

AVALIAÇÃO DO ALGORITMO WB-SLIM PARA A GERAÇÃO DE IMAGENS ACÚSTICAS

Farias, Johannes¹; Masiero, Bruno²

(1) Universidade estadual de Campinas, Cidade Universitária Zeferino Vaz - Barão Geraldo, Campinas - SP, johanneswarwick@gmail.com.

(2) Universidade estadual de Campinas, Cidade Universitária Zeferino Vaz - Barão Geraldo, Campinas - SP, masiero@unicamp.br.

RESUMO

Recentemente, técnicas de recuperação de sinais esparsos têm despertado interesse por sua aplicabilidade na área de localização de fontes acústicas. Uma dentre estas técnicas é a *Sparse Learning via Iterative Minimizațion* (SLIM), uma técnica de aprendizado esparso através de minimização iterativa. Entretanto, grande parte das técnicas propostas partem do pressuposto de que os sinais emitidos pelas fontes são de banda estreita, o que na prática não condiz com aplicações usuais em acústica. Surge então sua extensão para sinais de banda larga, a *Wideband SLIM* (WB-SLIM), capaz de lidar com fontes de banda larga e que promove a esparsidade espacial dos resultados. Este artigo busca avaliar o desempenho do WB-SLIM em diferentes cenários de imageamento acústico. Foram realizadas análises em duas regiões de varredura 2D, uma paralela e outra perpendicular ao arranjo de microfones. Verificou-se que a qualidade das imagens acústicas obtidas nos diferentes cenários propostos, avaliada através da taxa de erro de localização e pela comparação do sinal recuperado no tempo com o sinal original, diminui com o aumento do número de fontes geradoras e a com a redução da relação sinal-ruído presente nos sensores.

Palavras-chave: imageamento acústico, processamento de sinais, localização de fontes banda larga, recuperação de sinais esparsos.

ABSTRACT

In recent years, many techniques for sparse signal recovery have aroused interest due to their applicability in the area of acoustic source localization. Among them is the Sparse Learning via Iterative Minimization (SLIM) method. However, most of these techniques assume that the signals emitted by the sources have narrow band, which in practice does not fit usual acoustical applications. The wideband extension of SLIM (WB-SLIM) is able to handle broadband sources and promotes the spatial sparseness of the results. This paper aims to evaluate the performance of WB-SLIM in different acoustic imaging scenarios. We analyzed the performance of WB-SLIM in a 2D scanning region, both parallel and perpendicular to the microphone array. The results illustrate how the quality of the acoustic images in both proposed scenarios, evaluated as the source localization error rate and through the comparison between recovered and original time signals, decreases as the number of sources composing the sound field increases and the signal-to-noise ratio decreases.

Keywords: acoustic imaging, signal processing, wideband source localization, sparse signal recovery.

1. INTRODUÇÃO

Imageamento acústico é uma técnica que mapeia os níveis sonoros em uma dada região do espaço. Neste contexto, um arranjo de microfones espacialmente distribuídos é utilizado para captar o sinal das fontes sonoras e, em seguida, filtros espaciais (capazes de apontar o arranjo em diferentes posições de interesse) são aplicados para estimar os níveis sonoros em diferentes direções. Uma imagem acústica é resultado do mapeamento dos níveis acústicos em uma região 2D ou 3D [1].

Imagens acústicas podem ser sobrepostas a fotografias para auxiliar na identificação de fontes ativas e no controle e análise do ruído [2]. Desta forma, pode-se projetar meios de transporte [3, 4], máquinas industriais ou estruturas como salas de concerto com maior eficiência acústica, por exemplo.

Os primeiros trabalhos na área abordaram o uso do *beamforming* para a caracterização das fontes [5, 6]. Esta técnica baseia-se na informação da variação de fase em cada microfone. Apesar da versatilidade e rápido processamento, a informação resultante apresenta questões indesejadas como a má resolução em baixas frequências e presença de lóbulos laterais em altas frequências [2]. Estas limitações impulsionaram o desenvolvimento das chamadas técnicas de desconvolução, capazes de remover o efeito da *point spread function* [7], englobando técnicas como o DAMAS [8] e CLEAN [9].

