

Aprendizado de Dicionários para a Separação Semi-Cega de Fontes por Matching Pursuit

Diego Barreto Haddad^{1,2}, Mariane Rembold Petraglia², Lisandro Lovisoló³ e Paulo Bulkool Batalheiro³

Resumo—A Análise de Componentes Esparsos é uma das técnicas mais utilizadas para resolver o problema da separação de fontes. Neste artigo, uma técnica fundamentada em decomposições atômicas é empregada na identificação do sistema de mistura. Caso sejam acessíveis trechos isolados das fontes a separar, mostramos que recorrer a dicionários adaptados pode implicar uma identificação mais acurada (com um MSE inferior em aproximadamente 15 dB) e menos custosa em termos computacionais.

Palavras-Chave—Separação de Fontes, Matching Pursuit, Aprendizagem de Dicionários

Abstract—The Sparse Component Analysis is one of the most used techniques for handle the source separation problem. In this paper, a technique based in atomic decompositions, is used in the identification of a mixture system. If isolated excerpts of the sources are available, we show that using adapted dictionaries can imply a better (with a inferior MSE in, approximately, 15 dB) and less computationally intensive identification procedure.

Keywords—Source Separation, Matching Pursuit, Dictionary Learning

I. INTRODUÇÃO

Técnicas de separação de sinais (também conhecidos como fontes), empreendidas a partir de misturas dos mesmos, são objeto de pesquisa intensa desde meados da década de 90. Pode-se dizer que tais técnicas constituem um ramo maduro da área de processamento estatístico de sinais, embora haja diversos desafios (como a reverberação) que comprometem o desempenho das mais sofisticadas estratégias [1].

Duas das vertentes mais tradicionais utilizadas para tratar do problema da separação de fontes são a ICA (Análise de Componentes Independentes, do inglês *Independent Component Analysis*) e a SCA (Análise de Componentes Esparsos, do inglês *Sparse Component Analysis*) [2],[3].

Se temos acesso a M misturas e $s_i(k)$ denota a i -ésima fonte, com $i \in \{1, 2, \dots, N\}$, onde N é o número de fontes, um processo de mistura instantâneo e linear pode ser descrito como:

$$\underbrace{\begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_M(k) \end{bmatrix}}_{\triangleq \mathbf{x}_k} = \underbrace{\begin{bmatrix} h_{11} & h_{12} & \dots & h_{1N} \\ h_{21} & h_{22} & \dots & h_{2N} \\ \vdots & \ddots & \ddots & \vdots \\ h_{M1} & h_{M2} & \dots & h_{MN} \end{bmatrix}}_{\triangleq \mathbf{H}} \underbrace{\begin{bmatrix} s_1(k) \\ s_2(k) \\ \vdots \\ s_N(k) \end{bmatrix}}_{\triangleq \mathbf{s}_k}, \quad (1)$$

onde a matrix \mathbf{H} , desconhecida, é chamada de matrix de mistura, e $x_j(k)$ é a j -ésima mistura. Neste artigo supomos

que $h_{1i} = 1, \forall i \in \{1, 2, \dots, N\}$. Essa suposição não é restritiva, dada a inerente ambiguidade de escalamento das técnicas cegas de separação [2].

Este artigo emprega uma abordagem via SCA e, portanto, por não atuar no espaço de separação como as técnicas de ICA, pode se aplicar a um contexto subdeterminado (que se caracteriza por apresentar um número de fontes superior ao de sensores) o qual, dada a menor disponibilidade de informação, é mais desafiador.

As representações redundantes são exploradas neste artigo por meio da técnica *Matching Pursuit* (MP) [4]. A obtenção de representações esparsas que, apelando a conjuntos redundantes (ou dicionários) de formas de onda pré-definidas (também conhecidas como átomos), não raro permite uma representação compacta dos sinais de interesse.

Nos últimos anos, houve uma grande atividade no campo de aproximações esparsas. Isto é parcialmente motivado pelo potencial das técnicas relacionadas para tarefas típicas em processamento de sinais como análise, redução de dimensionalidade, retirada de ruído ou compressão, nas quais é necessária uma representação própria dos conjuntos de dados por meio de subespaços de dimensionalidade reduzida, os quais são adaptados tanto para as características dos sinais quanto para a tarefa de processamento em particular [5]. A principal vantagem desta classe de técnicas é a completa liberdade de projeto do dicionário, o qual pode ser eficientemente adaptado para emular as estruturas do sinal [6].

