf{G}^p}(c_{\text{ep}}) = -\frac{2\mathbf{G}^p}{(1 + \mathbf{G}^p \odot \mathbf{G}^p)}, \tag{4.2}$$

para $p = 0, 1, ..., \phi - 1$. A dedução da equação (4.2) pode ser vista na Seção B.1 do Apêndice B.

Conforme visto na Seção 3.6, as matrizes $\mathbf{B}^l$ também podem sofrer restrições. O conceito de esparsidade não se aplica apenas ao tempo, mas também à frequência: sinais esparsos na frequência são aqueles que preenchem apenas (poucos) canais do espectro de frequências, estando todo o restante dele vazio. Um exemplo de sinal desse tipo é uma nota emitida por um instrumento com *pitch* definido.

Mediante o exposto, deduz-se também ser possível aplicar critérios de esparsidade sobre as matrizes $\mathbf{B}^l$, que representam os vetores de base das frequências, de forma a tentar forçar a esparsidade na frequência como objetivo do algoritmo. Tanto o critério de esparsidade com normas L_γ quanto o critério de esparsidade com logaritmo poderiam restringir as matrizes $\mathbf{B}^l$ em vez das matrizes $\mathbf{G}^p$. Fazendo-se a modificação para o critério com normas L_γ por analogia com o que foi mostrado

para a SNMF2D na Seção 3.7, tem-se que

$$c_{\text{ep}}(\mathbf{B}) = \|\mathbf{B}\|_{\gamma} = \left(\sum_{l,n,m} \left| B^{l}_{n,m} \right|^{\gamma} \right)^{\frac{1}{\gamma}}, \tag{4.3}$$

em que $\mathbf{B}$ possui definição análoga à mostrada na equação (3.27) e $B^{l}_{n,m}$ representa cada elemento de $\mathbf{B}^{l}$. Assim, tem-se que

$$\nabla_{\mathbf{B}^{l}}(c_{\text{ep}}) = \frac{\mathbf{B}^{l^{\bullet(\gamma-1)}}}{\|\mathbf{B}\|_{\gamma}^{\gamma-1}}. \tag{4.4}$$

Para o critério com logaritmo, a equação que restringe as matrizes $\mathbf{B}^{l}$ seria análoga à equação (4.1)

4.2.2 Critérios de correlação

Ao fim do processo de separação, foi visto que as fontes sonoras obtidas podem ser sinais que correspondem a canais de frequências diferentes entre si (no caso da NMF comum), notas musicais diferentes entre si (no caso da NMFD) ou ainda instrumentos musicais inteiros e diferentes entre si (no caso da NMF2D). Entretanto, cada fonte pode ainda conter resíduos das outras fontes, já que a separação, por ser um processo computacional de aproximação, nunca será perfeita.

A fim de mitigar esse problema, pode-se utilizar o conceito de autocorrelação das fontes e correlação cruzada entre as fontes [5] [6] [31]. Ao se fazer $\mathbf{G}^{p}\mathbf{G}^{pT}$ para determinado valor de p, será obtida uma matriz de dimensão MxM (lembrando que M, para a NMF2D e a CNMF2D, corresponde ao número de instrumentos que se deseja separar), que pode ser escrita da forma

$$\mathbf{G}^{p}\mathbf{G}^{pT} = \begin{bmatrix} \sum_{t=0}^{T-1} g_1^p(t)g_1^p(t) & \sum_{t=0}^{T-1} g_1^p(t)g_2^p(t) & \cdots & \sum_{t=0}^{T-1} g_1^p(t)g_M^p(t) \\ \sum_{t=0}^{T-1} g_2^p(t)g_1^p(t) & \sum_{t=0}^{T-1} g_2^p(t)g_2^p(t) & \cdots & \sum_{t=0}^{T-1} g_2^p(t)g_M^p(t) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=0}^{T-1} g_M^p(t)g_1^p(t) & \sum_{t=0}^{T-1} g_M^p(t)g_2^p(t) & \cdots & \sum_{t=0}^{T-1} g_M^p(t)g_M^p(t) \end{bmatrix}, \tag{4.5}$$

em que T, novamente, é o número total de quadros do sistema, e

$$\mathbf{G}^p = \begin{bmatrix} g_1^p(0) & g_1^p(1) & \cdots & g_1^p(T-1) \\ g_2^p(0) & g_2^p(1) & \cdots & g_2^p(T-1) \\ \vdots & \vdots & \ddots & \vdots \\ g_M^p(0) & g_M^p(1) & \cdots & g_M^p(T-1) \end{bmatrix}, \tag{4.6}$$

onde cada linha corresponde aos ganhos de cada fonte em cada quadro de tempo, até a M-ésima fonte.

Pode-se observar que os M elementos da diagonal principal da matriz $\mathbf{G}^p\mathbf{G}^{pT}$ estão relacionados às autocorrelações de cada uma das M fontes a serem obtidas, enquanto que os demais elementos estão relacionados às correlações cruzadas entre essas fontes. O objetivo desse critério é reduzir a correlação cruzada entre as fontes, mantendo a autocorrelação de cada fonte. Isso pode ser feito com o auxílio de uma matriz $\mathbf{W}$ também de dimensão MxM, resultando em

$$c_{\text{cc}}(\mathbf{G}) = \sum_{p=0}^{\phi-1} \left| \mathbf{W} \odot \left(\mathbf{G}^p \mathbf{G}^{pT} \right) \right|. \tag{4.7}$$

A fim de penalizar somente as correlações cruzadas entre as fontes a serem obtidas do sinal de mistura, deve-se igualar a zero todos os elementos da diagonal principal de $\mathbf{W}$, enquanto os outros elementos são igualados à unidade. A derivada $\nabla_{\mathbf{G}^p}(c_{\text{cc}})$, por sua vez, é escrita como

$$\nabla_{\mathbf{G}^p}(c_{\text{cc}}) = 2\mathbf{W}\mathbf{G}^p, \tag{4.8}$$

para $p = 0, 1, ..., \phi - 1$. A dedução da equação (4.8) ser vista na Seção B.2 do Apêndice B.

O critério de correlação com o objetivo de cancelamento das correlações cruzadas entre as fontes também pode ser escrito com uma equação que não faz uso da matriz $\mathbf{W}$, subtraindo-se os termos de autocorrelação das fontes dos termos referentes às correlação cruzadas entre elas. Assim,

$$c_{cc}(\mathbf{G}) = \sum_{p=0}^{\phi-1} \left(\left| \mathbf{G}^{pT} \mathbf{G}^p \right| - \left| \mathbf{G}^p \odot \mathbf{G}^p \right| \right). \tag{4.9}$$

Para esta equação, a derivada $\nabla_{\mathbf{G}^p}(c_{\text{cc}})$ é escrita como

$$\nabla_{\mathbf{G}^p}(c_{\text{cc}}) = 2\mathbf{G}^p \left(\mathbf{1} - \mathbf{I} \right), \tag{4.10}$$

para $p = 0, 1, ..., \phi - 1$. Tem-se que $\mathbf{1}$ e $\mathbf{I}$ são uma matriz de 1s ("uns") e uma matriz identidade, respectivamente, ambas de dimensão TxT. A dedução da equação (4.10) pode ser vista na Seção B.3 do Apêdice B.

Outro critério de correlação, dessa vez com o objetivo de enfatizar as autocorrelações de cada fonte em vez de tentar cancelar as correlações cruzadas entre elas, pode ser escrito de maneira análoga ao critério da equação 4.7, ou seja,

$$c_{cc}(\mathbf{G}) = -\sum_{p=0}^{\phi-1} \left| \mathbf{Z} \odot \left(\mathbf{G}^p \mathbf{G}^{pT} \right) \right|, \tag{4.11}$$

com a diferença de que a matriz $\mathbf{Z}$ é Identidade, selecionando somente os elementos da diagonal principal de $\mathbf{G}^p\mathbf{G}^{pT}$, ou seja, aqueles que possuem as autocorrelações das fontes. Da mesma maneira, a derivada $\nabla_{\mathbf{G}^p}(c_{\mathrm{cc}})$ é escrita como

$$\nabla_{\mathbf{G}^p}(c_{\mathrm{cc}}) = -2\mathbf{Z}\mathbf{G}^p, \tag{4.12}$$

sendo a dedução similar à apresentada na Seção B.2.

4.2.3 Critérios de continuidade temporal

Um sinal possui continuidade (ou suavidade) temporal quando, ao se considerar um pequeno trecho contendo alguns quadros deste sinal, as características acústicas do sinal entre esses quadros variam pouco. Em outras palavras, se as correlações entre determinado quadro e quadros adjacentes for alta, este sinal possui continuidade (ou suavidade) temporal dentro do trecho que contém tais quadros.

A continuidade temporal é uma característica presente não apenas em sinais musicais, mas em grande parte dos sons naturais e cotidianos [10]. Por este motivo, trata-se de um critério interessante a ser adicionado à CNMF2D.

A equação do critério de continuidade temporal é semelhante à do critério de correlação cruzada com uso da matriz $\mathbf{W}$. Aqui, é feito o produto $\mathbf{G}^{pT}\mathbf{G}^p$, ou seja, com ordem invertida da que foi usada com o critério correlação cruzada. Este

produto gera uma matriz dimensão TxT, da forma

$$\mathbf{G}^{pT}\mathbf{G}^{p} = \begin{bmatrix} \sum_{m=0}^{M} g_m^p(1)g_m^p(1) & \sum_{m=0}^{M} g_m^p(1)g_m^p(2) & \cdots & \sum_{m=0}^{M} g_m^p(1)g_m^p(T) \\ \sum_{m=0}^{M} g_m^p(2)g_m^p(1) & \sum_{m=0}^{M} g_m^p(2)g_m^p(2) & \cdots & \sum_{m=0}^{M} g_m^p(2)g_m^p(T) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{m=0}^{M} g_m^p(T)g_m^p(1) & \sum_{m=0}^{M} g_m^p(T)g_m^p(2) & \cdots & \sum_{m=0}^{M} g_m^p(T)g_m^p(T) \end{bmatrix}. \quad (4.13)$$

Nesse caso, como se pode observar, é levada em consideração a correlação entre todos os T quadros, e não entre as fontes (para um mesmo quadro) a serem produzidas. Novamente, é necessário o uso de uma matriz de ajuste do critério, chamada aqui de $\mathbf{V}$. Assim, tem-se que [6]

$$c_{\text{ct}}(\mathbf{G}) = -\sum_{p=0}^{\phi-1} \left| \mathbf{V} \odot \left(\mathbf{G}^{pT}\mathbf{G}^{p} \right) \right|. \quad (4.14)$$

A matriz $\mathbf{V}$ é uma matriz de Toeplitz[1], de dimensão TxT, utilizada para fazer a ponderação das correlações entre cada quadro e os demais quadros. É importante lembrar que a equação (4.14) possui sinal negativo. Isso significa que o que se deseja é maximizar o somatório.

A matriz $\mathbf{V}$ é construída de forma que os valores dos elementos das diagonais sigam um padrão exponencial decrescente, partindo da diagonal principal em direção às diagonais à esquerda e à direita. Assim, garante-se que as correlações entre quadros próximos serão maiores que as correlações entre quadros distantes.

A derivada $\nabla_{\mathbf{G}^p}(c_{\text{ct}})$, por sua vez, é escrita como

$$\nabla_{\mathbf{G}^p}(c_{\text{ct}}) = -2\mathbf{G}^{p}\mathbf{V}, \quad (4.15)$$

para $p = 0, 1, ..., \phi - 1$. A dedução da equação (4.15) pode ser vista na Seção B.4 do Apêdice B.

Outro critério de continuidade temporal passível de ser adaptado à CNMF2D pode ser visto em [10]. Fazendo-se a devida adaptação, o critério pode ser escrito

[1]Também conhecida como *matriz de diagonais constantes*, é uma matriz cujas diagonais descendentes da esquerda para a direita possuem elementos iguais.

da forma

$$c_{\text{ct}}(\mathbf{G}) = \sum_{p=0}^{\phi-1} \sum_{m=1}^{M} \sum_{j=2}^{T} \left(G_{m,j}^{p} - G_{m,j-1}^{p}\right)^2, \tag{4.16}$$

em que está explícito o objetivo de minimizar as diferenças entre quadros consecutivos dos sinais, tentando torná-los mais contínuos. A obtenção da derivada $\nabla_{\mathbf{G}^p}(c_{\text{ct}})$, por sua vez, é trivial. Ela é escrita como

$$\nabla_{\mathbf{G}^p}(c_{\text{ct}}) = 2 \sum_{m=1}^{M} \sum_{j=2}^{T} \left(G_{m,j}^{p} - G_{m,j-1}^{p}\right), \tag{4.17}$$

para $p = 0, 1, ..., \phi - 1$.

Capítulo 5

Processamento, Síntese e Avaliação

5.1 Discussão inicial

Conforme visto na Figura 1.1 do Capítulo 1, o sistema de separação de fontes por NMF é constituído por cinco etapas. As primeiras duas etapas, que são a de Análise tempo-frequencial do sinal de mistura e a de Fatoração do espectrograma da mistura, já foram discutidas nos Capítulos 2 e 3, respectivamente. Aqui, neste capítulo, serão discutidas brevemente as três etapas restantes: Processamento dos espectrogramas separados, Síntese de sinais e Avaliação da qualidade da separação. Essas etapas não são o foco deste trabalho, mas são essenciais para a conclusão do estudo.

O Processamento dos espectrogramas separados tem duas finalidades: reconstruir cada espectrograma obtido a partir das matrizes não-negativas produzidas na separação e realizar uma filtragem a fim de realçar tais espectrogramas. A Síntese dos sinais, por sua vez, tem por objetivo transformar cada espectrogramas separado em um sinal no domínio do tempo através de um processo de estimação de fase. Com as fases estimadas, os sinais podem então passar pela etapa de Avaliação de qualidade, cuja finalidade é, como o nome já diz, aferir a qualidade da separação.

5.2 Processamento dos espectrogramas separados

Em um sistema de separação de fontes, existem dois tipos importantes de processamentos pós-separação que atuam sobre os espectrogramas separados: reconstrução dos espectrogramas a partir dos fatores, e filtragem desses espectrogramas.

A reconstrução do m-ésimo espectrograma separado (para $m = 1, 2, ..., M$), a partir dos fatores obtidos, se dá através da combinação entre seu correspondente vetor-coluna de base de frequências $\mathbf{b}_m$ com seu respectivo vetor-linha de base de

tempo $\mathbf{g}_m$ (ganhos), como mostrado nas equações (3.2) e (3.3). Para o caso da separação duplamente deconvolutiva (NMF2D e, em geral, CNMF2D), a reconstrução do espectrograma separado $\hat{\mathbf{X}}_m^{\log}$—que está relacionado ao instrumento m e se encontra com o eixo das frequências em escala logarítmica—é feita da forma

$$\hat{\mathbf{X}}_m^{\log} = \sum_{l=0}^{\tau-1} \sum_{p=0}^{\phi-1} \mathbf{B}_m^{\downarrow p} \mathbf{G}_m^{\rightarrow l} \,. \tag{5.1}$$

Para mapear cada espectrograma $\hat{\mathbf{X}}_m^{\log}$ para a escala linear, é necessária a realização de um procedimento inverso ao procedimento descrito pela equação (3.23). A solução adotada neste trabalho para essa transformação inversa é o procedimento iterativo

$$\hat{\mathbf{X}}_m = \hat{\mathbf{X}}_m \odot \frac{\mathbf{C}\hat{\mathbf{X}}_m^{\log}}{\mathbf{C}^T\mathbf{C}\hat{\mathbf{X}}_m^{\log}}, \tag{5.2}$$

que pode ser visto como o algoritmo básico da NMF aplicado à fatoração $\hat{\mathbf{X}}_m^{\log} = \hat{\mathbf{X}}\mathbf{C}$ para descobrir um dos fatores—no caso, a matriz $\hat{\mathbf{X}}$. Outras soluções existentes na literatura são as transformações pela *pseudoinversa* e pela *transposta*, ambas simples. Entretanto, a solução iterativa, apesar de ser computacionalmente mais complexa em relação as outras duas citadas, é a que consegue o melhor resultado no sentido dos mínimos quadrados (menor diferença entre o que é obtido e o que é desejado) [25] [7].

Já a filtragem do espectrograma é uma etapa subsequente à de reconstrução, ou seja, cada espectrograma separado pode passar por um processo de filtragem. Essa não é uma etapa obrigatória do processo geral de separação de fontes; porém, este procedimento atua no refinamento dos espectrogramas de forma a realçar as características desejadas de cada fonte. Um tipo usual de filtragem é a de Wiener, escrita como [7]

$$\hat{\mathbf{X}}_m^f = \frac{\hat{\mathbf{X}}_m^2}{\sum_{m'=1}^{M} \hat{\mathbf{X}}_{m'}^2} \odot \mathbf{X}, \tag{5.3}$$

em que $\hat{\mathbf{X}}_{m'}^2$ representa o espectro de potência do sinal referente à fonte m, e a razão $\dfrac{\hat{\mathbf{X}}_m^2}{\sum_{m'=1}^{M} \hat{\mathbf{X}}_{m'}^2}$ implementa o Filtro de Wiener. Dessa forma, dado o espectrograma da mistura $\mathbf{X}$, o sinal $\hat{\mathbf{X}}_m^f$ é uma nova estimativa sobre o espectrograma da fonte m. Outros tipos de filtragens existentes são o Mascaramento Espectral, o Cancelamento Cruzado e a Máscara Binária. Todos esses métodos são descritos

e comparados em [30]. Entretanto, o filtragem de Wiener foi a que apresentou os melhores resultados, e portanto foi adotada neste trabalho.