Os algoritmos mencionados anteriormente são amplamente utilizados e conhecidos na literatura, entretanto foram desenvolvidos assumindo que os sinais emitidos pelas fontes são de banda estreita, o que não condiz com a maioria das aplicações em acústica. Com isso, foi impulsionado o desenvolvimento de técnicas como o *wideband sparse learning via iterative minimization* (WB-SLIM) [10]. Resumidamente, seu funcionamento consiste em assumir que as fontes irradiam som de banda larga e quebrar o problema em diversos componentes de banda estreita, aproveitando a informação conjunta para localizar espacialmente as fontes, minimizando as demais regiões de forma iterativa.

De acordo com Xu *et al.* em [11], o WB-SLIM é capaz de localizar as fontes mesmo com poucos componentes em frequência, arranjos com geometria arbitrária e fontes coerentes ou não-coerentes, gerando resultados com alta resolução espacial.

Este artigo busca analisar o desempenho da técnica WB-SLIM em diferentes cenários. O objetivo é avaliar a capacidade do algoritmo em localizar corretamente diferentes números de fontes sonoras expostas a diferentes valores de relação sinal-ruído.

2. MODELO DO SINAL

Considerando um número *K* de fontes posicionadas em campo distante emitindo sinais de banda larga, que serão captados por um arranjo de *M* microfones, localizados em diversas posições $\theta_m = (x, y, z)$. Os sinais dos microfones são amostrados no tempo e usa-se um bloco de *L* amostras para, através da transformada rápida de Fourier, extrair *L* componentes de frequência. Neste caso, o sinal dos microfones na presença do ruído é expresso por

$$y_l = \sum_{k=1}^{K} a_l(\theta_k) x_{k,l} + n_l$$
, para $l = 1,...,L$, (1)

onde $y_l \in \mathbb{C}^{Mx1}$ representa o nível de pressão complexa adquirido em cada microfone para a faixa de frequência $l, a_l(\theta_k) \in \mathbb{C}^{Mx1}$ é o vetor de transferência que "aponta" o arranjo na direção de localização das fontes $\theta_k = (x, y, z), x_{k,l} \in \mathbb{C}$ representa a amplitude complexa do sinal irradiado pela fonte k na banda de frequência $l \in n_l \in \mathbb{C}^{Mx1}$ representa o ruído aditivo.

Considerando que a propagação é realizada por uma frente de onda plana, $a_l(\theta_k)$ pode ser calculado como

$$a_l(\theta_k) = e^{(-j\omega/c \|p_m - p_k\|)},\tag{2}$$

onde $||p_m - p_k||$ representa a distância euclidiana entre a posição da fonte *k* e o microfone *m*. ω é a frequência angular da onda e *c* a velocidade do som.

3. DELAY-AND-SUM BEAMFORMING

O *Delay-and-Sum beamforming* é a técnica clássica para identificar a localização de fontes. A metodologia desta técnica, como já diz o nome, é atrasar os sinais (ou deslocar em fase) e somar todos os sinais do arranjo em relação a um ponto observado no espaço. À medida em que a suposta localização coincide com o real posicionamento da fonte, o resultado é máximo. O sinal das fontes é estimado com *beamforming* por meio da expressão

$$x_{n,l} = \frac{a_l^H(\theta_n)y_l}{\|a_l(\theta_n)\|^2}, \text{ para } n = 1, \dots, N; \ l = 1, \dots, L,$$
(3)

em que $x_{n,l}$ representa o nível de pressão complexa em determinada posição de interesse e $a_l(\theta_n)$ é o vetor que aponta na direção θ_n . A variável $a_l(\theta_n)$ é definida como

$$a_l(\theta_n) = e^{(-j\omega/c\|p_m - p_n\|)},\tag{4}$$

onde $||p_m - p_n||$ representa a distância euclidiana entre a fonte e a posição de interesse. Interessante notar que cada ponto observado expressa uma suposição sobre o real posicionamento das fontes, e geralmente $N \gg K$ [11].

