



**Universidade Federal de Campina Grande**  
**Centro de Engenharia Elétrica e Informática**  
**Departamento de Engenharia Elétrica**

Trabalho de Conclusão de Curso

Modelamento Esparsos e Técnicas de Decomposição  
Matricial Aplicados à Detecção do Espectro

Frederico Carvalho Fontes do Amaral

Campina Grande - PB

Dezembro de 2019

Frederico Carvalho Fontes do Amaral

Trabalho de Conclusão de Curso

**Modelamento Esparsos e Técnicas de  
Decomposição Matricial Aplicados à  
Detecção do Espectro**

*Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande como parte dos requisitos necessários para a obtenção do grau de Bacharel em Engenharia Elétrica*

Área de Concentração: Telecomunicações

Orientador:

Professor Edmar Candeia Gurjão, D. Sc.

Campina Grande - PB

Dezembro de 2019

Frederico Carvalho Fontes do Amaral

Trabalho de Conclusão de Curso

# Modelamento Esparsos e Técnicas de Decomposição Matricial Aplicados à Detecção do Espectro

*Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande como parte dos requisitos necessários para a obtenção do grau de Bacharel em Engenharia Elétrica*

Área de Concentração: Telecomunicações

Aprovado em \_\_\_/\_\_\_/\_\_\_

**Professor Avaliador**

Universidade Federal de Campina Grande

Avaliador

**Edmar Candeia Gurjão**

Universidade Federal de Campina Grande

Orientador

## Agradecimentos

Primeiramente, agradeço aos meus pais, Luis e Maria, e ao meu irmão, Henrique. Muito obrigado por sempre estarem comigo nos momentos difíceis, por sempre acreditarem em mim quando eu mesmo não acreditava, e por sempre me ajudarem quando precisei. Serei eternamente grato por nunca pouparem esforços para que eu e meu irmão tivéssemos a melhor educação possível. É graças a esse gesto que sou quem sou hoje. Sem vocês, eu jamais teria chegado aqui.

Aos professores que foram importantes para a minha formação acadêmica. Vocês me ensinaram a beleza do ato de aprender, a importância do ato de questionar, e a necessidade de ser criativo. Guardarei esses ensinamentos pelo resto da minha vida. Em especial, gostaria de agradecer aos professores Edson Porto e Angilberto Muniz, por sempre se disporem a me ajudar durante a realização deste trabalho.

Aos meus amigos de Brasília, com os quais vivenciei os momentos mais lindos e felizes da minha vida. Eu sempre irei cultivar as lembranças que tenho de vocês com muito carinho. Aos meus colegas da UFCG, em especial àqueles do semestre 2014.1, com os quais sempre pude contar em momentos difíceis. Com vocês, aprendi lições muito importantes, que levarei para o resto da vida. Em especial, gostaria de agradecer aos meus colegas e amigos Kaio Nikelisson, Ulisses Gomes, João Pedro, Ravi Helon e Allender Vilar. Com vocês eu aprendi muito, dei muitas risadas, compartilhei vitórias e derrotas. Muito obrigado, por tudo.

Ao meu orientador, professor Edmar Candeia Gurjão, por todas as oportunidades e orientações concedidas ao longo desses últimos anos. Sempre serei grato por ter-me aceitado como orientando desde o início da minha trajetória acadêmica, e por ter me mostrado a beleza da área de pesquisa à qual nos dedicamos. Agradeço pela inestimável amizade, pela solicitude de todas as horas e por ter confiado no meu potencial desde o início. O senhor foi, é e sempre será um pilar fundamental de quem sou, e do profissional que virei a ser.

Por fim, agradeço aos que, de alguma forma, contribuíram para que eu chegasse até esse momento e por me ajudarem a me tornar quem sou hoje.

*"Man cannot remake himself without suffering,  
for he is both the marble and the sculptor."*

Alexis Carrel

# Resumo

A detecção do espectro é o primeiro passo para a utilização oportunista de frequências em Rádios Cognitivos (RC). A maioria das técnicas propostas anteriormente para detecção do espectro normalmente possuem um desempenho ruim em cenários com baixa SNR (*Signal-to-Noise Ratio*), e aquelas que possuem bom desempenho possuem, via de regra, um maior grau de complexidade. Neste trabalho, é apresentado um método inédito baseado em modelamento esparsos para detecção do espectro e identificação de canais livres para o espectro do sistema de televisão digital aberta brasileiro, considerando os efeitos do ruído aditivo gaussiano branco (*Additive White Gaussian Noise - AWGN*) e atenuação de sinal. A detecção é aprimorada por meio do uso de técnicas de decomposição matricial e os resultados mostram que o método proposto é capaz de identificar a ocupação dos canais com alta precisão, mesmo em cenários com SNR crítica.

**Palavras chave:** Espectro, Detecção, Modelamento esparsos, AWGN, Performance.

# Abstract

Spectrum sensing is the first step for enabling opportunistic frequency utilization in Cognitive Radios (CR). Most of the previous proposed techniques for spectrum sensing usually perform poorly in low SNR scenarios, and those having good performance normally have high complexity. In this paper, it is presented a method based on sparse modeling for spectrum sensing and free channel identification for the open digital television system spectrum in Brazil, considering the effects of additive white Gaussian noise and signal attenuation. The detection is enhanced through the use of matrix decomposition techniques and the results show that the proposed method can identify channel occupancy with high precision, even in critical SNR scenarios.

**Keywords:** Spectrum, Detection, Sparse modeling, AWGN, Performance.

## Lista de Figuras

1	Reta de quadrados mínimos de um certo conjunto de pontos . . . . .	4
2	Fluxograma do primeiro estágio do método implementado . . . . .	14
3	Processo de atualização de $\mathbf{M}_{sm}$ . . . . .	17
4	Fluxograma do segundo estágio do método implementado . . . . .	18
5	Espectro de uma amostra do espectro UHF . . . . .	19
6	Espectro da média espectral das amostras coletadas . . . . .	20
7	Magnitude das distâncias entre pares de valores singulares da matriz $\mathbf{H}$ . .	21
8	Espectro da média espectral após filtragem . . . . .	21
9	Espectro do modelo esparso construído . . . . .	22
10	Espectro do início do canal 21 da banda UHF . . . . .	23
11	Espectro do final do canal 21 da banda UHF . . . . .	24
12	Sinal de controle gerado no <i>software</i> MATLAB <sup>®</sup> . . . . .	25
13	Sinal de controle utilizado no GNU Radio . . . . .	26
14	Esfera unitária no $\mathbb{R}^3$ . . . . .	
15	Deformação de $\mathbf{S}$ por $\mathbf{A}$ . . . . .	
16	Deformação de $\mathbf{S}$ por $\mathbf{U}$ , $\mathbf{V}$ e $\mathbf{\Sigma}$ . . . . .	

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Fundamentação Teórica</b>	<b>2</b>
2.1	Decomposição em Valores Singulares . . . . .	2
2.2	Problema de Mínimos Quadrados . . . . .	3
2.3	Métodos Gananciosos . . . . .	5
2.3.1	Definições Básicas . . . . .	5
2.3.2	Definição . . . . .	6
2.3.3	Detalhamento do MP . . . . .	7
2.3.4	Etapas do MP . . . . .	7
2.3.5	Considerações sobre o MP . . . . .	9
2.3.6	Detalhamento do OMP . . . . .	10
2.4	Matrizes de Hankel . . . . .	12
<b>3</b>	<b>Materiais e Métodos</b>	<b>13</b>
3.1	Método Implementado . . . . .	13
3.1.1	Modelamento Esparso . . . . .	13
3.1.2	Identificação de Uso de Canal . . . . .	14
3.2	Simulações . . . . .	18
3.3	Experimentos . . . . .	22
<b>4</b>	<b>Resultados</b>	<b>27</b>
<b>5</b>	<b>Conclusão</b>	<b>28</b>

# 1 Introdução

Em diversos países, a utilização do espectro de frequências é regulamentada por agências regulatórias, que atribuem a cada usuário regulamentado (primário) uma certa faixa de frequências (canal). No Brasil, a entidade regulatória responsável por fazer essa regulamentação é a Agência Nacional de Telecomunicações (ANATEL), que estabelece o sistema de televisão aberta digital (SBTVD - Sistema Brasileiro de Televisão Digital) na banda UHF (*Ultra High Frequency*), que vai de 470 até 800 MHz. Essa banda é dividida em 55 canais, cada qual com 6 MHz de largura de faixa.

Em um mesmo sistema há, normalmente, muitos canais que não são utilizados, seja por falta de usuários interessados, ou mesmo porque usuários que receberam autorização de transmissão ainda não começaram a fazê-la, ou, ainda, interrompem a utilização em determinados instantes. A fim de melhorar o aproveitamento do espectro, foi proposto seu uso dinâmico, por vezes chamado de oportunista. Tal uso consiste em permitir a transmissão por usuários não regulamentados pela ANATEL, denominados oportunistas ou secundários, em canais que estejam livres em um dado momento. O uso oportunista do espectro foi um dos pilares para a proposição dos Rádios Cognitivos (RC), que são sistemas de transmissão que conseguem se adaptar ao ambiente em que se encontram.

Uma das tarefas básicas dos sistemas supracitados consiste na detecção do espectro para determinar os canais sem uso em um dado instante para, então, iniciar neles uma transmissão oportunista. Muitos métodos têm sido sugeridos para possibilitar a detecção do espectro em RC, tais como filtros casados e técnicas baseadas em autocorrelação, que envolvem um certo grau de complexidade e considerável custo computacional. Dentre os métodos sugeridos, pode-se destacar o método da energia, o qual identifica canais ocupados e livres por meio da comparação da energia dos mesmos com um certo limiar. Apesar de ser um método simples e de baixo custo computacional, sua eficiência é comprometida em cenários espectrais com baixa SNR devido à sua elevada sensibilidade ao ruído.

Assim, muitos dos métodos propostos para a detecção do espectro possuem desempenho insatisfatório em cenários com baixa SNR, e aqueles que possuem bom desempenho nestes normalmente envolvem um maior custo computacional, o que inviabiliza seu uso em muitas aplicações. Esse problema, porém, pode ser mitigado se as propriedades dos sistemas de comunicação forem levadas em consideração durante o processo de detecção.

Por exemplo, a maioria dos canais do SBTVD encontram-se livres em muitas regiões do país. Isso sugere que a ocupação das bandas de frequência desse sistema possam ser modeladas utilizando técnicas esparsas. Por naturalmente resultar na criação de modelos esparsos para sinais, que são mais compactos que sua representação convencional, o uso dessas técnicas implica no uso de uma menor quantidade de memória e poder computacional durante o processo de detecção espectral.

Nesse trabalho, é apresentado um método robusto e não-paramétrico para detecção espectral e identificação de canais livres na banda UHF, que explora a natureza esparsa da utilização dos canais do SBTVD. Esse método faz uso de técnicas baseadas em modelamento esparsos e técnicas de decomposição matricial para realizar a detecção do espectro. Ele leva em consideração os efeitos do AWGN e da atenuação na qualidade das amostras do espectro. Apesar desse método ter sido desenvolvido para o SBTVD, ele pode ser aplicado a outros países, com base no protocolo IEEE 802.22 [1].