5.3 Síntese dos sinais separados

A Síntese dos sinais separados corresponde ao quarto bloco do sistema geral de separação de fontes por NMF, como pode ser visto na Figura 1.1. Trata-se do processo de estimação das fases associadas aos espectrogramas de magnitude obtidos após a Fatoração. A separação de fontes por NMF descarta a informação de fase do sinal de mistura pelo fato de ser um processo que se dá no domínio da magnitude ou potência, devido à exigência de não-negatividade. Portanto, os espectrogramas obtidos ao final da etapa de Processamento, naturalmente, não terão fases associadas a eles.

A fase é importante para que se tenha na saída do sistema de separação sinais no domínio do tempo. Entretanto, em geral, os espectrogramas obtidos pelo processo de separação não são válidos, ou seja, não existem sinais reais (no domínio do tempo) que possuam tais espectrogramas de magnitude [32], já que foram obtidos através de um processo de estimação. Esses espectrogramas são conhecidos pelos métodos de estimação de fase como **espectrogramas modificados**. O sentido da palavra "modificado" nesse caso se refere ao fato de que esses espectrogramas seriam estimativas próximas do que seriam os espectrogramas válidos dos sinais reais.

Uma solução para o problema da falta de fases associadas aos espectrogramas modificados está na utilização da própria fase da mistura. É uma solução simples e bastante usada na literatura. Porém, quanto mais os espectrogramas forem modificados, ou seja, quanto mais eles estiverem distantes do que seriam os espectrogramas dos sinais reais, pior será o resultado da síntese. Para o processamento de sinais musicais, por exemplo, utilizar a própria fase da mistura em espectrogramas que correspondem a ataques rápidos é prejudicial, pois nesse caso é necessária a fase mais precisa possível para boa determinação do instante do ataque. Alguns métodos desenvolvidos na literatura tratam da estimação (reconstrução) das fases que estarão associadas aos espectrogramas modificados (separados).

O algoritmo de Griffin e Lim (G&L) foi um dos primeiros a serem desenvolvidos. O objetivo é encontrar um sinal no domínio do tempo $y(t)$ (para $t = 0, 1, 2..., T$) cujo espectrograma—aqui denominado de Magnitude da Transformada de Fourier de Curta Duração (do inglês, *Short-Time Fourier Transform Magnitude*, STFTM)—esteja o mais próximo possível, no sentido dos mínimos quadrados, do espectrograma

modificado—aqui denominado de Magnitude Modificada da Transformada de Fourier de Curta Duração (do inglês, *Modified Short-Time Fourier Transform Magnitude*, MSTFTM). A fase de $y(t)$ pode ser inicializada, por exemplo, com zeros, e ao longo das iterações vai sendo obtida junto com a STFTM através da minimização. Entretanto, a cada iteração, é a MSTFTM que é fixada para a composição do sinal $y(t)$ junto com a fase obtida. A STFTM é utilizada apenas para que se encontre a fase associada. O algoritmo G&L pode ser visto em [32].

No algoritmo G&L o sinal é processado como um todo, exigindo o cálculo do máximo de transformadas de Fourier (para todos os blocos do sinal). Isso significa que, na reconstrução de cada quadro, utiliza-se a informação dos quadros passados e futuros. Assim, o algoritmo G&L é inviável para aplicações em tempo real, por exigir o cálculo de transformadas de Fourier para todos os blocos do sinal.

Com o objetivo de contornar essas limitações, foi desenvolvido o algoritmo *Real-Time Iterative Spectrogram Inversion* (RTISI) [33], em que é estimado um quadro de cada vez através de um conjunto de iterações. Além disso, no RTISI, é possível utilizar uma inicialização melhor para a fase. Isso se deve à sobreposição dos quadros. Por exemplo: com uma sobreposição de $L = 4S$, o quadro a ser estimado possui contribuição das três amostras anteriores. Assim, a inicialização da fase não precisa ser feita com zeros; é possível começar com uma estimativa mais coerente.

O RTISI atende aos requisitos estruturais e computacionais para uma reconstrução em tempo real. O primeiro requisito diz respeito ao uso de somente informações provenientes dos quadros anteriores e do corrente para uma reconstrução quadro a quadro. Já o segundo diz respeito à pouca quantidade de computação requerida para que o algoritmo seja rápido o suficiente para ser aplicável em tempo real. O algoritmo RTISI pode ser visto em [33].

Embora o algoritmo RTISI atenda aos dois requisitos mencionados anteriormente, não é atendido o requisito de flexibilidade, que é: um bom algoritmo de reconstrução em tempo real deve reconstruir sinais com melhor qualidade se forem usados mais recursos computacionais. Dessa forma, o algoritmo deve ser adaptável (flexível) para funcionar em aplicações com diferentes demandas de qualidade *versus* velocidade de processamento. No RTISI, o desempenho em termos do erro entre as STFTMs (do sinal dado e do reconstruído) converge para uma assíntota e não é melhorado, mesmo que mais iterações sejam realizadas. Isso acontece devido ao fato de o RTISI somente utilizar informações proveninentes do quadro corrente e dos anteriores.

Visando a atender também o requisito de flexibilidade, desenvolveu-se o algoritmo denominado *Real-Time Iterative Spectrogram Inversion with Look-Ahead (RTISI-LA)* [34]. O RTISI-LA é uma extensão do RTISI, em que o termo adicional *look-ahead* (*olhar adiante*) indica uma estratégia em que um determinado número de quadros futuros influenciam na reconstrução do quadro corrente. Essa estratégia não compromete a adequação do algoritmo aos requisitos estruturais e computacionais necessários para aplicações em tempo real, pois, apesar de a carga computacional do RTISI-LA ser um pouco maior do que a do RTISI comum, geralmente apenas um pequeno número de quadros futuros são utilizados. Além do mais, a reconstrução continua sendo feita quadro a quadro.

O que deve ser levado em consideração é a flexibilidade que o RTISI-LA oferece. Em uma aplicação cuja preferência seja a qualidade da reconstrução (podendo haver certo retardo), basta serem utilizados mais quadros para o *look-ahead*. Por outro lado, em uma aplicação cuja meta principal seja a instantaneidade da reconstrução (em detrimento de certa qualidade), basta que poucos quadros (ou até mesmo nenhum) sejam utilizados para o *look-ahead*. O algoritmo do RTISI-LA pode ser visto em [34] [35]. A Tabela 5.1 mostra as principais diferenças entre os algoritmos aqui mencionados.

É interessante notar que, mesmo não havendo informação de fase nos sinais estimados resultantes da NMF, o fato de esta informação poder ser obtida a partir da STFTM, leva à reflexão de que cada STFTM guarda consigo o "molde" para a reconstrução da fase adequada. Daí, pode-se pensar que, de certa forma, há informação de fase contida na STFTM, que só precisa ser explicitada através de algum método de reconstrução.

Tabela 5.1: *Diferença entre os algoritmos de G&L, RTISI e RTISI-LA*

G&L	**RTISI**	**RTISI-LA**
Não pode ser usado em aplicações em tempo real	Ideal para aplicações em tempo real	Ideal para aplicações em tempo real
Estimativa de todos os *frames* a cada iteração	Estimativa *frame* a *frame*	Estimativa *frame* a *frame*
Não utiliza informação prévia	Utiliza informação de *frames* já reconstruídos	Utiliza informação de *frames* já reconstruídos e de *frames* futuros
Processamento mais lento	Processamento mais rápido	Flexível. Depende do número de *frames* futuros utilizados
Maior qualidade de reconstrução	Menor qualidade de reconstrução	Flexível. Depende do número de *frames* futuros utilizados

5.4 Avaliação de qualidade

5.4.1 Avaliação da separação

Para que se possa avaliar a qualidade de uma separação sob vários aspectos, é possível modelar um sinal de áudio separado, chamado $\hat{s}$ (sendo esta a estimativa de um sinal real s), da forma [36]

$$\hat{s} = s + e_{\mathrm{i}} + e_{\mathrm{a}} + e_{\mathrm{r}}, \tag{5.4}$$

onde s corresponde ao sinal original (o sinal "limpo"); e_{i} correponde à interferência causada por outras fontes (o resíduo de outro sinal presente no sinal que se deseja avaliar); e_{a} corresponde aos artefatos gerados pelo processo de separação; e e_{r} é o ruído (caso haja ruído na mistura).

Uma das medidas mais importantes utilizadas para avaliar quantitativamente a separação é denominada de Razão Fonte-Distorção (do inglês *Source-to-Distortion Ratio*, SDR), calculada como:

$$\mathrm{SDR} = 10 \log \frac{\|s\|^2}{\|e_i + e_a + e_r\|^2}. \tag{5.5}$$

Uma outra medida é denominada Razão Fonte-Interferência (do inglês *Source-to-Intererence Ratio*, SIR), e não considera o ruído e os artefatos gerados pelo método (caso haja). É calculada como:

$$\mathrm{SIR} = 10 \log \frac{\|s\|^2}{\|e_i\|^2}. \tag{5.6}$$

Há ainda uma terceira medida, que considera somente os artefatos gerados. É a Razão Fonte-Artefato (do inglês *Source-to-Artifact Ratio*, SAR):

$$\mathrm{SAR} = 10 \log \frac{\|s\|^2}{\|e_a\|^2}. \tag{5.7}$$

Existem também formas de avaliação da separação baseadas em Psicoacústica, em que o sinal de áudio é transformado do domínio do tempo para o domínio psicoacústico. Isso significa que se busca, através de um complexo processamento não-linear, simular a percepção sonora humana. O PEAQ (*Perceptual Evaluation of Audio Quality*) é um exemplo desse tipo de avaliador [7]. A saída desse medidor é a ODG (*Objective Difference Grade*), que é uma nota dada pelo sistema ao sinal de áudio separado com base no sinal original, podendo variar de -4 (degradação muito

incômoda) a 0 (degradação imperceptível).

É importante mencionar que, para todas essas medidas, é necessário haver os sinais de referência não-misturados. Além do mais, dentre elas, a SIR é a mais interessante para a separação de fontes, por ser a única que verifica os efeitos das interferências resíduais entre sinais. O PEAQ é muito sensível a pequenas distorções, não sendo bem aplicado neste trabalho. Portanto, a SIR foi a única avaliação objetiva considerada. Naturalmente, avaliações subjetivas informais também foram feitas com base na audição dos sinais, objetivando caracterizar qualitativamente os resultados e tentar interpretá-los face aos respectivos algoritmos.

Capítulo 6

Experimentos

6.1 Introdução

Neste capítulo são apresentadas as descrições dos experimentos realizados com os resultados obtidos através de cada um deles. Porém, antes disso, são descritas informações relativas ao *setup* dos testes, pois sendo a CNMF2D dotada de uma grande quantidade de parâmetros que poderiam ser variados (acarretando um volume enorme de testes), deve-se fixar alguns deles de forma a dar atenção somente aos parâmetros que estão sob maior foco, segundo a proposta deste trabalho. Naturalmente, antes dos testes, são também apresentados os sinais utilizados neste trabalho.

6.2 Parâmetros da CNMF2D

A Tabela 6.1 apresenta os principais parâmetros relativos à CNMF2D.

6.3 Fixação de parâmetros

Dentre os parâmetros apresentados na Tabela 6.1, os menos relevantes para a proposta deste trabalho (mostrados na Tabela 6.2) tiveram que ser fixados com os mesmos valores para todo o conjunto de testes de forma que os esses testes pudessem ser devidamente comparados. Como a Fatoração é a etapa principal de todo o sistema de separação estudado, somente seus parâmetros (mostrados na Tabela 6.3) foram deixados em aberto para serem modificados de acordo com os experimentos e sinais utilizados.

O número de pontos da STFT é o dobro do tamanho da janela pelo fato de que, no espectrograma, apenas a metade positiva dos canais de frequências obtidos pela

Tabela 6.1: *Parâmetros da CNMF2D.*

Etapa	**Parâmetros correspondentes**
Análise da mistura	• Número de pontos da FFT; • Tamanho da janela (L); • Tipo de janela; • Passo de sobreposição (S).
Fatoração	• Medida de distorção; • Número de fontes (M); • Número de quadros permitidos para caracterizar uma fonte (τ); • Número de canais de frequência permitidos para caracterizar uma fonte (ϕ); • Frequências mínima e máxima; • Resolução desejada (b); • Critérios associados à CNMF2D ; • Pesos dos critérios.
Processamento	• Método de mapeamento inverso; • Filtragem.
Síntese	• Algoritmo de síntese.

STFT é considerada. Assim, durante a transformação, o número de pontos restantes da STFT iguala-se ao tamanho da janela. A janela de Hamming é uma escolha tpica para a análise em blocos de sinais de áudio, e tem seus lobos secundários bastante atenuados em relação ao lobo principal, conforme visto no Capítulo 2. O passo de sobreposição para a análise da mistura também está intimamente ligado à etapa de Síntese dos sinais. As escolhas de 1/4 do tamanho da janela (ou 75 % de sobreposição) para o passo de sobreposição e do algoritmo RTISI-LA com *look-ahead* igual a 3 mostraram-se uma combinação com desempenho satisfatório em termos de velocidade de processamento *versus* qualidade da síntese. As escolhas dos parâmetros mencionados neste parágrafo são baseadas em [33], [35] e [34].

A medida de distorção escolhida é a divergência de Kullback-Leibler generalizada, por ser mais adequada para aplicações em áudio, conforme visto no Capítulo 3. As frequências mínimas e máximas consideradas para a separação são, naturalmente, os limites da audição humana. A resolução adotada para o mapeamento do espectrograma de mistura da escala linear para a logarítmica foi 24, ou seja, de um quarto de tom, considerada uma boa escolha para a aplicação [7].

Por fim, considerando a etapa de Processamento dos espectrogramas obtidos, para o mapeamento reverso dos espectrogramas da escala logarítmica para a escala linear foi adotado o método iterativo, e para a filtragem dos mesmos foi adotado o Filtro de Wiener, ambas escolhas justificadas no Capítulo 5.

Tabela 6.2: *Parâmetros da CNMF2D que foram fixados para a realização dos testes.*

Etapa	**Parâmetros correspondentes**
Análise da mistura	• Número de pontos da STFT: 2048; • Tamanho da janela (L): 1024; • Tipo de janela: Hamming; • Passo de sobreposição (S): 256.
Fatoração	• Medida de distorção: Kullback-Leibler; • Frequências mínima e máxima: 20 Hz a 20 kHz; • Resolução desejada (b): 24.
Processamento	• Método de mapeamento inverso: Algoritmo iterativo; • Filtragem: Wiener.
Síntese	• Algoritmo de síntese: RTISI-LA (*look-ahead*: 3).

Tabela 6.3: *Parâmetros da CNMF2D que foram variados.*

Parâmetros variáveis
Número de quadros permitidos para caracterizar uma fonte (τ);
Número de canais de frequência permitidos para caracterizar uma fonte (ϕ);
Número de fontes (M);
Critérios associados à CNMF2D;
Pesos dos critérios.

6.4 Banco de sinais

Todas as misturas utilizadas são mono[1] e estão no formato *wav*. As Tabelas 6.4 e 6.5 apresentam a lista das misturas que foram utilizadas, com uma descrição para cada uma, contendo o nome—que servirá para referenciar os sinais ao longo deste capítulo—, o número de fontes (M) nas quais se deseja fatorar a mistura, a natureza (*pitched*[2], P, ou *unpitched* [3], U), o número de quadros considerados para abranger uma nota (τ), o número de deslocamentos permitidos no eixo das frequências (ϕ), a frequência de amostragem em Hertz F_s, e a duração.

O sinal *piano_trompete* é composto por notas de trompete misturadas com notas de piano. Ao longo do trecho, as notas emitidas pelos dois instrumentos são as mesmas. Trata-se de uma mistura realizada sinteticamente para este trabalho, estando disponíveis seus sinais formadores originais.

O sinal *órgão_prato* é composto por uma nota longa de órgão misturada com três batidas de prato. Esta mistura tem a característica de ser formada por duas fontes com características duais: esparsa na frequência e longa no tempo (órgão); e esparsa no tempo e larga na frequência (prato). Essa também foi uma mistura realizada sinteticamente para este trabalho.

O sinal *Paganini* é composto por mistura entre violoncelo e piano, em que ambos emitem notas rápidas. As notas possuem menos sobreposição entre si. Trata-se de uma mistura natural, retirada de gravação comercial.

O sinal *Bach* é composto por mistura entre violoncelo e pianos, em que ambos emitem notas longas. As notas possuem mais sobreposição em si. Trata-se de uma mistura natural, retirada de gravação comercial.

O sinal *Far More Drums* é composto por mistura entre baixo, bateria e piano. O baixo possui notas graves, a bateria em questão usa caixa e pratos, e o piano possui emissão bem marcada. Trata-se de uma mistura natural, retirada de gravação comercial.

O sinal *Take*5 é composto por mistura entre baixo, bateria, saxofone e piano. Possui características parecidas com o sinal anterior, sendo que o saxofone emite

[1] Provenientes de um único canal.
[2] Com *pitch* definido.
[3] Com *pitch* indefinido.

notas melódicas longas e a bateria possui predomínio de pratos. Novamente, trata-se de mistura natural, retirada de gravação comercial.

Os valores de τ e ϕ variam de acordo com a misturas. Suas escolhas foram empíricas, procurando manter o equilíbrio entre tamanho das matrizes e adequabilidade às misturas, atendendo às características dos sinais que as compõem.