4. WIDEBAND SPARSE LEARNING VIA ITERATIVE MINIMIZATION (WB-SLIM)

A técnica SLIM, inicialmente proposta por Tan *et al.* em [10], é um método para solucionar um problema de minimização utilizando norma ℓ_q , onde $0 < q \le 1$, visando a recuperação de sinais esparsos, bem como a estimativa da energia do ruído. Já o modelo considerando sinais de banda larga, desenvolvido por Xu *et al.* em [11], extendeu o problema da área de recuperação de sinais esparsos para:

$$y_l = A_l x_l + n_l$$
, para $l = 1,...,L$, (5)

onde $A_l \in \mathbb{C}^{M_{xN}}$ representa a matriz que contém todos os pontos de interesse observados no espaço, $x_l \in \mathbb{C}^{N_{x1}}$ representa o vetor do sinal, $y_l \in \mathbb{C}^{M_{x1}}$ é o vetor de medidas resultante e $n_l \in \mathbb{C}^{M_{x1}}$ é o vetor de ruído. O índice *l* expressa a relação com a frequência.

Xu *et al.* propuseram em [11] duas variantes para o SLIM abordando sinais de banda larga: WB-SLIM-0 que busca minimizar a norma ℓ_0 e WB-SLIM-1 que busca minimizar a norma ℓ_1 . A principal diferença entre os dois métodos está na relação entre esparsidade e robustez do resultado. O primeiro algoritmo produz resultados mais esparsos, entretanto perde eficiência em situações de baixa relação sinal-ruído. Este artigo focará na avaliação do WB-SLIM-0, que pode ser definido como

$$\min_{\{x_l\}\{p_n\},\eta} g_{WB-SLIM-0}(\{x_l\}\{p_n\},\eta),$$
(6)

onde

$$g_{WB-SLIM-0} = LM \log \eta + \frac{1}{\eta} \sum_{l=1}^{L} \|y_l - A_l x_l\|^2 + L \sum_{n=1}^{N} \log p_n + \sum_{l=1}^{L} \sum_{n=1}^{N} \frac{|x_{n,l}|^2}{p_n}.$$
 (7)

Neste caso, o algoritmo obtém estimativas de x_l , $p_n \in \eta$ através da solução de um problema de estimação máximo a posteriori (MAP). Para isto foram utilizados os seguintes modelos bayesianos [11]:

$$f(\{y_l\} | \{x_l\}, \eta) = \prod_{l=1}^{L} \frac{1}{(\pi\eta)^M} e^{-(1/\eta) \|y_l - A_l x_l\|^2},$$
(8)

e

$$f(\{x_l\} | p_n) = \prod_{l=1}^{L} \frac{1}{\pi^N \prod_{n=1}^{N} p_n} e^{-x_l^H P^{-1} x_l}.$$
(9)

Estes modelos assumem que os vetores de ruído n_l são complexos, independentes e identicamente distribuídos (IID) e gaussianos circularmente simétricos, com média zero e matriz de covariância ηI , onde η representa a energia do ruído. Também é assumido que x_l é um vetor complexo, IID e gaussiano circularmente simétrico, com média zero e matriz de covariância diagonal $P \triangleq diag \{p_1, p_2, ..., p_N\}$ [11].

O problema de otimização da Equação (6) pode ser resolvido através da técnica de minimização cíclica [12], onde cada resultado é obtido através da simplificação do problema de otimização, derivando determinado parâmetro e assumindo os restantes como conhecidos. A partir deste ponto é possível obter soluções ótimas para p_n , $x_l \in \eta$. As equações resultantes e o processo de cálculo do algoritmo WB-SLIM-0 está resumido na Tabela 1.

 Tabela 1: Algoritmo para cálculo do WB-SLIM-0.

