

# Avaliação de algoritmos promotores de esparsidade para localização de fontes com arranjo esférico de microfones

Caldas, F.<sup>1</sup> ; Masiero, B. S.<sup>1</sup> 

<sup>1</sup> Departamento de Comunicações, Faculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas, SP, masiero@unicamp.br.

## Resumo

Arranjos esféricos de microfones são frequentemente utilizados para a detecção da direção de chegada (DOA) de sons vindos de todas as direções do espaço. No entanto, os algoritmos clássicos de estimação de DOA, como a decomposição em ondas planas e o *beamforming* esférico, apresentam baixa precisão para localizar fontes. Para melhorar a estimação da DOA, o algoritmo *compressive beamforming* (CB) foi proposto. O CB aplica a regularização promotora de esparsidade ao *beamforming* regular através do uso da minimização da norma  $\ell_1$ , assumindo, portanto, que cenas sonoras normalmente são compostas por apenas algumas fontes. Neste artigo, é comparado o desempenho de três algoritmos de regularização promotora de esparsidade em um modelo de decomposição em ondas planas: a minimização da norma  $\ell_1$  via *Disciplined Convex Program* (DCP), o método *Least Absolute Shrinkage and Selection Operator* (LASSO) e o *Orthogonal Matching Pursuit* (OMP). Verifica-se que os três algoritmos foram capazes de determinar acuradamente o número de fontes e suas direções para uma cena simulada. No entanto, o desempenho de todos algoritmos foi menos satisfatório quando aplicados a uma situação real com uma fonte gravada em ambiente quase anecoico. Neste caso, houve uma melhora do desempenho ao se combinar o LASSO para determinar o número de fontes com o OMP para refinar a estimativa da amplitude da onda.

**Palavras-chave:** localização de fonte sonora, regularização promotora de esparsidade, arranjos de microfones com geometria esférica.

**PACS:** 43.60.Fg, 43.60.Pt, 3.60.Vx, 43.66.Qp, 43.60.Jn.

## Evaluation of sparsity-promoting algorithms for sound source localization with a spherical array of microphones

### Abstract

Spherical microphone arrays are often used to detect the direction of arrival (DOA) of sounds coming from all directions in space. However, classical DOA algorithms, such as plane-wave decomposition or spherical beamforming, suffer from low localization accuracy. In an attempt to improve DOA estimation, the compressive beamforming (CB) algorithm is proposed. CB applies sparsity regularization to regular beamforming through the use of  $\ell_1$ -norm minimization, therefore assuming that common sound scenes are usually composed of only a handful of sound sources. This paper compares the performance of three sparsity regularization algorithms on a plane-wave decomposition model: the  $\ell_1$ -norm minimization via Disciplined Convex Program (DCP), the Least Absolute Shrinkage and Selection Operator (LASSO) method, and the Orthogonal Matching Pursuit (OMP). It is shown that the three algorithms were able to accurately estimate the number of sources and their directions for an artificial sound scene. However, the performance of the three algorithms presents less satisfactory results in a practical situation with a recorded sound source in an anechoic chamber with a reflective ceiling. In this case, there was an improvement with the combination of the LASSO to estimate the number of sound sources and the OMP to refine the wave amplitude.

**Keywords:** sound source localization, sparsity regularization, spherical geometry of microphone arrays.

## 1. INTRODUÇÃO

Ao se realizar o processamento do áudio espacial de cenas sonoras, costuma ser necessário primeiro detectar as fontes sonoras presentes na cena, o que é comumente feito com o auxílio de um arranjo de microfones e um algoritmo detector de direção de chegada (DOA). A determinação da quantidade e direção das fontes presentes em um campo sonoro pode ser feita, a princípio, usando-se filtros espaciais para processar os sinais adquiridos por um arranjo de microfones [1]. Caso o sistema deva analisar o som vindo de todas as direções possíveis, então deve-se utilizar um arranjo esférico de microfones. O resultado deste processamento é frequentemente apresentado na forma de imagens acústicas. Esta técnica, no entanto, costuma apresentar baixa resolução espacial devido ao número relativamente reduzido de microfones que constituem os arranjos [2].

Diversos métodos para aumentar a qualidade da estimativa da imagem acústica sem aumentar o número de sensores foram propostos. Uma classe de algoritmos que apresentou bons resultados aplica técnicas de desconvolução para eliminar o efeito da função de espalhamento do arranjo [3–7]. Uma outra alternativa é a regularização promotora de esparsidade [8,9], que usa uma regularização de norma  $\ell_1$  na solução do problema de determinação dos objetos sonoros. Esta segunda classe de algoritmos funciona privilegiando cenas compostas por um pequeno número de fontes sonoras (o que, na prática, costuma ser uma suposição válida), prometendo uma melhor separação dos objetos sonoros presentes na cena. Algoritmos que aplicam a regularização promotora de esparsidade também são conhecidos por *compressive beamforming* (CB).

As técnicas mencionadas previamente são frequentemente usadas com arranjos uni- ou bidimensionais — respectivamente, arranjos de microfones lineares [9] ou planos [10]. No entanto, estas geometrias de arranjo permitem a detecção de fontes em uma região restrita do espaço. Para permitir a busca por fontes em todas as direções, faz-se necessário o uso de arranjos com geometria tridimensional, sendo o

arranjo esférico a geometria mais comumente usada [11]. Apenas recentemente surgiram trabalhos aplicando CB para arranjos esféricos de microfones [12–14].

Os algoritmos promotores de esparsidade têm atraído interesse crescente em razão da sua capacidade de representar sinais com um número reduzido de medições e por conta de sua flexibilidade de uso. No entanto, não existe uma única maneira de se implementar esta restrição. Diversas formas de implementação destes algoritmos foram propostas, como a decomposição da norma atômica [15], reconstrução via FOCUSS (*FOcal Underdetermined System Solver*) [16], algoritmos evolutivos para reconstrução esparsa [17], algoritmos de regressão, algoritmos gulosos (*greedy algorithms*) e métodos de compressão iterativa (*iterative-shrinkage*) [18].

O objetivo deste artigo é comparar o desempenho da regularização promotora de esparsidade com um algoritmo de *beamforming* esférico usando três métodos distintos de promoção de esparsidade. Os métodos avaliados foram escolhidos por terem sido utilizados em artigos recentemente publicados sobre este assunto, sendo eles: *Disciplined Convex Programming* (DCP) [19], o algoritmo de regressão que usa a norma  $l_1$  como penalização *Least Absolute Shrinkage and Selection Operator* (LASSO) [20] e o algoritmo de regressão guloso *Orthogonal Matching Pursuit* (OMP) [13].

O restante deste artigo é estruturado iniciando-se com uma descrição do problema a ser tratado. Em seguida, são apresentados os fundamentos teóricos relacionados à Transformada de Fourier Esférica, que servem como base para o modelo do sinal. Segue-se com uma revisão da teoria de *beamforming* esférico, cujas imagens acústicas resultantes são usadas como base de comparação. Então é apresentado o conceito de *compressive beamforming* com implementação por DCP, LASSO e OMP. Por fim, o desempenho destes algoritmos é avaliado por meio de resultados de simulação e medição. A simulação e o processamento da gravação foram programados em Python [21].

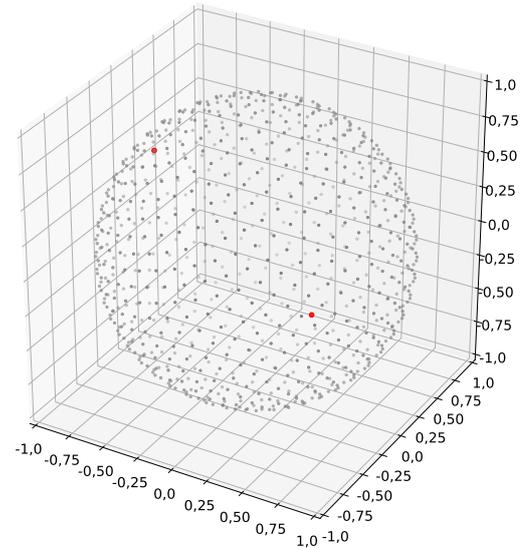
## 2. DESCRIÇÃO DO PROBLEMA

Um campo sonoro é composto pela superposição de ondas geradas por diversas fontes e pode ser modelado como uma somatória de ondas planas vindas de diversas direções. Utiliza-se aqui um arranjo esférico de microfones para amostrar o campo sonoro nas posições dos microfones e, a partir de informações de amplitude e fase dos sinais captados pelos microfones determinar a direção e intensidade das ondas planas que constituem o campo sonoro, podendo-se então aplicar filtros espaciais para produzir uma saída com características desejadas—como, por exemplo, realçar sinais vindos de uma determinada direção e atenuar sinais fora da região de interesse. Um tipo de filtro com tal função é conhecido como *beamforming* esférico e uma das suas aplicações mais comuns é a geração de imagens acústicas, que permitem estimar a direção e a amplitude de fontes em um campo sonoro. No entanto, este método carece de precisão, especialmente quando há fontes muito próximas entre si [22]. Uma alternativa para aumentar a resolução espacial destes algoritmos é usar a suposição de esparsidade da cena sonora.

Se uma cena sonora for composta por poucas fontes distribuídas no espaço, ela é dita esparsa e pode, portanto, ser representada por um número pequeno de medições [23]. Matematicamente, isto significa uma representação vetorial com um grande número de entradas nulas e algumas poucas não-nulas. Um exemplo desta representação encontra-se na Figura 1, na qual os pontos vermelhos representam duas fontes sonoras “ligadas” (entradas não-nulas) entre várias outros pontos cinzas que representam fontes sonoras “desligadas” (entradas nulas). O conjunto de posições dos pontos da Figura 1 é chamado de malha de busca. Ao juntar a estrutura da malha de busca com o sinal captado pelo arranjo de microfones, um algoritmo promotor de esparsidade é capaz de informar onde estão localizadas as fontes e quais são suas amplitudes.

## 3. FUNDAMENTOS TEÓRICOS

Nesta seção, apresentaremos a notação usada no modelo do sinal para a decomposição em



**Figura 1:** Exemplo de uma cena sonora esparsa: Em vermelho, as posições de duas fontes. Em cinza, os pontos de uma malha de busca densa.

ondas planas usando um arranjo esférico de microfones [24].