Este artigo investiga se a informação lateral oriunda de trechos isolados das fontes (exógenos às misturas) permite-nos utilizar técnicas de adaptação de dicionários, de sorte a melhorar o desempenho de técnicas de identificação da matrix de mistura \mathbf{H} .

II. OBTENÇÃO DE REPRESENTAÇÕES ESPARSAS VIA MP

Em expansões redundantes, o sinal é escrito em termos de uma combinação linear de um subconjunto de vetores extraídos de um dicionário redundante. Tais vetores não são linearmente independentes, o que acarreta perda da unicidade da expansão e aumento da complexidade computacional. Encontrar a representação mais esparsa de um sinal em um dado dicionário redundante é, em geral, um difícil problema NP (do inglês *Nondeterministic Polynomial Time*). Mesmo algoritmos subótimos (ou que empregam heurísticas subótimas) implicam alto custo computacional. O emprego dessas heurísticas permite encontrar representações que, sendo suficientemente bem-sucedidas, não se distanciam em demasia da otimalidade. Essas estratégias reduzem a complexidade computacional por

¹Coordenação de Telecomunicações, CEFET-RJ UnED Nova Iguaçu, Nova Iguaçu-RJ, Brasil. ²Programa de Engenharia Elétrica (PEE), COPPE/UFRJ, Rio de Janeiro-RJ, Brasil. ³Departamento de Engenharia Eletrônica e de Telecomunicações/PROSAICO, UERJ, Rio de Janeiro-RJ, Brasil. E-mails: diego@pads.ufrj.br, mariane@pads.ufrj.br, lisandro@uerj.br. Este trabalho foi parcialmente financiado pelo CNPq e pela FAPERJ.

meio do relaxamento do problema de otimização original [6], o que permite esquivar-se da elevada complexidade oriunda de uma busca combinatorial.

Uma dessas estratégias mais populares é o *Matching Pursuit* (MP) [4] o qual, partindo de um dicionário redundante, define uma decomposição adaptativa de um sinal. Esta decomposição progressivamente isola as estruturas do sinal coerentes com o dicionário adotado por meio de projeções ortogonais nos átomos do dicionário, o que provê uma representação adaptativa na qual os coeficientes mais significativos são extraídos primeiramente [7]. Tal propriedade adaptativa do MP é adequada para a SCA, pois motiva uma técnica de identificação do sistema de mistura que, sem apelo à inversão do sistema, pode reconstruir as fontes mesmo quando o sistema é degenerado, como no caso em que o número de fontes supera o de sensores.

Seja um dicionário $\mathcal{D} = \{\phi^{(m)}\}$, com $m \in \{1, \dots, \#D\}$ composto de átomos de comprimento L e norma unitária, com o m -ésimo átomo denotado por $\phi^{(m)}$. Cumpra ao MP decompor um determinado vetor-coluna $\bar{\mathbf{x}}_i$ (de comprimento L), de forma iterativa de modo que, findo o processo, o vetor $\bar{\mathbf{x}}_i$ seja decomposto como:

$$\bar{\mathbf{x}}_i = \sum_{j=1}^J \langle \mathcal{R}^{(j-1)}, \phi^{(m_j)} \rangle \phi^{(m_j)} + \mathcal{R}^{(J)}, \quad (2)$$

onde $\mathcal{R}^{(J)}$ é o resíduo obtido após J iterações.

III. EMPREGO DO MP NA IDENTIFICAÇÃO DO SISTEMA DE MISTURA

Por simplicidade, doravante consideramos o caso estéreo ($M = 2$). Para um dado dicionário (e portanto, para um L específico), o k -ésimo bloco das misturas é formado por trechos sincronizados $\{\bar{\mathbf{x}}_1(k), \bar{\mathbf{x}}_2(k)\}$, onde $\bar{\mathbf{x}}_i$ é um bloco de comprimento L da i -ésima mistura. Neste contexto multicanal, pode-se adaptar o MP de modo a selecionar um átomo $\phi^{(m)}$ que maximize $|\langle \phi^{(m)}, \bar{\mathbf{x}}_1 \rangle|^2 + |\langle \phi^{(m)}, \bar{\mathbf{x}}_2 \rangle|^2$. Este procedimento é adotado neste artigo. Sendo as fontes (por hipótese) descorrelacionadas e os átomos bem localizados no espaço tempo \times frequência, é comum na literatura a suposição de que cada componente (átomo escalado) esteja majoritariamente associado a uma das fontes, de sorte que