Inicializar o algoritmo estimando
$$x_{n,l} \in \eta$$
 através do *Delay-and-Sum*:
 $x_{n,l}^{(0)} = a_l^H(\theta_n) y_l / ||a_l(\theta_n)||^2$, para $n = 1,...,N; \ l = 1,...,L;$
 $\eta^{(0)} = max \left\{ \frac{1}{10LM} \sum_{l=1}^{L} ||x_l^{(0)}||^2, \varepsilon \right\}$
Repetir os seguintes cálculos para $t = 0, 1, 2, ...:$
 $p_n^{(t+1)} = max \left\{ \frac{1}{L} \sum_{l=1}^{L} |x_{n,l}^{(t)}|^2, \varepsilon \right\}$, para $n = 1,...,N;$
 $x_l^{(t+1)} = P^{(t+1)} A_l^H (A_l P^{(t+1)} A_l^H + \eta^{(t)} I)^{-1} y_l$, para $l = 1,...,L;$
 $\eta^{(t+1)} = max \left\{ \frac{1}{LM} \sum_{l=1}^{L} ||y_l - A_l x_l^{(t+1)}||^2, \varepsilon \right\}$
até atingir a convergência.

5. Metodologia

Para avaliar o WB-SLIM-0 serão realizadas simulações em uma região de varredura 2D, paralela e perpendicular ao arranjo de microfones, utilizando as seguintes configurações:

5.1 Varredura paralela ao arranjo de microfones

No primeiro cenário as fontes foram posicionadas de forma paralela ao arranjo de microfones, utilizando os eixos (x, y). Foram utilizadas duas fontes descorrelacionadas emitindo ruído gaussiano branco, com nível sonoro de 10 *dB*. As fontes estão localizadas em (0, 132; -1, 585; 14, 000) e (1,959; 1,947; 14,000) m.

A área varrida para o cálculo das imagens acústicas varia de -2,500 a 2,500 m nos eixos (x,y). A altura foi mantida em z = 14 m. Foi utilizado um espaçamento de 0,062 m entre os pontos, totalizando 6561 pontos observados. Estas dimensões foram escolhidas para avaliar se o algoritmo seria capaz de identificar fontes próximas no espaço, já que a área descrita apresenta uma variação de aproximadamente 20° em relação ao centro do arranjo. A Figura 1 ilustra a configuração geral do sistema.



Figura 1: Configuração do sistema para varredura paralela ao arranjo, onde os pontos pretos ilustram as fontes, os pontos vermelhos ilustram as direções de visada e os pontos azuis ilustram os microfones.

O arranjo possui área de 1 m^2 , sendo composto por 64 microfones que estão localizados nos eixos (x, y), variando entre -0,500 e 0,500 m em ambas as coordenadas, e mantendo z = 0 m. Os microfones foram posicionados com geometria aleatória e espaçamento aleatório. Esta abordagem busca gerar padrões de arranjo cuja a matriz de transformação A_l atenda à propriedade de isometria restrita requerida em algoritmos promotores de esparsidade [13].

Durante a aquisição em cada microfone os sinais foram corrompidos por ruído branco aditivo. Neste caso foi usada uma relação sinal ruído de 17 dB. Finalmente, é importante ressaltar que tanto as fontes como o ruído foram pré-processados através da transformada de Fourier, decompondo as informações em diversos componentes de frequência. Na simulação foram utilizados 63 componentes, e os espectros variam entre 100 Hz e 5 kHz.

5.2 Varredura perpendicular ao arranjo de microfones

O segundo caso aborda uma configuração perpendicular com relação ao arranjo de microfones, utilizando os eixos (x,z). Novamente foram posicionadas duas fontes descorrelacionadas emitindo ruído branco, com nível sonoro de 10dB. As fontes estão localizadas em (0,818; -13,274) e (-0,121; 13,362) m, com y = 0 m.

A área coberta no cálculo das imagens acústicas varia de -2,500 a 2,500 m no eixo x, e de 11,500 a 16,500 m no eixo z, com y = 0 m. Foi utilizado um espaçamento de 0,062 m, totalizando 6561 pontos observados. A Figura 2 ilustra a configuração geral do sistema.

O arranjo possui a mesma configuração do problema anterior, mantendo a geometria aleatória e o espaçamento aleatório. Também foi adicionado ruído ao problema, porém foi usada uma relação sinal ruído de 80 dB para que fosse obtido algum resultado satisfatório. Manteve-se o uso das informações no domínio da frequência, com 63 componentes utilizados, e espectros variando entre 100 e 5 kHz.



Figura 2: Configuração do sistema para varredura perpendicular ao arranjo, onde os pontos pretos ilustram as fontes, os pontos vermelhos ilustram as direções de visada e os pontos azuis ilustram os microfones.