## 2 Fundamentação Teórica

O método apresentado neste trabalho faz uso de uma grande quantidade de ferramentas matemáticas e computacionais. Nesta seção, serão expostas as fundamentações teóricas de algumas das ferramentas mais importantes utilizadas pelo método implementado.

A Decomposição em Valores Singulares possui grande importância teórica e prática. Uma explicação mais detalhada dessa decomposição, bem como da sua interpretação geométrica, encontram-se disponíveis no apêndice desse trabalho. Essa decomposição, juntamente com as matrizes de Hankel, são utilizadas para atenuar o ruído gaussiano branco presente nos sinais utilizados pelo método.

O método ganancioso denominado *Orthogonal Matching Pursuit*, que tem como uma de suas etapas a resolução de um problema de mínimos quadrados, é utilizado para construir os modelos esparsos para a banda UHF.

### 2.1 Decomposição em Valores Singulares

Seja uma matriz  $\mathbf{A} \in \mathbb{K}^{m \times n}$  (sendo  $\mathbb{K} = \mathbb{R}$  ou  $\mathbb{K} = \mathbb{C}$ ) com *rank*  $k$ . A decomposição em valores singulares (SVD - Singular Value Decomposition) de  $\mathbf{A}$  será

dada por

$$\mathbf{A}_{m \times n} = \mathbf{U}_{m \times m} \mathbf{\Sigma}_{m \times n} \mathbf{V}_{n \times n}^*,$$

onde

- $\mathbf{U}$ : matriz ortonormal cujos vetores coluna são denominados *vetores singulares esquerdos* de  $\mathbf{A}$ . Esses vetores formam uma base para o espaço coluna de  $\mathbf{A}$ ;
- $\mathbf{V}$ : matriz ortonormal cujos vetores coluna são denominados *vetores singulares esquerdos* de  $\mathbf{A}$ . Esses vetores formam uma base para o espaço linha de  $\mathbf{A}$ ;
- $\mathbf{\Sigma}$ : matriz diagonal cujos elementos da diagonal principal são tais que  $\sigma_i \in \mathbb{R}_+^*$ ,  $i \in \{1, \dots, k\}$ . Esses elementos são denominados *valores singulares* de  $\mathbf{A}$ , e estão dispostos em ordem decrescente de magnitude.

## 2.2 Problema de Mínimos Quadrados

O problema de mínimos quadrados essencialmente consiste na escolha de uma função  $y = f(x)$  que minimize o erro quadrático médio

$$e = \frac{1}{n} \sum_{n=1}^N [y_n - f(x_n)]^2.$$

Caso se deseje que  $y = f(x)$  seja uma reta, o problema se torna um problema regressão linear. Essa reta, denominada reta de quadrados mínimos, tem como objetivo representar um conjunto de  $N$  pontos da melhor forma possível (i.e., minimizar o erro médio quadrático). Depreende-se da definição do problema que, dependendo da quantidade de pontos a ser representada pela reta, sua resolução pode consumir uma quantidade apreciável de tempo computacional.

Uma solução conceitualmente simples e de fácil implementação (embora não seja a mais eficiente, tampouco a menos custosa do ponto de vista computacional) para o problema de regressão linear é o método de quadrados mínimos. Esse método, apesar da sua ineficiência, é, dada sua simplicidade, uma ferramenta de grande valor teórico para a resolução do problema supracitado.

De forma sucinta, o método de quadrados mínimos consiste na determinação da reta

$$y = \hat{a} + \hat{b}.x,$$

onde

$$\hat{a} = \bar{y} - \hat{b}.\bar{x}$$

e

$$\hat{b} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Uma breve análise das equações acima revela o porquê da simplicidade do método: a determinação dos parâmetros da reta de quadrados mínimos são determinados por simples somatórios. Na Figura 1, é ilustrado um exemplo de uma reta de quadrados mínimos para 20 pontos.

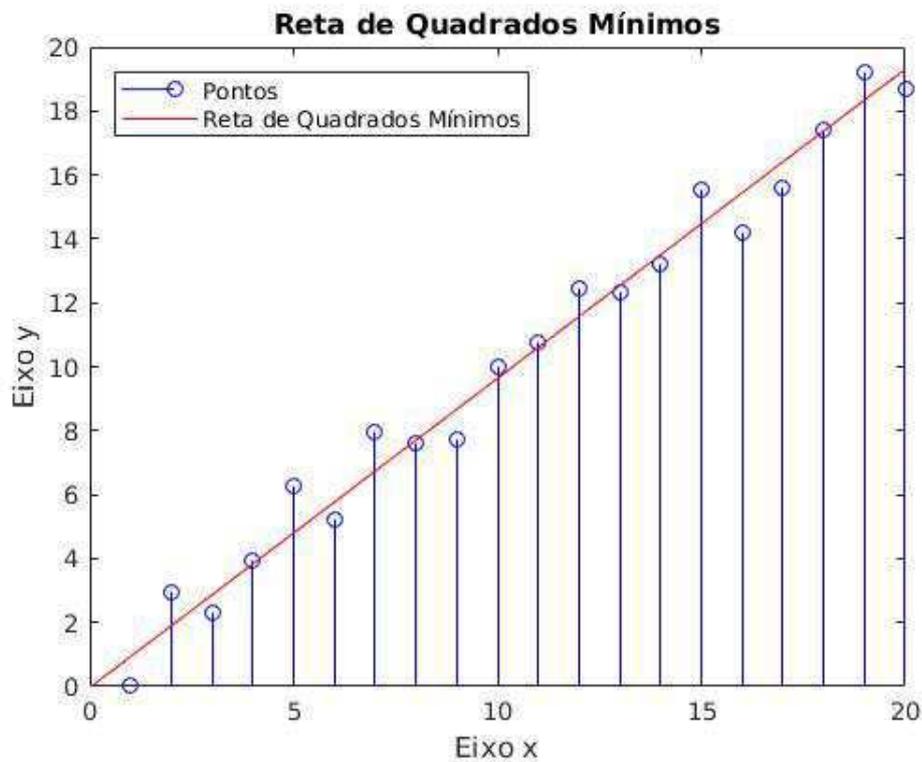


Figura 1: Reta de quadrados mínimos de um certo conjunto de pontos

Como ilustrado na Figura 1, conjuntos de dados são, por vezes, constituídos por uma grande quantidade de amostras aparentemente descorrelacionadas entre si. Isso não somente pode dificultar a análise desses conjuntos, como também a visualização das

informações mais importantes neles contidas. A reta de quadrados mínimos exemplificada na Figura 1 é uma reta tal que o somatório da distância dela até cada um dos pontos do conjunto (i.e., resíduos) é a mínima possível. Ou seja, ela representa da melhor forma possível o conjunto como um todo. Ela nos permite visualizar os pontos mais relevantes do conjunto (mais próximos da reta) e os menos (mais afastados) de maneira mais clara.

## 2.3 Métodos Gananciosos

### 2.3.1 Definições Básicas

**Definição 1** *Um espaço de sinais é um espaço euclidiano (i.e., com produto interno) vetorial complexo de dimensão finita  $\mathbb{C}^d$ . Os elementos desse espaço são denominados sinais.*

**Definição 2** *Um dicionário para um espaço de sinais consiste em um conjunto finito denotado por  $\mathcal{D}$ , constituído por uma coleção finita de vetores unitários elementares, denominados átomos. Cada átomo do dicionário é denotado por  $\varphi_\omega$ , onde  $\omega \in \Omega$ , o conjunto de indexadores. Os índices dos átomos funcionam como uma espécie de endereço para os mesmos, tendo grande importância nos algoritmos aqui abordados. De forma mais rigorosa, definimos um dicionário como sendo  $\mathcal{D} = \{\varphi_\omega \mid \omega \in \Omega\}$ . Doravante, denotaremos por  $N$  a cardinalidade do dicionário. Assim,  $|\mathcal{D}| = |\Omega| = N$ . Se o dicionário é um conjunto gerador para o espaço de sinais, então dizemos que o dicionário é completo. Nesse caso, cada sinal do espaço de sinais pode ser aproximado com erro zero por meio da combinação linear de átomos do dicionário. Se existirem átomos linearmente dependentes nesse dicionário, dizemos que o mesmo é redundante. Para que um dicionário seja completo,  $N \geq d$ . Para que ele seja redundante, basta que  $N > d$ .*

**Definição 3** *Um problema de aproximação esparsa consiste na busca de uma aproximação para um determinado sinal que pode ser representada com custo baixo (ou seja, como a combinação linear de uma quantidade mínima de átomos).*

**Definição 4** *A esparsidade de um vetor de coeficientes  $\mathbf{c}$  é igual ao número de entradas nulas do mesmo.*

**Definição 5** *A diversidade de um vetor de coeficientes  $\mathbf{c}$  é igual ao número de entradas não-nulas do mesmo, sendo, portanto, o contrário da esparsidade.*

### 2.3.2 Definição

Os métodos gananciosos são métodos numéricos utilizados em problemas de otimização. Eles tentam encontrar a escolha ótima (i.e., o máximo ou o mínimo) em cada iteração, com o objetivo de encontrar o máximo global. Esses métodos tomam decisões com base apenas nas informações disponíveis em cada iteração, sem se preocupar com os efeitos futuros dessas decisões (razão pela qual esses métodos são conhecidos como métodos míopes).

Esses métodos são largamente utilizados na resolução de problemas que envolvem a construção de aproximações esparsas de sinais. Objetivamente, o uso de tais métodos visa construir gradualmente uma aproximação esparsa para um dado sinal, a partir de funções (ou vetores) mais elementares, denominadas átomos.

Se, ao tentarmos resolver um problema de aproximação esparsa, dispomos de um dicionário ortonormal, é possível resolvê-lo por meio da escolha dos átomos cuja correlação com o sinal (no sentido do produto interno) seja o maior possível. Uma forma de se atingir esse objetivo é por meio da seleção do átomo mais correlacionado com o sinal, seguida da subtração daquele deste, e, então, repetir o procedimento (i.e., iterar) para os demais átomos. Doravante, esse procedimento será referido como algoritmo básico para um dicionário ortonormal. Métodos gananciosos para aproximações esparsas refinam esse procedimento, de forma que ele possa ser aplicado a uma maior quantidade de dicionários, e não apenas aos ortonormais. Entretanto, seja qual for o dicionário utilizado, seus vetores coluna devem ser unitários. Isso se deve ao fato de que tais vetores não devem possuir nenhuma informação além de sua direção e sentido. Caso sejam não-unitários, os métodos gananciosos potencialmente irão divergir.

Dentre os métodos gananciosos, podem-se destacar o MP (do inglês *Matching Pursuit*) e o OMP (do inglês *Orthogonal Matching Pursuit*). O MP é, resumidamente, uma técnica que advém imediatamente do algoritmo básico para um dicionário ortonormal. O OMP, por sua vez, é muito semelhante ao MP. A diferença entre eles reside no fato de que o OMP tem como uma de suas etapas a resolução de um problema de mínimos quadráticos, o que aumenta seu desempenho. Como tanto o MP como o OMP são métodos iterativos, deve-se estabelecer um critério de parada. Esse critério irá depender do tipo de problema de aproximação esparsa a ser resolvido, além da precisão desejada.