$$\lambda_h(k) = \frac{|\langle \phi^{(m)}, \bar{\mathbf{x}}_2(k) \rangle|}{|\langle \phi^{(m)}, \bar{\mathbf{x}}_1(k) \rangle|} \approx h_{2i}, \quad (3)$$

onde $i \in \{1, 2, \dots, N\}$. Assim, o cálculo da razão $\lambda_h(k)$ num dado bloco fornece uma estimativa pontual de um dos coeficientes da segunda linha de \mathbf{H} . Aplicando a decomposição MP em diversos blocos e efetuando a extração de picos de uma função potencial (que é uma espécie de histograma), é possível estimar a segunda linha de \mathbf{H} (vide [8] para maiores detalhes).

Uma escolha popular de dicionários consiste no emprego de dicionários parametrizados [4] obtidos via versões moduladas, transladadas e escalonadas de uma função $\bar{\phi}(\cdot)$, de modo que:

$$\phi^{(m)}(k) = c_m \bar{\phi} \left(\frac{k - \vartheta_1}{\vartheta_2} \right) \cos(2\pi\vartheta_3(k - \vartheta_1) + \vartheta_4), \quad (4)$$

onde ϑ_1 é responsável pelo deslocamento temporal, ϑ_2 pelo escalonamento, ϑ_3 pela frequência de modulação, ϑ_4 pela fase,

c_m é uma constante de normalização escolhida de modo a tornar unitária a norma do átomo $\phi^{(m)}$ e m é o índice do átomo que define os valores de ϑ_i , para $i = 1, \dots, 4$.

IV. APLICAÇÃO DA APRENDIZAGEM DE DICIONÁRIOS A TÉCNICAS SEMI-CEGAS DE SEPARAÇÃO

Embora constituam uma escolha popular, dicionários paramétricos são oriundos de formulações analíticas artificiais que frequentemente não garantem uma boa modelagem (ou uma decomposição eficiente) dos sinais sob análise. Ainda que tal ineficiência intrínseca de modelagem possa em parte ser contornada por meio do aumento da cardinalidade do dicionário, o aumento do custo computacional e uma possível perda de qualidade nas estimativas dos parâmetros do sistema de mistura são possíveis consequências do emprego de dicionários paramétricos.

Conquanto tradicionalmente as pesquisas em representações esparsas tenham se concentrado na decomposição de sinais com relação a um dado dicionário, recentemente o projeto de dicionários adaptados aos sinais (área também conhecida como *aprendizagem de dicionários*) tem sido uma preocupação relevante na literatura [9],[10]. O emprego de dicionários adaptados é motivado pelo fato de que uma compreensão mais adequada das causas subjacentes ao fenômeno físico capturado pelos sensores, ou dos subespaços onde os dados se encontram, é crucial para a extração da informação relevante que está contida em dados multidimensionais [5].

Se no problema de separação de fontes temos previamente acesso a trechos isolados das fontes que não constam na mistura a separar, podemos empregar o conhecimento oriundo desses trechos para modelar cada fonte, de modo a promover uma melhora do processo de separação. Este procedimento, por tornar semi-cego o problema de separação, é menos geral do que as técnicas cegas. Todavia, apresenta maior potencial de separação, o que justifica o contínuo interesse por ele [11],[12].

Uma modelagem possível consiste em adaptar um dicionário para cada fonte. Embora a adaptação de dicionários em geral seja feita no domínio do tempo, não há dificuldade técnica em estendê-la para o domínio da frequência. A principal vantagem do uso de dicionários no domínio da frequência consiste na possibilidade de se trabalhar com representações redundantes dos blocos de sinais a decompor, ao se utilizar, por exemplo, diversos quadros adjacentes e com sobreposição, no domínio da STFT (do inglês *Short-Time Fourier Transform*).

O emprego de dicionários adaptados poderia obter um desempenho superior na etapa de identificação do sistema de mistura em contextos de misturas instantâneas. Além disso, esse desempenho superior pode se traduzir num menor número de iterações, reduzindo o custo computacional do processo de identificação.