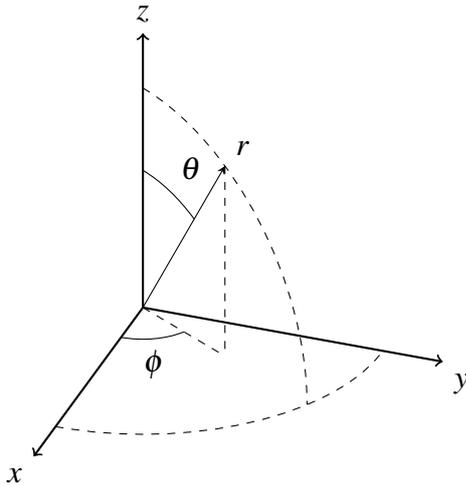
### 3.1 Funções esféricas

A partir do sistema de coordenadas Cartesianas  $\mathbf{x} \equiv (x, y, z) \in \mathbb{R}^3$ , é definido o sistema de coordenadas esféricas  $\mathbf{r} \equiv (r, \theta, \phi)$ , no qual:  $r \in [0, \infty[$  é o raio medido a partir da origem,  $\phi \in [0, 2\pi]$  é o ângulo azimutal medido do eixo  $x$  ao eixo  $y$  e  $\theta \in [0, \pi]$  é o ângulo de elevação medido do eixo  $z$  ao plano  $xy$  como mostra a Figura 2. A relação entre as coordenadas dos dois sistemas é dada por  $(x, y, z) = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta)$ .

Em um arranjo linear uniforme de microfones, o comportamento da onda sonora pode ser processado em função da Transformada de Fourier. Já em arranjos esféricos, é possível usar os Harmônicos Esféricos como base para descrever a Transformada de Fourier Esférica (TFE). Os Harmônicos Esféricos são definidos como

$$Y_n^m(\theta, \phi) \equiv \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos \theta) e^{jm\phi}, \quad (1)$$

em que  $P_n^m(\cdot)$  são as funções de Legendre associadas,  $m \in \mathbb{Z}$  é o grau da função,  $n \in \mathbb{N}$  é sua ordem e  $j = \sqrt{-1}$ . Uma função  $f(\theta, \phi)$



**Figura 2:** Relação entre o Sistema de Coordenadas Cartesiano e o Sistema de Coordenadas Esférico.

que pertença ao espaço de Hilbert (conjunto de funções quadraticamente integráveis em uma esfera unitária) pode ser representada como a soma ponderada de harmônicos esféricos

$$f(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n f_{nm} Y_n^m(\theta, \phi), \quad (2)$$

em que  $f_{nm}$  são os coeficientes resultantes da TFE de  $f(\theta, \phi)$ , tais que

$$f_{nm} = \int_0^{2\pi} \int_0^{\pi} f(\theta, \phi) [Y_n^m(\theta, \phi)]^* \sin \theta \, d\theta \, d\phi, \quad (3)$$

ou seja, a Equação (3) é a TFE e a Equação (2) a sua inversa. Na prática, a somatória infinita é truncada em uma ordem  $N$ .

### 3.2 Amostragem pelo arranjo

A TFE é definida como a integral sobre uma superfície esférica. No entanto, em aplicações práticas não é possível saber o valor do campo sonoro em toda a superfície da esfera, mas somente nos pontos onde se encontram microfones. Neste caso, o campo sonoro é amostrado por um arranjo de microfones distribuídos sobre a superfície (rígida ou aberta) de uma esfera de raio  $r_a$ . Os  $Q$  microfones estão localizados em  $(r_a, \theta_q, \phi_q) \equiv (r_a, \Omega_q)$ ,  $q = 1, \dots, Q$ .

Para minimizar erros de rebatimento (*aliasing*), é recomendável que a distribuição dos microfones sobre a esfera seja uniforme, o que só é

possível para alguns valores de  $Q$  até  $Q = 20$ , obtidos posicionando-se os microfones nos vértices dos sólidos platônicos [24]. Para outras configurações, busca-se uma distribuição quase-uniforme, que pode ser obtida, por exemplo, dividindo-se a esfera em partes com mesma área e diâmetro mínimo [25].

Uma vez escolhida a distribuição dos microfones no arranjo esférico rígido, a TFE é calculada aproximando-se a integral sobre a superfície esférica por uma somatória dos pontos de amostragem, ou seja

$$\int_0^{2\pi} \int_0^{\pi} h(\theta, \phi) \sin \theta \, d\theta \, d\phi \approx \frac{4\pi}{Q} \sum_{q=1}^Q h(\theta_q, \phi_q). \quad (4)$$

Esta aproximação só é válida para distribuições uniformes ou quase-uniformes [25]. Conforme apresentado na Seção 5.1, a vantagem de se usar estas distribuições é que a “energia” dos pontos não fica concentrada em algumas regiões, como nos polos da esfera em uma distribuição angular uniforme, por exemplo.

### 3.3 Modelo do sinal

Assume-se uma cena sonora composta por  $S$  ondas planas e amostrada por um arranjo esférico de microfones distribuídos sobre uma esfera rígida. Neste caso, como mostra Rafaely [26], a pressão sonora em cada um dos  $Q$  microfones é dada, em coordenadas esféricas, por

$$p(k, r_a, \Omega_q) = \sum_{s=1}^S \sum_{n=0}^N \sum_{m=-n}^n d_s(k) R_n(kr_a) [Y_n^m(\Omega_s)]^* Y_n^m(\Omega_q), \quad (5)$$

em que  $k$  é o número de onda,  $d_s(k)$  é a amplitude complexa e  $\Omega_s = (\theta_s, \phi_s)$  a direção (vista da origem) de cada onda  $s$  e

$$R_n(kr_a) = 4\pi j^n \left[ J_n(kr_a) - \frac{J_n'(kr_a)}{H_n^{(2)'}(kr_a)} H_n^{(2)}(kr_a) \right]. \quad (6)$$

O primeiro termo de  $R_n(kr_a)$  caracteriza a onda plana incidente, enquanto o segundo termo caracteriza a difração causada pela esfera rígida na onda plana.  $J_n(x)$  é a função esférica de Bes-

sel e  $H_n^{(2)}(x)$  é a função esférica de Hankel de segundo grau.

Arranjos de microfones possuem limitações práticas que influenciam na forma que o campo sonoro é amostrado. Uma delas é a garantia de bom funcionamento apenas em uma faixa de frequência, a qual depende da geometria do arranjo. No caso de um arranjo esférico rígido de microfones, a maior frequência ( $f_{\max}$ ) permitida é obtida a partir da relação [27]

$$f_{\max} = \frac{c}{2r_a\Theta}, \quad (7)$$

na qual  $c$  é a velocidade do som no ar e  $\Theta$  é o menor ângulo entre dois microfones do arranjo. Segundo [24], a frequência mínima ( $f_{\min}$ ) suportada pelo arranjo esférico rígido é tal que  $kr_a > 3$ , ou seja,

$$f_{\min} = \frac{3c}{2\pi r_a}. \quad (8)$$

### 3.3.1 Matriz de sensibilidade

A pressão sonora medida por cada microfone depende da direção e da frequência da onda plana e da influência do próprio arranjo no campo sonoro. Logo, para que a busca feita pelos algoritmos de recuperação seja bem sucedida, é necessário fornecer a ele a estrutura da malha de busca e como cada ponto dela influencia na forma que o sinal da fonte chega até o arranjo. Essa é a função da matriz de sensibilidade, que é fixa para uma malha de busca pré-definida.

A partir da Equação (5), é possível definir uma malha de busca com  $L$  direções e reescrevê-la como o somatório da influência de cada direção na cena sonora. Isto é feito transformando a Equação (5) em uma multiplicação matricial. Define-se  $\mathbf{p}_{Q \times 1} = [p(\Omega_1), \dots, p(\Omega_Q)]^T$  como o vetor de pressão sonora no qual  $p(\Omega_q)$  é a pressão medida no microfone  $q$ . Logo, tem-se

que

$$\mathbf{p} = \mathbf{A}\mathbf{b} = \begin{bmatrix} g(\Omega_1, \Omega_1) & \dots & g(\Omega_1, \Omega_L) \\ \vdots & g(\Omega_q, \Omega_l) & \vdots \\ g(\Omega_Q, \Omega_1) & \dots & g(\Omega_Q, \Omega_L) \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_L \end{bmatrix}, \quad (9)$$

sendo

$$g(\Omega_q, \Omega_l) = \sum_{n=0}^N R_n(kr_a) \sum_{m=-n}^n Y_n^m(\Omega_l)^* Y_n^m(\Omega_q). \quad (10)$$

Define-se  $\Omega_l = (\theta_l, \phi_l)$ ,  $l = 0, \dots, L$  como a direção  $l$  na malha de busca — que também pode ser entendida como direção de visada/escuta do arranjo — e  $g(\Omega_q, \Omega_l)$  é o elemento  $(q, l)$  da matriz de sensibilidade  $\mathbf{A}_{Q \times L}$ .  $\mathbf{b}_{L \times 1} = [b_1, \dots, b_L]^T$  é o vetor que contém a amplitude complexa de cada direção na malha de busca e possui entradas não-nulas apenas quando as direções  $\Omega_l$  e  $\Omega_s$  coincidem. Nestas entradas não-nulas,  $b_l = d_s(k)$ . Quando  $L \gg S$ , a cena sonora é dita **esparsa** e possibilita a aplicação de algoritmos de regularização promotora de esparsidade.

## 4. BEAMFORMING ESFÉRICO

Para avaliar os algoritmos promotores de esparsidade, será usado como base de comparação o *beamforming* esférico tradicional, algoritmo frequentemente usado em aplicações de imageamento acústico [24].

Um filtro espacial esférico é definido como a integral da pressão sonora sobre uma região esférica ponderada por uma função  $w(k, \theta, \phi)$  que irá maximizar o resultado da integral para uma onda plana incidente da direção  $(\theta, \phi)$ , ao passo que atenua o resultado da integral para ondas planas provenientes de outras direções. A saída do *beamforming* é escrita como

$$y = \int_0^{2\pi} \int_0^\pi [w(k, \theta, \phi)]^* p(k, r, \theta, \phi) \sin \theta \, d\theta \, d\phi. \quad (11)$$

O mesmo cálculo pode ser realizado no domínio da TFE. Neste caso, a saída do *beamforming* é definida como uma soma ponderada dos coefi-

cientes da TFE da pressão sonora, ou seja,

$$y = \sum_{n=0}^N \sum_{m=-n}^n [w_{nm}]^* p_{nm}, \quad (12)$$

na qual  $w_{nm}$  são os pesos escolhidos e  $p_{nm}$  são os coeficientes da TFE da pressão sonora sobre a esfera, tais que

$$p_{nm}(k, r_a) = \frac{4\pi}{Q} \sum_{q=1}^Q p(k, r_a, \Omega_q) [Y_n^m(\Omega_q)]^*. \quad (13)$$

O valor absoluto da saída do *beamforming* para cada direção  $(\theta, \phi)$  permite gerar uma imagem acústica, como a apresentada na Figura 4.