### 2.3.3 Detalhamento do MP

Vamos, primeiramente, expor uma descrição formal do procedimento MP, e em seguida analisaremos cada uma das etapas do mesmo de maneira detalhada. Inicialmente, vamos estabelecer um critério de parada e fixar um dicionário  $\mathcal{D}$ .

- Entrada: um sinal  $\mathbf{s} \in \mathbb{C}^d$ ;
- Saída: um vetor de coeficientes  $\mathbf{c} \in \mathbb{C}^\Omega$ .
- Etapas:
  1.  $\mathbf{c} \leftarrow \mathbf{0}$ ,  $\mathbf{r}_0 \leftarrow \mathbf{s}$  e  $t \leftarrow 1$ ;
  2. Determinar um índice  $\lambda_t$  tal que

$$\lambda_t \in \arg \max_{\omega} |\langle \mathbf{r}_{t-1}, \boldsymbol{\varphi}_{\omega} \rangle|;$$

3. Atualizar o vetor de coeficientes:

$$\mathbf{c}(\lambda_t) \leftarrow \mathbf{c}(\lambda_t) + \langle \mathbf{r}_{t-1}, \boldsymbol{\varphi}_{\lambda_t} \rangle;$$

4. Calcular o novo resíduo:

$$\mathbf{r}_t \leftarrow \mathbf{r}_{t-1} - \langle \mathbf{r}_{t-1}, \boldsymbol{\varphi}_{\lambda_t} \rangle \boldsymbol{\varphi}_{\lambda_t};$$

5. Incremente o indexador de *loop*:

$$t \leftarrow t + 1;$$

6. Se o critério de parada ainda não foi atingido, retornar à segunda etapa.

### 2.3.4 Etapas do MP

Na primeira etapa, as "condições iniciais" do problema serão especificadas. Ela consiste em atribuir o vetor nulo ao vetor de coeficientes ( $\mathbf{c} \leftarrow \mathbf{0}$ ), o sinal original ao resíduo ( $\mathbf{r}_0 \leftarrow \mathbf{s}$ ) e fixar o indexador de *loop* em 1 ( $t \leftarrow 1$ ). Essa etapa será realizada uma única vez, independente da quantidade de *loops* que sejam realizados.

A explicação da segunda etapa (também conhecida como seleção gananciosa) requer a definição do operador  $\arg \max$ . De forma rigorosa, pode-se definir esse operador como segue.

**Definição 6** *Seja  $f : X \rightarrow Y$ . O  $\arg \max$  sobre algum subconjunto  $S \subseteq X$  é*

$$\arg \max_{x \in S \subseteq X} f(x) := \{x | x \in S \wedge \forall y \in S : f(y) \leq f(x)\}.$$

*Caso  $S = X$  ou  $S$  seja evidenciado pelo contexto,*

$$\arg \max_x f(x) := \{\forall y \in S : f(y) \leq f(x)\}.$$

Ou seja,  $\arg \max$  é o conjunto de pontos  $x$  em  $X$  ou em  $S \subseteq X$  onde  $f(x)$  alcança seu valor máximo. O referido conjunto pode ser vazio, ser constituído por um único elemento ou ser composto por vários elementos. É importante destacar que o operador  $\arg \max$  não sugere, em nenhuma hipótese, um algoritmo computacional para se determinar os argumentos que maximizam o valor de  $f(x)$ . A elaboração de tal algoritmo depende da natureza do problema, sendo desejável, evidentemente, que ele seja o mais eficiente possível.

No algoritmo acima explicitado,  $f(x)$  é o produto interno usual. Mais especificamente, o produto interno do resíduo naquela iteração com um átomo do dicionário,  $|\langle \mathbf{r}_{t-1}, \boldsymbol{\varphi}_\omega \rangle|$ . Dessa forma, essa etapa consiste em uma busca que objetiva encontrar, dentre todos os átomos que compõe o dicionário, aquele que possui um maior grau de correlação com o resíduo naquela iteração. Esse grau de correlação é quantificado pelo valor de  $|\langle \mathbf{r}_{t-1}, \boldsymbol{\varphi}_\omega \rangle|$ . Como explicado preteritamente, o índice  $\omega$  em  $\boldsymbol{\varphi}_\omega$  tem como finalidade identificar cada átomo do dicionário, funcionando como um identificador individual. Via de regra, a busca consiste em um *loop*, que irá, a cada etapa, calcular  $|\langle \mathbf{r}_{t-1}, \boldsymbol{\varphi}_\omega \rangle|$  e iterar.

Para melhor compreender esse procedimento, analisemos o seguinte exemplo. Seja  $t = 3$ . O referido *loop* irá inicialmente calcular o produto interno entre  $\mathbf{r}_2$  e  $\boldsymbol{\varphi}_1$ . Em seguida, irá realizar a mesma operação, mas com  $\boldsymbol{\varphi}_2$ , e assim sucessivamente, que todos os átomos do dicionário sejam testados. Suponhamos que, ao final do *loop*, tenhamos constatado que o átomo  $\boldsymbol{\varphi}_{56}$  tenha maximizado  $|\langle \mathbf{r}_2, \boldsymbol{\varphi}_\omega \rangle|$ . Assim,  $\lambda_t \leftarrow 56$ .

Na terceira etapa, ocorre a construção do vetor de coeficientes  $\mathbf{c}$  propriamente dita. Cada posição desse vetor corresponde a um átomo do dicionário. Por exemplo, a

primeira posição de  $\mathbf{c}$  (denotada por  $c_1$ ) corresponde ao átomo  $\varphi_1$  do dicionário, e assim por diante. Na primeira etapa, o vetor nulo foi atribuído a  $\mathbf{c}$ . Logo, inicialmente, todas as entradas de  $\mathbf{c}$  são iguais a zero. Na segunda etapa, determinamos o indexador  $\lambda_t$  do átomo que, para o valor de  $t$  naquela iteração, maximiza o valor de  $|\langle \mathbf{r}_{t-1}, \varphi_\omega \rangle|$ . Agora, o valor que quantifica o quanto  $\varphi_{\lambda_t}$  está correlacionado com o resíduo naquela iteração será atribuído à posição  $\lambda_t$  de  $\mathbf{c}$  (posição esta correspondente ao indexador  $\omega$  do átomo que maximizou  $|\langle \mathbf{r}_{t-1}, \varphi_\omega \rangle|$  naquela iteração). Assim, o vetor  $\mathbf{c}$  funciona como um vetor que armazena as informações dos átomos que contribuem de maneira mais significativa com a informação contida no resíduo em cada iteração.

Na quarta etapa, o resíduo será atualizado. Após a realização da terceira etapa, sabemos qual é o átomo que, naquela iteração, tem maior correlação com o resíduo, e temos essa informação armazenada em  $\mathbf{c}$ . Uma vez que detemos essa informação, ela passa a ser irrelevante para o procedimento. Eliminamos, pois, a componente do sinal (i.e., a parcela da sua informação) na direção de  $\varphi_{\lambda_t}$ . Essa eliminação resulta em um novo resíduo, que será utilizado na próxima iteração do procedimento. Na quinta etapa, ocorre tão somente o incremento do indexador de *loop*.

Na sexta etapa, será decidido, baseado no critério de parada especificado, se o procedimento continuará sendo executado, ou se o mesmo deve ser interrompido naquele ponto. Apesar da sua aparente trivialidade, essa etapa é fundamental para o procedimento em estudo. Como dito anteriormente, o MP pode não oferecer soluções aproximadamente ótimas para problemas de aproximação esparsa. Evidentemente, o critério de parada depende do problema a ser abordado, bem como do grau de precisão desejado para uma possível solução do mesmo. Dessa forma, deve-se atentar para a possibilidade de não convergência do algoritmo (caso especifique um erro mínimo como critério de parada) e a possibilidade de encontrar uma solução extremamente discrepante daquela esperada (caso especifique um certo número de *loops* como critério de parada).

### 2.3.5 Considerações sobre o MP

De forma resumida, o MP busca construir, a partir de sinais mais simples (i.e., átomos), uma aproximação para o sinal original. Essa aproximação será tão melhor quanto mais completo e menos redundante for o dicionário.

Por fim, destacamos uma deficiência computacional do MP. A primeira é que ele

pode selecionar o mesmo indexador diversas vezes, caso o dicionário não seja ortonormal. Isso se dá pelo fato de que o produto interno entre o átomo e o resíduo não leva em consideração as contribuições dos outros átomos para o resíduo.

O fato supracitado possui um impacto consideravelmente negativo na eficiência do MP. Isso porque, quando da resolução de problemas mais robustos, que envolvam sinais mais complexos e dicionários maiores e mais redundantes (i.e., linearmente dependentes), o MP gasta uma grande quantidade de tempo computacional na resolução da segunda etapa.

### 2.3.6 Detalhamento do OMP

Tal qual foi feito com o MP primeiramente vamos expor uma descrição formal do procedimento MP, e em seguida analisaremos cada uma das etapas do mesmo detalhadamente. Inicialmente, vamos estabelecer um critério de parada e fixar um dicionário  $\mathcal{D}$ .

- Entrada: um sinal  $\mathbf{s} \in \mathbb{C}^d$ ;
- Saída: um vetor de coeficientes  $\mathbf{c} \in \mathbb{C}^\Omega$ ;
- Etapas:
  1. Faça  $\Lambda_0 = \emptyset$ ,  $\mathbf{r}_0 \leftarrow \mathbf{s}$  e  $t \leftarrow 1$ .
  2. Determine um índice  $\lambda_t$  tal que

$$\lambda_t \in \arg \max_{\omega} |\langle \mathbf{r}_{t-1}, \boldsymbol{\varphi}_{\omega} \rangle|.$$

3. Atualize o conjunto de índices:

$$\Lambda_t \leftarrow \Lambda_{t-1} \cup \{\lambda_t\}.$$

4. Determine a solução  $\mathbf{c}$  para o problema de mínimos quadrados:

$$\min_{\mathbf{c} \in \mathbb{C}^{\Lambda_t}} \left\| \mathbf{s} - \sum_{j=1}^t \mathbf{c}(\lambda_j) \boldsymbol{\varphi}_{\lambda_j} \right\|_2.$$

5. Calcule o novo resíduo por meio do uso dos coeficientes determinados na etapa anterior:

$$\mathbf{r}_t \leftarrow \mathbf{s} - \sum_{j=1}^t \mathbf{c}(\lambda_j) \varphi_{\lambda_j}.$$

6. Incremente o indexador de *loop*:

$$t \leftarrow t + 1.$$

7. Se o critério de parada ainda não foi atingido, retorne à segunda etapa.

A primeira etapa do OMP é análoga àquela do MP, sendo a etapa onde as "condições iniciais" do problema serão especificadas. Entretanto, não fazemos  $\mathbf{c} \leftarrow 0$ , mas sim  $\Lambda_0 = \emptyset$ . A importância desse fato ficará evidente mais adiante. A segunda etapa do OMP é idêntica àquela do MP.