A. Algoritmo K -SVD

Seja $\mathbf{D}_{L \times \#D}$ a matriz que contém, em suas colunas, os átomos (de comprimento L) do dicionário \mathcal{D} . A partir de um conjunto de Z exemplos de sinais de comprimento L $\{\bar{\mathbf{x}}_i\}_{i=1}^Z$, cabe ao algoritmo de aprendizado de dicionário

buscar uma matriz \mathbf{D} de modo a maximizar a esparsidade da representação. Em termos de otimização, tal tarefa pode ser expressa, idealmente, como [13]:

$$\begin{aligned} \arg \min_{\mathbf{D}, \mathbf{A}} \mathcal{J}(\mathbf{X}, \mathbf{D}, \mathbf{A}) &= \|\mathbf{X} - \mathbf{DA}\|_F^2, \\ \text{sujeito a } \|\mathbf{a}_i\|_0 &\leq k_0, \forall 1 \leq i \leq Z, \end{aligned} \quad (5)$$

onde \mathbf{a}_i é a i -ésima coluna de \mathbf{A} (matriz responsável pelo escalamento dos átomos), $\|\cdot\|_F$ é a norma de Frobenius associada a uma métrica de distância entre \mathbf{X} e \mathbf{DA} e $k_0 \ll L < \#D$.

Uma das técnicas associada ao estado da arte em aprendizado de dicionários é o K -SVD [14], cujo nome advém da junção da estratégia do algoritmo K -means com a fatoração por decomposição em valores singulares (SVD, do inglês *Singular Value Decomposition*). Esta técnica é utilizada neste artigo para adaptação de dicionários, sendo portanto resumidamente descrita a seguir.

A primeira etapa da técnica K -SVD é o estágio de codificação. Sendo esta etapa empreendida a partir de um dicionário fixo, o problema de otimização (5) torna-se:

$$\begin{aligned} \arg \min_{\mathbf{A}} \mathcal{J}(\mathbf{X}, \mathbf{D}, \mathbf{A}) &= \sum_{i=1}^Z \|\bar{\mathbf{x}}_i - \mathbf{D}\mathbf{a}_i\|_2^2, \\ \text{sujeito a } \|\mathbf{a}_i\|_0 &\leq k_0, \forall 1 \leq i \leq Z. \end{aligned} \quad (6)$$

Como $k_0 \ll L$, tal problema de otimização admite apenas matrizes \mathbf{A} que apresentem colunas esparsas. Tal restrição pode ser garantida caso recorramos a técnicas MP, que automaticamente associam a cada bloco de dados $\bar{\mathbf{x}}_i$ um conjunto de átomos.

A seguir, atualiza-se o dicionário. Se na etapa de codificação esparsa cada vetor $\bar{\mathbf{x}}_i$ é aproximado por uma combinação linear de k_0 átomos, tem-se:

$$\bar{\mathbf{x}}_i = \sum_{j=1}^{k_0} a_{\mathcal{I}^0(i,j),i} \phi^{\mathcal{I}^0(i,j)} + \bar{e}_i, \quad (7)$$

onde \bar{e}_i é o erro residual, $a_{i,j}$ é o elemento correspondente da matriz \mathbf{A} e $\mathcal{I}^0(i,j)$ é o índice do j -ésimo átomo usado para descrever $\bar{\mathbf{x}}_i$. Para atualizar o k -ésimo átomo $\phi^{(k)}$, pode-se minimizar a norma quadrática

$$\min_{\phi^{(k)}, \bar{\mathbf{a}}_k} \sum_{i=1}^Z \left\| \bar{\mathbf{x}}_i - \sum_{j=1, \mathcal{I}^0(i,j) \neq k}^{k_0} a_{\mathcal{I}^0(i,j),i} \phi^{\mathcal{I}^0(i,j)} - a_{k,i} \phi^{(k)} \right\|_2^2, \quad (8)$$