Tabela 6.4: *Sinais utilizados: descrição.*

Nome	**Descrição**
piano_trompete	Trompete misturado com piano, emitindo as mesmas notas
órgão_prato	Nota longa de órgão com três batidas de prato
Paganini	Violoncelo misturado com piano em notas rápidas, com menos sobreposição
Bach	Violoncelo misturado com piano em notas longas, com mais sobreposição
Far_More_Drums	Mistura entre baixo, bateria e piano
*Take*5	Mistura entre baixo, bateria, saxofone e piano

Tabela 6.5: *Sinais utilizados: características.*

Nome	$\boldsymbol{M}$	**P/U**	τ	ϕ	F_s (Hz)	**Duração (s)**
piano_trompete	2	P	3	10	8000	4,1
órgão_prato	2	P/U	3	20	44100	3,8
Paganini	2	P	3	10	44100	4,0
Bach	2	P	3	10	44100	4,0
Far_More_Drums	3	P/U	3	10	44100	5,6
*Take*5	4	P/U	5	20	44100	3,0

6.5 Inicialização e convergência

A convergência dos algoritmos de separação tem sido um problema a ser contornado. Constatou-se nos testes preliminares que nem sempre o algoritmo converge para a solução desejada. Quando a convergência não ocorre, o erro passa a crescer a partir de uma determinada iteração e, para não ultrapassar o limite de representação do computador, o algoritmo é parado. Para todos os experimentos, o número máximo de iterações da fatoração escolhido foi de 1000, e a inicialização das matrizes $\mathbf{B}^l$ e $\mathbf{G}^p$ foi feita de forma aleatória com distribuição uniforme entre 0 e 1.

A fatoração é considerada correta (ou seja, converge para o resultado esperado) quando cada sinal separado obtido corresponde a somente uma das fontes da mistura, e quando o valor da SIR possui valores positivos (lembrando que os valores estão em dB), nos casos em que é possível utilizar avaliação objetiva.

6.6 NMF2D *versus* CNMF2D

6.6.1 Objetivo

Este foi um experimento prévio realizado com o objetivo de comparar a CNMF2D proposta neste trabalho com a NMF2D comum proposta em [21]. Para se realizar uma comparação coerente, a CNMF2D deve estar com as restrições inativas, ou seja, em teoria funcionando como a NMF2D comum. Entretanto, como proposto em [19] e considerado neste trabalho, a CNMF2D possui a normalização vista na equação (3.30), o que acaba por incluir em suas equações de atualização termos que não existem nas equações de atualização da NMF2D. Portanto, o objetivo deste teste é verificar se há algum impacto resultante dessa diferença.

6.6.2 Descrição

Para este experimento, utilizou-se a mistura *piano_trompete*. O maior impacto obervado foi na convergência, cuja curva, para o caso da CNMF2D, decresce mais lenta e monotonicamente durante algumas iterações, até apresentar uma queda final brusca. A maior lentidão na convergência é um efeito esperado, tendo em vista a maior complexidade das equações de atualização da CNMF2D com relação à NMF2D. Em termos de qualidade da separação, não foram observadas diferenças. A Figura 6.1 apresenta o gráfico de convergência da NMF2D e da CNMF2D, respectivamente. Utilizando como limiar o valor de $0,1$ para a divergência de Kullback-Leibler, observou-se que a NMF2D convergiu com 137 iterações enquanto a CNMFD convergiu com 567 iterações.

6.7 Teste de pesos

6.7.1 Objetivo

Este foi um experimento prévio realizado com o objetivo de verificar a influência dos pesos dos critérios (restrições) na fatoração. Em teoria, um peso muito pequeno em determinada restrição faz com que tal restrição não seja tão contemplada pelo algoritmo, tendendo a fatoração a produzir um resultado mais próximo do da

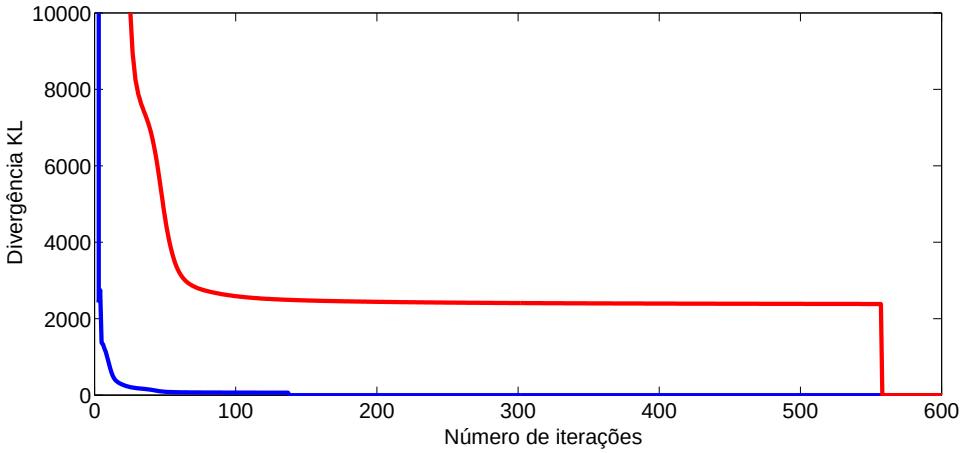

Figura 6.1: Comparação entre as convergências da NMF2D (linha azul) e da CNMFD (linha vermelha).

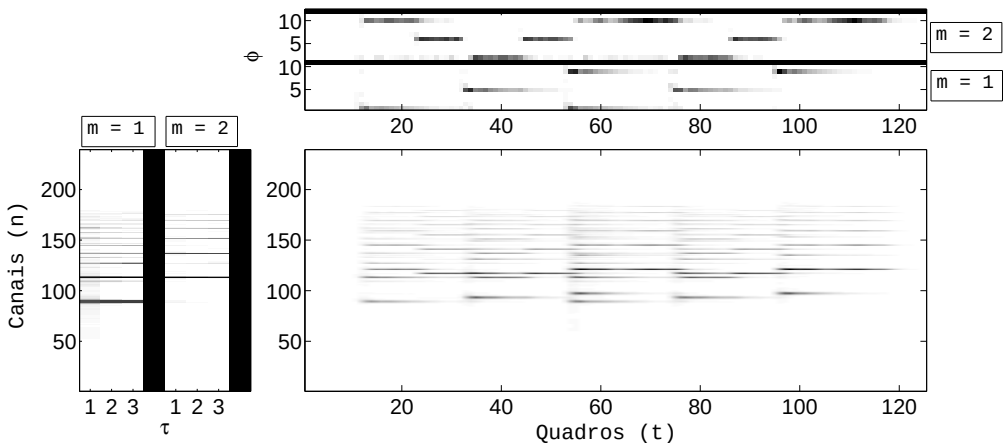

Figura 6.2: Resultado da separação por CNMF2D somente com critério de esparsidade ativado e peso igual a 1. n representa cada canal (raia) de frequência de um total de N canais.

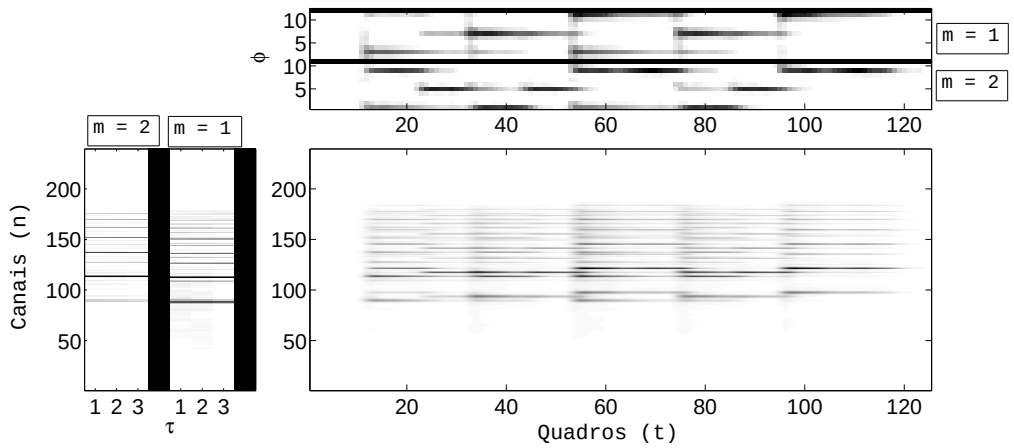

Figura 6.3: Resultado da separação por CNMF2D somente com critério de esparsidade ativado e peso igual a 100. É importante notar que, comparando com a figura anterior, a ordem das fontes separadas ($m = 1$ e $m = 2$) foi trocada. Esse é um problema inerente à separação cega de fontes, em que não é possível determinar a ordem em que elas serão entregues.

NMF2D comum. Em contrapartida, houve a dúvida quanto ao impacto de pesos muito elevados na fatoração.

6.7.2 Descrição

Para este experimento, utilizou-se a mistura *piano_ trompete*. O critério escolhido como exemplo foi o de esparsidade com norma 2. Não foi levada em consideração a eficácia da esparsidade diante da separação executada sem ela como restrição; este experimento está documentado mais adiante. Dentre os pesos analisados, estão mostrados nas Figuras 6.2 e 6.3 os resultados relativos aos pesos de valores 1 e 100, respectivamente. Em cada uma dessas figuras, o gráfico à esquerda apresenta as bases espectrais para cada instrumento, o gráfico acima apresenta os ganhos de cada base ao longo dos quadros de tempo com seu *pitch* (ϕ) associado, e o gráfico à direita abaixo apresenta a espectrograma que resulta da associação entre as matrizes, ou seja, a soma dos produtos das matrizes $\mathbf{B}^l$ e $\mathbf{G}^p$ para cada instrumento [4]. Observa-se que o peso de valor 100 atrapalha a fatoração (esse fato também foi constatado auditivamente), enquanto o peso unitário mostra-se uma escolha bem mais adequada.

[4] Os diagramas que mostram matrizes $\mathbf{B}$ e $\mathbf{G}$ em geral, quando inspecionados visualmente, fornecem informações qualitativas sobre aspectos como o vazamento de uma nota em outra (por exemplo, pela ocorrência de harmônicos de uma nota no padrão da outra nas matrizes $\mathbf{B}$, ou sobreposição de ganhos nas matrizes $\mathbf{G}$).

Assim como os valores de τ e ϕ, os pesos dos critérios contemplados ao longo dos experimentos também foram escolhidos de forma empírica, com base em diversas execuções preliminares de cada combinação mistura-algoritmo.

6.8 Experimentos com os critérios da CNMF2D

A partir deste ponto serão apresentados os experimentos que foram realizados com os critérios da CNMF2D, com o intuito de comparar seus efeitos em relação aos resultados obtidos quando os critério estavam desativados. Os critérios que fizeram parte desse conjunto de experimentos foram testados isoladamente (ou seja, estando os demais critérios desativados) para duas misturas artificiais simples, a fim de verificar os efeitos de cada um na separação. Depois, utilizando sinais mais complexos, os critérios foram combinados, de forma a serem verificados seus efeitos conjuntos.

Em um experimento, se determinado(s) critério(s) estava(m) sendo testado(s), os demais tiveram seus pesos ajustados para o valor "zero".

Para os três experimentos seguintes, foram escolhidas duas misturas sintéticas simples: *piano_trompete* e *órgão_prato*. São misturas que, por possuírem características diferentes entre si, como visto na Secção 6.4, possibilitam certa abrangência no experimento.

6.9 CMF2D com critérios de esparsidade

6.9.1 Objetivo

Este experimento foi realizado com o objetivo de verificar os efeitos dos critérios de esparsidade na frequência e no tempo (ou seja, respectivamente sobre as matrizes $\mathbf{B}^l$ e sobre as matrizes $\mathbf{G}^p$) um de cada vez, na separação.

6.9.2 Descrição

Os critérios de esparsidade utilizados foram os das equações (3.28) e (4.3)—que utilizam norma 2—por serem bem menos sensíveis ao peso, como visto em [10], e confirmado em testes preliminares. Tais critérios foram chamados aqui de *esparsidade 1* e *esparsidade 2*, respectivamente.

Para o sinal *piano_trompete*, considerando a avaliação subjetiva tanto para o uso da *esparsidade 1* quanto para o uso de *esparsidade 2*, não foram observadas vantagens ou pioras sobre o experimento sem a atividade desses critérios. Isso se justifica pelo fato de que os critérios de esparsidades exigem que, na mistura, os sinais que se desejam separar possuam ao menos um trecho em que eles toquem sozinhos, no caso da esparsidade no tempo, e que, por um instante, emitam notas diferentes, no caso da esparsidade da frequência. Entretanto, analisando a mistura, observa-se que, ao longo do trecho, o piano e o trompete emitem as mesmas notas. Além disso, não há um momento no trecho em que os dois instrumentos emitam sozinhos. Portanto, as características dessa mistura não favorecem os efeitos que os critérios de esparsidade na frequência e no tempo poderiam aplicar. Entretanto, em análise objetiva, verifica-se que valores da SIR mostram vantagem para o critério *esparsidade 1*. Tais valores podem ser vistos nas Tabelas 6.6, 6.7 e 6.8.

Para o sinal *órgão_prato* foi observado que, considerando a avaliação subjetiva tanto para o uso da *esparsidade 1* quanto para o uso de *esparsidade 2*, a componente relativa ao prato ficou mais íntegra do que se comparada com o resultado em que os critérios estavam desativados, sendo que com *esparsidade 1*, a integridade foi maior do que com *esparsidade 2*, apresentando SIR consideravelmente maior, conforme se pode verificar nas Tabelas 6.9, 6.10 e 6.11. Apesar de a SIR do órgão ter apresentado pequena desvantagem para *esparsidade 1*, o alto valor da SIR para o prato faz novamente com que *esparsidade 1* tivesse maior vantagem geral. A característica dessa mistura, em que o órgão é esparso na frequência e preenche todo o tempo e o prato é esparso no tempo e preenche melhor os canais de frequência, favoreceram este experimento.

Tabela 6.6: *SIR para separação do sinal piano_trompete com critérios desativados.*

SIR (dB)	**trompete separado**	**piano separado**
trompete original	−5,16	–
piano original	–	18,29

Tabela 6.7: *SIR para separação do sinal piano_trompete com esparsidade 1 ativada.*

SIR (dB)	**trompete separado**	**piano separado**
trompete original	23,25	–
piano original	–	17,37

Tabela 6.8: *SIR para separação do sinal piano_ trompete com esparsidade 2 ativada.*

SIR (dB)	trompete separado	piano separado
trompete original	14,63	–
piano original	–	8,02

Tabela 6.9: *SIR para separação do sinal órgão_ prato com critérios desativados.*

SIR (dB)	órgão separado	prato separado
órgão original	32,64	–
prato original	–	22,64

Tabela 6.10: *SIR para separação do sinal órgão_ prato com esparsidade 1 ativada.*

SIR (dB)	órgão separado	prato separado
órgão original	30,81	–
prato original	–	90,66

Tabela 6.11: *SIR para separação do sinal órgão_ prato com esparsidade 2 ativada.*

SIR (dB)	órgão separado	prato separado
órgão original	37,54	–
prato original	–	32,53

6.10 NMF2D com critério de correlação

6.10.1 Objetivo

Este experimento foi realizado com o objetivo de verificar os efeitos dos critérios de correlação (correlação cruzada e autocorrelação), um de cada vez, na separação.

6.10.2 Descrição

Os critérios de correlação utilizados foram os das equações (4.7), (4.9) e (4.11), em que os dois primeiros referem-se ao cancelamento da correlação cruzada e o terceiro refere-se à enfatização da autocorrelação de cada fonte. Para facilitar as referências, esses critérios serão chamados de *correlação 1*, *correlação 2* e *correlação 3*, respectivamente.

Para o sinal *piano_ trompete* foi observado que, no caso em que os critérios estavam desligados, houve vazamento das notas de piano na fonte relativa ao trompete, bem como a presença do trompete na fonte relativa ao piano. Ao realizar o experimento com a *correlação 2*, houve redução do vazamento das notas de piano na fonte relativa ao trompete, entretanto percebeu-se a presença de trechos de baixa

amplitude (brancos) nesse sinal, além de mistura maior na fonte relativa ao piano. Ao realizar o experimento com a *correlação 1*, foi observada melhora subjetiva nas duas fontes em relação aos dois testes anteriores, apesar da permanência do vazamento das notas de piano. O uso da *correlação 3*, perceptivamente, não apresentou vantagens. Os valores da SIR estão apresentados nas Tabelas 6.6, 6.12, 6.13 e 6.14. Tais valores mostram que a *correlação 1* gerou o melhor equibílibrio na separação das duas fontes.

Para o sinal *órgão_prato* foi observado que, o uso dos três critérios melhorou a integridade dos ataques de prato, se comparado com o teste realizado com os critérios desligados. Entretanto, observou-se que houve supressão do órgão nos trechos em que o prato ocorreria com o uso da *correlação 2*. Com o uso da *correlação 1*, foi observada menor interferência dos ataques de prato no órgão, sendo este critério novamente vantajoso sobre o outro. Os valores da SIR estão apresentados nas Tabelas 6.9, 6.15, 6.16 e 6.17.