Na terceira etapa reside grande parte da diferença entre o MP e o OMP no que concerne eficiência computacional. O conjunto  $\Lambda_t$ , ao qual inicialmente foi atribuído o conjunto vazio, é denominado conjunto de indexadores. Isso porque ele será utilizado para armazenar cada um dos indexadores obtidos na segunda etapa. Nesta, o átomo  $\varphi_{\lambda_t}$  mais correlacionado com o resíduo naquela iteração é identificado, e seu indexador é armazenado em  $\Lambda_t$ . Isso é feito por meio da operação de união entre  $\Lambda_{t-1}$  (o conjunto de indexadores armazenados em  $t - 1$ ) e  $\{\lambda_t\}$  (o conjunto formado exclusivamente pelo indexador  $\lambda_t$  obtido em  $t$ ). Essa união resulta no novo conjunto  $\Lambda_t$  que contém o indexador de todos os átomos mais correlacionados com o sinal determinados até o presente instante.

A quarta etapa envolve um problema de mínimos quadrados. Ele consiste em um problema de minimização da norma euclidiana do vetor  $\mathbf{s} - \sum_{j=1}^t \mathbf{c}(\lambda_j) \varphi_{\lambda_j}$ . O somatório representa a combinação linear de átomos do dicionário multiplicados pelos seus respectivos elementos de ponderação (determinados na segunda etapa). Temos, ainda, que  $\mathbf{c} \in \mathbb{C}^{\Lambda_t}$ . Ou seja, estamos buscando o vetor de coeficientes  $\mathbf{c}$  cujas entradas não-nulas tenham como indexador elementos de  $\Lambda_t$ .

A quinta etapa do OMP é análoga àquela do MP, sendo nela onde ocorre a atualização do resíduo. O somatório nada mais é que a combinação linear de todos átomos cujos indexadores foram armazenados em  $\Lambda_t$  até o presente momento. Os coeficientes dessa operação são justamente as entradas do vetor  $\mathbf{c}$  encontrado anteriormente. Após a construção da referida combinação, subtraímos a mesma do sinal original  $\mathbf{s}$ . Ou seja,

eliminamos do sinal original a contribuição dos átomos mais correlacionados com o mesmo encontrados até o momento. O resultado dessa subtração é o novo resíduo, que será utilizado posteriormente. A sexta e a sétima etapas do OMP são análogas à quinta e à sétima do MP, respectivamente.

## 2.4 Matrizes de Hankel

Seja o vetor de sinal ruidoso

$$\mathbf{x} = [x_1 \ x_2 \ \dots \ x_N]^T$$

composto por  $N$  amostras. Se considerarmos que o ruído presente em  $\mathbf{x}$  é aditivo e não possui correlação com o sinal propriamente dito, podemos reescrever  $\mathbf{x}$  como

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{n},$$

onde  $\bar{\mathbf{x}}$  representa a componente do sinal e  $\mathbf{n}$ , o ruído.

A partir de  $\mathbf{x}$ , pode-se construir a seguinte matriz de Hankel  $\mathbf{H}_{L \times K}$  tal que  $K + L = N + 1$  e  $L \geq M$ :

$$\mathbf{H} = \begin{bmatrix} x_1 & x_2 & \dots & x_{K-1} & x_K \\ x_2 & x_3 & \dots & x_K & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_L & x_{L+1} & \dots & x_{N-1} & x_N \end{bmatrix}.$$

A matriz  $\mathbf{H}$  construída, também conhecida como matriz de trajetória, é amplamente utilizada no âmbito Análise Espectral Singular (*Singular Spectrum Analysis - SSA*) para, dentre outras finalidades, promover a extração de sinal, atenuação de ruído e filtragem.

É importante destacar a relação entre as matrizes de Hankel e as de Toeplitz, cujo uso é amplamente difundido no meio científico. Seja  $\mathbf{J}_n$  uma matriz de reflexão, isto é, tal que  $\mathbf{J}_{i,j} = 1$ ,  $j = n - i + 1$  e  $\mathbf{J}_{i,j} = 0$  caso contrário. Se  $\mathbf{H}_{L \times K}$  é uma matriz de Hankel, então  $\mathbf{H}_{L \times K} = \mathbf{T}_{L \times K} \mathbf{J}_K$ , onde  $\mathbf{T}_{L \times K}$  é uma matriz de Toeplitz.

## 3 Materiais e Métodos

### 3.1 Método Implementado

Como dito anteriormente, o método proposto neste trabalho consiste em uma ferramenta robusta e não-paramétrica para detecção espectral e identificação de uso de canal. Ele consiste essencialmente em dois estágios, denominados modelamento esparso e identificação de uso de canal. O primeiro utiliza técnicas de atenuação de ruído, juntamente com o OMP (*Orthogonal Matching Pursuit*) para obter um modelo esparso para o espectro da banda UHF. O segundo, por sua vez, tem como escopo identificar as bandas e os canais ocupados, bem como os usuários primários e secundários realizando essa ocupação.

#### 3.1.1 Modelamento Esparso

Inicialmente, são coletadas amostras da banda UHF, cada qual contendo  $n_{\text{entradas}}$  elementos (majoritariamente não nulos, devido à presença de ruído). Essas amostras, então, são armazenadas nas colunas da matriz  $\mathbf{M}_{\text{amostras}}$ . Esse processo é repetido até que  $\mathbf{M}_{\text{amostras}}$  possua  $N$  vetores coluna de amostras. É imediato, pois, que os valores de  $n_{\text{entradas}}$  e  $N$  impactam tanto a resolução das amostras coletadas como o custo computacional do método.

Depois da construção de  $\mathbf{M}_{\text{amostras}}$ , calcula-se a matriz  $\mathbf{V}_{\text{média}}$ , que consiste na média aritmética das colunas de  $\mathbf{M}_{\text{amostras}}$ . Essa abordagem foi adotada com o objetivo de reduzir o efeito de AWGN. Em seguida, uma matriz de Hankel  $\mathbf{H}$  é criada a partir de  $\mathbf{V}_{\text{média}}$ , e seu SVD é calculado. A distância entre os valores singulares de  $\mathbf{H}$  é, então, utilizada para determinar o *rank* da aproximação de *rank*  $k$  de  $\mathbf{H}$ ,  $\mathbf{H}_k$ , a partir da qual a matriz  $\mathbf{V}_{\text{filtrada}}$ , que consiste na média aritmética filtrada das colunas de  $\mathbf{M}_{\text{amostras}}$ , é obtida. Esse procedimento mostrou-se muito eficaz para a redução dos efeitos do ruído AWGN presente em  $\mathbf{V}_{\text{média}}$ , melhorando ainda mais o desempenho do método em cenários espectrais com SNR crítica.

Uma vez que  $\mathbf{V}_{\text{filtrada}}$  é obtida, o dicionário de rádio-frequências (RF)  $\mathbf{D}_{\text{RF}}$  é criado. As colunas de  $\mathbf{D}_{\text{RF}}$  consistem em pares alternados de senos e cossenos. Cada par possui uma frequência pertencente à banda UHF, e as frequências dos pares são igualmente espaçadas entre si.  $\mathbf{V}_{\text{filtrada}}$  e  $\mathbf{D}_{\text{RF}}$  são utilizadas juntamente com o OMP para criar

o modelo esparso para  $V_{filtrada}$ , denominado  $V_{esparsa}$ . Para tanto, o OMP realiza a projeção de  $V_{filtrada}$  sobre os vetores coluna de  $D_{RF}$ , com o objetivo de determinar a maior correlação (no sentido do produto escalar) em cada iteração. Os resultados são, então, utilizados para identificar quais frequências do  $D_{RF}$  mais provavelmente pertencem a um canal ocupado. O critério de parada do OMP é a esparsidade  $n_S$  desejada para o modelo esparso, que consiste em um inteiro positivo correspondente ao arredondamento de uma fração de  $n_{entradas}$ . Após  $V_{esparsa}$  ser obtida, o método calcula sua transformada de Fourier (FT) por meio do uso da transformada rápida de Fourier (FFT), obtendo, assim, a representação esparso do espectro no domínio da frequência. Todos os procedimentos descritos encontram-se resumidos no fluxograma da Figura 2.

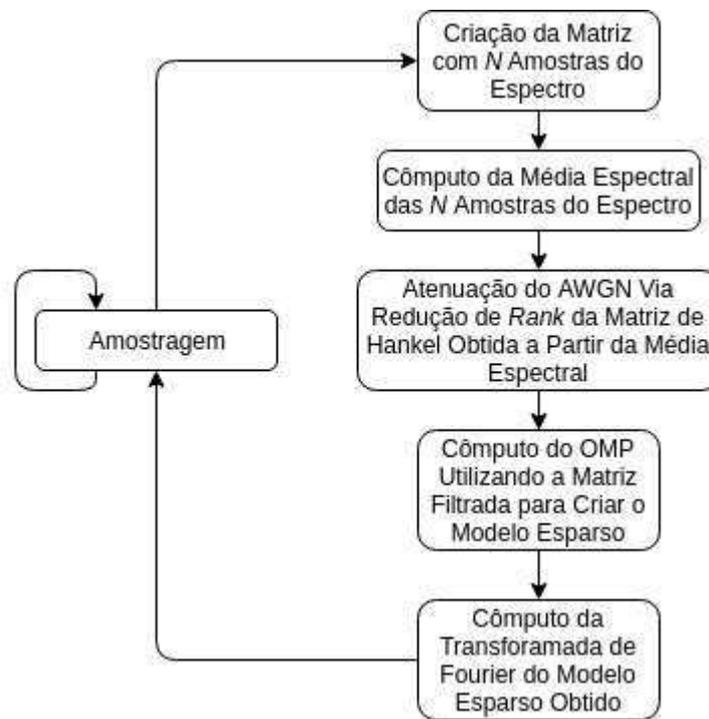


Figura 2: Fluxograma do primeiro estágio do método implementado

### 3.1.2 Identificação de Uso de Canal

Como foi dito anteriormente, esse estágio tem como objetivo identificar as bandas de frequência e os canais do espectro de interesse que se encontram ocupados, bem como identificar os usuários primários e secundários responsáveis por essa ocupação. Para realizar essa tarefa, o método precisa de vários modelos esparsos do espectro, para que ele possa compreender como suas frequências são utilizadas pelos usuários. Dessa forma, o primeiro estágio do método é repetido  $n_{sm}$ . Em cada repetição, um modelo esparso

para o espectro é criado a partir de  $N$  amostras e, então, armazenado nas colunas da matriz  $\mathbf{M}_{sm}$ . Esse processo é repetido até que  $\mathbf{M}_{sm}$  contenha  $n_{sm}$  colunas. Os elementos não-nulos destas correspondem às frequências ocupadas, e mostram como o espectro da banda UHF estava sendo utilizado durante o processo de amostragem. Entradas dos modelos esparsos associadas com frequências que se encontravam constantemente ocupadas durante o processo de amostragem irão possuir valores majoritariamente não-nulos. Por sua vez, entradas dos modelos esparsos associadas com frequências que se encontravam apenas momentaneamente ocupadas durante o processo de amostragem irão possuir valores majoritariamente nulos. Para melhor compreender essas observações, consideremos

$$M_{sm} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ a_L & b_L & c_L & \dots & d_L & e_L & f_L \\ a_{L+1} & b_{L+1} & c_{L+1} & \dots & d_{L+1} & e_{L+1} & f_{L+1} \\ a_{L+2} & b_{L+2} & c_{L+2} & \dots & d_{L+2} & e_{L+2} & f_{L+2} \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & b_K & 0 & \dots & 0 & e_K & 0 \\ 0 & b_{K+1} & 0 & \dots & 0 & e_{K+1} & 0 \\ 0 & b_{K+2} & 0 & \dots & 0 & e_{K+2} & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}.$$