onde $\bar{\mathbf{a}}_k$ é a k -ésima linha de \mathbf{A} . O problema de otimização (8) é facilmente tratável, porém não é adequado para o aprendizado de dicionários, porque preenche $\bar{\mathbf{a}}_k$ com elementos não-nulos. Isto equivale a impor que cada átomo $\phi^{(k)}$ modele todos os blocos de dados. Tal imposição não é coerente com representações esparsas, pois nessas apenas alguns átomos do dicionário descrevem cada vetor $\bar{\mathbf{x}}_i$. Seja $\mathcal{I}^1(i)$ o conjunto dos índices dos átomos utilizados na etapa pregressa de decomposição MP para descrever o i -ésimo bloco $\bar{\mathbf{x}}_i$. A técnica K -SVD restringe a otimização do k -ésimo átomo aos blocos $\bar{\mathbf{x}}_i$ nos quais $k \in \mathcal{I}^1(i)$. Na etapa de adaptação de dicionário, o K -SVD implementa portanto a otimização:

$$\min_{\phi^{(k)}, \bar{\mathbf{a}}_k} \sum_{i=1, k \in \mathcal{I}^1(i)}^Z \left\| \bar{\mathbf{x}}_i - \sum_{j=1, \mathcal{I}^0(i,j) \neq k}^{k_0} a_{\mathcal{I}^0(i,j),i} \phi^{\mathcal{I}^0(i,j)} - a_{k,i} \phi^{(k)} \right\|_2^2, \quad (9)$$

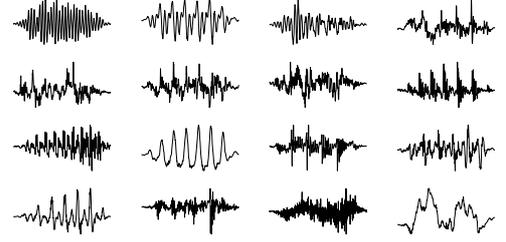


Fig. 1. Exemplos de átomos extraídos do dicionário adaptado para o idioma us, com $Q = 3$ e $N = 8$.

a qual pode ser eficientemente resolvida via técnica SVD [14]. Na etapa de atualização do dicionário, o algoritmo K -SVD apresenta a particularidade de também atualizar os coeficientes não-nulos de \mathbf{A} (o que não ocorre com outras técnicas de aprendizagem de dicionários, como o MOD¹[15]).

V. RESULTADOS

Nesta seção, utilizaremos sinais de voz extraídos da base de dados *The Open Speech Repository*², amostrados a 8 kHz. O emprego desta base de dados específica se deve ao fato de que contempla diferentes idiomas, cada qual com poucos locutores. Cada locutor apresenta diversas locuções as quais, caso reunidas, fornecem informação suficiente para a adaptação de dicionários. Nos ateremos aos idiomas fr e uk, que correspondem a diversas falas em francês e inglês britânico, respectivamente. Cabe ressaltar que os trechos de fala de cada idioma são constituídos majoritariamente por um único locutor (o mesmo utilizado nos testes), embora possa haver sinais oriundos de locutores diferentes. A existência de diferentes locutores torna o problema mais difícil, pois deve polarizar a adaptação rumo ao aprendizado de um determinado idioma (independente de locutor), o que é uma tarefa mais difícil do que apreender as particularidades de um locutor específico. Nas simulações desta seção, adapta-se um dicionário distinto para cada idioma.

Para explorar a redundância dos sinais de voz, optamos por adaptar dicionários no domínio da STFT, com janela de Hanning, sobreposição de 50% entre quadros adjacentes, de comprimento $F = 256$. Definindo Q como o número de quadros adjacentes utilizados para formar um átomo, o número de amostras de cada átomo é $L = (\frac{Q+1}{2})F$. A cardinalidade dos dicionários é linearmente dependente do fator de redundância \mathcal{N} , segundo a fórmula $\#D = \mathcal{N}L$.

Quando $\mathcal{N} = 1$, o dicionário não é redundante, apresentando cardinalidade equivalente ao comprimento dos átomos. Em nossos testes, a adaptação de dicionários foi efetuada para combinações $Q \times \mathcal{N}$, com $Q \in \{1, 2, 3\}$ e $\mathcal{N} \in \{1, 2, 4, 8\}$. As Figs. 1 e 2 apresentam alguns exemplos de átomos extraídos dos dicionários adaptados para os idiomas us e fr.

Nesta seção, comparamos a qualidade das estimativas dos coeficientes de atenuação do sistema de mistura em contextos

¹Do inglês *Method of Optimal Directions*.