Tabela 6.12: *SIR para separação do sinal piano_trompete com correlação 1 ativada.*

SIR (dB)	**trompete separado**	**piano separado**
trompete original	17,07	–
piano original	–	16,91

Tabela 6.13: *SIR para separação do sinal piano_trompete com correlação 2 ativada.*

SIR (dB)	**trompete separado**	**piano separado**
trompete original	24,41	–
piano original	–	5,93

Tabela 6.14: *SIR para separação do sinal piano_trompete com correlação 3 ativada.*

SIR (dB)	**trompete separado**	**piano separado**
trompete original	8,91	–
piano original	–	24,86

Tabela 6.15: *SIR para separação do sinal órgão_prato com correlação 1 ativada.*

SIR (dB)	**órgão separado**	**prato separado**
órgão original	55,35	–
prato original	–	33,96

Tabela 6.16: *SIR para separação do sinal órgão_prato com correlação 2 ativada.*

SIR (dB)	órgão separado	prato separado
órgão original	48,25	–
prato original	–	20,77

Tabela 6.17: *SIR para separação do sinal órgão_prato com correlação 3 ativada.*

SIR (dB)	órgão separado	prato separado
órgão original	37,15	–
prato original	–	25,17

6.11 NMF2D com critérios de continuidade temporal

6.11.1 Objetivo

Este experimento foi realizado com o objetivo de verificar os efeitos dos critérios de continuidade temporal na separação.

6.11.2 Descrição

Os critérios utilizados foram os das equações (4.14) e (4.16), e serão chamados de *continuidade 1* e *continuidade 2*, respectivamente.

Para o sinal *piano_trompete*, considerando tanto o uso da *continuidade 1* quanto o da *continuidade 2*, por avaliação subjetiva, foi observada uma leve redução do ataque do piano no trompete, quase imperceptível. Entretanto, a SIR mostrou-se melhor com o uso da *continuidade 2*, atráves das Tabelas 6.6, 6.18 e 6.19.

Para o sinal *órgão_prato*, foi observado, por avaliação subjetiva, que o uso dos critérios melhorou a integridade dos ataques de prato se comparado com o teste realizado com os critérios desligados, assim como nos casos anteriores. No caso do órgão, ocorrem tons curtos associados às batidas dos pratos; na *continuidade 1* esse efeito é mais localizado, enquanto que na *continuidade 2* ele é mais espalhado. As Tabelas 6.9, 6.20 e 6.21 mostram os valores da SIR, que corroboram a avaliação subjetiva.

Com base nos valores da SIR, apesar de *continuidade 1* ter apresentado maior afinidade com o sinal *órgão_prato* e *continuidade 2* ter apresentado maior afinidade com *piano_trompete*, considerou-se que *continuidade 1* teve melhor desempenho

geral, pois, em avaliação subjetiva, não foi verificada desvantagem deste critério em relação ao critério *continuidade 2* para nenhum dos dois sinais. Além do mais, o sinal *órgão_prato* é um sinal mais abrangente do que o sinal *piano_trompete* no sentido de ser composto por dois sinais com características distintas entre si (conforme visto anteriormente), e neste sinal a *continuidade 1* funcionou melhor.

Tabela 6.18: *SIR para separação do sinal piano_trompete com o critério continuidade 1 ativado.*

SIR (dB)	trompete separado	piano separado
trompete original	9,17	–
piano original	–	0,98

Tabela 6.19: *SIR para separação do sinal piano_trompete com o critério continuidade 2 ativado.*

SIR (dB)	trompete separado	piano separado
trompete original	24,93	–
piano original	–	17,66

Tabela 6.20: *SIR para separação do sinal órgão_prato com o critério continuidade 1 ativado.*

SIR (dB)	órgão separado	prato separado
órgão original	42,13	–
prato original	–	24,22

Tabela 6.21: *SIR para separação do sinal órgão_prato com o critério continuidade 2 ativado.*

SIR (dB)	órgão separado	prato separado
órgão original	36,93	–
prato original	–	18,29

6.12 CNMF2D com critérios combinados

6.12.1 Objetivo

Esta seção abriga uma série de experimentos realizados com quatro sinais distintos e mais complexos do que o *piano_trompete* e o *órgão_prato*. O objetivo deste conjunto de testes foi o de encontrar quais critérios de restrição poderiam ser combinados para cada um desses quatro sinais, e combiná-los, de forma que se pudesse verificar seus efeitos conjuntos.

6.12.2 Descrição

De forma a dar objetividade aos testes, foi estabelecido o seguinte procedimento para a realização dos experimentos:

1− Foi escolhido apenas um critério de cada natureza (esparsidade, correlação e continuidade), de forma que fossem combinados apenas critérios de naturezas diferentes.

2− Cada critério escolhido foi testado isoladamente;

3− Se determinado critério isolado melhorou ou manteve a qualidade da fatoração (se comparado com o experimento sem critérios ativados), então tal critério tornou-se candidato a ser combinado com outro critério na mesma situação.

A escolha dos critérios foi baseada nos resultados dos experimentos anteriores realizados com as misturas artificiais simples para cada critério isolado. Nesses experimentos foi observado e comentado que os critérios *esparsidade 1* (esparsidade no tempo), *correlação 1* e *continuidade 1* (um de cada natureza) foram os que apresentaram vantagens sobre os outros de mesma natureza. Portanto, esses foram os critérios escolhidos, sendo aqui chamados simplesmente de *esparsidade*, *correlação* e *continuidade*, respectivamente.

Os sinais escolhidos para esta bateria de experimentos foram *Paganini*, *Bach*, *Far More Drums* e *Take5*, cujas características podem ser vistas na Seção 6.4. Conforme já mencionado, todos esses sinais são misturas naturais, e, portanto, não estão disponíveis as fontes originais referentes a cada instrumento que compõe tais misturas. Portanto, para esses experimentos, não foi possível realizar avaliação objetiva através da medida da SIR. Assim, os resultados serão apresentados através dos gráficos ϕ *versus* $\mathbf{G}^p$, pois as matrizes $\mathbf{G}^p$ são as matrizes sobre as quais os critérios escolhidos atuam. Os resultados dos experimentos serão apresentados por mistura e, dentro das misturas, por critérios.

Ainda, nos testes apresentados a seguir, um sinal separado é considerado como referente a determinado instrumento que compõe a mistura quando a presença deste instrumento é maior do que a presença dos demais. E, por padrão, as considerações feitas sobre as separações foram realizadas com base em avaliações subjetivas. Entretanto, está explicitado no texto quando alguma consideração feita foi extraída dos gráficos obtidos.

6.12.3 *Paganini*

Critérios desativados: Foi observado que o sinal separado relativo ao piano encontrou-se ainda muito misturado com o violoncelo. Entretanto, as notas de piano permaneceram íntegras nesse sinal. De forma complementar, para o caso do sinal relativo ao violoncelo, apenas parte de suas notas o compôs. Entretanto, este esteve com pouca interferência do piano. Em outras palavras, houve muita interferência do violoncelo no piano e pouca interfência do piano no violoncelo.

Com esparsidade: Em comparação com o experimento com os critérios desativados, verificou-se piora na separação. O sinal relativo ao violoncelo recebeu mais interferência do piano, além de ter-se percebido alteração em seu *pitch*. Já o sinal relativo ao piano ainda permanceceu bastante misturado ao violoncelo e, dessa vez, estando menos íntegro.

Com correlação: Neste experimento, a fatoração produziu sinais com características similares ao experimento com os critérios desativados, porém com melhora. Verificou-se que, no sinal relativo ao piano, o violoncelo ficou mais reduzido.

Com continuidade: Tanto para o sinal relativo ao violoncelo quanto para o sinal relativo ao piano, as características da separação foram muito similares aos do experimento com os critérios desativados, sem vantagens ou desvantagens perceptivas.

Com correlação e continuidade: Devido ao fato de os experimentos com correlação e continuidade terem produzidos, respectivamente, resultados melhores e resultados muito similares ao do experimento com os critérios desativados (tendo o critério de esparsidade apresentado resultado pior), tais critérios foram ativados ao mesmo tempo neste experimento a fim de se verificar seus efeitos combinados para o sinal em teste. Para o sinal relativo ao violoncelo, houve significativa melhora em comparação aos experimentos anteriores. As notas de violoncelo ficaram mais íntegras e quase não sofreram interferência das notas de piano. Consequentemente, para o piano, verificou-se também a integridade em suas notas, embora ainda misturadas com resquícios do violoncelo.

O experimento com critérios de correlação e continuidade combinados resultou em separação melhor do que quando empregados isoladamente. É interessante notar que os efeitos dos dois critérios não foram apenas somados, afinal o critério de continuidade não apresentou vantagem perceptiva quando testado sozinho. Percebeu-se que

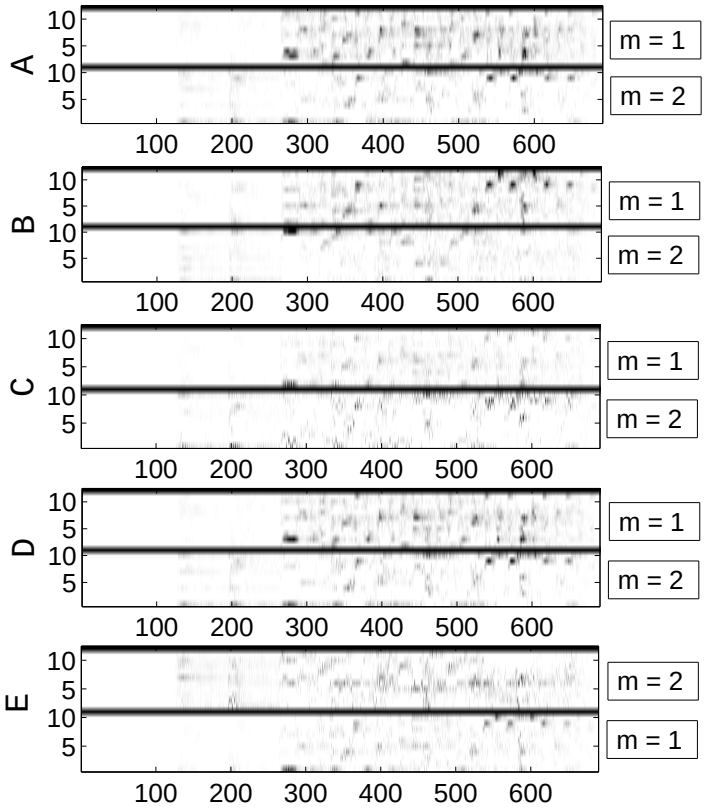

Figura 6.4: Fatoração com *Paganini*, em que $m = 1$ corresponde ao violoncelo e $m = 2$ corresponde ao piano. Os gráficos A, B, C, D e E correspondem, respectivamente, à separação sem o uso de critérios, com esparsidade, com correlação, com continuidade e com correlação e continuidade combinados. Os gráficos possuem no eixo das abscissas os quadros de tempo e no eixo das ordenadas os valores de ϕ.

um critério melhorou o efeito do outro, produzindo, em conjunto, uma separação melhor. Os gráficos dos experimentos estão apresentados na Figura 6.4.

6.12.4 *Bach*

Critérios desativados: Para o sinal relativo ao violoncelo, verificou-se boa integridade do instrumento, ainda que com indícios das notas de piano misturados a ele. Verificou-se também um efeito parecido com eco. O sinal relativo ao piano apresentou mistura com resquícios das frequências do violoncelo, estando presente um efeito parecido com o do vibrato[5]. Na Figura 6.5, o gráfico A mostra o resultado desta separação.

[5]Vibrato é uma técnica que consiste na modificação aproximadamente periódica da altura da nota emitida, pelo instrumentista. É um recurso expressivo comum em instrumentos de corda e sopro.

Com esparsidade: Em comparação ao experimento com critérios desativados, o sinal separado relativo ao violoncelo ficou com interferência um pouco menor do piano. Ainda foi verificado um efeito parecido com eco. De forma complementar, o sinal relativo ao piano ficou mais íntegro e um pouco menos misturado com o violoncelo, estando o efeito de vibrato reduzido. Na Figura 6.5, o gráfico B mostra a melhor separação obtida para este critério em comparação ao experimento com o critério desativado.

Com correlação: Em comparação tanto ao experimento com critérios desativados quanto ao experimento com critério de esparsidade, a separação dos dois instrumentos resultou em qualidade pior. Enquanto para o sinal relativo ao violoncelo houve aumento no efeito parecido com eco, com a nota de piano ao fundo assemelhando-se ao toque de um sino, para o sinal relativo ao piano observou-se presença distorcida do violoncelo.

Com continuidade: Em comparação ao experimento com critérios desativados, para o sinal relativo ao violoncelo, o critério de continuidade apresentou vantagem similar ao critério de esparsidade, o que pode ser constatado pelo gráfico D. Para o sinal relativo ao piano, verificou-se presença um pouco menor do violoncelo. Através do gráfico D, observa-se que a execução das notas de piano (uma nota grave seguida de duas notas mais agudas) ficou melhor definida.

Com esparsidade e continuidade: Devido ao fato de os experimentos com esparsidade e continuidade terem produzidos, isoladamente, resultados melhores do que o experimento com os critérios desativados (tendo o critério de correlação apresentado resultado pior), tais critérios foram ativados ao mesmo tempo a fim de verificar-se seus efeitos combinados para o sinal em teste. Neste caso, tanto para o sinal relativo ao violoncelo quanto para o sinal relativo ao piano, verificou-se separação similar ao experimento com critério de continuidade isolado, ou seja, superior ao experimento com os critérios desativados. Entretanto, através do gráfico E, verificou-se que as notas de piano ficaram ainda melhor definidas do que no experimento anterior.

O experimento com critérios de esparsidade e continuidade combinados resultou em separação mais próxima ao resultado com critério de continuidade isolado. Os gráficos dos experimentos estão apresentados na Figura 6.5.

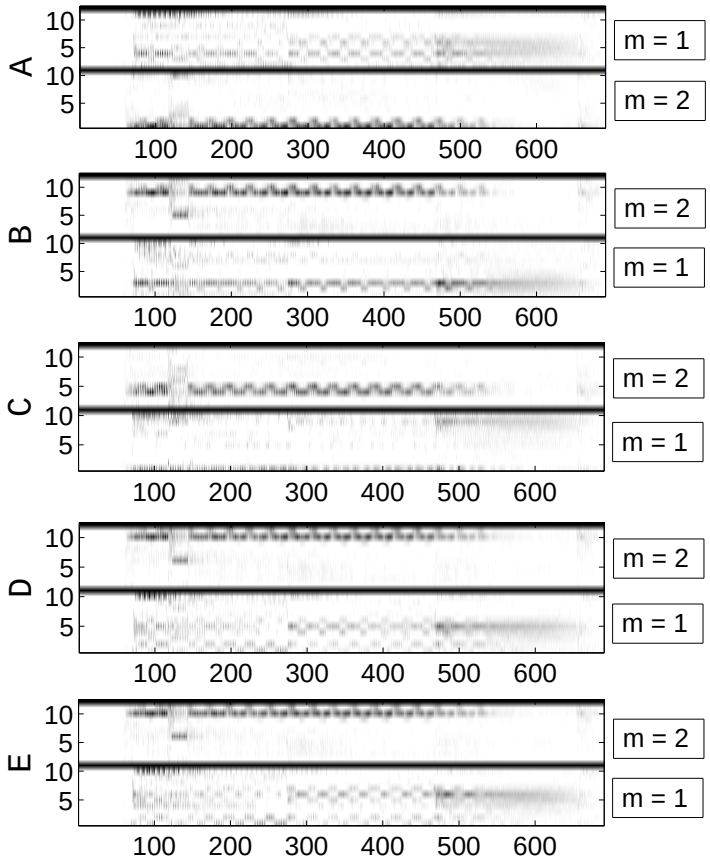

Figura 6.5: Fatoração com *Bach*, em que $m = 1$ corresponde ao piano e $m = 2$ corresponde ao violoncelo. Os gráficos A, B, C, D e E correspondem, respectivamente, à separação sem o uso de critérios, com esparsidade, com correlação, com continuidade e com esparsidade e continuidade combinados. Os gráficos possuem no eixo das abscissas os quadros de tempo e no eixo das ordenadas os valores de ϕ.

6.12.5 *Far More Drums*

Critérios desativados: Verificou-se a predominância de um instrumento diferente em cada sinal separado, o que significa que a separação, a despeito das interferências, funcionou. Verificou-se que o sinal relativo ao baixo sofreu interferência do piano e, por sua vez, o sinal relativo ao piano sofreu interferência da bateria no início do trecho. O sinal relativo à bateria foi o único instrumento que ficou melhor isolado dos demais, entretanto com interferência bastante discreta do piano.

Com esparsidade: Em comparação ao experimento com os critérios desativados, a qualidade da separação foi um pouco pior. O sinal relativo ao baixo sofreu ainda mais interferência do piano. O sinal relativo ao piano sofreu a mesma interferência da bateria no início, porém ficou mais distorcido após a metade do trecho. A sinal relativa à bateria permaneceu com interferência discreta do piano.

Com correlação: Também apresentou separação com qualidade geral um pouco pior do que no caso do experimentos com critérios desativados.

Com continuidade: Apresentou separação similar ao experimento com os critérios desativados, como pode ser verificado através do gráfico D da Figura 6.6.

Devido ao fato de os três critérios não terem trazido melhoras para a separação, nenhuma combinação entre eles foi realizada. Os gráficos dos experimentos estão mostrados na Figura 6.6. Vale mencionar que a bateria é a que melhor se separa dos demais instrumentos. Isso pode ser explicado pelo fato de a bateria ser um instrumento sem *pitch* definido, diferente dos demais, o que acaba por destacá-la para o algoritmo.

6.12.6 *Take*5

Critérios desativados: Para este sinal, verificou-se que nenhum instrumento foi isolado dos demais. Entretanto, houve predominância do baixo e da bateria em cada um entre dois dos sinais separados obtidos. No sinal com predominância do baixo, verificaram-se sombras do saxofone e do piano e, para o caso da bateria, os pratos ficaram totalmente isolados em um dos sinais. Entretanto, nos dois outros sinais, houve distribuição do saxofone. Em um deles, o saxofone ficou misturado com as caixas da bateria, e no outro, ficou misturado com o restante do piano.