Como a maioria das entradas entre as linhas  $L$  e  $L + 2$  são não-nulas, supõe-se que as frequências a elas relacionadas são ocupadas por um usuário primário. A maioria das entradas entre as linhas  $K$  e  $K + 2$ , por sua vez, são não-nulas apenas em algumas colunas. Dessa forma, supõe-se que as frequências a elas relacionadas são ocupadas por um usuário secundário. Essas suposições são quantificadas por um parâmetro denominado persistência, que é associado à cada linha de  $M_{sm}$  (i.e., a cada frequência da banda UHF considerada), conforme mostrado na equação abaixo. A persistência de cada linha é armazenada na matriz  $P_{er}$ , e é igual à quantidade de elementos não-nulos de cada linha de  $\mathbf{M}_{sm}$ , denominada  $\alpha_i$ ,  $i = 1, \dots, n_{entradas}$ , dividida por  $n_{sm}$ :

$$\mathbf{P}_{er} = \left[ P_{er_i} \right], \quad P_{er_i} = \frac{\alpha_i}{n_{sm}}, \quad \text{onde } 0 \leq \alpha_i \leq n_{sm} \therefore 0 \leq P_{er_i} \leq 1.$$

Se  $P_{er_i} \approx 1$ , é muito provável que a  $i$ -ésima frequência esteja sendo utilizada por um usuário primário, visto ela está sendo identificada como estando ocupada durante a maior parte do processo de amostragem. Em contrapartida, quanto menor for o valor de  $P_{er_i}$ , mais provável será a possibilidade de que a  $i$ -ésima frequência esteja sendo utilizada por um usuário secundário, uma vez que ela está sendo identificada como estando ocupada apenas em parte do processo de amostragem. Por fim, se  $P_{er_i} = 0$ , é muito provável que a frequência esteja disponível. Defini-se, então, a persistência de banda  $P_{er_B}$  como

$$P_{er_B} = P_{er_{max}},$$

onde  $P_{er_{max}}$  é o maior valor de  $P_{er_i}$  dentre as frequências que a compõe. É imediata, pois, a necessidade de um limiar  $\zeta$  para determinar o quanto  $P_{er_B}$  de uma banda deve estar próxima de 1 para que ela seja considerada como ocupada por um espaço primário. Outros parâmetros das bandas ocupadas também são obtidos de forma a aumentar a confiabilidade do processo de detecção, como, por exemplo, a energia bandas, sua largura de faixa e sua componente de frequência mais energética, denominada  $\mathcal{P}$ . O primeiro passo para o processo de identificação de uso de canal é identificar as bandas para as quais  $P_{er_B} \geq \zeta$ , visto que a possibilidade das mesmas estarem sendo ocupadas por usuários primários é maior. O segundo passo é calcular a média aritmética das componentes de frequência mais energéticas dessas bandas, denominada  $\mathcal{P}_{média}$ , e atribuir à cada banda o valor

$$\beta = \frac{|\mathcal{P}_{média} - \mathcal{P}|}{\mathcal{P}_{média}}.$$

O valor de  $\beta$  indica o quão longe o valor  $\mathcal{P}$  da banda se encontra de  $\mathcal{P}_{média}$ . Novamente, o estabelecimento de um limiar  $\xi$  mostra necessário para que se possa avaliar o quão próximo de  $\mathcal{P}_{média}$  o do  $\mathcal{P}$  da banda deve estar para que ela seja considerada como ocupada por um espaço primário. O objetivo principal dessa abordagem é levar em consideração o efeito da atenuação dos sinais. Bandas de frequências ocupadas que satisfaçam  $P_{er_B} \geq \zeta$  e  $\beta \leq \xi$  são consideradas como sendo ocupadas por usuários primários. Supõe-se, então, que as demais bandas ocupadas são ocupadas por usuários secundários.

Uma vez que o processo de identificação é concluído pela primeira vez, o primeiro estágio do método é repetido, de forma a criar um novo modelo esparso para o espectro. Este é, então, utilizado para atualizar a matriz  $\mathbf{M}_{sm}$ . Esse processo, ilustrado na Figura 3, consiste na remoção da coluna mais à esquerda de  $\mathbf{M}_{sm}$ , correspondente ao modelo esparso mais antigo, e na concatenação do novo modelo, que passa a corresponder à coluna mais à direita de  $\mathbf{M}_{sm}$ .

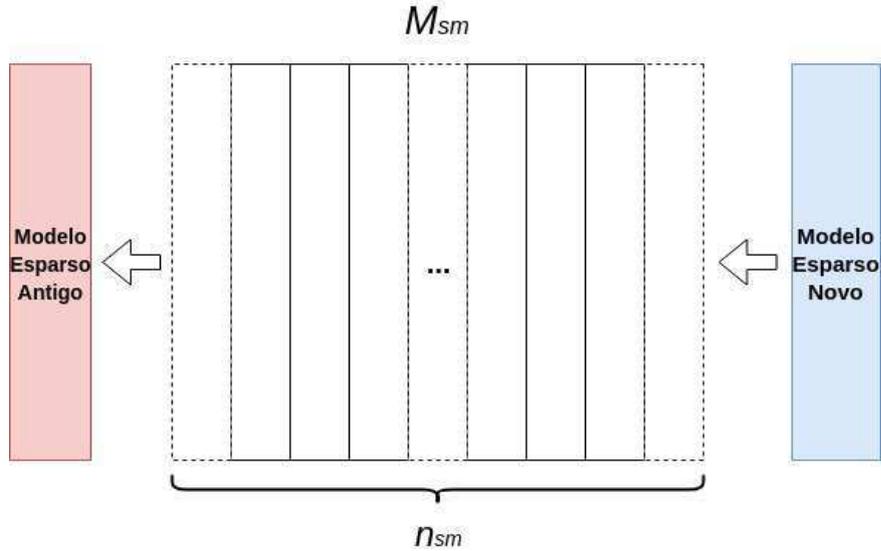


Figura 3: Processo de atualização de  $\mathbf{M}_{sm}$

Após concluído o processo de atualização, o processo de identificação de uso de canal é repetido utilizando a matriz  $\mathbf{M}_{sm}$  atualizada. Esse procedimento possibilita ao método utilizar os  $n_{sm}$  modelos esparsos mais recentemente obtidos no processo de identificação. Caso um usuário primário interrompa sua transmissão, os valores de  $P_{er_i}$  associados com as frequências por ele ocupadas irão diminuir cada vez que  $\mathbf{M}_{sm}$  for atualizada. Quando  $P_{er_B} \leq \zeta$ , o canal não mais será considerado como sendo ocupado por um usuário primário. Por outro lado, caso um usuário primário inicie uma transmissão nesse canal, os valores de  $P_{er_i}$  associados com as frequências por ele ocupadas irão aumentar cada vez que  $\mathbf{M}_{sm}$  for atualizada. Quando  $P_{er_B} \geq \zeta$ , o canal será considerado como estando ocupado por um usuário primário. Caso um usuário secundário tenha  $P_{er_B} \leq \zeta$ , a condição  $\beta \leq \xi$  será utilizada para distingui-lo dos usuários primários. Todos os procedimentos descritos encontram-se resumidos no fluxograma da Figura 4.

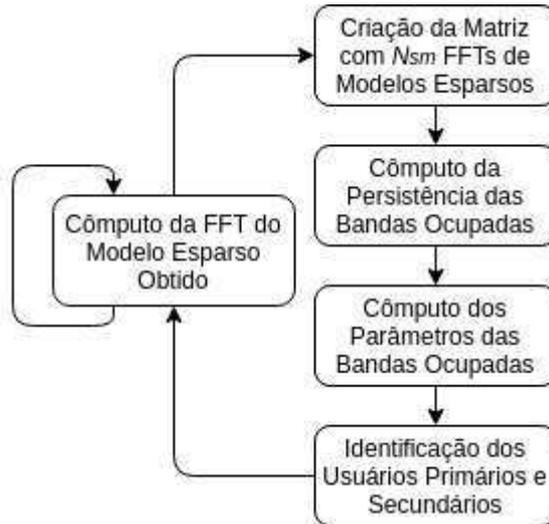


Figura 4: Fluxograma do segundo estágio do método implementado

### 3.2 Simulações

O estudo foi realizado por meio de simulações em computador, nas quais foi considerado um conjunto de usuários primários e secundários transmitindo e uma certa quantia de canais vazios, levando-se os efeito do ruído gaussiano e da atenuação dos sinais. As simulações foram realizadas por meio da implementação de um algoritmo no *software* MATLAB<sup>®</sup>, em um computador Dell<sup>®</sup> Inspiron 14R 5437-A20 com processador Intel<sup>®</sup> Core<sup>™</sup> i7-4500U CPU @ 1,80GHz × 4. A realização dessas simulações teve como objetivo primário averiguar o impacto da aplicação das decomposições matriciais estudadas para atenuação do ruído gaussiano no algoritmo de detecção do espectro simulado.

Os parâmetros, os limiares e as dimensões das matrizes foram determinados empiricamente. As amostras utilizadas nas simulações possuíam  $n_{entries} = 8192$  elementos majoritariamente não-nulos. As amostras foram contaminadas com AWGN de maneira que a SNR fosse igual a -40 dB. Essa contaminação foi realizada por meio do uso da função *awgn*, do MATLAB<sup>®</sup>. Essa função recebe como argumentos o sinal a ser contaminado, a potência desse sinal em dBW, e o valor da SNR desejada. A partir desses argumentos, ela produz um sinal contaminado com ruído gaussiano branco com a SNR desejada. O espectro de uma dessas amostras pode ser visto na Figura 5, e mostra o quão crítica era a SNR nos cenários espectrais simulados.