²Acessível em http://www.voiptroubleshooter.com/open_speech.

instantâneos (sem atrasos), utilizando a técnica MP. Havendo um dicionário adaptado para cada fonte, deve-se eleger um critério para definir qual dicionário fornecerá o átomo utilizado em uma iteração MP de modo a aproximar o resíduo de um dado bloco \bar{x}_i . Optamos por adotar uma estratégia competitiva, na qual os dicionários competem para apresentar o átomo mais correlacionado com o bloco de sinal. A partir da seleção do átomo e do dicionário, é possível gerar uma estimativa pontual $\gamma_h(n)$ para algum coeficiente h_{ji} .

Espera-se que o MSE das estimativas advindas com o emprego do dicionário adaptado seja inferior ao de outros dicionários. Para efetuarmos tal comparação, escolhemos empregar como dicionários alternativos o paramétrico e um dicionário cuja formação dos átomos se dá por meio da extração aleatória de trechos correspondentes da base de dados de treinamento. A variação de Q (e, por conseguinte, de L) é fácil de se impor a estes dicionários alternativos. Porém, por razões de justiça na comparação, cumpre igualar a cardinalidade dos diferentes dicionários, o que exige um critério de seleção de átomos do dicionário paramétrico, já que a cardinalidade deste dicionário (para um dado valor de L) pode ser muito maior do que a cardinalidade do dicionário aprendido.

A formação de um novo dicionário paramétrico de menor cardinalidade pode se dar a partir da extração aleatória de átomos do dicionário paramétrico padrão. Porém, esse procedimento pode prejudicar o desempenho, porque os átomos selecionados podem não modelar razoavelmente os sinais. Tal problema pode ser contornado mediante a seleção de átomos que descrevam melhor os sinais envolvidos. A análise do impacto dos parâmetros dos átomos na confiabilidade dos mesmos (empreendida em [16]) concorre para embasar o procedimento seletivo, a partir da constatação de que átomos que apresentam escalas altas e frequências baixas são menos confiáveis. A partir dessa constatação, adotamos o procedimento de selecionar apenas átomos que apresentem escala $\log_2(\vartheta_2) < 4$ e frequência $\vartheta_3 \geq 0,3$. Se após este procedimento a cardinalidade do dicionário paramétrico ainda superar a cardinalidade do dicionário adaptado, optamos por efetuar uma retirada aleatória de átomos do dicionário paramétrico, até que se estabeleça a equivalência do número de átomos entre diferentes dicionários.

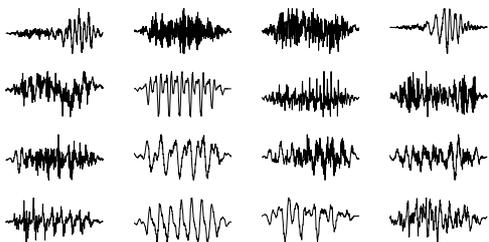


Fig. 2. Exemplos de átomos extraídos do dicionário adaptado para o idioma fr, com $Q = 3$ e $N = 8$.

A partir da matriz de mistura

$$\mathbf{H} = \begin{bmatrix} 1 & 1 \\ -0.5 & 0.7 \end{bmatrix}, \quad (10)$$

gerou-se uma mistura a partir dos sinais de teste oriundos dos idiomas fr e uk. Todas as locuções de um mesmo idioma foram concatenadas em um mesmo arquivo, com os primeiros 20 s do mesmo reservados para os testes de separação e os 500 s (um pouco mais de 8 minutos) restantes destinados ao aprendizado de dicionário.