É necessário, então, calcular os pesos  $w_{nm}$  — no domínio dos harmônicos esféricos — para que o *beamforming* tenha um comportamento desejado. A seguinte formulação é proposta em [24]

$$[w_{nm}(k)]^* = \frac{Y_n^m(\theta_l, \phi_l)}{R_n(kr)}. \quad (14)$$

A divisão por  $R_n(kr)$  garante que a imagem acústica não dependa do comportamento físico do campo sonoro ao redor do arranjo, pois remove o efeito de espalhamento da esfera rígida. Rafaely [24] comenta que as direções  $\Omega_l$  não se relacionam com os pontos de amostragem  $\Omega_q$  e que, quando é usado o *beamforming* com objetivo de imageamento acústico, é recomendável que  $\Omega_l$  seja extraído de um conjunto de pontos denso.

Apesar de cumprirem a mesma função teórica — representar o campo sonoro em um número finito de posições —, as direções  $\Omega_l$  escolhidas para a malha de busca do *compressive beamforming* podem ou não ser iguais às direções  $\Omega_l$  escolhidas para o *beamforming* esférico.

## 5. COMPRESSIVE BEAMFORMING ESFÉRICO

O objetivo do *Compressive Beamforming* é resolver o sistema linear  $\mathbf{p} = \mathbf{A}\mathbf{b}$  — em que  $\mathbf{p}$  e  $\mathbf{A}_{Q \times L}$  são conhecidos e  $\mathbf{b}$  é a variável que se deseja determinar — de forma que  $\mathbf{b}$  seja esparso. Como o tamanho do vetor  $\mathbf{b}$  é muito maior do

que o tamanho do vetor  $\mathbf{p}$ , as colunas da matriz  $\mathbf{A}$  são linearmente dependentes e, por isso, o sistema é subdeterminado e possui infinitas soluções. Algumas soluções, devido à sua característica esparsa, podem ser úteis e são obtidas quando determinadas condições são impostas ao problema.

É possível obter uma solução única incluindo uma restrição como, por exemplo, a minimização da norma  $\ell_p$  do vetor  $\mathbf{b}$  ( $\|\mathbf{b}\|_p = (\sum_i |b_i|^p)^{1/p}$ ). Diferentes normas promovem diferentes soluções para o problema. Neste trabalho, são utilizadas as normas  $\ell_1$  e  $\ell_2$ . Mas, primeiro, é necessário avaliar a “qualidade” da malha de busca.

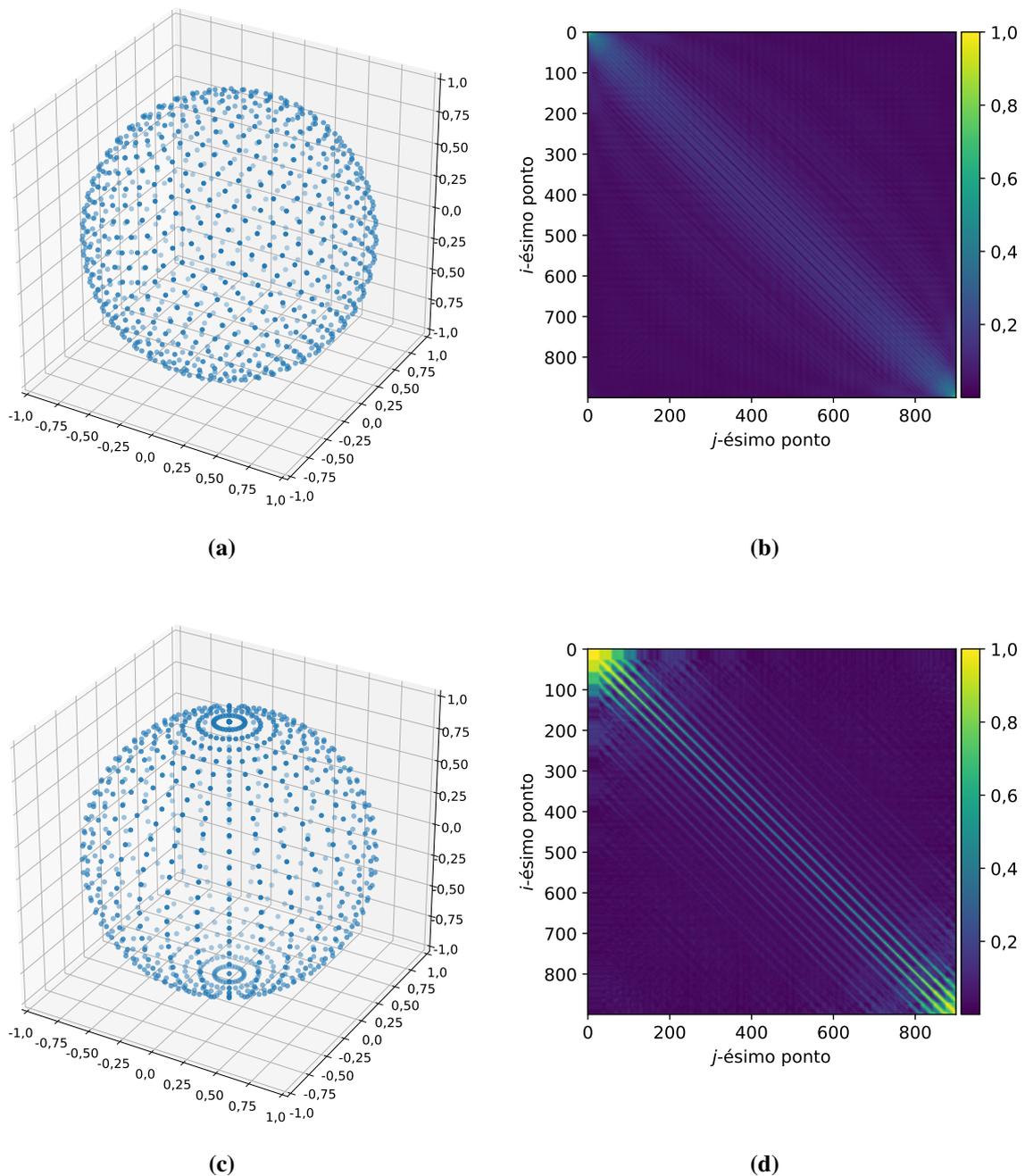
### 5.1 Coerência

Para que o CB esférico tenha bom desempenho, a distribuição dos pontos na malha de busca deve gerar uma matriz de sensibilidade com colunas que possuem baixa coerência entre si. Em [13], os autores utilizam a máxima dependência linear entre duas colunas da matriz de sensibilidade para avaliar a distribuição dos pontos. Isto é conhecido como coerência mútua e é definido como o elemento com valor máximo fora da diagonal principal da matriz de Gram normalizada de  $\mathbf{A}$ , ou seja,

$$\mu(\mathbf{A}) = \max_{i \neq j} G_{ij} = \max_{i \neq j} \frac{|\mathbf{a}_i^H \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2}, \quad (15)$$

no qual  $G_{ij}$  é o elemento  $(i, j)$  da matriz de Gram normalizada e  $\mathbf{a}_i$  é a  $i$ -ésima coluna de  $\mathbf{A}$ . A coerência mútua varia de 0 a 1, sendo 1 o mais coerente.

No entanto, apesar de ser uma medida comum em *compressive sensing*, a coerência mútua só analisa dois pontos com máxima dependência. Aqui, é preciso analisar como a coerência está distribuída em cada malha de busca para avaliar qual permite um melhor desempenho do CB esférico. Para isto, julga-se ser mais apropriado calcular a dependência linear média entre os pontos [28], ou seja, a média, aqui denotada por  $\bar{g}$ , de todos os elementos fora da diagonal principal da matriz normalizada de Gram.



**Figura 3:** Comparação entre (a) Distribuição quase-uniforme sobre uma esfera de raio fixo com (b) sua matriz de Gram e (c) Distribuição angular uniforme com (d) sua matriz Gram.

Na Figura 3 são comparadas as matrizes de Gram normalizadas de duas malhas de busca com 900 pontos: uma com distribuição quase-uniforme [25] dos pontos sobre uma esfera de raio fixo (Figura 3 (a)) e outra com distribuição angular uniforme ( $\phi \in [0; 360^\circ]$  com passo  $\Delta\phi = 12^\circ$  e  $\theta \in [0; 180^\circ]$  com passo  $\Delta\theta = 6^\circ$ ) também sobre uma esfera de raio fixo (Figura 3 (c)).

Analisando apenas as matrizes de Gram normalizadas nas Figuras 3 (b) e 3 (d), é possível assumir que a distribuição quase-uniforme permitirá um melhor desempenho do CS esférico. Esta é uma suposição válida, mas é necessário calcular a média  $\bar{g}$  em ambos os casos para confirmarmos. Na distribuição quase-uniforme,  $\mu(\mathbf{A}) = 0,93$  e  $\bar{g} = 0,07$ . Na distribuição angular uniforme,  $\mu(\mathbf{A}) = 1,00$  e  $\bar{g} = 0,08$ . Logo, a dis-

tribuição quase-uniforme pode trazer melhores resultados. É notável também que a distribuição angular uniforme faz com que os pontos da região dos polos tenham maior coerência entre si quando comparada com a distribuição quase-uniforme, já que tais pontos estão próximos da diagonal principal na Figura 3 (d).

## 5.2 Algoritmo de recuperação não-esparso

Um método clássico de regularização de sistemas subdeterminados é a minimização da norma  $\ell_2$ , também conhecida como mínimos quadrados, descrita por

$$\min_{\mathbf{b}} \|\mathbf{b}\|_2 \quad \text{s. a.} \quad \mathbf{p} = \mathbf{A}\mathbf{b}. \quad (16)$$

Quando se lida com sinais práticos, é necessário levar em consideração o efeito do ruído no sistema, o que pode ser modelado por

$$\tilde{\mathbf{p}} = \mathbf{A}\mathbf{b} + \mathbf{n}, \quad (17)$$

em que  $\mathbf{n}_{Q \times 1} = [n_1, \dots, n_Q]^T$  é o vetor que representa o ruído aditivo. A minimização da norma  $\ell_2$  pode então ser relaxada por

$$\min_{\mathbf{b}} \|\mathbf{b}\|_2 \quad \text{s. a.} \quad \|\tilde{\mathbf{p}} - \mathbf{A}\mathbf{b}\|_2 \leq \varepsilon, \quad (18)$$

em que  $\varepsilon$  é a energia do ruído e equivale a  $\|\mathbf{n}\|_2$ . Uma forma de estimar  $\varepsilon$  é por meio da razão sinal-ruído, estimada por  $\text{SNR} = 10 \log_{10}(\|\mathbf{p}\|_2 / \|\mathbf{n}\|_2)$ .