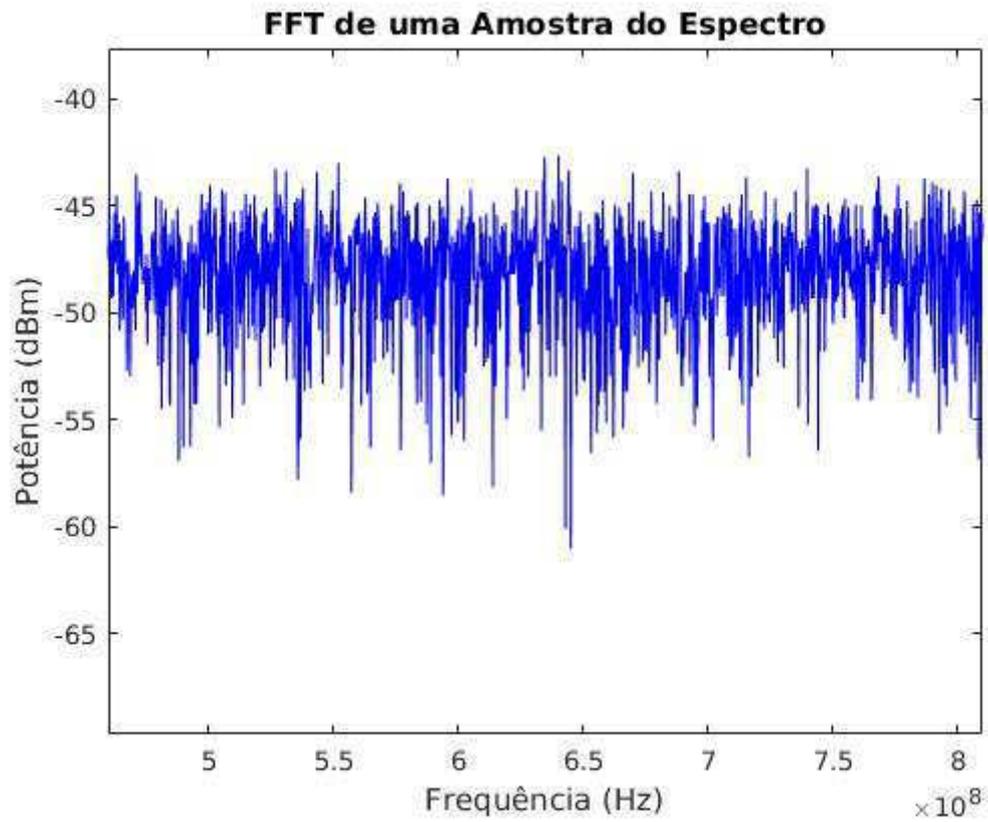


Figura 5: Espectro de uma amostra do espectro UHF

O cálculo da média espectral que resulta na matriz  $\mathbf{V}_{média}$  envolveu  $N = 10000$  amostras. O espectro de um exemplo de  $\mathbf{V}_{média}$  pode ser vista na Figura 6, e mostra o impacto positivo desse procedimento para a SNR.

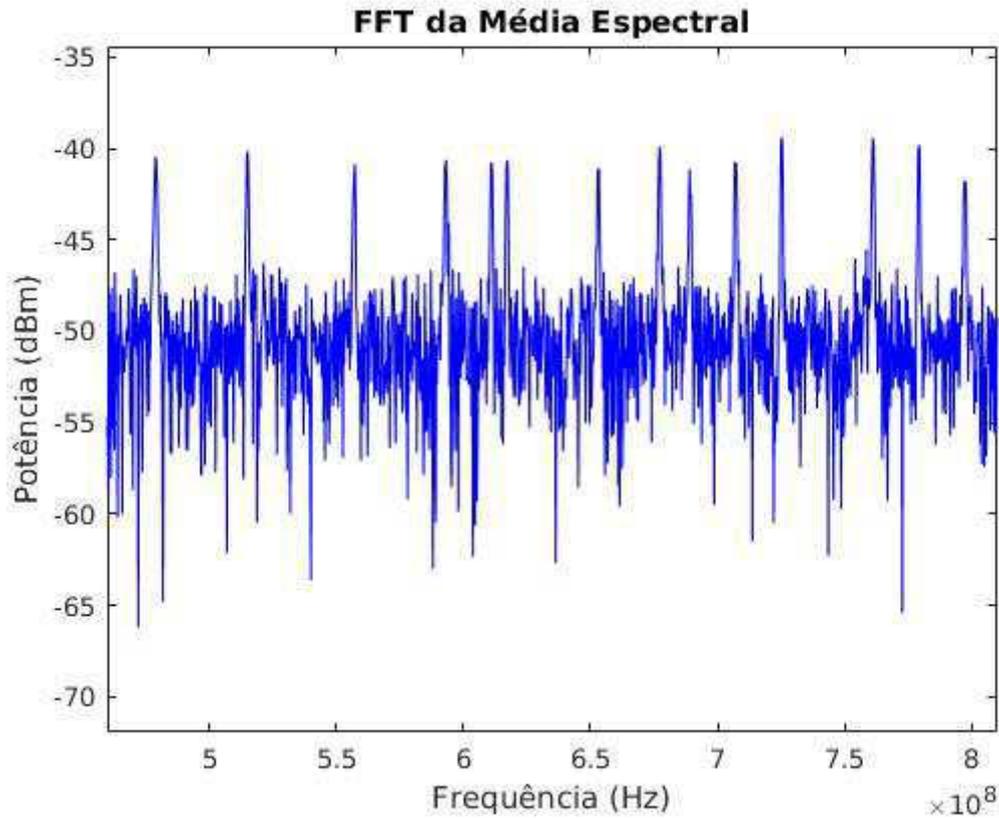


Figura 6: Espectro da média espectral das amostras coletadas

Em seguida, conforme explicitado anteriormente, uma atenuação ainda maior do AWGN presente em  $\mathbf{V}_{média}$  foi obtida por meio do uso da aproximação de *rank k* da matriz  $\mathbf{H}$ , resultando na matriz  $\mathbf{V}_{filtrada}$ . Na Figura 7, o eixo das ordenadas corresponde à magnitude da distância entre dois valores singulares vizinhos de  $\mathbf{H}$ , que compõe um par. O eixo das abscissas, por sua vez, corresponde ao índice do par (e.g., o primeiro par é constituído pelo primeiro e pelo segundo valores singulares de  $\mathbf{H}$ ). O gráfico evidencia o fato de que as magnitudes das distâncias entre os valores singulares de  $\mathbf{H}$  se torna menor à medida que o índice do par aumenta. Esse fato é utilizado para realizar o procedimento de filtragem, que consiste em eliminar as informações contidas nos valores singulares menos relevantes de  $\mathbf{H}$ . A FTT de um exemplo de  $\mathbf{V}_{filtrada}$  pode ser vista na Figura 8, e mostra o aumento da SNR devido ao uso desse procedimento.

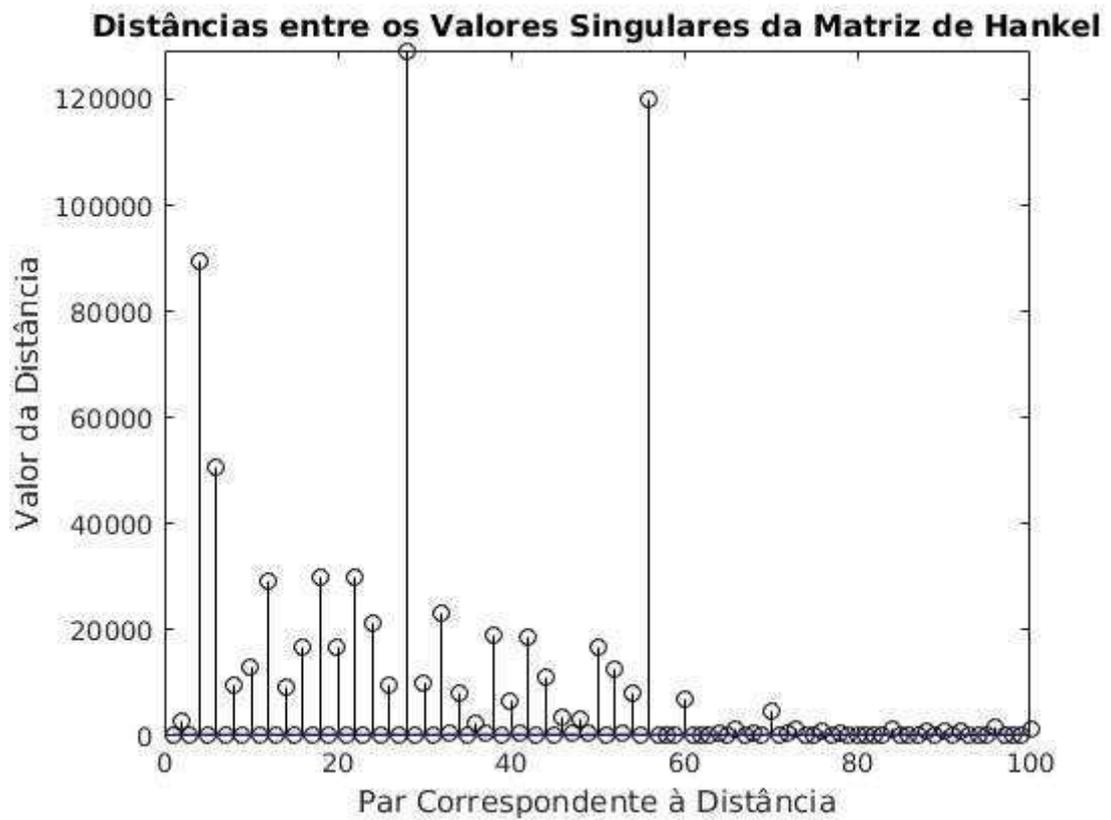


Figura 7: Magnitude das distâncias entre pares de valores singulares da matriz  $H$

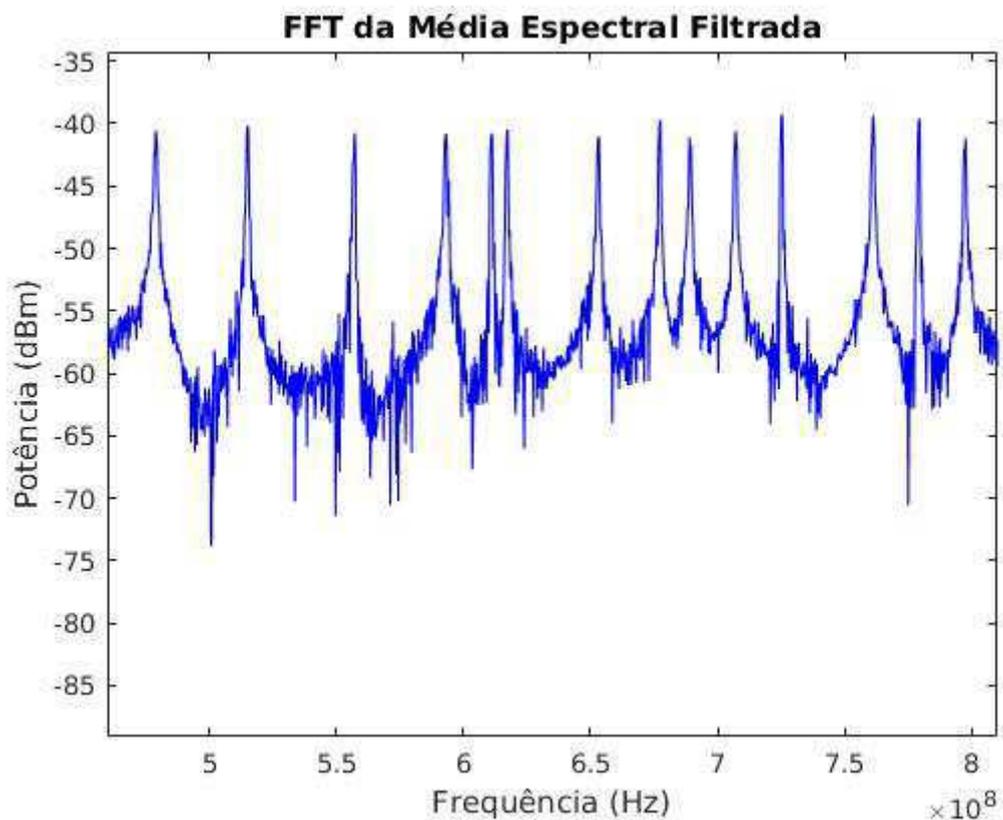


Figura 8: Espectro da média espectral após filtragem

O dicionário  $\mathbf{D}_{RF}$  criado possuía 1100 pares de senos e cossenos, cada qual separado por uma frequência de 0.3 MHz. O valor de  $n_S$  foi definido como sendo 82, o que significa que o modelo esparso criado pelo OMP possuía apenas 82 posições não-nulas (o que corresponde à aproximadamente 1% do valor de  $n_{entradas}$ ). A FTT de um exemplo de  $\mathbf{V}_{esparsa}$  pode ser visto na Figura 9. Esse procedimento é repetido até que a matriz  $\mathbf{M}_{sm}$  contenha  $n_{sm} = 200$  colunas. Os parâmetros das bandas dos modelos esparsos ocupadas são, então, determinados, e a identificação de uso de canais é realizada. Após o processo de identificação ser concluído, o primeiro e o segundo estágio do método são repetidos para atualizar as colunas da matriz  $\mathbf{M}_{sm}$  e continuamente identificar os usuários primários e secundários, bem como as bandas de frequência disponíveis.