A partir da mistura, diversos blocos de comprimento idêntico à dimensionalidade dos dicionários foram aleatoriamente selecionados. Em cada bloco, aplicou-se apenas uma iteração MP, capaz de gerar uma estimativa pontual de um dos coeficientes de atenuação da segunda linha de \mathbf{H} . As possibilidades testadas para o número de blocos pertencem ao conjunto $\{100, 200, 300, 400, 500\}$, enquanto que o número de quadros adjacentes Q variou no intervalo $\{1, 2, 3\}$. Diferentes fatores de redundância N , que controlam a cardinalidade dos dicionários, foram testados, dentro do conjunto $\{1, 2, 4, 8\}$. Para cada combinação (número de blocos, Q, N) foram efetuados 200 procedimentos de estimação. As diferenças entre um determinado procedimento de estimação e os demais são: blocos utilizados para estimativa (aleatoriamente selecionados), dicionário paramétrico utilizado (pois os átomos aleatoriamente retirados do mesmo para impor a igualdade no número de átomos entre diferentes dicionários variou em cada procedimento de estimação) e dicionário cujos átomos consistem em trechos dos sinais de treinamento empregado (aleatoriamente escolhidos). Para cada procedimento de estimação, os parâmetros extraídos da aplicação do MP aos diversos blocos foram utilizados como entrada para a varredura de Zibulevsky [8], com parâmetro $\lambda = 1$ e 10000 ângulos de teste θ_t , uniformemente espaçados no intervalo $[0, \pi]$. Para cada procedimento de estimação, calculou-se a norma quadrática dos erros entre os parâmetros reais e os estimados. Em alguns poucos casos (menos de 0,5% das ocorrências) o procedimento de estimação falhou na seleção de picos, gerando erros muito grandes. Para evitar que essas falhas comprometam a interpretação dos resultados médios, os resultados que apresentaram norma quadrática dos erros superior a 0,1 foram descartados. O MSE foi calculado a partir da média da norma quadrática dos erros ao longo dos 200 procedimentos de estimação, a menos dos resultados descartados.

As Figs. 3-5 apresentam os MSEs para as configurações testadas, os quais evidenciam a melhora do MSE obtida por dicionários aprendidos. É possível, por vezes, angariar um desempenho melhor com dicionários adaptados com 100 blocos do que com 500 blocos do dicionário paramétrico.

Utilizaremos a seguir o MSE do processo de identificação dos parâmetros do sistema de mistura como medida de qualidade das estimativas oriundas de diferentes dicionários. Sejam $\mathcal{D}_{\text{learned}}$, $\mathcal{D}_{\text{param}}$ e $\mathcal{D}_{\text{samples}}$ o dicionário aprendido, o paramétrico e o obtido por trechos dos sinais de treinamento. Nos testes efetuados, $\mathcal{D}_{\text{samples}}$ apresenta desempenho superior ao dicionário paramétrico, o que não é surpreendente, porque este deriva de uma formulação *a priori* artificial. Tal melhora é de, em média, 10,6 dB. $\mathcal{D}_{\text{learned}}$ supera $\mathcal{D}_{\text{param}}$ em 15,4 dB, na

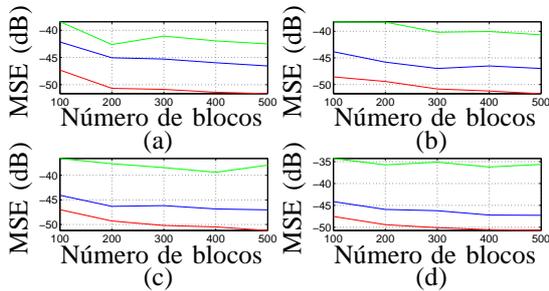


Fig. 3. MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para $Q = 1$. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$.

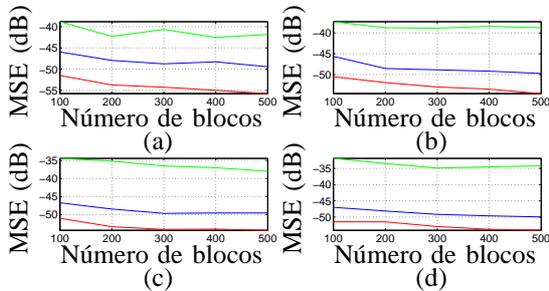


Fig. 4. MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para $Q = 2$. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$.

média. Já o dicionário $\mathcal{D}_{\text{samples}}$ é superado, na média, em 4,7 dB, por $\mathcal{D}_{\text{learned}}$. Isso significa que, em contextos onde há uma base de dados para treinamento das fontes disponíveis e não se dispõe de muitos recursos computacionais para adaptar um dicionário, é mais interessante formar um dicionário a partir de trechos da base de dados do que empregar um dicionário paramétrico. Havendo maior disponibilidade de recursos, a adaptação de dicionários é uma opção mais adequada.

Observa-se que, nos dicionários $\mathcal{D}_{\text{learned}}$ e $\mathcal{D}_{\text{samples}}$, o desempenho aumenta com o número de quadros modelados por cada átomo.