Apesar de fornecer uma solução única para o problema, a minimização da norma  $\ell_2$  não fornece a localização das direções das ondas planas com exatidão, como veremos na Seção 6. Isto pode ser explicado pelo fato de a norma  $\ell_2$  priorizar a minimização da energia do sinal em vez da sua esparsidade. Ao resolver através da norma  $\ell_2$ , a solução possui baixa resolução espacial (obtemos vários pontos em torno da direção das fontes com aproximadamente a mesma energia). Para tentar melhorá-la, podemos utilizar outras normas.

## 5.3 Algoritmos de recuperação esparsos

Os algoritmos promotores de esparsidade aplicados aqui buscam um vetor solução  $\hat{\mathbf{b}}$  composto por um pequeno número de coeficientes não-nulos. A forma como cada algoritmo encontra a solução  $\hat{\mathbf{b}}$  é detalhada a seguir.

### 5.3.1 Minimização da Norma $\ell_1$ via DCP

Idealmente, uma restrição de minimização da norma  $\ell_0$  deve promover uma solução esparsa, já que a norma  $\ell_0$  contabiliza a quantidade de entradas não nulas em  $\mathbf{x}$ . No entanto, a minimização da norma  $\ell_0$  é um problema não-convexo, que se torna computacionalmente intratável [9]. Uma boa aproximação da norma  $\ell_0$  é feita através da norma  $\ell_1$ , que resulta em um problema de otimização convexo—o que promove a convergência para um mínimo global—e possui a tendência de gerar uma solução esparsa [18]. A minimização da norma  $\ell_1$  é definida como

$$\min_{\mathbf{b}} \|\mathbf{b}\|_1 \quad \text{s. a.} \quad \|\tilde{\mathbf{p}} - \mathbf{A}\mathbf{b}\|_2 \leq \varepsilon. \quad (19)$$

Neste artigo esta minimização é implementada com ajuda da biblioteca CVXPY, que usa a programação convexa disciplinada (DCP) para garantir que os problemas de otimização especificados sejam convexos [29].

### 5.3.2 Minimização da Norma $\ell_1$ via LASSO

A Equação (19) pode ser reescrita sem a presença de restrições através do *least absolute shrinkage and selection operator* (LASSO, em tradução livre: operador de mínima compactação e seleção absoluta) com o uso do regularizador  $\lambda \geq 0$  [20],

$$\min_{\mathbf{b}} \frac{1}{2} \|\tilde{\mathbf{p}} - \mathbf{A}\mathbf{b}\|_2^2 + \lambda \|\mathbf{b}\|_1. \quad (20)$$

O problema acima é de minimização da norma  $\ell_2$  regularizada pela norma  $\ell_1$  da solução  $\hat{\mathbf{b}}$ . Esta formulação busca, ao mesmo tempo, uma solução que seja esparsa e que minimize a energia do erro de aproximação.  $\lambda$  é um fator de compromisso entre o erro de aproximação e a esparsidade do vetor solução. Quanto maior o  $\lambda$ , mais esparsa é a solução. À medida que  $\lambda$  diminui,

valores não-nulos são gradativamente “liberados” no vetor  $\hat{\mathbf{b}}$  até chegar à solução do problema de mínimos quadrados (quando  $\lambda = 0$ ).

Para cada  $\varepsilon$ , existe um  $\lambda$  tal que (19) e (20) sejam iguais. O máximo valor que  $\lambda$  assume é  $\|\mathbf{A}^H \tilde{\mathbf{p}}\|_\infty$ , e, acima desse valor,  $\hat{\mathbf{b}} = \mathbf{0}$  permanece como a solução ótima [18]. Quando  $\lambda = \|\mathbf{A}^H \tilde{\mathbf{p}}\|_\infty$ , a primeira entrada não-nula de  $\hat{\mathbf{b}}$  é liberada. À medida que  $\lambda$  diminui, novas entradas são liberadas e suas respectivas amplitudes aumentam. Então, se quisermos gerar uma solução  $S$ -esparsa ótima, encontramos o valor mínimo de  $\lambda$  tal que o vetor  $\hat{\mathbf{b}}$  tenha  $S$  entradas não-nulas. No entanto, na prática, este pode não ser o caso. Como será visto na Seção 6, o  $\lambda$  é escolhido com base no aumento das amplitudes à medida que as entradas não-nulas são liberadas. É possível visualizar  $S$  amplitudes em destaque, as quais correspondem às direções das ondas planas presentes na cena sonora. Resta, então, escolher um  $\lambda$  que proporcione uma solução  $\hat{\mathbf{b}}$  esparsa e, ao mesmo tempo, com altos valores das amplitudes das  $S$  ondas planas.

O LASSO permite encontrar o melhor valor para  $\lambda$  a partir de uma análise gráfica com base no valor inicial  $\lambda = \|\mathbf{A}^H \tilde{\mathbf{p}}\|_\infty$  e, por isso, não será calculado previamente o valor ideal de  $\lambda$  para que sejam encontradas com máxima exatidão as  $S$  ondas planas. Esta é a desvantagem do caminho escolhido para o LASSO, o qual servirá como um refinamento da solução obtida através da minimização da norma  $\ell_1$  tal qual (19).

### 5.3.3 Orthogonal Matching Pursuit

O *Orthogonal Matching Pursuit* faz parte da classe dos algoritmos gulosos (*greedy algorithms*), cujo princípio básico é encontrar sequencialmente as colunas não-nulas presentes em uma medição por meio de um processo iterativo na matriz de sensibilidade. O algoritmo OMP [30] consiste em duas etapas: seleção da coluna não-nula e atualização do resíduo.

Como o vetor  $\mathbf{b}$  possui  $S$  elementos não-nulos, o vetor da pressão  $\tilde{\mathbf{p}}$  é uma combinação linear de  $S$  colunas de  $\mathbf{A}$ . A cada iteração, o algoritmo

escolhe a coluna que mais se relaciona com o resíduo, o qual é inicializado com o valor de  $\tilde{\mathbf{p}}$ , por meio do produto interno entre cada coluna e o resíduo e guarda o índice dessa coluna. Calcula-se então um vetor  $\hat{\mathbf{x}}$  que contém o peso de cada coluna de  $\mathbf{A}$  selecionada utilizando mínimos quadrados. Sua contribuição é então subtraída de  $\tilde{\mathbf{p}}$  e um novo resíduo é encontrado. O número de iterações é igual ao nível de esparsidade da solução, ou seja,  $S$ . A saída é a solução  $\hat{\mathbf{b}}$  cujas entradas não-nulas estão nas posições das colunas de  $\mathbf{A}$  selecionadas e possuem amplitudes definidas pelo vetor  $\hat{\mathbf{x}}$  calculado na última iteração. Abaixo, o funcionamento do algoritmo OMP [30].

**Entrada:** Matriz de sensibilidade  $\mathbf{A}$ , pressão medida  $\tilde{\mathbf{p}}$ , nível de esparsidade  $S$ .

**Inicialização:** Resíduo  $\mathbf{r}_0 = \tilde{\mathbf{p}}$ , conjunto  $\Upsilon_0 = \emptyset$  dos índices das entradas não-nulas, matriz  $\tilde{\mathbf{A}}_0 = \emptyset$  composta pelas colunas de  $\mathbf{A}$  selecionadas, vetor  $\hat{\mathbf{b}}$  contendo apenas entradas nulas, contador  $t = 1$ .

**Enquanto**  $t \leq S$ :

1. Encontre, por meio do produto interno ( $\langle \cdot, \cdot \rangle$ ), o índice  $v_t$  da coluna de  $\mathbf{A}$  que possui maior dependência linear com o resíduo

$$v_t = \operatorname{argmax}_{j=1, \dots, L} |\langle \mathbf{r}_{t-1}, \mathbf{a}_j^* \rangle|; \quad (21)$$

2. Aumente o conjunto  $\Upsilon_t = \Upsilon_{t-1} \cup v_t$  e a matriz  $\tilde{\mathbf{A}}_t = [\tilde{\mathbf{A}}_{t-1} \ \mathbf{a}_{v_t}]$ ;

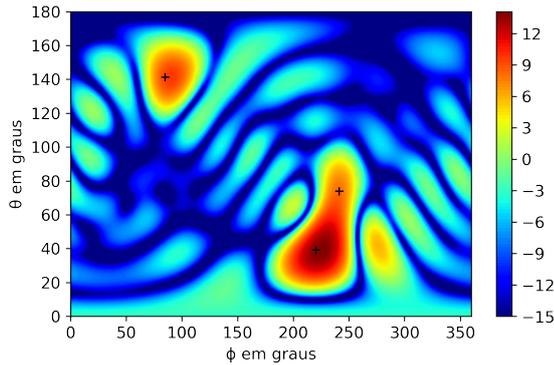
3. Estime uma saída  $\hat{\mathbf{x}}$  resolvendo o problema de mínimos quadrados

$$\min_{\mathbf{x}} \|\tilde{\mathbf{p}} - \tilde{\mathbf{A}}_t \mathbf{x}\|_2; \quad (22)$$

4. Calcule o novo resíduo  $\mathbf{r}_t = \tilde{\mathbf{p}} - \tilde{\mathbf{A}}_t \hat{\mathbf{x}}$  excluindo a influência da coluna da matriz  $\mathbf{A}$  selecionada;

5. Incremente  $t$ .

**Saída:**  $\hat{\mathbf{b}}$  possui índices não-nulos listados em  $\Upsilon_S$  e o vetor estimado  $\hat{\mathbf{b}}_{v_t}$  é o  $i$ -ésimo elemento de  $\hat{\mathbf{x}}_S$ .

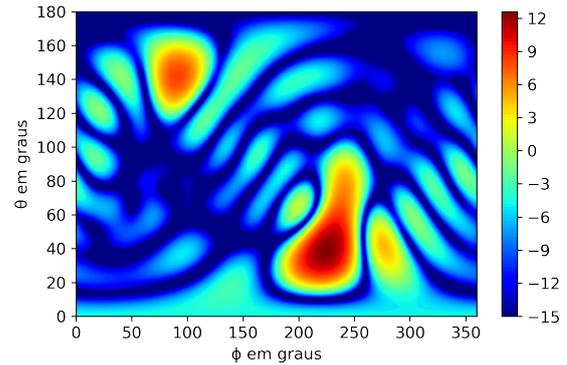


**Figura 4:** Imagem acústica (colorido, em dB) de uma cena sonora composta por 3 (+) ondas planas de 3500 Hz com amplitudes  $[1, 0, 0; 0, 70; 0, 50]$  vindas das direções  $[(\Omega_{100}, \Omega_{325}, \Omega_{800})] = [(39, 04^\circ; 220, 47^\circ), (73, 94^\circ; 241, 22^\circ), (141, 16^\circ; 85, 04^\circ)]$ .