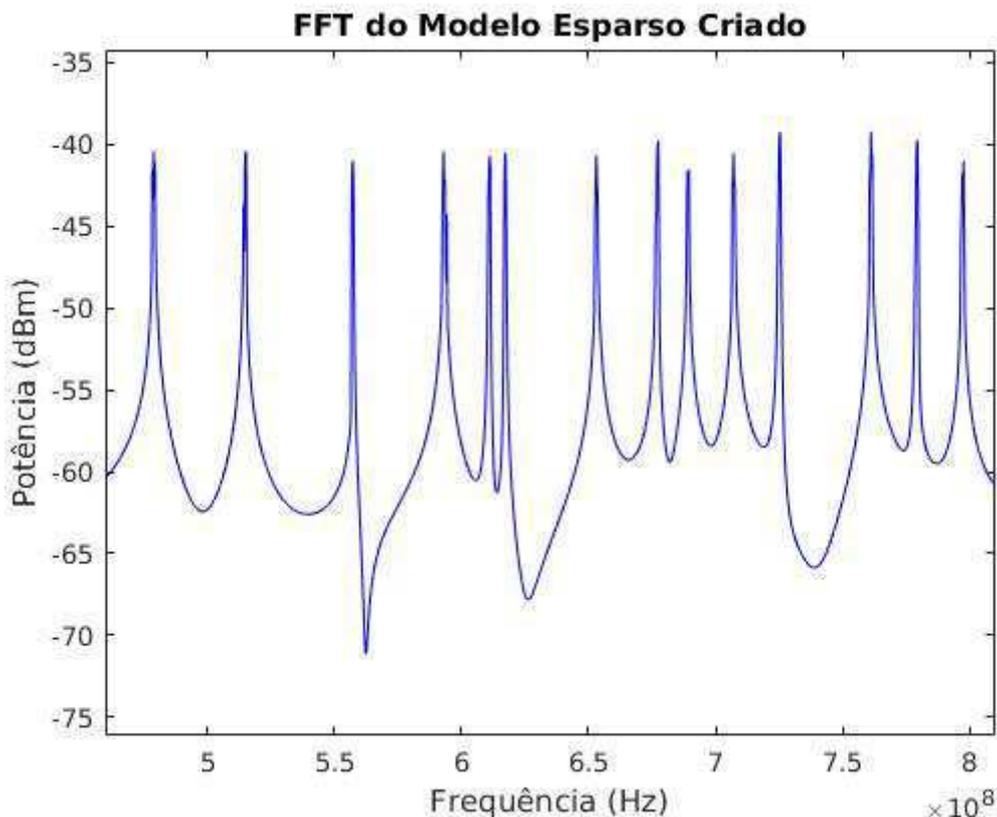


Figura 9: Espectro do modelo esparso construído

### 3.3 Experimentos

Foram realizados vários experimentos para obtenção de amostras da faixa UHF do espectro de frequências. Nestes, fez-se uso da placa USRP (*Universal Software Radio Peripheral*), produzida pela empresa Ettus<sup>®</sup>, bem como do *software* GNU Radio Companion 3.7.11.

No início de cada experimento, a placa USRP era alimentada e conectada à uma antena direcional adequada e ao computador utilizado. Em todos os experimentos, fez-se uso do bloco "UHD: USRP Source", nativo do GNU Radio, para configurar os parâmetros da placa utilizada. As Figuras 10 e 11 mostram, respectivamente, o início e o final do espectro do canal 21 da banda UHF, correspondente à faixa de frequências entre 512 e 518 MHz, atribuído à TV Paraíba (Globo). O sinal observado próximo à frequência 512 MHz é, provavelmente, um sinal de controle utilizado pelo sistema do usuário primário, visto que ele se encontra imediatamente acima da frequência inicial do canal. Essas imagens são um exemplo do espectro de trechos da banda UHF que puderam ser observados por meio do uso do arranjo experimental realizado e de blocos nativos do *software* GNU Radio.

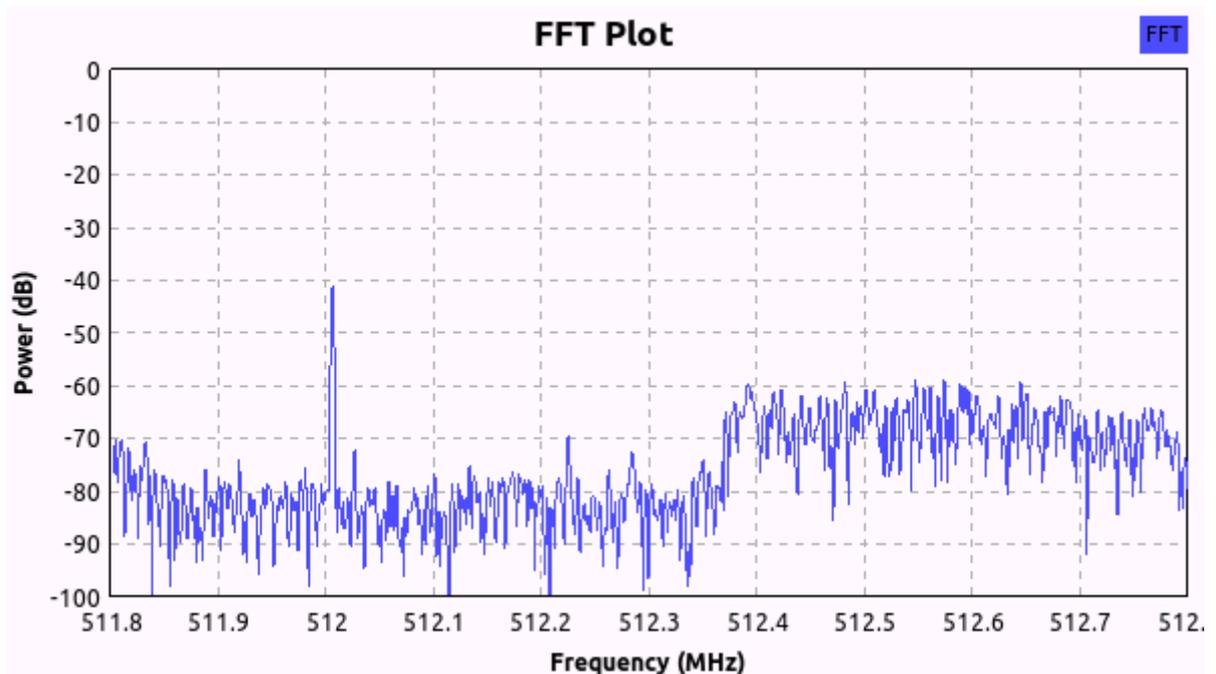


Figura 10: Espectro do início do canal 21 da banda UHF

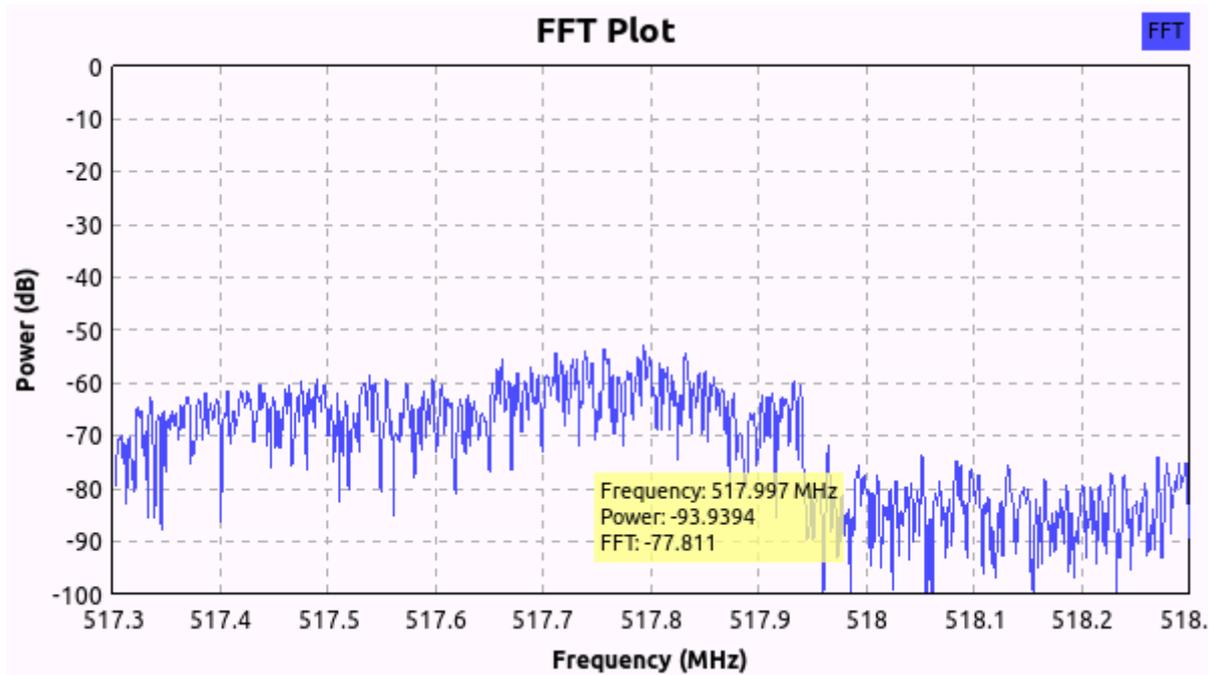


Figura 11: Espectro do final do canal 21 da banda UHF

Devido às limitações da placa e do *software* utilizados, os mesmos não eram capazes de abranger toda a banda UHF, visto que a mesma é muito larga. Dessa forma, foi feita uma busca por uma maneira de viabilizar a coleta de amostras dessa banda. Mais especificamente, foram pesquisadas formas de se modificar a frequência central da placa durante a execução do programa. Isso faria com que a banda UHF fosse varrida continuamente, obtendo uma amostra do espectro a cada varredura.

A alternativa estudada foi o uso dos blocos "Probe Signal" e "Probe Function". O primeiro amostra o sinal recebido na sua entrada, enquanto o segundo atribui esse valor a uma variável. Dessa forma, tentou-se utilizar esses blocos para, a partir de um sinal de controle, modificar o valor da frequência central da placa utilizada periodicamente. O sinal de controle utilizado consiste em um sinal escada com 165 degraus e período igual a 66 segundos, conforme ilustrado na Figura 12. Esse sinal consiste em um sinal de áudio no formato .wav, gerado pelo *software* MATLAB®.