6. RESULTADOS

Serão apresentados os resultados do WB-SLIM-0 resultante de simulações em diferentes cenários, através de um sistema contendo fontes pontuais e um arranjo de microfones situados em campo livre. As simulações foram realizadas utilizando o software *Matlab*.

6.1 Varredura paralela ao arranjo de microfones

A imagem acústica após o cálculo do *Delay-and-Sum beamforming* é apresentada na Figura 3. É possível constatar que para baixas frequências torna-se inviável a localização das fontes devido à baixa resolução da imagem. A medida em que a frequência aumenta verifica-se a formação dos lóbulos principais que correspondem à região das fontes, entretanto a localização continua imprecisa e a imagem sofre com os efeitos dos lóbulos laterais e artefatos.

Após o cálculo do WB-SLIM-0, cujos resultados são mostrados na Figura 4, verificou-se que



Figura 3: Imagem acústica através do *Delay-and-Sum beamforming* para direções de visada paralelas ao arranjo, em diferentes frequências. Os círculos azuis indicam o local onde as fontes estão posicionadas. O padrão de cores está em dB.

o algoritmo conseguiu localizar as fontes dentro das configurações propostas, melhorando de forma significativa a resolução das imagens para todas as frequências. Também verificou-se o alto poder de convergência, pois o resultado foi obtido com menos de 15 iterações, conforme mencionado por Tan *et al.* em [10].



Figura 4: Imagem acústica através do WB-SLIM-0 para direções de visada paralelas ao arranjo, em diferentes frequências. Os círculos azuis indicam o local onde as fontes estão posicionadas. O padrão de cores está em dB.

Outra informação relevante é o tempo de processamento para a geração das imagens acústicas. Para as configurações descritas na Seção 5.1, o *Delay-and-Sum* gerou imagens levando em média 2 s de processamento, enquanto o WB-SLIM-0 levou 8 s de processamento. As simulações foram realizadas em um computador com processador quad-core (clock de 3,40 GHz) e 8 GBytes de memóra RAM.

Conforme mencionado na Seção 3, a variável $x_{n,l}$ representa a pressão complexa estimada para dada localização e frequência, e de posse de um conjunto de medições é possível criar o

chamado pseudo-espectro para um ponto θ_n no espaço, obtendo o vetor $\{x_n\}_{l=1}^L$. Deste modo, através da transformada inversa de Fourier, é possível avaliar os sinais no tempo após o cálculo do WB-SLIM-0 e compará-los com os sinais inicialmente emitidos pelas fontes. Verifica-se, analisando a Figura 5, que neste caso as amostras recuperadas são idênticas ao sinais emitidos, indicando que o algoritmo é capaz de lidar com as distorções geradas pelo ruído.



Figura 5: Comparação entre as amostras no tempo de dois sinais descorrelacionados emitidos por duas fontes em posições distintas no espaço, ilustrados em azul, e das estimativas, ilustrados em vermelho, após o cálculo do WB-SLIM-0.

A Figura 6 faz um apanhado geral sobre as capacidades do WB-SLIM-0. Neste caso foram executadas diversas simulações, variando o número de fontes e a relação sinal-ruído. Ao final de cada execução o resultado foi analisado para avaliar se o algoritmo fora capaz de detectar as fontes corretamente. Além disso, cada execução gera posicionamentos aleatórios tanto para as fontes pontuais quanto para os microfones, respeitando a área descrita na configuração do cenário. Verifica-se que o algoritmo possui desempenho satisfatório até os 20 dB de relação sinal-ruído. A medida em que o SNR diminui, o algoritmo perde a capacidade de produzir resultados esparsos, apresentando dificuldades em identificar fontes presentes no espaço varrido.

6.2 Varredura perpendicular ao arranjo de microfones

A imagem acústica após o cálculo do *Delay-and-Sum beamforming* é apresentada na Figura 7. Neste caso, a identificação das fontes torna-se ainda mais difícil, pois os lóbulos principais possuem maior extensão. Além disso, para as configurações descritas anteriormente, existe o agravante de que os lóbulos se fundiram devido à proximidade das fontes.