Com esparsidade: Apresentou resultado similar ao do experimento com os critérios desligados, porém com piora de qualidade no sinal relativo ao saxofone misturado as caixas da bateria.

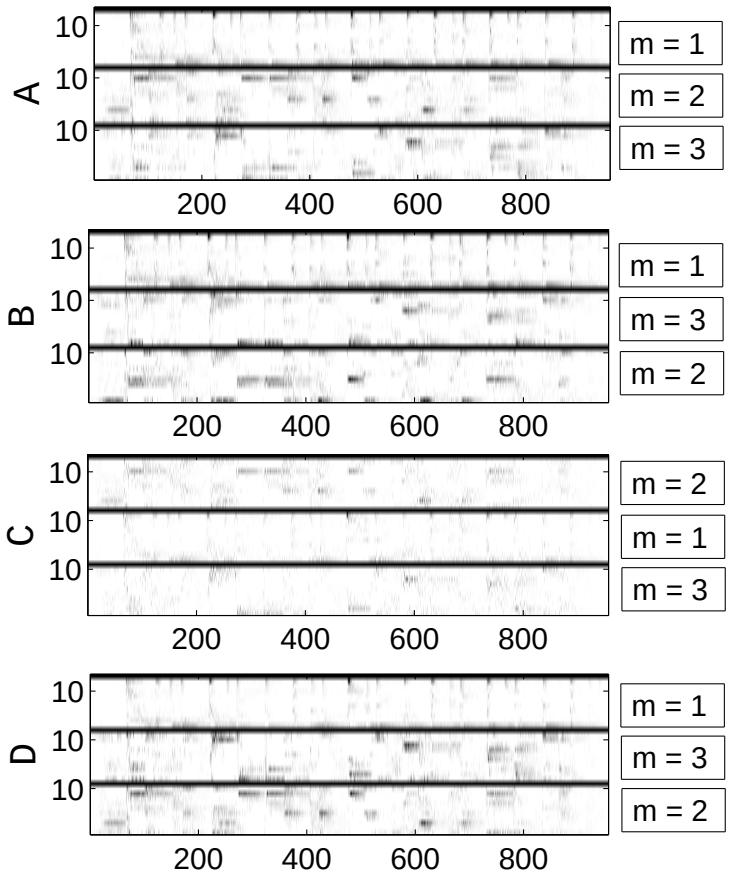

Figura 6.6: Fatoração com *Far More Drums*, em que $m = 1$ corresponde à bateria, $m = 2$ corresponde ao piano e $m = 3$ corresponde ao baixo. Os gráficos A, B, C e D correspondem, respectivamente, à separação sem o uso de critérios, com esparsidade, correlação e com continuidade. Os gráficos possuem no eixo das abscissas os quadros de tempo e no eixo das ordenadas os valores de ϕ.

Com correlação: Em comparação com o experimento com os critérios desativados, observou-se melhor separação na bateria, ficando esta composta dos pratos e da caixa. O baixo ficou misturado com um resíduo irregular. Uma parte do saxofone permaneceu misturada com um pouco da caixa da bateria e com o piano. A outra parte do saxofone ficou misturado com o restante do piano.

Com continuidade: Em comparação com o experimento com os critérios desativados, verificou-se mistura do baixo e de parte do saxofone em um dos sinais. Outra parte do saxofone ficou misturada ao piano em outro sinal. Os pratos da bateria também ficaram isolados, e com qualidade levemente superior. O restante do saxofone ficou misturado com as caixas da bateria.

Com correlação e continuidade: Mesmo com a dificuldade em analisar e comparar os experimentos com os critérios isolados entre si (devido ao elevado número de instrumentos, o que acarretou em sinais ainda muito misturados), decidiu-se realizar as combinações entre os critérios de correlação e continuidade. O resultado foi similar ao com o critério de correlação isolado. Em comparação com o experimento com os critérios desativados, houve melhora na separação da bateria, que ficou mais íntegra, contendo os pratos e as caixas. O sinal relativo ao baixo ficou com o resíduo irregular apresentado no critério de correlação isolada. O saxofone se dividiu com o piano em uma fonte e com o restante dos pratos da bateria em outra. Os gráficos dos resultados estão apresentados em 6.7.

Mais uma vez, verificou-se que a bateria (pratos) foi o instrumento que melhor se separou dos demais, pelo mesmo motivo exposto anteriormente: ser o único com ausência de *pitch* definido. Entretanto, também vale mencionar que, neste sinal, a bateria é também composta por caixas, e que, ao longo dos experimentos, houve a tendência de que caixas e pratos fossem separados. Isso leva a uma reflexão sobre o que o algoritmo considera como instrumento. No caso dos pratos e das caixas da bateria, elas possivelmente foram tratadas como instrumentos diferentes.

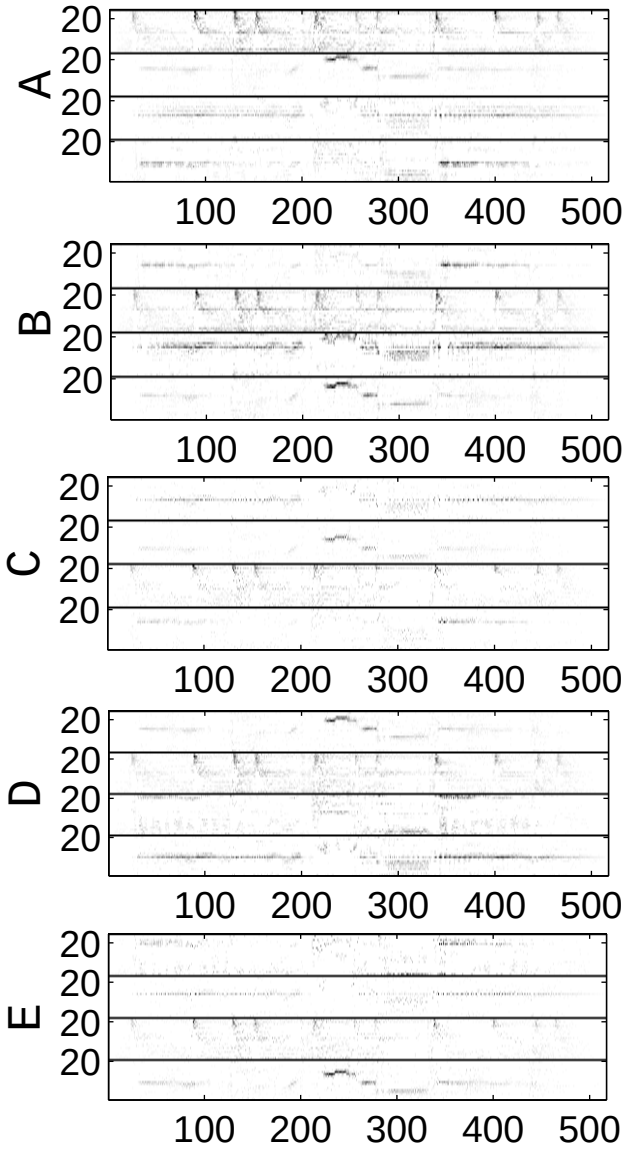

Figura 6.7: Fatoração com *Take5*. Os instrumentos não foram sinalizados devido à grande mistura que permaneceu entre eles para todos os sinais separados. Os gráficos A, B, C, D e E correspondem, respectivamente, à separação sem o uso de critérios, com esparsidade, com correlação, com continuidade e com correlação e continuidade combinados. Os gráficos possuem no eixo das abscissas os quadros de tempo e no eixo das ordenadas os valores de ϕ.

Capítulo 7

Conclusões e trabalhos futuros

7.1 Contribuição desta dissertação

Esta dissertação apresentou o sistema completo de separação de fontes sonoras (mais especificamente, de sinais musicais) de uma linha de algoritmos com base na NMF. Os algoritmos foram mostrados em ordem cronológica de desenvolvimento até a última conhecida da literatura, a SNMF2D. Esta, além de ser aplicada na separação de instrumentos musicais inteiros (como a NMF2D), faz uso de uma restrição de esparsidade, cujo objetivo é condicionar as matrizes de ganhos no tempo $\mathbf{G}^p$ a serem esparsas.

Entretanto, a restrição de esparsidade aplicada à NMF2D que a transforma na SNMF2D é somente uma das restrições possíveis que já foram aplicadas à NMF em outro trabalhos [10], [6] e [5], tais como restrições de correlação e de continuidade temporal. A proposta desta dissertação foi a de considerar também a adaptação de outros critérios de restrição para a inserção na NMF2D. Portanto, deu-se origem não à SNMF2D, mas a uma versão genérica, que vislumbra englobar mais restrições: a CNMF2D.

Tendo-se em mente que a NMF2D é equivalente à NMF quando os valores de τ e ϕ são iguais a zero, e que a CNMF2D é equivalente à NMF2D com os critérios desativados (ou seja, sendo seus pesos iguais a zero), pode-se considerar a CNMF2D como um algoritmo que engloba e descreve toda a linha de algoritmos com base na NMF sob uma notação comum. Assim, mais do que uma genereralização da SNMF2D, a CNMF2D é uma generalização da NMF básica, que possibilita, por meio da escolha de seus parâmetros, o *setup* para os mais diversos tipos de experimentos, sinais e objetivos almejados. A CNMF2D pode funcionar como a NMF básica, como a NMF2D, como a SNMF2D, e vai mais além, com a escolha de outros critérios e/ou combinações entre eles.

É importante notar que a CNMF2D não é um algoritmo fechado no que diz respeito às restrições: não existe um conjunto fixo de restrições que devam fazer parte da CNMF2D. Tantos critérios quanto forem desejados e desenvolvidos podem compor ou deixar de compor o algoritmo, podem estar ativos em determinado experimento ou não. Tantas combinações entre eles quanto forem imaginadas podem ser testadas. Este trabalho considerou os critérios de restrição mais populares encontradas na literatura: critérios de esparsidade, correlação e continuidade, comentados abaixo:

Esparsidade: Verificou-se que os critérios de esparsidade possuem afinidade com misturas entre instrumentos já esparsos. Quanto mais esparsos entre si são os instrumentos, maior é a chance de se obter um bom resultado. Observou-se também a preferência da esparsidade por misturas entre sinais mais suaves, de transição lenta entre as notas;

Correlação Verificou-se que os critérios de correlações são os mais propensos a degradarem os sinais. Como eles tentam eliminar a interferência de uma fonte em outra, existe a possibilidade de que eles incidam até mesmo no instrumento que deveria permancecer, nos trechos em que a interferência aparece. Um ajuste ideal de pesos é fundamental para equilibrar os efeitos desses critérios;

Continuidade São os critérios que sempre melhoram (ou pelo menos mantêm) a qualidade da separação. Em todos os experimentos realizados neste trabalho, a continuidade temporal mostrou-se vantajosa em relação às fatorações com os critérios desativados. Isso se deve ao fato de que, conforme já explicado, o princípio da continuidade temporal baseia-se no fato de que a mistura seja composta por sinais que não possuem transições bruscas entre quadros de tempo próximos, o que é razoável de se assumir para todos os sinais musicais.

Por fim, esta dissertação também contribuiu com a demonstrações dos algoritmos estudados, em especial com a demonstração das equações de atualização da SNMF2D (feita na Seção A.5), cujo desenvolvimento é mais complexo e está oculto na literatura correspondente [19], constituindo um espaço em branco que precisava ser preenchido e compreendido para dar consistência ao estudo e à própria CNMF2D.

7.2 Trabalhos futuros

A motivação para trabalhos futuros relativos à CNMF2D vem do fato de ser grande o número de sinais, restrições e combinações que ainda podem ser testados.

Por ser um algoritmo que engloba outros algoritmos da NMF existentes na literatura, a CNMF2D possui um grande número de parâmetros a serem explorados, conforme já mencionado, incluindo aqueles que mantiveram-se fixos para os experimentos realizados neste trabalho. Pode-se considerar que a separação de fontes sonoras por NMF e seus algoritmos derivados não se encontra em estagnação.

Alguns pontos relativos à separação de fontes ainda são problemas em aberto. Um deles é a questão da escolha do número de fontes. É desejável que algoritmo consiga identificar o número de fontes apropriado para cada mistura de maneira automática, já que a NMF é classificada como método não-supervisionado. Para o caso da NMF2D (e, mais genericamente, CNMF2D), a situação é agravada com a presença dos parâmetros τ e ϕ, que também deveriam ser escolhidos automaticamente pelo algoritmo.

Seguindo a problemática do automatismo da CNMF2D, tem-se ainda a questão da escolha das restrições. Seria interessante que o algoritmo conseguisse identificar se determinada mistura seria melhor fatorada aplicando-se determinado(s) tipo(s) de critério(s) de separação, realizando uma breve análise preliminar do sinal e fazendo os ajustes de pesos adequados.

Um problema mais complexo é o fato de que, por mais que os algoritmos de separação venham sendo refinados, eles ainda estão longe de realizar o trabalho que a audição humana faz em identificar sinais que compõem misturas. Por isso, soluções baseadas em psicoacústica poderiam ser desenvolvidas.

Referências Bibliográficas

[1] CHERRY, E. C. "Some experiments on the recognition of speech, with one and two ears", *Journal of the Acoustical Society of America*, v. 25, n. 5, pp. 975–979, Setembro 1953.

[2] HYVÄRINEN, A., OJA, E. "Independent Component Analysis: Algorithms and Applications", *Neural Networks*, v. 13, n. 4-5, pp. 411–430, Maio-Junho 2000.

[3] VIRTANEN, T. *Sound Source Separation in Monoaural Music Signals*. Ph.D. thesis, Tampere University of Technology, Finlândia, Novembro 2006.

[4] LEE, D. D., SEUNG, H. S. "Learning the parts of objects by non-negative matrix factorization", *Nature*, v. 401, n. DOI 10.1038/44565, pp. 788–791, Outubro 1999.

[5] RACZYŃSKI, S., ONO, N., SAGAYAMA, S. "Extending Nonnegative Matrix Factorization - A Discussion in the Context of Multiple Frequency Estimation of Musical Signals". In: *Proceedings of the 2009 European Signal Processing Conference (EUSIPCO 2009)*, pp. 934–938, Glascow, Escócia, Agosto 2009. EURASIP.

[6] RACZYŃSKI, S., ONO, N., SAGAYAMA, S. *Multipitch Analysis with Harmonic Nonnegative Matrix Approximation*. Relatório técnico, Graduate School of Information Science and Engineering, The University of Tokyo, Tóquio, Japão, 2007.

[7] TYGEL, A. F. *Métodos de Fatoração de Matrizes Não-negativas Para Separação de Sinais Musicais*. Tese de Mestrado, PEE/COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil, Dezembro 2009.

[8] DINIZ, P. S. R., SILVA, E. A., NETTO, S. L. *Processamento Digital de Sinais: Projeto e Análise de Sistemas*. Porto Alegre, Brasil, Bookman, 2004.

[9] VIRTANEN, T. "Unsupervised Learning Methods for Source Separation in Monoaural Music Signals". In: Klapuri (Ed.), *Signal Processing Methods for*

Music Transcription, Signal Processing Methods for Music Transcription, Springer, cap. 9, pp. 267–296, Nova Iorque, EUA, 2006.

[10] VIRTANEN, T. "Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria", *IEEE Transactions on Audio, Speech, and Language Processing*, v. 15, n. 3, pp. 1066–1074, Março 2007.

[11] JOLLIFE, I. T. *Principal Component Analysis*. Nova Iorque, EUA, Springer, 2002.

[12] HYVÄRINEN, A., KARHUNEN, J., OJA, E. *Independent Component Analysis*. Nova Iorque, EUA, John Wiley, 2001. ISBN: 0-471-22131-7.

[13] CASEY, M. A. *Separation of Mixed Audio Sources by Independent Subspace Analysis*. Technical Report TR-2001-31, Mitsubishi Eletric Research Labs, Massachusetts, EUA, Setembro 2001.

[14] OLSHAUSEN, B. A., FIELD, D. J. "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?" *Vision Research*, v. 37, n. 23, pp. 3311–3325, Dezembro 1996.

[15] LEE, D. D., SEUNG, H. S. "Algorithms for Non-negative Matrix Factorization", *Neural Information Processing Systems*, v. 13, pp. 556–562, Abril 2001.

[16] SMARAGDIS, P., BROWN, J. C. "Non-negative matrix factorization for polyphonic music transcription". In: *Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, EUA, 2003. IEEE.

[17] ANTONIOU, A., LU, W. S. *Practical Optimization: Algorithms and Engineering Applications*. Nova Iorque, EUA, Springer, 2007.

[18] SMARAGDIS, P. "Non-negative Matrix Factor Deconvolution; Extraction of Multiple Sound Sources from Monophonic Inputs". In: *Proceedings of the 5th International Congress on Independent Component Analysis and Blind Signal Separation*, v. 3195, pp. 494–499, Grenada, Setembro 2004.

[19] MORUP, M., SCHMIDT, M. N. *Sparse Non-negative Matrix Factor 2-D Deconvolution*. Technical Report DK-2800, Informatics and Mathematical Modeling, Technical University of Denmark, Lyngby, Dinamarca, 2006.

[20] SMARAGDIS, P. *Convolutive Speech Bases and their Application to Supervised Speech Separation*. Technical Report TR2007-002, Mitsubishi Electric Research Laboratories, Cambridge, EUA, Janeiro 2007.