## 6. SIMULAÇÃO

Foi simulada uma cena sonora composta por três ondas planas de 3500 Hz e amplitudes normalizadas que valem, respectivamente, 1,00, 0,70 e 0,50, localizadas nos pontos de números 100, 325 e 800 (sendo os pontos 100 e 325 próximos entre si e o ponto 800 distante destes) de uma malha de busca que contém 900 pontos, que equivalem às direções  $\Omega_{100} = (39, 04^\circ; 220, 47^\circ)$ ,  $\Omega_{325} = (73, 94^\circ; 241, 22^\circ)$  e  $\Omega_{800} = (141, 16^\circ; 85, 04^\circ)$  na distribuição quase-uniforme, como discutido na Seção 5.1. O ruído aditivo está centrado em 3500 Hz, com amplitude 20 dB menor que a amplitude da pressão sonora gerada pela fonte sonora, e fase aleatoriamente selecionada para cada microfone. A captação do campo sonoro é feita por um arranjo de microfones esférico rígido com  $Q = 64$  microfones e raio  $r_a = 0,10$  m. As características do arranjo na simulação (incluindo as posições dos microfones) são as mesmas do arranjo que será usado na gravação (Seção 7) e foram fornecidas pelo fabricante do arranjo (fabricante: Visisonics, modelo do arranjo: 5/64 Audio Visual Camera). A aproximação da TFE foi feita com  $N = 7$ . A faixa de frequência permitida para estas características está entre, aproximadamente, 2220 Hz e 3700 Hz. A imagem acústica gerada pelo *beamforming* esférico por esta cena encontra-se na Figura 4.

Como a imagem acústica é uma representação



**Figura 5:** Minimização da norma  $\ell_2$ : Imagem acústica (colorida, em dB) reconstruída de uma cena sonora com ruído.

em duas dimensões da distribuição das fontes em três dimensões, o uso de uma malha de pontos equiespaçados iria resultar em distorção da imagem, principalmente nos polos. Para contornar isso, os ângulos de visada da Equação (14) não são os mesmos ângulos das posições da malha de busca utilizados para calcular a pressão sonora na Equação (10). Para o primeiro caso é usada uma distribuição angular uniforme contendo 40.000 posições com 200 valores diferentes de  $\theta$  e 200 diferentes valores de  $\phi$ , equidistantes entre si. É importante ressaltar que isto não afeta o desempenho dos algoritmos de recuperação, já que só serve para facilitar a visualização dos resultados.

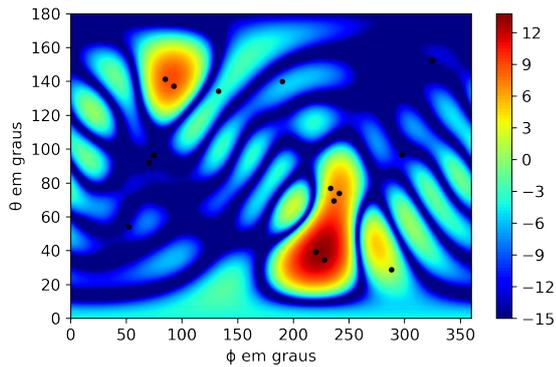
A biblioteca CVXPY [29] foi utilizada para resolver os problemas de minimização e considerou-se um valor em módulo superior a  $10^{-6}$  como o limiar para tratar esta amplitude como não-nula.

### 6.1 Norma $\ell_2$

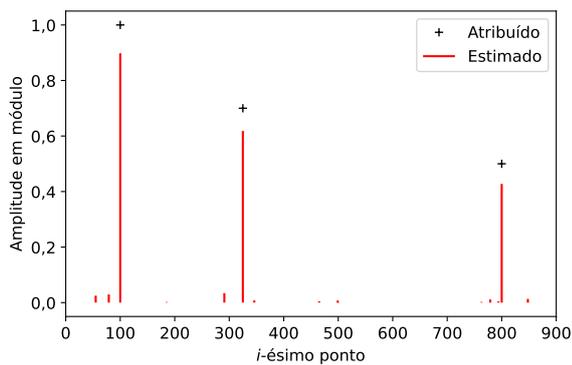
O resultado por meio da minimização da norma  $\ell_2$  (ver Figura 5) gerou um vetor  $\hat{\mathbf{b}}$  com norma  $\|\hat{\mathbf{b}}\|_0 = 900$ . Na prática, esta solução corresponde ao *beamforming* esférico regular (ver Apêndice A).

### 6.2 Norma $\ell_1$ via DCP

O resultado por meio da minimização da norma  $\ell_1$  gerou um vetor  $\hat{\mathbf{b}}$  com norma  $\|\hat{\mathbf{b}}\|_0 = 15$  — Figuras 6 (a) e 6 (b). Foi obtido um vetor es-



(a)



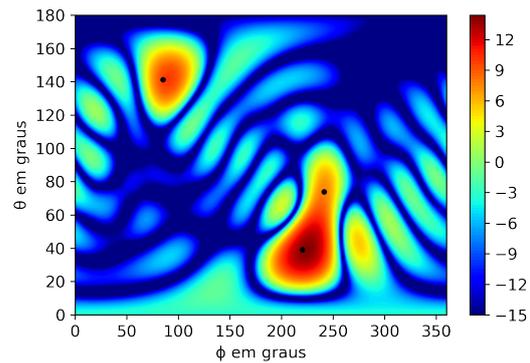
(b)

**Figura 6:** Minimização da norma  $\ell_1$  via DCP com ruído: (a) Imagem acústica reconstruída (colorido, em dB) e pontos de amplitudes não-nulas (preto, 15 pontos) (b) Amplitudes dos pontos da malha de busca atribuídas (+) na simulação e estimadas (vermelho).

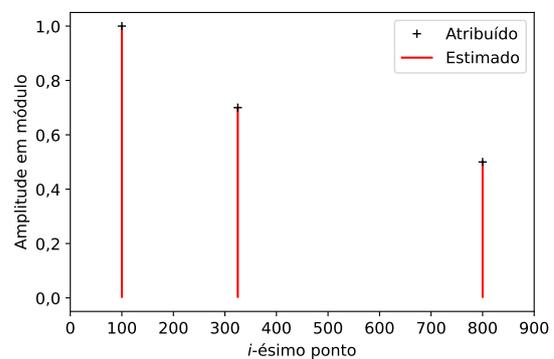
parso cujas entradas não-nulas estão nos pontos da vizinhança das 3 ondas planas. É notável também que as 3 entradas de maior valor estão exatamente nos pontos atribuídos: 100, 325, 800. Aqui, o vetor não possui exatamente 3 entradas não-nulas por causa da presença do ruído, já que o caso sem ruído (Figuras 7 (a) e 7 (b)) conseguiu localizar as fontes com exatidão. Na sequência, é avaliado o comportamento da minimização da norma  $\ell_1$  via LASSO.

### 6.3 Norma $\ell_1$ via LASSO

O primeiro passo é obter o  $\lambda$  que gera o resultado mais esparsa ( $\|\hat{\mathbf{b}}\|_0 = 1$ ) acima do qual  $\hat{\mathbf{b}} = \mathbf{0}$ :  $\lambda_{max} = \lambda_0 = \|\mathbf{A}^H \tilde{\mathbf{p}}\|_\infty = 98,91$  (ver curva azul, correspondente ao ponto de maior amplitude  $l = 100$ , da Figura 8). À medida que o  $\lambda$  diminui, a amplitude desta pri-



(a)

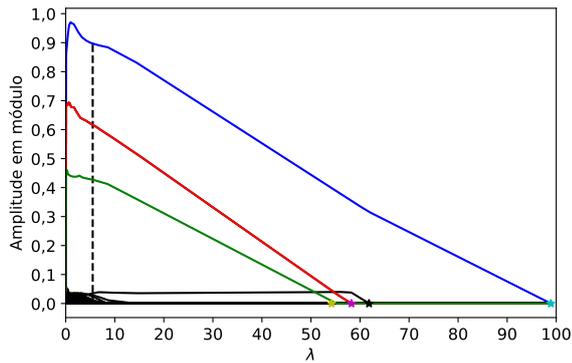


(b)

**Figura 7:** Minimização da norma  $\ell_1$  via DCP sem ruído: (a) Imagem acústica reconstruída (colorido, em dB) e pontos de amplitudes não-nulas (preto, 3 pontos) (b) Amplitudes dos pontos da malha de busca atribuídas (+) na simulação e estimadas (vermelho).

meira entrada não-nula aumenta e mais entradas não-nulas são liberadas. Quando  $\lambda_1 = 61,84$ , a segunda entrada não-nula é liberada. Como será visto na Figura 8, esta entrada não corresponde a um dos pontos atribuídos, mas encontra-se na vizinhança da posição atribuída  $\Omega_{325} = (73, 94^\circ; 241, 22^\circ)$ , localizando-se em  $\Omega_{291} = (69, 38^\circ; 236, 49^\circ)$ . Diminuindo  $\lambda$  até  $\lambda_2 = 58,25$  é liberada a terceira entrada não-nula, que corresponde ao ponto 325 (curva vermelha). A quarta entrada, correspondente a  $l = 800$ , é liberada quando  $\lambda_3 = 54,20$  (curva verde).

Quando  $\lambda_4 = 5,50$ ,  $\|\hat{\mathbf{b}}\|_0 = 15$  e o nível de esparsidade da solução através do LASSO se iguala ao da solução via DCP, então deve-se escolher um  $\lambda > 5,50$  para garantir uma solução mais esparsa. Como penalidade, não seria possível



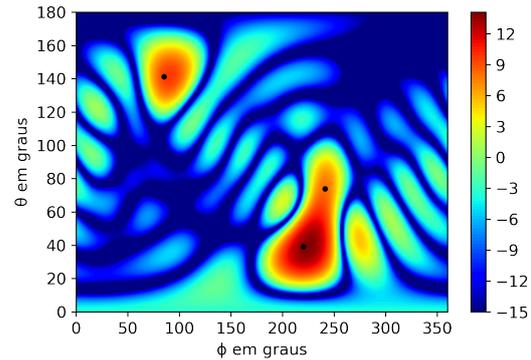
**Figura 8:** Caminho LASSO: Valor das entradas não-nulas em função do  $\lambda$ . Em azul, vermelho e verde estão, respectivamente, os pontos 100, 325 e 800. Em preto estão todos os outros 897 pontos. As linhas tracejadas indicam os possíveis valores de  $\lambda$  a serem escolhidos.

recuperar as amplitudes das ondas planas como vemos na Figura 8. Apesar de não ter gerado uma solução  $S$ -esparsa com as amplitudes recuperadas, o caminho LASSO forneceu a informação sobre o número de ondas planas presentes na cena sonora, já que há apenas três pontos em destaque na Figura 8 cujas amplitudes crescem quase linearmente. Em uma situação sem o conhecimento prévio sobre esta informação, o LASSO mostrou-se um método eficaz para estipular o número de ondas planas. Tal informação é utilizada no algoritmo OMP, pois este necessita do número de entradas não-nulas da solução  $\hat{\mathbf{b}}$ .