A imagem acústica obtida para o caso de varredura perpendicular com o WB-SLIM-0 está ilustrada na Figura 8. O algoritmo foi capaz de apresentar um resultado esparso, entretanto a localização das fontes não foi precisa. Apesar disto, ainda é possível estimar a direção das fontes corretamente no eixo x.

7. CONCLUSÃO

Este artigo apresentou a técnica WB-SLIM-0 e avaliou o seu desempenho no problema de imageamento acústico envolvendo sinais de banda larga, através de simulações em diferentes



Figura 6: Análise geral do desempenho do WB-SLIM-0. O eixo *x* representa o nível de relação sinal ruído, e o eixo *y* representa o número de fontes. O código de cores expressa a porcentagem de acerto na localização das fontes.



Figura 7: Imagem acústica através do *Delay-and-Sum beamforming* para direções de visada perpendiculares ao arranjo, em diferentes frequências. Os círculos azuis indicam o local onde as fontes estão posicionadas. O padrão de cores está em dB.

configurações espaciais. Verificou-se que o algoritmo é capaz de localizar fontes e produzir imagens esparsas, entretanto exibe certa sensibilidade com relação ao nível de ruído presente no problema. Além disso, verificou-se que o algoritmo apresentou bons resultados nas configurações com varredura paralela ao arranjo, porém possui dificuldades no caso em que a varredura é perpendicular, exibindo imagens acústicas menos precisas.

Novos estudos abordariam o uso do WB-SLIM-1 no problema de imageamento acústico, bem como sugerir adaptações entre os algoritmos com o objetivo de produzir melhores imagens, explorando as relações entre a qualidade da resolução e a tolerância ao nível de ruído.



Figura 8: Imagem acústica através do WB-SLIM-0 para direções de visada perpendiculares ao arranjo, em diferentes frequências. Os círculos azuis indicam o local onde as fontes estão posicionadas. O padrão de cores está em dB.

REFERÊNCIAS

- [1] Padois, T. e Berry, A. Two and three-dimensional sound source localization with beamforming and several deconvolution techniques. *Acta Acustica united with Acustica*, 103(3):392–400, 2017.
- [2] Nascimento, V. H.; Masiero, B. S.; Ribeiro, F. P.; Coelho, R.; Queiroz, R.; Romano, J. e Cavalcante, C. Acoustic imaging using the Kronecker array transform. *Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing*, pages 153–178, 2015.
- [3] Meyer, A. e Döbler, D. Noise source localization within a car interior using 3d-microphone arrays. *Proceedings* of the BeBeC, pages 1–7, 2006.
- [4] Siller, H.; Drescher, M.; Saueressig, G. e Lange, R. Fly-over source localisation on a boeing 747-400. Em *Berlin Beamforming Conference (BeBeC)*, page 3. sn, 2010.
- [5] Billingsley, J. e Kinns, R. The acoustic telescope. Journal of Sound and Vibration, 48(4):485–510, 1976.
- [6] Van Veen, B. D. e Buckley, K. M. Beamforming: A versatile approach to spatial filtering. *IEEE assp magazine*, 5(2):4–24, 1988.
- [7] Johnson, D. H. e Dudgeon, D. E. Array signal processing: concepts and techniques: PTR Prentice Hall Englewood Cliffs, 1993.
- [8] Brooks, T. F. e Humphreys, W. M. A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. *Journal of Sound and Vibration*, 294(4-5):856–879, 2006.
- [9] Högbom, J. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement Series*, 15:417, 1974.
- [10] Tan, X.; Roberts, W.; Li, J. e Stoica, P. Sparse learning via iterative minimization with application to mimo radar imaging. *IEEE Transactions on Signal Processing*, 59(3):1088–1101, 2011.
- [11] Xu, L.; Zhao, K.; Li, J. e Stoica, P. Wideband source localization using sparse learning via iterative minimization. *Signal Processing*, 93(12):3504–3514, 2013.
- [12] Stoica, P. e Selén, Y. Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher. *IEEE Signal Processing Magazine*, 21(1):112–114, 2004.
- [13] Candes, E. J. e Tao, T. Decoding by linear programming. *IEEE transactions on information theory*, 51(12): 4203–4215, 2005.