[21] SCHMIDT, M. N., MORUP, M. "Non-negative Matrix Factor 2-D Deconvolution for Bling Single Channel Source Separation". In: *Proceedings of the 6th International Congress on Independent Component Analysis and Blind Signal Separation*, pp. 700–707, Charleston, EUA, Março 2006.

[22] BROWN, J. C. "Calculation of a Constant Q Spectral Transform", *Journal of the Acoustical Society of America*, v. 89, pp. 425–434, Janeiro 1991.

[23] BROWN, J. C., PUCKETTE, M. S. "An Efficient Algorithm for the Calculation of a Constant Q Transform", *Journal of the Acoustical Society of America*, v. 92, pp. 2698–2701, Novembro 1992.

[24] FITZGERALD, D., GRANITCH, M., CYCHOWSKI, M. "Towards an Inverse Constant Q Transform". In: *120th AES Convention*, Paris, 2006.

[25] FITZGERALD, D., GRANITCH, M., COYLE, E. "Resynthesis methods for Sound Source Separation using shifted Non-negative Factorisation Models". In: *Proceedings of the Irish Signals and Systems Conference*, pp. 1–5, Derry, Irlanda, Setembro 2007.

[26] COMON, P., JUTTEN, C. *Handbook of Blind Source Separation*. Oxford, EUA, Elsevier, 2010.

[27] HOYER, P. O. "Non-negative Matrix Factorization with Sparseness Constraints", *Journal of Machine Learning Research*, v. 5, pp. 1457–1469, Janeiro 2004.

[28] FITZGERALD, D., CRANITCH, M., COYLE, E. "Extended Nonnegative Tensor Factorisation Models for Musical Sound Source Separation", *Computational Intelligence and Neuroscience*, v. 2008, n. ID 872425, pp. 1–15, Abril 2008.

[29] KAMEOKA, H., ONO, N., KASHINO, K., et al. "Complex NMF: a new sparse representation for acoustic signals". In: *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing 2009 (ICASSP 2009)*, pp. 3437–3440, Taipei, Taiwan, Abril 2009. IEEE.

[30] TYGEL, A., BISCAINHO, L. W. P. "Sound Source Separation via Nonnegative Matrix Factor 2-D Deconvolution Using Linearly Sampled Spectrum". In: *Anais do 7o Congresso Nacional da AES Brasil*, pp. 58–65, São Paulo, SP, Maio 2009.

[31] CHEN, Z., CICHOCKI, A. *Nonnegative Matrix Factorization with Temporal Smoothness and/or Spatial Decorrelation Constraints*. Relatório técnico, RIKEN Brain Science Institute, Saitama, Japão, 2005.

[32] GRIFFIN, D. W., LIM, J. S. "Signal Estimation from Modified Short-Time Fourier Transform", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 32, n. 2, pp. 236–243, Abril 1984.

[33] ZHU, X., BEAUREGARD, G. T., WYSE, L. L. "An Efficient Algorithm For Real-Time Spectrogram Inversion". In: *Proceedings of the 8th Conference on Digital Audio Effects (DAFX-05)*, pp. 116–121, Madri, Espanha, Setembro 2005.

[34] ZHU, X., BEAUREGARD, G. T., WYSE, L. L. "Real-Time Signal Estimation From Modified Short-Time Fourier Transform Magnitude Spectra", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 15, n. 5, pp. 1645–1653, Julho 2007.

[35] ZHU, X., BEAUREGARD, G. T., WYSE, L. L. "Real-Time Iterative Spectrum Inversion With Look-Ahead". In: *Proceedings of the IEEE International Conference on Multimedia & Expo (ICME 2006)*, pp. 229–232, Ontario, Canadá, Julho 2006.

[36] VICENT, E., GRIBONVAL, R., FÉVOTTE, C. "Performance measurement in blind audio source separation", *IEEE Transactions on Audio, Speech, and Language Processing*, v. 14, n. 4, pp. 1462–1469, Julho 2006.

Apêndice A

Demonstrações das equações de atualização das versões da NMF

A.1 Para o algoritmo básico da NMF (Equações (3.6), (3.7), (3.8) e (3.9))

Primeiramente, a demonstração será feita considerando o quadrado da distância Euclidiana como função-custo, ou seja,

$$D_{\text{euc}} = \frac{1}{2}\left\|\mathbf{X} - \hat{\mathbf{X}}\right\|_F^2 = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{T}\left(X_{i,j} - \hat{X}_{i,j}\right)^2. \tag{A.1}$$

A equação de atualização de $\mathbf{B}$, segundo o método do gradiente descendente, é

$$\mathbf{B} = \mathbf{B} - \eta\frac{\partial D_{\text{euc}}}{\partial \mathbf{B}}, \tag{A.2}$$

onde η indica o passo de atualização, que será adequadamente escolhido depois. Agora, deve-se calcular $\dfrac{\partial D_{\text{euc}}}{\partial B_{n,d}}$ (onde $B_{n,d}$ indica cada elemento de $\mathbf{B}$), resultando em

$$\frac{\partial D_{\text{euc}}}{\partial B_{n,d}} = \frac{\partial}{\partial B_{n,d}}\left(\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{T}\left(X_{i,j} - \hat{X}_{i,j}\right)^2\right) = -\sum_{j=1}^{T}\left(X_{n,j} - \hat{X}_{n,j}\right)\frac{\partial \hat{X}_{n,j}}{\partial B_{n,d}}. \tag{A.3}$$

Ao se calcular $\dfrac{\partial \hat{X}_{n,j}}{\partial B_{n,d}}$, tem-se

$$\frac{\partial \hat{X}_{n,j}}{\partial B_{n,d}} = \frac{\partial}{\partial B_{n,d}}\sum_{m=1}^{M} B_{n,m}G_{m,j} = G_{d,j}, \tag{A.4}$$

para o caso em que $m = d$. Logo,

$$\frac{\partial D_{\text{euc}}}{\partial B_{n,d}} = -\sum_{j=1}^{T} \left(X_{n,j} - \hat{X}_{n,j} \right) G_{d,j}. \tag{A.5}$$

Na forma matricial, para todos os elementos de $\mathbf{B}$, tem-se

$$\frac{\partial D_{\text{euc}}}{\partial \mathbf{B}} = -\left(\mathbf{X} - \hat{\mathbf{X}} \right) \mathbf{G}^T. \tag{A.6}$$

Portanto, a equação de atualização para $\mathbf{B}$, apresentada inicialmente na equação (A.2), torna-se

$$\mathbf{B} = \mathbf{B} + \eta \left(\mathbf{X} - \hat{\mathbf{X}} \right) \mathbf{G}^T = \mathbf{B} + \eta \left(\mathbf{X}\mathbf{G}^T - \hat{\mathbf{X}}\mathbf{G}^T \right). \tag{A.7}$$

Essa equação de atualização não garante a não-negatividade requerida de $\mathbf{B}$. Uma forma de se garantir isso é transformar a equação (A.7) em uma equação puramente multiplicativa (visto que esta equação possui três termos), com apenas um termo sendo multiplicado pela matriz $\mathbf{B}$ da iteração anterior. Assim, caso a matriz $\mathbf{B}$ seja inicializada com elementos não-negativos, ela se manterá não-negativa ao longo das iterações. Para que isso seja feito, deve-se escolher um valor de η que cancele dois termos, restando apenas um termo multiplicado por $\mathbf{B}$. Mediante o exposto, ao se fazer

$$\eta = \frac{\mathbf{B}}{\hat{\mathbf{X}}\mathbf{G}^T}, \tag{A.8}$$

onde η tem a mesma dimensão de $\mathbf{B}$ e a divisão entre matrizes e a multiplicação de η com os termos devem ser feitas ponto-a-ponto, obtém-se

$$\begin{aligned} \mathbf{B} &= \mathbf{B} + \frac{\mathbf{B}}{\hat{\mathbf{X}}\mathbf{G}^T} \odot \left(\mathbf{X}\mathbf{G}^T - \hat{\mathbf{X}}\mathbf{G}^T \right) \\ &= \mathbf{B} + \frac{\mathbf{B}}{\hat{\mathbf{X}}\mathbf{G}^T} \odot \mathbf{X}\mathbf{G}^T - \frac{\mathbf{B}}{\hat{\mathbf{X}}\mathbf{G}^T} \odot \hat{\mathbf{X}}\mathbf{G}^T. \end{aligned} \tag{A.9}$$

Cancelando o primeiro e o terceiro termos da equação anterior, e fazendo $\hat{\mathbf{X}} = \mathbf{B}\mathbf{G}$, obtém-se finalmente a equação de atualização de $\mathbf{B}$ na forma desejada,

$$\mathbf{B} = \mathbf{B} \odot \frac{\mathbf{X}\mathbf{G}^T}{\mathbf{B}\mathbf{G}\mathbf{G}^T}. \tag{A.10}$$

O desenvolvimento da equação de atualização para a matriz $\mathbf{G}$ é análogo. Basta considerar a decomposição de $\mathbf{X}^T$, em que as matrizes $\mathbf{B}$ e $\mathbf{G}$ são substituídas por $\mathbf{G}^T$ e $\mathbf{B}^T$, respectivamente, e repetir todo o desenvolvimento para $\mathbf{G}^T$. Assim, ao

final, obtém-se

$$\begin{aligned}\mathbf{G} &= \left(\mathbf{G}^T \odot \frac{\mathbf{X}^T\mathbf{B}}{\mathbf{G}^T\mathbf{B}^T\mathbf{B}}\right)^T \\ &= \mathbf{G} \odot \frac{\mathbf{B}^T\mathbf{X}}{\mathbf{B}^T\mathbf{B}\mathbf{G}}.\end{aligned} \tag{A.11}$$

Agora, será apresentado o desenvolvimento da equação de atualização de $\mathbf{B}$ considerando a divergência de Kullback-Leibler generalizada como função-custo, ou seja,

$$D_{\text{kl}} = \left|\mathbf{X} \odot \ln\frac{\mathbf{X}}{\hat{\mathbf{X}}} - \mathbf{X} + \hat{\mathbf{X}}\right| = \sum_{i=1}^{N}\sum_{j=1}^{T}\left(X_{i,j}\ln\frac{X_{i,j}}{\hat{X}_{i,j}} - X_{i,j} + \hat{X}_{i,j}\right). \tag{A.12}$$

Partindo da equação (A.2), fazendo a devida substituição de D_{euc} por D_{kl}, ou seja,

$$\mathbf{B} = \mathbf{B} - \eta\frac{\partial D_{\text{kl}}}{\partial \mathbf{B}}, \tag{A.13}$$

calcula-se o valor $\frac{\partial D_{\text{kl}}}{\partial B_{n,d}}$, de onde se obtém

$$\begin{aligned}\frac{\partial D_{\text{kl}}}{\partial B_{n,d}} &= \frac{\partial}{\partial B_{n,d}}\left(\sum_{i=1}^{N}\sum_{j=1}^{T} X_{i,j}\ln\frac{X_{i,j}}{\hat{X}_{i,j}} - X_{i,j} + \hat{X}_{i,j}\right) \\ &= \frac{\partial}{\partial B_{n,d}}\left[\sum_{i=1}^{N}\sum_{j=1}^{T} X_{i,j}\left(\ln X_{i,j} - \ln\hat{X}_{i,j}\right) - X_{i,j} + \hat{X}_{i,j}\right] \\ &= \sum_{j=1}^{T}\left[X_{n,j}\left(-\frac{1}{\hat{X}_{n,j}}\right)\frac{\partial\hat{X}_{n,j}}{\partial B_{n,d}} + \frac{\partial\hat{X}_{n,j}}{\partial B_{n,d}}\right] \\ &= \sum_{j=1}^{T}\left(1 - \frac{X_{n,j}}{\hat{X}_{n,j}}\right)\frac{\partial\hat{X}_{n,j}}{\partial B_{n,d}}.\end{aligned} \tag{A.14}$$

O termo $\dfrac{\partial\hat{X}_{n,j}}{\partial B_{n,d}}$ da equação anterior é então substituído por seu valor mostrado na equação (A.4),

$$\frac{\partial D_{\text{kl}}}{\partial B_{n,d}} = \sum_{j=1}^{T}\left(1 - \frac{X_{n,j}}{\hat{X}_{n,j}}\right)G_{d,j}. \tag{A.15}$$

Na forma matricial, para todos os elementos de $\mathbf{B}$, tem-se

$$\frac{\partial D_{\text{kl}}}{\partial \mathbf{B}} = \left(\mathbf{1} - \frac{\mathbf{X}}{\hat{\mathbf{X}}}\right)\mathbf{G}^T, \tag{A.16}$$

onde o primeiro $\mathbf{1}$ denota uma matriz de elementos unitários de dimensão NxN que faz o somatório na variável i, e o segundo possui a mesma dimensão de $\mathbf{X}$ e $\hat{\mathbf{X}}$, ou seja, NxT. Assim, a equação de atualização de $\mathbf{B}$ será

$$\mathbf{B} = \mathbf{B} - \eta \left(\mathbf{1}\mathbf{G}^T - \frac{\mathbf{X}}{\hat{\mathbf{X}}}\mathbf{G}^T \right). \tag{A.17}$$

Fazendo-se

$$\eta = \frac{\mathbf{B}}{\mathbf{1}\mathbf{G}^T}, \tag{A.18}$$

e utilizando-se as mesmas considerações feitas anteriormente a respeito das multiplicações e divisões ponto-a-ponto, obtém-se

$$\mathbf{B} = \mathbf{B} - \frac{\mathbf{B}}{\mathbf{1}\mathbf{G}^T} \odot \mathbf{1}\mathbf{G}^T + \frac{\mathbf{B}}{\mathbf{1}\mathbf{G}^T} \odot \frac{\mathbf{X}}{\hat{\mathbf{X}}}\mathbf{G}^T, \tag{A.19}$$

Ao se cancelar o primeiro e o segundo termos, é obtida finalmente a equação de atualização multiplicativa para $\mathbf{B}$,

$$\mathbf{B} = \mathbf{B} \odot \frac{\frac{\mathbf{X}}{\hat{\mathbf{X}}}\mathbf{G}^T}{\mathbf{1}\mathbf{G}^T}, \tag{A.20}$$

que assim preserva a não-negatividade da matriz.

Como apresentado anteriormente para o caso do quadrado da distância Euclidiana como função-custo, pode-se obter também de forma análoga a equação de atualização para a matriz $\mathbf{G}$ considerando a divergência de Kullback-Leibler como função-custo, bastando fazer as devidas substituições. Logo,

$$\mathbf{G} = \left(\mathbf{G}^T \odot \frac{\frac{\mathbf{X}^T}{\hat{\mathbf{X}}^T}\mathbf{B}}{\mathbf{1}^T\mathbf{B}} \right)^T, \tag{A.21}$$

$$\mathbf{G} = \mathbf{G} \odot \frac{\mathbf{B}^T\frac{\mathbf{X}}{\hat{\mathbf{X}}}}{\mathbf{B}^T\mathbf{1}}. \tag{A.22}$$

A.2 Para a NMFD (Equações (3.14) e (3.15))

Considerando a divergência de Kullback-Leibler generalizada, pode-se partir da equação (A.14), reescrevendo-a da forma

$$\frac{\partial D_{\text{kl}}}{\partial B^l_{n,d}} = \sum_{i=1}^{N}\sum_{j=1}^{T}\left(1 - \frac{X_{i,j}}{\hat{X}_{i,j}}\right)\frac{\partial \hat{X}_{i,j}}{\partial B^l_{n,d}}. \tag{A.23}$$

Sabendo-se que na NMFD cada elemento $\hat{X}_{i,j}$ é escrito como

$$\hat{X}_{i,j} = \sum_{l=0}^{\tau-1}\sum_{m=1}^{M} B^l_{i,m} G_{m,j-l}, \tag{A.24}$$

ao se calcular $\dfrac{\partial \hat{X}_{i,j}}{\partial B^l_{n,d}}$ e depois substituir-se o resultado obtido na equação (A.23), tem-se

$$\frac{\partial D_{\text{kl}}}{\partial B^l_{n,d}} = \sum_{i=1}^{N}\sum_{j=1}^{T}\left(1 - \frac{X_{i,j}}{\hat{X}_{i,j}}\right) G_{d,j-l}, \tag{A.25}$$

que, na forma matricial, para todos os elementos de $\mathbf{B}^l$, torna-se igual a

$$\frac{\partial D_{\text{kl}}}{\partial \mathbf{B}^l} = \left(\mathbf{1} - \frac{\mathbf{X}}{\hat{\mathbf{X}}}\right)\overrightarrow{\mathbf{G}}^{l^T}. \tag{A.26}$$

A partir daí, substituindo-se a equação (A.26) na equação (A.13) (e substituindo-se $\mathbf{B}$ por $\mathbf{B}^l$) e seguindo-se o mesmo raciocício feito na seção anterior quanto à escolha de η para o cancelamento de dois termos, obtém-se a equação de atualização de $\mathbf{B}^l$ na forma multiplicativa:

$$\mathbf{B}^l = \mathbf{B}^l \odot \frac{\dfrac{\mathbf{X}}{\hat{\mathbf{X}}}.\overrightarrow{\mathbf{G}}^{l^T}}{\mathbf{1}.\overrightarrow{\mathbf{G}}^{l^T}}, \tag{A.27}$$

que deve ser calculada para cada valor de l.