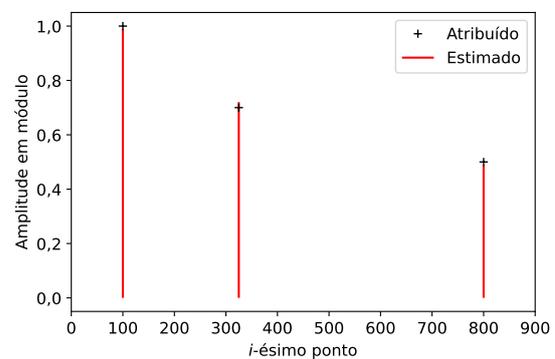
#### 6.4 Orthogonal Matching Pursuit

É possível implementar o OMP em Python de, pelo menos, duas formas: através da função própria da biblioteca Scikit-learn [31] ou através do pseudocódigo em 5.3.3. A desvantagem da função da Scikit-learn é que ela só aceita matrizes reais como entrada, e isto distorce o resultado. Por esta razão, foi implementado o algoritmo com base no pseudocódigo da Seção 5.3.3. No OMP, o nível de esparsidade deve ser fornecido como parâmetro de entrada do algoritmo e isto permite encontrar o vetor com o número de entrada não-nulas desejado.

A partir da Figura 9 (b), nota-se que o OMP conseguiu encontrar as direções atribuídas às ondas planas e suas amplitudes reconstruídas



(a)



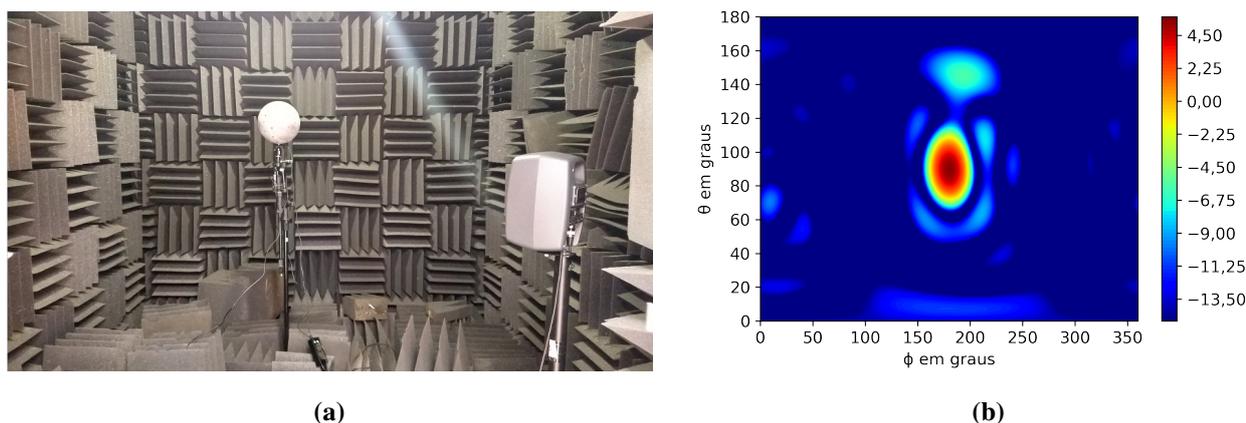
(b)

**Figura 9:** Aplicação do OMP: (a) Imagem acústica reconstruída de uma cena sonora com ruído (colorido, em dB) e pontos de amplitudes não-nulas (preto, 3 pontos) (b) Amplitudes dos pontos da malha de busca atribuídas (+) na simulação e estimadas (vermelho) pelo *Orthogonal Matching Pursuit* em uma cena sonora com ruído.

estão bem próximas das atribuídas. O sucesso do OMP pode ser explicado pelo conhecimento do nível de esparsidade de  $\hat{\mathbf{b}}$ , já que o algoritmo força uma saída que contenha exatamente, neste caso, 3 entradas não-nulas. No entanto, isto é uma desvantagem caso estejamos lidando com uma gravação sem nenhum conhecimento prévio sobre o número de ondas planas na cena sonora.

## 7. RESULTADOS EXPERIMENTAIS

Uma gravação foi realizada com um arranjo esférico rígido de microfones com as mesmas especificações da Seção 6 em uma câmara anecoica na qual havia uma área reflexiva no teto. Um alto-falante distante 2 m do arranjo emitiu



**Figura 10:** (a) Montagem da gravação de um sinal emitido por um alto-falante a 2 m de um arranjo esférico de microfones em uma câmara anecoica (b) Imagem acústica gerada pela gravação para uma frequência de 3500 Hz.

uma varredura senoidal de 20 Hz a 20 kHz. A Figura 10 (a) mostra a montagem feita para esta gravação. A frequência escolhida para aplicarmos os algoritmos de esparsidade foi 3500 Hz. A energia medida do ruído nos microfones é  $\varepsilon = 0,12$ . Para estimar a direção e a amplitude da onda plana, os resultados dos três algoritmos de esparsidade seguiram a mesma lógica da simulação. Na Figura 10 (b) está a imagem acústica desta medição gerada usando o *beamforming* esférico.

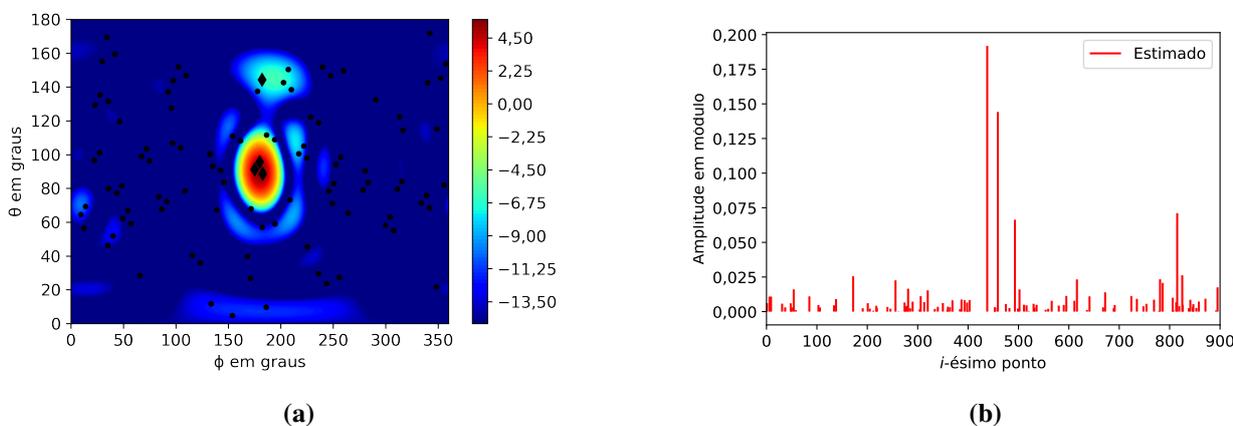
O primeiro algoritmo aplicado foi a minimização da norma  $\ell_1$  via DCP (cuja solução gerou a imagem acústica da Figura 11 (a)) que gerou uma solução  $\hat{\mathbf{b}}$  com norma  $\|\hat{\mathbf{b}}\|_0 = 102$ . As

quatro direções de maior amplitude são:

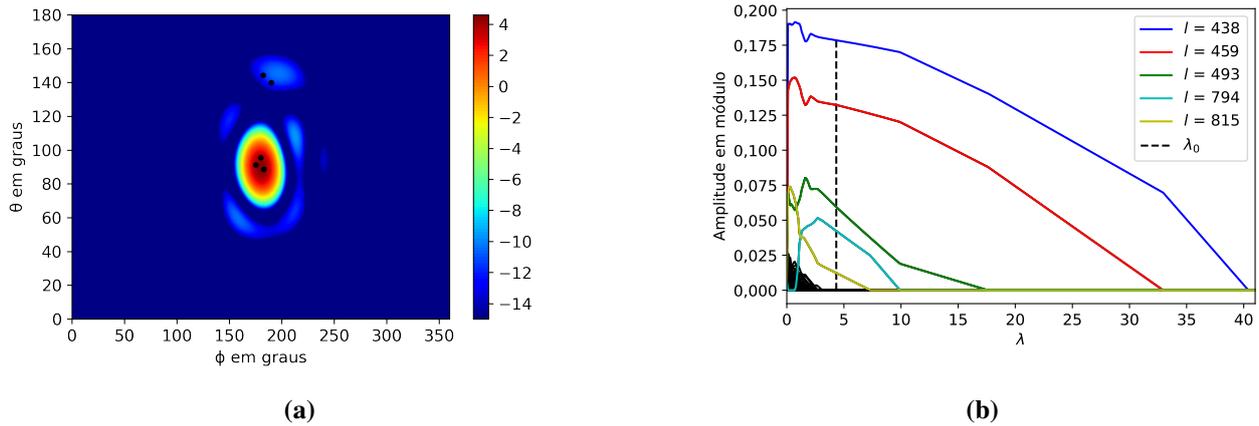
$$\begin{aligned} \Omega_{438} &= (88, 54^\circ; 182, 85^\circ), \\ \Omega_{459} &= (91, 21^\circ; 175, 18^\circ), \\ \Omega_{494} &= (95, 68^\circ; 42, 41^\circ) \text{ e} \\ \Omega_{815} &= (144, 31^\circ; 182, 42^\circ). \end{aligned}$$

Não é possível afirmar com exatidão quantas ondas planas estão presentes na cena sonora, mas já é possível notar duas regiões de onde vêm ondas com maior amplitude. A seguir, é avaliada a minimização da norma  $\ell_1$  via LASSO.

Na Figura 12 (b), é apresentada a variação da amplitude das entradas não-nulas em função de  $\lambda$  gerada pelo caminho do LASSO. Cinco



**Figura 11:** Minimização da norma  $\ell_1$  via DCP: (a) Imagem acústica reconstruída de uma cena sonora com ruído (colorido, em dB) e pontos de amplitudes não-nulas (preto, 102 pontos ao total). Os pontos representados por  $\blacklozenge$  são os quatro de maior amplitude (b) Amplitudes em módulo dos pontos da malha de busca estimadas (vermelho).