VI. CONCLUSÕES

Este artigo estuda o efeito que o conhecimento *a priori* acerca das fontes pode, via adaptação de dicionários, apresentar na estimativa (baseada em MP) dos coeficientes de um sistema de mistura instantâneo. Observa-se que a adaptação gera resultados superiores até mesmo com relação a um dicionário que extrai aleatoriamente trechos dos sinais de treinamento, o qual também incorpora conhecimento *a priori*. Caso comparado a um dicionário paramétrico, é possível se obter ganhos na identificação superiores a 15 dB. Descartado o custo computacional da adaptação, é possível verificar que, em nossas simulações, a estimativa advinda do emprego de dicionários aprendidos é menos pesada computacionalmente, além de mais acurada.

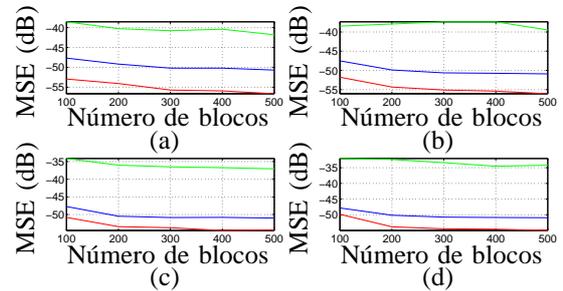


Fig. 5. MSE das estimativas dos coeficientes do sistema de mistura geradas a partir do dicionário aprendido (em vermelho), do dicionário paramétrico (em verde) e do dicionário formado por trechos aleatoriamente extraídos dos sinais de treinamento (em azul), para $Q = 3$. (a) $\mathcal{N} = 1$; (b) $\mathcal{N} = 2$; (c) $\mathcal{N} = 4$ e (d) $\mathcal{N} = 8$.

REFERÊNCIAS

- [1] R. Mazur and A. Mertins, "On separation performance enhancement in convolutive blind source separation," *Proc. of the 32nd Asilomar Conference on Signals, Systems and Computers (ACSSC'08)*, pp. 1718–1721, Oct. 2008.
- [2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley & Sons, 1 ed., 2001.
- [3] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined sparse source separation of convolutive mixtures with observation vector clustering," *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS 2006)*, pp. 3594–3597, May 2006.
- [4] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.
- [5] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, pp. 27–38, Mar. 2011.
- [6] P. Jost, P. Vandergheynst, and P. Frossard, "Tree-based pursuit: Algorithm and properties," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4685–4697, Dec. 2006.
- [7] C. D. Vleeschouwer and B. Macq, "Subband dictionaries for low-cost matching pursuit of video residues," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 984–993, Oct. 1999.
- [8] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, pp. 2353–2362, Nov. 2001.
- [9] M. G. Jafari and M. D. Plumbley, "Fast dictionary learning for sparse representations of speech signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, pp. 1025–1031, Sept. 2011.
- [10] T. P. R. Rubinstein and M. Elad, "Analysis k-svd: A dictionary-learning algorithm for the analysis sparse model," *IEEE Transactions on Signal Processing*, vol. 61, pp. 661–677, Feb. 2013.
- [11] L. Benaroya, F. Bimbot, and R. Gribonval, "Audio source separation with a single sensor," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 191–199, Jan. 2006.
- [12] K. Itoyama, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, "Simultaneous processing of sound source separation and musical instrument identification using bayesian spectral modeling," *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2011)*, pp. 3816–3819, May 2011.
- [13] L. N. Smith and M. Elad, "Improving dictionary learning: Multiple dictionary updates and coefficient reuse," *IEEE Signal Processing Letters*, vol. 20, pp. 79–82, Jan. 2013.
- [14] M. E. M. Aharon and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4311–4322, Nov. 2006.
- [15] K. Engan, S. O. Aase, and J. H. Hakon-Husoy, "Method of optimal directions for frame design," *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 1999)*, vol. 5, pp. 2443–2446, Mar. 1999.
- [16] D. B. Haddad, M. R. Petraglia, L. Lovisollo, P. B. Batalheiro, and J. C. P. Filho, "Emprego de técnicas de matching pursuit para a identificação de um sistema de mistura anecoico," *XI Congresso de Áudio da AES Brasil*, pp. 55–62, Maio 2013.