Da mesma forma, para o desenvolvimento da equação de atualização de $\mathbf{G}$, pode-se partir de

$$\frac{\partial D_{\text{kl}}}{\partial G_{d,m}} = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(1 - \frac{X_{i,t+l}}{\hat{X}_{i,t+l}}\right) B_{i,d}, \tag{A.28}$$

em que $j = t + l$. Fazendo considerações semelhantes às anteriores acerca do valor de η, obtém-se no final

$$\mathbf{G} = \mathbf{G} \odot \frac{\mathbf{B}^{l^T} . \left(\frac{\overleftarrow{\mathbf{X}}^{l}}{\hat{\mathbf{X}}} \right)}{\mathbf{B}^{l^T} . \mathbf{1}}. \tag{A.29}$$

A.3 Para a NMF2D (Equações(3.21) e (3.22))

Considerando a divergência de Kullback-Leibler generalizada, pode-se partir da equação (A.14), reescrevendo-a da forma

$$\frac{\partial D_{\text{kl}}}{\partial B^l_{n,d}} = \sum_{i=1}^{N} \sum_{j=1}^{T} \left(1 - \frac{X_{i,j}}{\hat{X}_{i,j}} \right) \frac{\partial \hat{X}_{n,j}}{\partial B^l_{n,d}}. \tag{A.30}$$

Sabendo-se que na NMF2D cada elemento $\hat{X}_{i,j}$ é escrito como

$$\hat{X}_{i,j} = \sum_{l=0}^{\tau-1} \sum_{p=0}^{\phi-1} \sum_{m=1}^{M} B^l_{i-p,m} G^p_{m,j-l}, \tag{A.31}$$

calcula-se $\dfrac{\partial \hat{X}_{i,j}}{\partial B^l_{n,d}}$, que resulta em

$$\begin{aligned} \frac{\partial \hat{X}_{i,j}}{\partial B^l_{n,d}} &= \frac{\partial}{\partial B^l_{n,d}} \sum_{l'=0}^{\tau-1} \sum_{p=0}^{\phi-1} \sum_{m=1}^{M} B^{l'}_{i-p,m} G^p_{m,j-l'} \\ &= \frac{\partial}{\partial B^l_{n,d}} \sum_{p=0}^{\phi-1} B^l_{i-p,d} G^p_{d,j-l}. \end{aligned} \tag{A.32}$$

Substituindo-se a equação (A.32) na equação (A.30), tem-se que

$$\begin{aligned} \frac{\partial D_{\text{kl}}}{\partial B^l_{n,d}} &= \sum_{p=0}^{\phi-1} \sum_{i=1}^{N} \sum_{j=1}^{T} \left(1 - \frac{X_{i,j}}{\hat{X}_{i,j}} \right) \frac{\partial}{\partial B^l_{n,d}} B^l_{i-p,d} G^p_{d,j-l} \\ &= \sum_{p=0}^{\phi-1} \sum_{j=1}^{T} \left(1 - \frac{X_{n+p,j}}{\hat{X}_{n+p,j}} \right) G^p_{d,j-l}, \end{aligned} \tag{A.33}$$

para o caso em que $i = n + p$. Quando $i \neq n + p$, o valor $\frac{\partial}{\partial B^l_{n,d}}$ é nulo.

Na forma matricial, para todos os elementos de $\mathbf{B}^l$, a equação torna-se igual a

$$\frac{\partial D_{\mathrm{kl}}}{\partial \mathbf{B}^l} = \sum_{p=0}^{\phi-1} \left(\mathbf{1} - \left(\frac{\overset{\uparrow p}{\mathbf{X}}}{\hat{\mathbf{X}}} \right) \right) \overset{\rightarrow l}{\mathbf{G}^p}^T . \tag{A.34}$$

Substituindo-se a equação (A.34) na equação (A.13), considerando-se $\mathbf{B}^l$ no lugar de $\mathbf{B}$, tem-se

$$\mathbf{B}^l = \mathbf{B}^l - \eta \left(\sum_{p=0}^{\phi-1} \mathbf{1}.\overset{\rightarrow l}{\mathbf{G}^p}^T - \sum_{p=0}^{\phi-1} \left(\frac{\overset{\uparrow p}{\mathbf{X}}}{\hat{\mathbf{X}}} \right) \overset{\rightarrow l}{\mathbf{G}^p}^T \right). \tag{A.35}$$

Fazendo-se escolha similar às que foram feitas antes para η de forma a cancelar o primeiro e o segundo termos da equação (A.35), a equação de atualização de $\mathbf{B}^l$ se torna

$$\mathbf{B}^l = \mathbf{B}^l \odot \frac{\sum_{p=0}^{\phi-1} \left(\frac{\overset{\uparrow p}{\mathbf{X}}}{\hat{\mathbf{X}}} \right) .\overset{\rightarrow l}{\mathbf{G}^p}^T}{\sum_{p=0}^{\phi-1} \mathbf{1}.\overset{\rightarrow l}{\mathbf{G}^p}^T} \tag{A.36}$$

para cada valor de l. De forma análoga, a equação de atualização para $\mathbf{G}^p$ também pode ser demonstrada. Para isso, novamente, basta fazer a decomposição de $\mathbf{X}^T$ em vez de $\mathbf{X}$, em que se substituem as matrizes $\mathbf{B}^l$ e $\mathbf{G}^p$ por $\mathbf{G}^{pT}$ e $\mathbf{B}^{lT}$ respectivamente, os operadores $\overset{\rightarrow l}{(.)}$ e $\overset{\leftarrow l}{(.)}$ por $\overset{\downarrow p}{(.)}$ e $\overset{\uparrow p}{(.)}$, respectivamente, e o somatório em p pelo somatório em l. Logo, tem-se no final

$$\mathbf{G}^p = \left(\mathbf{G}^{pT} \odot \frac{\sum_{l=0}^{\tau-1} \left(\frac{\overset{\leftarrow l}{\mathbf{X}^T}}{\hat{\mathbf{X}}^T} \right) . \overset{\downarrow p}{\mathbf{B}^l}}{\sum_{l=0}^{\tau-1} \overset{\downarrow p}{\mathbf{B}^l} .\mathbf{1}} \right)^T , \tag{A.37}$$

$$\mathbf{G}^p = \mathbf{G}^p \odot \frac{\sum_{l=0}^{\tau-1} \overset{\downarrow p}{\mathbf{B}^l}^T . \left(\frac{\overset{\leftarrow l}{\mathbf{X}}}{\hat{\mathbf{X}}} \right)}{\sum_{l=0}^{\tau-1} \overset{\downarrow p}{\mathbf{B}^l}^T .\mathbf{1}}. \tag{A.38}$$

A.4 Para o algoritmo da NMF básica com restrição (Equações (3.25) e (3.26))

O desenvolvimento é muito similar ao mostrado na Seção A.1 deste apêndice. Considerando a divergência de Kullback-Leibler generalizada como medida de distorção e o caso em que há apenas uma restrição c_B com peso α e que ela se dá na matriz $\mathbf{B}$, tem-se que

$$\frac{\partial \mathrm{D}_{\text{custo}}}{\partial B_{n,d}} = -\sum_{i=1}^{N}\sum_{j=1}^{T}\left(1 - \frac{X_{i,j}}{\hat{X_{i,j}}}\right) G_{d,j} + \frac{\partial c_B}{\partial B_{n,d}}, \tag{A.39}$$

em analogia à equação (A.15), que, na forma matricial, para todos os elementos de $\mathbf{B}$, torna-se

$$\frac{\partial \mathrm{D}_{\text{custo}}}{\partial B_{n,d}} = -\sum_{i=1}^{N}\sum_{j=1}^{T}\left(\mathbf{1} - \frac{\mathbf{X}}{\hat{\mathbf{X}}}\right) \mathbf{G}^T + \nabla_{\mathbf{B}} c_B. \tag{A.40}$$

Comparando com o que é mostrado na Seção A.2, o valor de η deverá ser igual a

$$\eta = \frac{\mathbf{B}}{\mathbf{1}\mathbf{G}^T + \nabla_{\mathbf{B}} c_B}, \tag{A.41}$$

de forma a restar somente um termo multiplicativo na regra de atualização de $\mathbf{B}$. Portanto, tem-se que

$$\mathbf{B} = \mathbf{B} \odot \frac{\frac{\mathbf{X}}{\hat{\mathbf{X}}}\mathbf{G}^T}{\mathbf{1}\mathbf{G}^T + \nabla_{\mathbf{B}} c_B}. \tag{A.42}$$

Se houver restrição, a equação de atualização de $\mathbf{G}$ é análoga à equação (A.42), com o operador ∇ incidindo sobre o critério correspondente. Se houver mais do que um critério, basta somar cada operador ∇ ao denominador da equação de atualização da matriz sobre a qual cada critério incide. As equações de atualização para as matrizes $\mathbf{B}$ e $\mathbf{G}$ quando não há restrições sobre elas são exatamente iguais às do algoritmo básico, o que significa que a NMF com restrições nada mais é do que uma generalização da NMF básica.

A.5 Para a SNMF2D (Equações (3.33) e (3.34))

Na SNMF2D, cada elemento $\hat{X}_{i,j}$ é escrito como

$$\hat{X}_{i,j} = \sum_{m=1}^{M}\sum_{l=0}^{\tau-1}\sum_{p=0}^{\phi-1} \tilde{B}^{l}_{i-p,m} G^{p}_{m,j-l}. \tag{A.43}$$

Considerando a equação (3.30), pode-se rescreever cada elemento $\hat{X}_{i,j}$ como sendo obtido da forma

$$\hat{X}_{i,j} = \sum_{m=1}^{M} \sum_{l=0}^{\tau-1} \sum_{p=0}^{\phi-1} \left(\frac{B_{i-p,m}^{l} G_{m,j-l}^{p}}{\sqrt{\sum_{l'} \sum_{p'} \left(B_{i-p',m}^{l'} \right)^2}} \right). \tag{A.44}$$

O denominador da equação (A.44) pode ser colocado fora dos somatórios em l e p, já que está em função de outros dois somatórios em variáveis que representam os mesmos índices, mas que ali são chamadas de l' e p' para diferenciá-las de l e p. Portanto,

$$\hat{X}_{i,j} = \sum_{m=1}^{M} \left(\frac{\sum_{l=0}^{\tau-1} \sum_{p=0}^{\phi-1} B_{i-p,m}^{l} G_{m,j-l}^{p}}{\sqrt{\sum_{l'} \sum_{p'} \left(B_{i-p',m}^{l'} \right)^2}} \right). \tag{A.45}$$

Substituindo-se a equação anterior na equação (A.30) (que considera a divergência de Kullback-Leibler como medida de distorção), tem-se que

$$\frac{\partial D_{\mathrm{kl}}}{\partial B_{n,m_0}^{l_0}} = \sum_{i=1}^{N} \sum_{j=1}^{T} \left\{ \left(1 - \frac{X_{i,j}}{\hat{X}_{i,j}} \right) \frac{\partial}{\partial B_{n,m_0}^{l_0}} \left[\sum_{m=1}^{M} \left(\frac{\sum_{l=0}^{\tau-1} \sum_{p=0}^{\phi-1} B_{i-p,m}^{l} G_{m,j-l}^{p}}{\sqrt{\sum_{l'} \sum_{p'} \left(B_{i-p',m}^{l'} \right)^2}} \right) \right] \right\}. \tag{A.46}$$

Como se deseja fazer a derivada em relação a um elemento $B_{n,m_0}^{l_0}$ específico, pode-se eliminar o somatório em m da equação anterior, substituindo-se m por m_0. Então,

$$\frac{\partial D_{\mathrm{kl}}}{\partial B_{n,m_0}^{l_0}} = \sum_{i=1}^{N} \sum_{j=1}^{T} \left\{ \left(1 - \frac{X_{i,j}}{\hat{X}_{i,j}} \right) \frac{\partial}{\partial B_{n,m_0}^{l_0}} \left[\frac{\sum_{l=0}^{\tau-1} \sum_{p=0}^{\phi-1} B_{i-p,m_0}^{l} G_{m_0,j-l}^{p}}{\sqrt{\sum_{l'} \sum_{p'} \left(B_{i-p',m_0}^{l'} \right)^2}} \right] \right\}. \tag{A.47}$$

Fazendo-se

$$u = \sum_{l=0}^{\tau-1} \sum_{p=0}^{\phi-1} B_{i-p,m_0}^{l} G_{m_0,j-l}^{p} \tag{A.48}$$

e

$$v = \sqrt{\sum_{l'} \sum_{p'} \left(B^{l'}_{i-p',m_0} \right)^2} = \left\| B_{m_0} \right\|_2 , \tag{A.49}$$

e aplicando-se a regra do quociente para derivação, em que

$$\left(\frac{u}{v} \right)' = \frac{u'v - uv'}{v^2}, \tag{A.50}$$

tem-se que

$$\begin{aligned} \frac{\partial D_{\text{kl}}}{\partial B^{l_0}_{n,m_0}} &= \sum_{i=1}^{N} \sum_{j=1}^{T} \left[\left(1 - \frac{X_{i,j}}{\hat{X}_{i,j}} \right) \left(\frac{v \dfrac{\partial u}{\partial B^{l_0}_{n,m_0}} - u \dfrac{\partial v}{\partial B^{l_0}_{n,m_0}}}{v^2} \right) \right] \\ &= \sum_{i=1}^{N} \sum_{j=1}^{T} \left[\left(1 - \frac{X_{i,j}}{\hat{X}_{i,j}} \right) \left(\frac{\frac{\partial u}{\partial B^{l_0}_{n,m_0}}}{v} - \frac{u \frac{\partial v}{\partial B^{l_0}_{n,m_0}}}{v^2} \right) \right]. \end{aligned} \tag{A.51}$$

Agora, calcula-se cada uma das derivadas. Considerando primeiro $\dfrac{\partial u}{\partial B^{l_0}_{n,m_0}}$, obtém-se

$$\begin{aligned} \frac{\partial u}{\partial B^{l_0}_{n,m_0}} &= \frac{\partial}{\partial B^{l_0}_{n,m_0}} \sum_{l=0}^{\tau-1} \sum_{p=0}^{\phi-1} B^{l}_{i-p,m_0} G^{p}_{m_0,j-l} \\ &= \frac{\partial}{\partial B^{l_0}_{n,m_0}} \sum_{p=0}^{\phi-1} B^{l}_{i-p,m_0} G^{p}_{m_0,j-l}. \end{aligned} \tag{A.52}$$

Considerando $\dfrac{\partial v}{\partial B_{n,m_0}^{l_0}}$ e utilizando a Regra da Cadeia para derivada, obtém-se

$$\begin{aligned}
\frac{\partial v}{\partial B_{n,m_0}^{l_0}} &= \frac{\partial}{\partial B_{n,m_0}^{l_0}} \sqrt{\sum_{l'=0}^{\tau-1}\sum_{p'=0}^{\phi-1}\left(B_{i-p',m_0}^{l'}\right)^2} \\
&= \frac{1}{2}\left(\sum_{l'=0}^{\tau-1}\sum_{p'=0}^{\phi-1}\left(B_{i-p',m_0}^{l'}\right)^2\right)^{-\frac{1}{2}} \frac{\partial}{\partial B_{n,m_0}^{l_0}} \sum_{l'=0}^{\tau-1}\sum_{p'=0}^{\phi-1}\left(B_{i-p',m_0}^{l'}\right)^2 \\
&= \frac{1}{2}\left(\sum_{l'=0}^{\tau-1}\sum_{p'=0}^{\phi-1}\left(B_{i-p',m_0}^{l'}\right)^2\right)^{-\frac{1}{2}} \frac{\partial}{\partial B_{n,m_0}^{l_0}} \sum_{p'=0}^{\phi-1}\left(B_{i-p',m_0}^{l_0}\right)^2 \\
&= \frac{\frac{1}{2}\frac{\partial}{\partial B_{n,m_0}^{l_0}} \sum_{p'=0}^{\phi-1}\left(B_{i-p',m_0}^{l_0}\right)^2}{\|B_{m_0}\|_2}
\end{aligned} \tag{A.53}$$

Agora, substituem-se u, v, $\dfrac{\partial u}{\partial B_{n,m_0}^{l_0}}$ e $\dfrac{\partial v}{\partial B_{n,m_0}^{l_0}}$ por seus respectivos valores na equação (A.51). Assim, tem-se que

$$\frac{\partial D_{\mathrm{kl}}}{\partial B_{n,m_0}^{l_0}} = \sum_{p=0}^{\phi-1}\sum_{j=1}^{T}\left[\left(1-\frac{X_{n+p,j}}{\hat{X}_{n+p,j}}\right)\left(\frac{K_1}{\|B_{m_0}\|_2} - \frac{K_2}{\|B_{m_0}\|_2^{\,3}}\right)\right], \tag{A.54}$$

onde

$$\begin{aligned}
K_1 &= \frac{\partial}{\partial B_{n,m_0}^{l_0}} B_{i-p,m_0}^{l} G_{m_0,j-l}^{p} \\
&= G_{m_0,j-l}^{p},
\end{aligned} \tag{A.55}$$

quando $i = n + p$ (sendo $K_1 = 0$ quando $i \neq n + p$), e

$$\begin{aligned}
K_2 &= \frac{1}{2}\left(\sum_{l=0}^{\tau-1} B_{i-p,m_0}^{l} G_{m_0,j-l}^{p}\right) \frac{\partial}{\partial B_{n,m_0}^{l_0}} \sum_{p'=0}^{\phi-1}\left(B_{i-p',m_0}^{l_0}\right)^2 \\
&= \left(\sum_{l=0}^{\tau-1} B_{i-p,m_0}^{l} G_{m_0,j-l}^{p}\right) \sum_{p'=0}^{\phi-1} B_{n,m_0}^{l_0},
\end{aligned} \tag{A.56}$$

quando $i = n + p'$ (sendo $K_2 = 0$ quando $i \neq n + p'$). Portanto, para que tanto K_1 quanto K_2 não sejam nulos, p deve ser igual a p'. Nota-se que os somatórios em p que estariam presentes em K_1 e K_2, devidos às expressões de $\frac{\partial u}{\partial B_{n,m_0}^{l_0}}$ e u respectivamente, estão representados em apenas um somatório em p no início da equação (A.54). Como p' deve sempre acompanhar o valor de p, não faz sentido

manter o somatório em p' na expressão de K_2 (que, aliás, deixa de estar em função de p' após o cálculo da derivada). Portanto,

$$K_2 = \left(\sum_{l=0}^{\tau-1} B^l_{i-p,m_0} G^p_{m_0,j-l} \right) B^{l_0}_{n,m_0}. \tag{A.57}$$

Nota-se também que o somatório em i que havia no início da equação (A.51) também foi descartado devido à substituição $i = n + p$.