**Figura 12:** Minimização da norma  $\ell_1$  via LASSO: (a) Imagem acústica reconstruída (colorido, em dB) com  $\lambda_0 = 4,33$ . (b) O caminho da variação das amplitudes dos pontos em função de  $\lambda$ . Os gráficos em azul, vermelho, verde, ciano e amarelo correspondem, respectivamente a 438, 459, 493, 794 e 815 (pontos destacados sobre a imagem acústica).  $\lambda_0$  é o menor  $\lambda$  tal que a saída  $\hat{\mathbf{b}}$  tenha  $\|\hat{\mathbf{b}}\|_0 = 5$ .

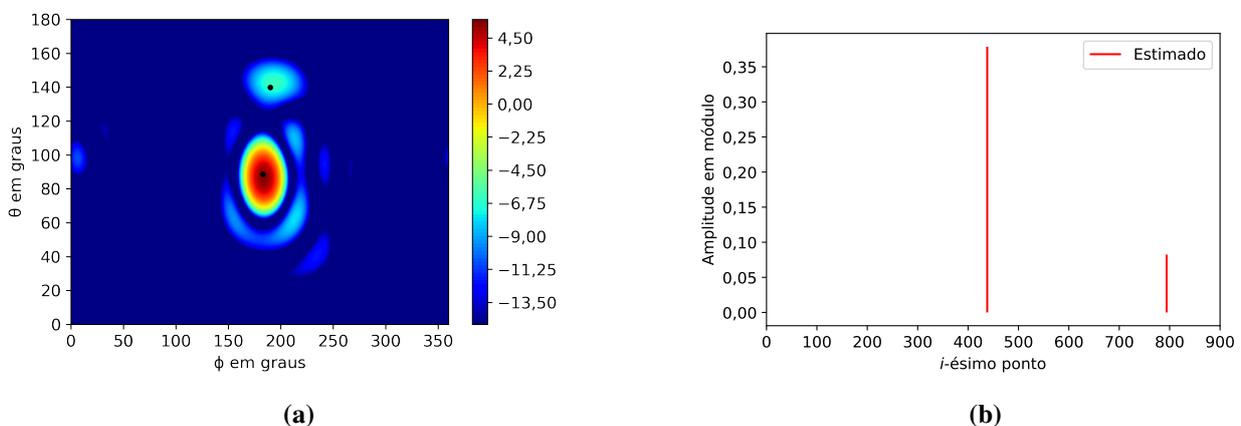
direções se destacam:

$$\begin{aligned} \Omega_{438} &= (88, 54^\circ; 182, 85^\circ), \\ \Omega_{459} &= (91, 21^\circ; 175, 18^\circ), \\ \Omega_{493} &= (95, 55^\circ; 179, 92^\circ), \\ \Omega_{794} &= (139, 96^\circ; 190, 08^\circ) \text{ e} \\ \Omega_{815} &= (144, 31^\circ; 182, 42^\circ). \end{aligned}$$

A aplicação do LASSO com  $\lambda_0 = 4,33$ —menor  $\lambda$  tal que a saída  $\hat{\mathbf{b}}$  tenha  $\|\hat{\mathbf{b}}\|_0 = 5$ —gerou a imagem acústica da Figura 12 (a). Devido à proximidade destas direções e ao fato de a cena sonora ser prática, ou seja, com presença de ruído, é possível separar os pontos em duas regiões: uma formada por  $\Omega_{438}$ ,  $\Omega_{459}$  e  $\Omega_{493}$  e outra for-

mada por  $\Omega_{794}$  e  $\Omega_{815}$ . Assim, são definidas duas ondas planas para aplicar o OMP.

Por meio do OMP (Figura 13), foram obtidas as direções  $\Omega_{438} = (88, 54^\circ; 182, 85^\circ)$  e  $\Omega_{794} = (139, 96^\circ; 190, 08^\circ)$ , cujas amplitudes valem, respectivamente,  $(-0,0035 - j0,38)$  e  $(0,05 - j0,06)$ . A primeira direção corresponde ao local da fonte e a segunda à reflexão presente no teto.



**Figura 13:** Aplicação do OMP: (a) Imagem acústica reconstruída (colorido, em dB) e pontos de amplitudes não-nulas (preto, 2 pontos) (b) Amplitudes em módulo dos pontos da malha de busca estimadas (vermelho).

## 8. CONCLUSÃO

Neste artigo, foram utilizados três algoritmos promotores de esparsidade para determinar a direção de fontes em uma cena sonora utilizando um arranjo esférico de microfones, cuja vantagem é conseguir captar o sinal em todas as direções. Verificou-se que para uma cena simulada, tanto com quanto sem ruído aditivo, todos os algoritmos foram capazes de determinar acuradamente o número de fontes e suas direções.

Tanto simulação quanto o processamento da gravação foram feitas em um computador com o sistema operacional Ubuntu 20.04 e processador Intel Core i5-5200U@2,20 GHz. Quanto à implementação dos algoritmos, a minimização da norma  $\ell_1$  via DCP e o LASSO são relativamente simples graças à biblioteca CVXPY, que possui uma função pronta que resolve o problema de minimização. Para a implementação do OMP também usou-se esta mesma biblioteca na etapa do algoritmo em que há a resolução do problema de mínimos quadrados. Os tempos de processamento da minimização da norma  $\ell_1$  via DCP e do OMP foram, respectivamente, 3,98 s e 0,46 s. O LASSO foi dividido em duas etapas: escolha de um  $\lambda$  por meio do caminho apresentado na Figura 8 e a minimização dada pela Equação (20) para o  $\lambda$  escolhido. A etapa mais demorada é primeira, a qual levou 1h20min21s. Isto faz com que o LASSO seja computacionalmente muito custoso quando comparado aos outros dois algoritmos. A segunda etapa tem um tempo de processamento comparável à minimização da norma  $\ell_1$  via DCP: 5,95 s. A vantagem da minimização da norma  $\ell_1$  via DCP frente ao LASSO está, então, no tempo de processamento.

Já para uma situação real com uma fonte gravada em ambiente quase anecoico, observou-se um resultado menos satisfatório dos algoritmos quando comparado ao resultado da simulação. Isto pode ser explicado pela complexidade da cena sonora prática, a qual apresenta condições que não estão presentes na simulação, uma vez que não há garantia de o espaço seja totalmente anecoico ou que o ruído seja aleatório, além de que a fonte não é uma fonte pontual, podendo

gerar reflexões e difrações vindas de outras direções do espaço. No entanto, a combinação do LASSO para determinar o número de fontes com o OMP para refinar a estimativa da amplitude da onda apresentou resultados animadores.

Para trabalhos futuros, recomenda-se repetir a simulação com outros tipos de fontes não pontuais e repetir a medição com mais de uma fonte, emitindo sinais correlacionados ou não, com diferentes ângulos de separação entre as fontes. Além disso, sugere-se analisar o comportamento destes algoritmos para a tarefa de separação de fontes.

## 9. AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processos nº 2017/08120-6 e nº 2018/26531-6. Agradecemos pelo apoio recebido.

## REFERÊNCIAS

1. BENESTY, J.; CHEN, J.; HUANG, Y. *Microphone Array Signal Processing*. [S.l.]: Springer Berlin Heidelberg, 2008. (Springer Topics in Signal Processing). doi: [10.1007/978-3-540-78612-2](https://doi.org/10.1007/978-3-540-78612-2). ISBN 978-3540786122.
2. NASCIMENTO, Vítor H.; MASIERO, Bruno S.; RIBEIRO, Flávio P. Acoustic imaging using the Kronecker array transform. In: COELHO, Rosangela Fernandes; NASCIMENTO, Vitor Heloiz; QUEIROZ, Ricardo Lopes de; ROMANO, Joao Marcos Travassos; CAVALCANTE, Charles Casimiro (Ed.). *Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing*. [S.l.]: CRC Press, 2015. p. 153–178. ISBN 978-1498722360.
3. HÖGBOM, J. A. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement*, v. 15, p. 417–426, 1974. ISSN 0365-0138. Dis-

ponível em: <https://ui.adsabs.harvard.edu/abs/1974A%26AS...15..417H/>.

4. WANG, Yanwei; LI, Jian; STOICA, Petre; SHEPLAK, Mark; NISHIDA, Toshikazu. Wideband relax and wideband clean for aeroacoustic imaging. *The Journal of the Acoustical Society of America*, v. 115, n. 2, p. 757–767, 2004. doi: [10.1121/1.1639906](https://doi.org/10.1121/1.1639906).

5. DOUGHERTY, Robert P. Extensions of DAMAS and benefits and limitations of deconvolution in beamforming. In: *11th AIAA/CEAS Aeroacoustics Conference (26th AIAA Aeroacoustics Conference)*. [S.l.: s.n.], 2005. p. 1–13. ISBN 156-3477300. ISSN 0146-3705. doi: [10.2514/6.2005-2961](https://doi.org/10.2514/6.2005-2961).

6. YAN, Shefeng; MA, Yuanliang; HOU, Chaohuan. Optimal array pattern synthesis for broadband arrays. *The Journal of the Acoustical Society of America*, v. 122, n. 5, p. 2686–2696, 2007. ISSN 00014966. doi: [10.1121/1.2785037](https://doi.org/10.1121/1.2785037).

7. EHRENFRIED, Klaus; KOOP, Lars. Comparison of iterative deconvolution algorithms for the mapping of acoustic sources. *AIAA Journal*, v. 45, n. 7, p. 1584–1595, 2007. ISSN 0001-1452. doi: [10.2514/1.26320](https://doi.org/10.2514/1.26320).

8. YARDIBI, Tarik; LI, Jian; STOICA, Petre; CATTAFESTA, Louis N. Sparsity constrained deconvolution approaches for acoustic source mapping. *The Journal of the Acoustical Society of America*, v. 123, n. 5, p. 2631–2642, 2008. doi: [10.1121/1.2896754](https://doi.org/10.1121/1.2896754).

9. XENAKI, Angeliki; GERSTOFT, Peter; MOSEGAARD, Klaus. Compressive beamforming. *The Journal of the Acoustical Society of America*, v. 136, n. 1, p. 260–271, 2014. ISSN 1520-8524. doi: [10.1121/1.4883360](https://doi.org/10.1121/1.4883360).

10. Shi, J.; Hu, G.; Zhang, X.; Sun, F.; Zhou, H. Sparsity-based two-dimensional DOA estimation for coprime array: From sum–difference coarray viewpoint. *IEEE Transactions on Signal Processing*, v. 65, n. 21, p. 5591–5604, 2017. doi: [10.1109/TSP.2017.2739105](https://doi.org/10.1109/TSP.2017.2739105).

11. EPAIN, Nicolas; JIN, Craig; SCHAIK, André van. The application of compressive sam-

pling to the analysis and synthesis of spatial sound fields. In: *127th AES Convention*. New York, USA: [s.n.], 2009. p. 1–12. Disponível em: <http://www.aes.org/e-lib/browse.cfm?elib=15052>.