Ao se considerar a notação normalizada e substituindo K_1 e K_2 por suas respectivas expressões, a equação (A.54) pode ser reescrita da forma

$$\begin{aligned} \frac{\partial D_{\text{kl}}}{\partial B^{l_0}_{n,m_0}} &= \sum_{p=0}^{\phi-1} \sum_{j=1}^{T} \left[\left(1 - \frac{X_{n+p,j}}{\hat{X}_{n+p,j}} \right) \left(\frac{G^p_{m_0,j-l}}{\|B_{m_0}\|_2} - \frac{\tilde{B}^{l_0}_{n,m_0} \sum_{l=0}^{\tau-1} \tilde{B}^l_{n,m_0} G^p_{m_0,j-l}}{\|B_{m_0}\|_2} \right) \right] \\ &= \frac{1}{\|B_{m_0}\|_2} \sum_{p=0}^{\phi-1} \sum_{j=1}^{T} \left[\left(1 - \frac{X_{n+p,j}}{\hat{X}_{n+p,j}} \right) \left(G^p_{m_0,j-l} - \tilde{B}^{l_0}_{n,m_0} \sum_{l=0}^{\tau-1} \tilde{B}^l_{n,m_0} G^p_{m_0,j-l} \right) \right] \end{aligned} \tag{A.58}$$

Agora, deve-se expandir a expressão que está entre colchetes. Portanto,

$$\frac{\partial D_{\text{kl}}}{\partial B^{l_0}_{n,m_0}} = \frac{1}{\|B_{m_0}\|_2} \sum_{p=0}^{\phi-1} \sum_{j=1}^{T} (A + B - C - D), \tag{A.59}$$

onde

$$A = G^p_{m_0,j-l_0}, \tag{A.60}$$

$$B = \frac{X_{n+p,j}}{\hat{X}_{n+p,j}} \tilde{B}^{l_0}_{n,m_0} \sum_{l=0}^{\tau-1} \tilde{B}^l_{n,m_0} G^p_{m_0,j-l}, \tag{A.61}$$

$$C = \tilde{B}^{l_0}_{n,m_0} \sum_{l=0}^{\tau-1} \tilde{B}^l_{n,m_0} G^p_{m_0,j-l}, \tag{A.62}$$

$$D = \frac{X_{n+p,j}}{\hat{X}_{n+p,j}} G^p_{m_0,j-l_0}. \tag{A.63}$$

Inserindo a equação (A.59) na regra de atualização para o método do gradiente descendente mostrada na equação (A.13), e substituindo inicialmente os dois termos

$\mathbf{B}$ por $\tilde{\mathbf{B}}^l$, ou seja,

$$\tilde{\mathbf{B}}^l = \tilde{\mathbf{B}}^l - \eta \frac{1}{\|B_m\|_2} \sum_{p=0}^{\phi-1} \sum_{j=1}^{T} (A + B - C - D), \tag{A.64}$$

deve-se escolher o valor de η de tal forma que se tenha uma equação multiplicativa que garanta a não-negatividade de $\tilde{\mathbf{B}}^l$. Fazendo

$$\eta = \frac{\tilde{\mathbf{B}}^l \|B_m\|_2}{\sum_{p=0}^{\phi-1} \sum_{j=1}^{T} (A + B)}, \tag{A.65}$$

e rearranjando os termos, obtém-se

$$\tilde{\mathbf{B}}^l = \tilde{\mathbf{B}}^l \odot \frac{\sum_{p=0}^{\phi-1} \sum_{j=1}^{T} (D + C)}{\sum_{p=0}^{\phi-1} \sum_{j=1}^{T} (A + B)}, \tag{A.66}$$

$$\mathbf{B}^l = \tilde{\mathbf{B}}^l \odot \frac{\sum_{p=0}^{\phi-1} \sum_{j=1}^{T} \left(\frac{X_{n+p,j}}{\hat{X}_{n+p,j}} G^p_{m_0,j-l_0} + \tilde{B}^{l_0}_{n,m_0} \sum_{l=0}^{\tau-1} \tilde{B}^l_{n,m_0} G^p_{m_0,j-l} \right)}{\sum_{p=0}^{\phi-1} \sum_{j=1}^{T} \left(G^p_{m_0,j-l_0} + \frac{X_{n+p,j}}{\hat{X}_{n+p,j}} \tilde{B}^{l_0}_{n,m_0} \sum_{l=0}^{\tau-1} \tilde{B}^l_{n,m_0} G^p_{m_0,j-l} \right)}. \tag{A.67}$$

Nota-se que no lado esquerdo da equação (A.67), a normalização de $\mathbf{B}^l$ foi retirada. Isso se deve ao fato de que, na realidade, como o valor de $\mathbf{B}^l$ é atualizado, deve-se realizar nova normalização sobre o novo valor, pois a atualização não garante sua normalização automática.

Reescrevendo a expressão anterior na forma matricial, para todos os elementos $\mathbf{B}^l$ $(l = 0, 1, 2, ..., \tau - 1)$, tem-se que a fórmula de atualização é dada por

$$\mathbf{B}^l = \tilde{\mathbf{B}}^l \odot \frac{\sum_{p=0}^{\phi-1} \left[\left(\frac{\overset{\uparrow p}{\mathbf{X}}}{\hat{\mathbf{X}}} \right) . \overset{\rightarrow l}{\mathbf{G}^p}^T + \tilde{\mathbf{B}}^l \mathrm{diag} \left(\sum_{l=0}^{\tau-1} \mathbf{1}. \left(\left(\mathbf{1}. \overset{\rightarrow l}{\mathbf{G}^p}^T \right) \odot \tilde{\mathbf{B}}^l \right) \right) \right]}{\sum_{p=0}^{\phi-1} \left[\mathbf{1}. \overset{\rightarrow l}{\mathbf{G}^p}^T + \tilde{\mathbf{B}}^l \mathrm{diag} \left(\sum_{l=0}^{\tau-1} \mathbf{1}. \left(\left(\left(\frac{\overset{\uparrow p}{\mathbf{X}}}{\hat{\mathbf{X}}} \right) . \overset{\rightarrow l}{\mathbf{G}^p}^T \right) \odot \tilde{\mathbf{B}}^l \right) \right) \right]}. \tag{A.68}$$

Observa-se que, na equação (A.68), os primeiros termos da soma dentro dos colchetes tanto no numerador quanto no denominador são idênticos aos termos da NMF2D. Os termos adicionais no numerador e no denominador são devidos à normalização que foi proposta para $\mathbf{B}^l$ na SNMF2D, mostrada na equação (3.30).

A correspondência entre as equações (A.67) e (A.68) não é trivial e carece de explicação. Tomando apenas o termo adicional no numerador para que se possa analisar a correspondência entre as equações (A.67) e (A.68), observa-se que não é possível simplesmente realizar a multiplicação direta entre $\tilde{\mathbf{B}}^l$ e $\overrightarrow{\mathbf{G}^p}^l$. Portanto, à matriz $\overrightarrow{\mathbf{G}^p}^{l\,T}$ (TxD), multiplica-se uma matriz de elementos unitários (NxT), de forma que se possa multiplicar $\mathbf{1}_{\cdot}^{\rightarrow l}\,\mathbf{G}^{p\,T}$ ponto-a-ponto com $\tilde{\mathbf{B}}^l$ (NxD) e também realizar o somatório em j. O somatório em l mantém-se explícito na equação (A.68). O operador diag(.) tem a função de inserir seus argumentos (que foram vetorizados pelo matriz unitária de dimensão 1xN) na diagonal de uma matriz quadrada de dimensão igual ao número de argumentos (M) e com o restante ds elementos iguais a zero. O objetivo é evitar que haja somatório na dimensão M quando todo o conjunto é multiplicado por $\tilde{\mathbf{B}}^l$ novamente. Para o denominador, a análise é análoga.

O desenvolvimento para se obter a equação de atualização da matriz $\mathbf{G}^p$ na SNMF2D é muito parecido com o da NMF2D, visto que $\mathbf{G}^p$ não sofre nenhum tipo de normalização que possa alterar a minimização da função-custo. A única diferença é que a equação carrega consigo um termo adicional ao denominador referente ao critério de esparsidade inserido, conforme visto na Secção (3.6). Portanto,

$$\mathbf{G}^p = \tilde{\mathbf{G}}^p \odot \frac{\sum_{l=0}^{\tau-1} \overrightarrow{\mathbf{B}^l}^{p\,T} \cdot \left(\frac{\overleftarrow{\mathbf{X}}^l}{\hat{\mathbf{X}}}\right)}{\sum_{l=0}^{\tau-1} \left(\overrightarrow{\mathbf{B}^l}^{p\,T} \cdot \mathbf{1}\right) + \alpha \nabla_{\mathbf{G}^p} c_{esp}}, \tag{A.69}$$

para $p = 0, 1, ..., \phi - 1$.

Apêndice B

Demonstrações das derivadas dos critérios para a CNMF2D

B.1 Demonstração da equação (4.2)

Partindo da equação(4.1), o objetivo é encontrar sua derivada em relação a $\mathbf{G}^p$, ou seja,

$$\nabla_{\mathbf{G}^p}(c_{\text{ep}}) = \frac{\partial c_{\text{ep}}(\mathbf{G})}{\partial \mathbf{G}^p} = -\frac{\partial}{\partial \mathbf{G}^p} \sum_p \left|\log\left(1 + \mathbf{G}^p \odot \mathbf{G}^p\right)\right|. \tag{B.1}$$

Para determinado elemento $G^p_{m,t}$ de $\mathbf{G}^p$, tem-se que

$$\frac{\partial c_{\text{ep}}(\mathbf{G})}{\partial G^p_{m,t}} = -\frac{\partial}{\partial G^p_{m,t}} \sum_{d,j,p} \log\left(1 + G^p_{d,j} G^p_{d,j}\right). \tag{B.2}$$

Utilizando a Regra da Cadeia para derivadas,

$$\frac{\partial c_{\text{ep}}(\mathbf{G})}{\partial G^p_{m,t}} = -\frac{\partial}{\partial G^p_{m,t}} \sum_{d,j,p} \log\left(1 + G^p_{d,j} G^p_{d,j}\right) = -\frac{1}{1 + G^p_{m,t} G^p_{m,t}} 2G^p_{m,t}. \tag{B.3}$$

Colocando-se na forma matricial,

$$\frac{\partial c_{\text{ep}}(\mathbf{G})}{\partial \mathbf{G}^p} = -\frac{2\mathbf{G}^p}{1 + \mathbf{G}^p \odot \mathbf{G}^p}, \tag{B.4}$$

para $p = 0, 1, ..., \phi - 1$.

B.2 Demonstração da equação (4.8)

Partindo da equação (4.7), o objetivo é encontrar sua derivada em relação a $\mathbf{G}^p$, ou seja,

$$\nabla_{\mathbf{G}^p}(c_{\text{cc}}) = \frac{\partial c_{\text{cc}}(\mathbf{G})}{\partial \mathbf{G}^p} = \frac{\partial}{\partial \mathbf{G}^p} \sum_p \left| \mathbf{W} \odot \left(\mathbf{G}^p \mathbf{G}^{pT} \right) \right|. \tag{B.5}$$

Para determinado elemento $G^p_{m,t}$ de $\mathbf{G}^p$, tem-se que

$$\frac{\partial c_{\text{cc}}(\mathbf{G})}{\partial G^p_{m,t}} = \frac{\partial}{\partial G^p_{m,t}} \sum_{d,k,j,p} W_{d,k} G^p_{d,j} G^p_{k,j} = \sum_d W_{d,m} G^p_{d,t} + \sum_k W_{m,k} G^p_{k,t}. \tag{B.6}$$

Pode-se fazer $d = k$, já que são índices que representam as mesmas variáveis. Portanto,

$$\frac{\partial c_{\text{cc}}(\mathbf{G})}{\partial G^p_{m,t}} = \sum_d \left(W_{d,m} + W_{m,d} \right) G^p_{d,t}. \tag{B.7}$$

Colocando-se na forma matricial,

$$\frac{\partial c_{\text{cc}}(\mathbf{G})}{\partial \mathbf{G}^p} = \left(\mathbf{W} + \mathbf{W}^T \right) \mathbf{G}^p, \tag{B.8}$$

para $p = 0, 1, ..., \phi - 1$. Sabendo que $\mathbf{W}$ é uma matriz circulante, tem-se que $\mathbf{W} = \mathbf{W}^T$. Logo,

$$\frac{\partial c_{\text{cc}}(\mathbf{G})}{\partial \mathbf{G}^p} = 2\mathbf{W}\mathbf{G}^p. \tag{B.9}$$

B.3 Demonstração da equação (4.10)

Partindo da equação(4.9), o objetivo é encontrar sua derivada em relação a $\mathbf{G}^p$, ou seja,

$$\nabla_{\mathbf{G}^p}(c_{\text{cc}}) = \frac{\partial c_{\text{cc}}(\mathbf{G})}{\partial \mathbf{G}^p} = \sum_{p=0}^{\phi-1} \left(\left| \mathbf{G}^{pT} \mathbf{G}^p \right| - \left| \mathbf{G}^p \odot \mathbf{G}^p \right| \right) \tag{B.10}$$

Para determinado elemento $G^p_{m,t}$ de $\mathbf{G}^p$, tem-se que

$$\begin{aligned} \frac{\partial c_{\text{cc}}(\mathbf{G})}{\partial G^p_{m,t}} &= \frac{\partial}{\partial G^p_{m,t}} \sum_{d,i,j,p} G^p_{d,i} G^p_{d,j} - \frac{\partial}{\partial G^p_{m,t}} \sum_{d,j,p} G^p_{d,j} G^p_{d,j} \\ &= \frac{\partial}{\partial G^p_{m,t}} \sum_{d,i,j,p} G^p_{d,i} G^p_{d,j} - \frac{\partial}{\partial G^p_{m,t}} \sum_{d,j,p} \left(G^p_{d,j} \right)^2 \\ &= \sum_i G^p_{m,i} + \sum_j G^p_{m,j} - 2G^p_{m,t}. \end{aligned} \tag{B.11}$$

Pode-se fazer $i = j$, já que são índices que representam as mesmas variáveis. Portanto,

$$\frac{\partial c_{\text{cc}}(\mathbf{G})}{\partial G^p_{m,t}} = 2\sum_i G^p_{m,i} - 2G^p_{m,t}. \tag{B.12}$$

Colocando-se na forma matricial,

$$\frac{\partial c_{\text{cc}}(\mathbf{G})}{\partial \mathbf{G}^p} = 2\mathbf{G}^p\left(\mathbf{1} - \mathbf{I}\right), \tag{B.13}$$

para $p = 0, 1, ..., \phi - 1$, em que $\mathbf{1}$ é uma matriz unitária e $\mathbf{I}$ é uma matriz identidade, ambas de dimensão TxT.

B.4 Demonstração da equação (4.12)

Partindo da equação (4.11), o objetivo é encontrar sua derivada em relação a $\mathbf{G}^p$, ou seja,

$$\nabla_{\mathbf{G}^p}(c_{\text{ct}}) = \frac{\partial c_{\text{ct}}(\mathbf{G})}{\partial \mathbf{G}^p} = -\frac{\partial}{\partial \mathbf{G}^p}\sum_p \left|\mathbf{V} \odot \left(\mathbf{G}^{pT}\mathbf{G}^p\right)\right|. \tag{B.14}$$

Para determinado elemento $G^p_{m,t}$ de $\mathbf{G}^p$, tem-se que

$$\frac{\partial c_{\text{ct}}(\mathbf{G})}{\partial G^p_{m,t}} = -\frac{\partial}{\partial G^p_{m,t}}\sum_{d,i,j,p} V_{i,j}G^p_{d,i}G^p_{d,j} = -\sum_i V_{i,t}G^p_{m,i} + \sum_j V_{t,j}G^p_{m,j}. \tag{B.15}$$

Pode-se fazer $i = j$, já que são índices que representam as mesmas variáveis. Portanto,

$$\frac{\partial c_{\text{ct}}(\mathbf{G})}{\partial G^p_{m,t}} = -\sum_i \left(V_{i,m} + V_{m,i}\right) G^p_{m,t}. \tag{B.16}$$

Colocando-se na forma matricial,

$$\frac{\partial c_{\text{ct}}(\mathbf{G})}{\partial \mathbf{G}^p} = -\left(\mathbf{V} + \mathbf{V}^T\right)\mathbf{G}^p. \tag{B.17}$$

Sabendo que $\mathbf{V}$ é uma matriz circulante, tem-se que $\mathbf{V} = \mathbf{V}^T$. Logo,

$$\frac{\partial c_{\text{ct}}(\mathbf{G})}{\partial \mathbf{G}^p} = -2\mathbf{V}\mathbf{G}^p, \tag{B.18}$$

para $p = 0, 1, ..., \phi - 1$.