12. FERNANDEZ-GRANDE, Efren; XENAKI, Angeliki. Compressive sensing with a spherical microphone array. *The Journal of the Acoustical Society of America*, v. 139, n. 2, p. EL45–EL49, 2016. doi: [10.1121/1.4942546](https://doi.org/10.1121/1.4942546).

13. PING, Guoli; CHU, Zhigang; YANG, Yang. Compressive spherical beamforming for acoustic source identification. *Acta Acustica united with Acustica*, v. 105, n. 6, p. 1000–1014, 2019. ISSN 1610-1928. doi: [doi:10.3813/AAA.919406](https://doi.org/10.3813/AAA.919406).

14. PAN, J. Spherical harmonic atomic norm and its application to DOA estimation. *IEEE Access*, v. 7, p. 156555–156568, 2019. doi: [10.1109/ACCESS.2019.2950016](https://doi.org/10.1109/ACCESS.2019.2950016).

15. CHEN, Scott Shaobing; DONOHO, David L.; SAUNDERS, Michael A. Atomic decomposition by basis pursuit. *SIAM Review*, v. 43, n. 1, p. 129–159, 2001. doi: [10.1137/S003614450037906X](https://doi.org/10.1137/S003614450037906X).

16. GORODNITSKY, I. F.; RAO, B. D. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, v. 45, n. 3, p. 600–616, 1997. doi: [10.1109/78.558475](https://doi.org/10.1109/78.558475).

17. ERKOC, M. E.; KARABOGA, N. Evolutionary algorithms for sparse signal reconstruction. *Signal, Image and Video Processing*, p. 1863–1711, 2019. doi: [10.1007/s11760-019-01473-w](https://doi.org/10.1007/s11760-019-01473-w).

18. ELAD, M. *Sparse and Redundant Representations*. [S.l.]: Springer-Verlag New York, 2010. (Springer Topics in Signal Processing). doi: [10.1007/978-1-4419-7011-4](https://doi.org/10.1007/978-1-4419-7011-4). ISBN 978-1441970114.

19. FERNANDEZ-GRANDE, Efren; XENAKI, Angeliki. Compressive sensing with a spherical microphone array. *The Journal of the Acoustical Society of America*, v. 139, p. EL45–EL49, 02 2016. doi: [10.1121/1.4942546](https://doi.org/10.1121/1.4942546).

20. GERSTOFT, Peter; XENAKI, Angeliki; MECKLENBRÄUKER, Christoph. Multiple and single snapshot compressive beamforming. *The Journal of the Acoustical Society of America*, v. 138, n. 4, p. 2003–2014, 2015. doi: [10.1121/1.4929941](https://doi.org/10.1121/1.4929941).

21. ROSSUM, Guido Van; DRAKE, Fred L. *Python 3 Reference Manual*. Scotts Valley, USA: CreateSpace, 2009. ISBN 978-1441412690.

22. TREES, Harry L. Van. *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 2002. ISBN 9780470542965. Disponível em: <https://books.google.com.br/books?id=J5TZDwAAQBAJ>.

23. CANDÈS, Emmanuel J.; WAKIN, M. B. An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine*, v. 25, n. 2, p. 21–30, 2008. ISSN 1053-5888. doi: [10.1109/MSP.2007.914731](https://doi.org/10.1109/MSP.2007.914731).

24. RAFAELY, B. *Fundamentals of Spherical Array Processing*. [S.l.]: Springer International Publishing, 2018. (Springer Topics in Signal Processing). doi: [10.1007/978-3-662-45664-4](https://doi.org/10.1007/978-3-662-45664-4). ISBN 978-3319995618.

25. SAFF, E.B.; KUIJLAARS, A.B.J. Distributing many points on a sphere. *The Mathematical Intelligencer*, v. 19, p. 5–11, 1997. doi: [10.1007/BF03024331](https://doi.org/10.1007/BF03024331).

26. RAFAELY, Boaz. *Fundamentals of Spherical Array Processing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015. v. 8. (Springer Topics in Signal Processing, v. 8). doi: [10.1007/978-3-662-45664-4](https://doi.org/10.1007/978-3-662-45664-4). ISSN 09250042. ISBN 978-3662456637.

27. BERTET, Stéphanie; DANIEL, Jerome; PARIZET, Etienne; WARUSFEL, Olivier. Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources. *Acta Acustica united with Acustica*, v. 99, n. 4, p. 642 – 657, 2013. doi: [10.3813/AAA.918643](https://doi.org/10.3813/AAA.918643).

28. ELAD, M. Optimized projections for compressed sensing. *IEEE Transactions on Signal*

*Processing*, v. 55, n. 12, p. 5695–5702, 2007. doi: [10.1109/TSP.2007.900760](https://doi.org/10.1109/TSP.2007.900760).

29. DIAMOND, Steven; BOYD, Stephen. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, v. 17, n. 1, p. 2909–2913, 2016. ISSN 1532-4435. doi: [doi/10.5555/2946645.3007036](https://doi.org/10.5555/2946645.3007036).

30. TROPP, J. A.; GILBERT, A. C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, v. 53, n. 12, p. 4655–4666, 2007. doi: [10.1109/TSP.2015.2413384](https://doi.org/10.1109/TSP.2015.2413384).

31. PEDREGOSA, Fabian; VAROQUAUX, Gaël; GRAMFORT, Alexandre; MICHEL, Vincent; THIRION, Bertrand; GRISEL, Olivier; BLONDEL, Mathieu; PRETTENHOFER, Peter; WEISS, Ron; DUBOURG, Vincent; VANDERPLAS, Jake; PASSOS, Alexandre; COURNAPEAU, David; BRUCHER, Matthieu; PERROT, Matthieu; DUCHESNAY Édouard. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, n. oct, p. 2825–2830, 2011. Disponível em: <http://jmlr.org/papers/v12/pedregosa11a.html>.

## A. REPRESENTAÇÃO MATRICIAL

A matriz de sensibilidade  $\mathbf{A}$ , cujo elemento  $(q, l)$  é calculado pela Equação (10), pode ser decomposta em  $\mathbf{A} = \mathbf{Y}_Q \mathbf{R} \mathbf{Y}_L^H$ , em que  $\mathbf{R}_{(N+1)^2 \times (N+1)^2}$  é uma matriz diagonal definida por

$$\mathbf{R} = \text{diag}[R_0(kr_a), R_1(kr_a), R_1(kr_a), R_1(kr_a), \dots, R_N(kr_a)]. \quad (23)$$

Para representar os Harmônicos Esféricos das posições dos microfones e das posições dos pontos da malha de busca, tem-se, respectivamente,  $\mathbf{Y}_Q$  de dimensão  $Q \times (N+1)^2$  e  $\mathbf{Y}_L$  de dimensão  $L \times (N+1)^2$  tais que:

$$\mathbf{Y}_Q = \begin{bmatrix} Y_0^0(\Omega_1) & \dots & Y_N^N(\Omega_1) \\ \vdots & \ddots & \vdots \\ Y_0^0(\Omega_Q) & \dots & Y_N^N(\Omega_Q) \end{bmatrix} \quad (24)$$

e

$$\mathbf{Y}_L = \begin{bmatrix} Y_0^0(\Omega_1) & \dots & Y_N^N(\Omega_1) \\ \vdots & \ddots & \vdots \\ Y_0^0(\Omega_L) & \dots & Y_N^N(\Omega_L) \end{bmatrix}. \quad (25)$$

para  $\mathbf{Y}_Q$ ,

$$\mathbf{y} = \mathbf{Y}_L \mathbf{R}^{-1} \mathbf{Y}_Q^{-1} \tilde{\mathbf{p}}. \quad (33)$$

Logo,  $\hat{\mathbf{b}}_{\ell_2} = \frac{4\pi}{L} \mathbf{y}$ .

O mesmo pode ser feito com a TFE da pressão nos microfones do arranjo na Equação (13). Seja  $\mathbf{p}_{nm} = [p_{00}, p_{1(-1)}, p_{10}, p_{11}, \dots, p_{NN}]^T$  o vetor que contém os coeficientes  $p_{nm}$  da TFE; então,

$$\mathbf{p}_{nm} = \frac{4\pi}{Q} \mathbf{Y}_Q^H \mathbf{p}. \quad (26)$$

A saída do *beamforming*, neste caso, pode ser escrita como

$$\mathbf{y} = \mathbf{Y}_L \mathbf{R}^{-1} \mathbf{p}_{nm} = \frac{4\pi}{Q} \mathbf{Y}_L \mathbf{R}^{-1} \mathbf{Y}_Q^H \mathbf{p}. \quad (27)$$

A solução da minimização da norma  $\ell_2$  (Equação (18)) está diretamente relacionada com a saída  $\mathbf{y}$  do *beamforming*. Esta solução é conhecida e obtida através da pseudo inversa de  $\mathbf{A}$ :

$$\hat{\mathbf{b}}_{\ell_2} = \mathbf{A}^+ \tilde{\mathbf{p}} = \mathbf{A}^H (\mathbf{A} \mathbf{A}^H)^{-1} \tilde{\mathbf{p}}. \quad (28)$$

Substituindo  $\mathbf{A} = \mathbf{Y}_Q \mathbf{R} \mathbf{Y}_L^H$  na Equação (28) tem-se

$$\hat{\mathbf{b}}_{\ell_2} = \mathbf{Y}_L \mathbf{R}^H \mathbf{Y}_Q^H (\mathbf{Y}_Q \mathbf{R} \mathbf{Y}_L^H \mathbf{Y}_L \mathbf{R}^H \mathbf{Y}_Q^H)^{-1} \tilde{\mathbf{p}}. \quad (29)$$

Como  $\mathbf{Y}_L$  possui distribuição dos pontos quase-uniforme [24],

$$\frac{4\pi}{L} \mathbf{Y}_L^H \mathbf{Y}_L = \mathbf{I}. \quad (30)$$

Então,

$$\hat{\mathbf{b}}_{\ell_2} = \frac{4\pi}{L} \mathbf{Y}_L \mathbf{R}^H \mathbf{Y}_Q^H (\mathbf{Y}_Q \mathbf{R} \mathbf{R}^H \mathbf{Y}_Q^H)^{-1} \tilde{\mathbf{p}}. \quad (31)$$

Aqui, é usado  $Q = (N + 1)^2$  permitindo que, além de  $\mathbf{R}$  que é uma matriz quadrada,  $\mathbf{Y}_Q$  seja também inversível. Assim,

$$\hat{\mathbf{b}}_{\ell_2} = \frac{4\pi}{L} \mathbf{Y}_L \mathbf{R}^{-1} \mathbf{Y}_Q^{-1} \tilde{\mathbf{p}}. \quad (32)$$

Solução similar é obtida a partir da Equação (27) pois, utilizando a igualdade da Equação (30)