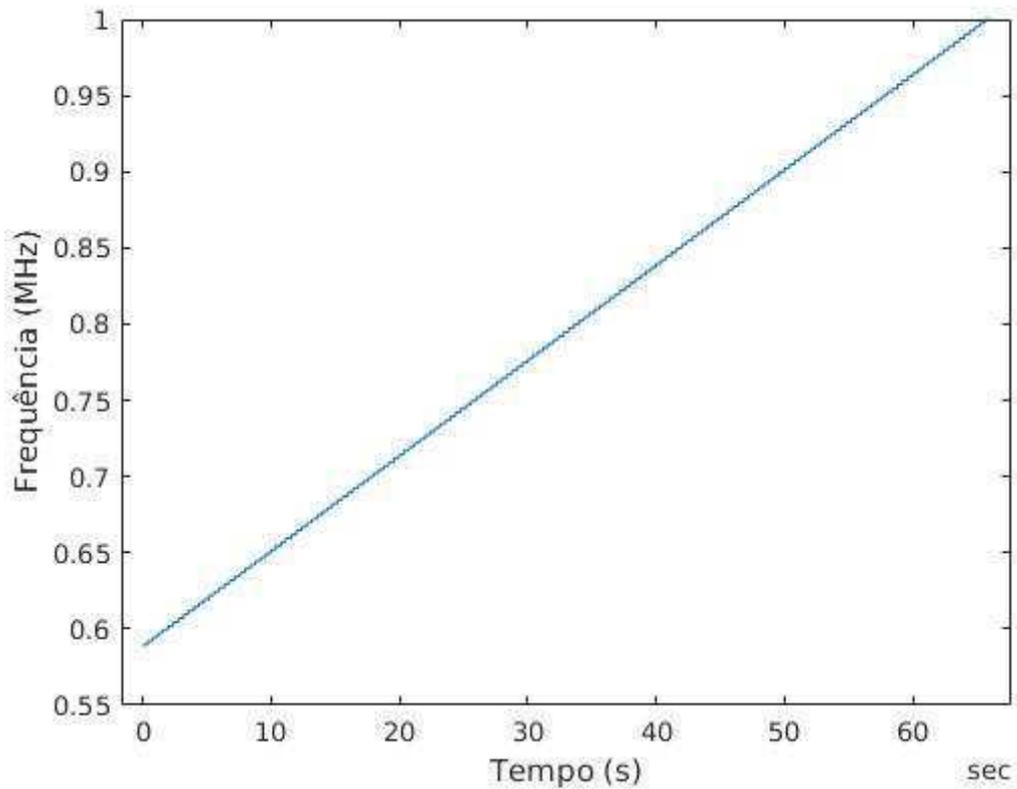


Figura 12: Sinal de controle gerado no *software* MATLAB®

No GNU Radio, esse sinal era gerado pelo bloco "Wav File Source" e, então, multiplicado pelo fator de reescalonamento  $799 \cdot 10^6$  conforme ilustrado na Figura 13. Dessa forma, cada degrau desse sinal possuía duração de  $1/66$  segundos, e correspondia à frequência central de uma faixa de frequências de 2 MHz da banda UHF.

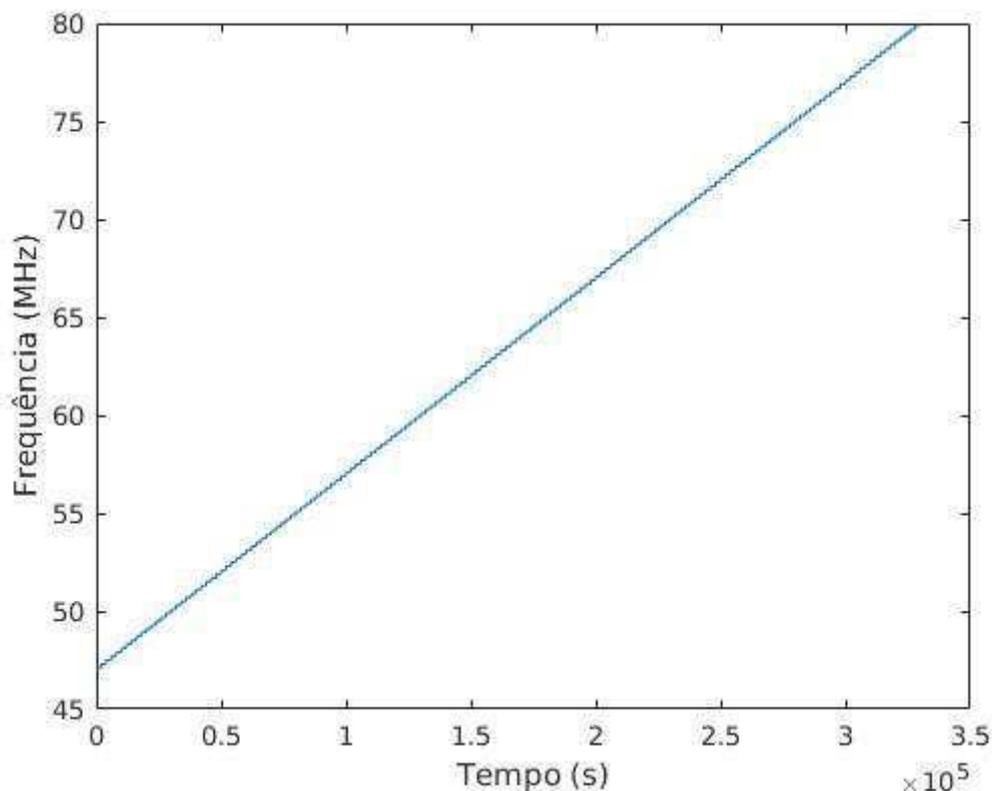


Figura 13: Sinal de controle utilizado no GNU Radio

Os parâmetros desse sinal foram especificados de forma que, em cada degrau do sinal (i.e., enquanto a frequência central da USRP fosse igual ao valor do degrau), fossem coletadas 1000 amostras daquela faixa. Dessa forma, em 66 segundos, seria obtido uma amostra da banda UHF total.

Entretanto, a implementação desse procedimento falhou. Isso porque, para que o procedimento descrito pudesse ser realizado, o bloco "Probe Signal" deveria estar em sincronia com o sinal de controle utilizado. Entretanto, apesar de não o ser especificado na sua documentação, o bloco "Probe Signal" é assíncrono ao fluxo de dados do programa. Dessa forma, o mesmo não amostrava o sinal de controle de forma adequada, o que inviabilizou o controle correto da frequência central da placa USRP.

Um dos maiores obstáculos enfrentados nos experimentos foi a documentação insuficiente e, não raramente, precária dos blocos do *software* utilizado. Como o GNU Radio não foi projetado tendo em vista a modificação automática de variáveis durante sua execução, a correta documentação de blocos que a possibilitem era crucial para a obtenção das amostras desejadas. Dessa forma, a documentação deficiente prejudicou enormemente os resultados dos esforços feitos.

Uma alternativa considerada foi a criação de blocos OOT *Out of tree*, capazes de realizar o processo de amostragem da banda UHF. Entretanto, devido às limitações de tempo, essa alternativa não foi explorada.

## 4 Resultados

Para uma SNR de -40 dB, que consiste em um cenário com SNR crítica, o método desenvolvido é capaz de identificar corretamente as bandas de frequência ocupadas por usuários primários com erro praticamente nulo. Isso é fundamental para um método dessa natureza, visto que os usuários primários possuem, por definição, uma maior prioridade de acesso às frequências dos canais a eles concedidos. Aproximadamente 67% das demais bandas ocupadas foi identificada corretamente, ocorrendo erros quando as componentes de ruído eram erroneamente identificadas como usuários secundários. Apesar dos procedimentos de atenuação de ruído realizados para evitar a ocorrência de erros dessa natureza, eles ainda mostrou-se presente.

Como esperado, para cenários com valores ainda mais críticos de SNR, o desempenho do método diminuiu. Para uma SNR de -45dB, o método proposto identifica erroneamente 7% dos usuários primários. Esse resultado, apesar de inviabilizar o uso do método em aplicações práticas, ainda é relativamente positivo, visto que a SNR foi reduzida enquanto todos os demais parâmetros foram mantidos constantes. Para uma SNR de -50dB, o método proposto identifica erroneamente 23% dos usuários primários.

A inadequação dos blocos nativos do *software* GNU Radio para os experimentos realizados inviabilizou a obtenção de amostras reais da banda UHF em tempo hábil. Entretanto, os resultados positivos obtidos para os cenários com SNR crítica simulados sugerem que o método obteria resultados satisfatórios a partir das amostras coletadas durante os experimentos.

## 5 Conclusão

Considerando que o uso das frequências da banda UHF no Brasil é feito de maneira esparsa, foi proposto um método de detecção espectral capaz de tirar proveito dessa característica. Esse método foi criado com o objetivo de realizar a detecção do

espectro em cenários espectrais com SNR crítica, levando em consideração o efeito de atenuação dos sinais.

Diferente de outros métodos propostos, o método proposto neste trabalho é não-paramétrico, robusto frente a cenários com SNR crítica e possui baixa complexidade computacional. Ele também não requer nenhum conhecimento prévio do espectro (e.g., canais atribuídos a usuários primários em uma dada região).

Como proposta para trabalhos futuros, está a realização de novos experimentos utilizando o *software* GNU Radio para coletar amostras da banda UHF. Ao longo destes, recomenda-se a criação de blocos OOT adequados para a realização dos experimentos, uma vez que os blocos nativos do *software* utilizado não foram adequados para a mesma.

## Referências

- [1] SHELLHAMMER, Stephen J. *Spectrum sensing in IEEE 802.22. IAPR Wksp. Cognitive Info. Processing*, Páginas 9-10, 2008.
- [2] HAYKIN, Simon. *Cognitive radio: brain-empowered wireless communications. IEEE journal on selected areas in communications*, Volume 23, Número 2, Páginas 201-220, 2005.
- [3] YUCEK, Tevfik; ARSLAN, Huseyin. *A survey of spectrum sensing algorithms for cognitive radio applications. IEEE communications surveys & tutorials*, Volume 11, Número 1, Páginas 116-130, 2009.
- [4] ARJOUNE, Youness; et al. *Spectrum sensing: Enhanced energy detection technique based on noise measurement. 2018 IEEE 8th Annual Computing and Communication Workshop and Conference*, Páginas 828-834, 2018.
- [5] JENSEN, Søren Holdt; et al. *Reduction of broad-band noise in speech by truncated QSVD. IEEE Transactions on Speech and Audio Processing*, Volume 3, Número 6, Páginas 439-448, 1995.
- [6] NGUYEN, Lam H.; TRAN, Trac D. *Efficient and robust RFI extraction via sparse recovery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Volume 9, Número 6, Páginas 2104-2117, 2016.
- [7] MALLAT, Stéphane; ZHANG, Zhifeng. *Matching pursuit with time-frequency dictionaries. IEEE Transactions on signal processing*, Volume 41, Número 12, Páginas 3397-3415, 1993.
- [8] MITOLA, Joseph; MAGUIRE, Gerald Q. *Cognitive radio: making software radios more personal. IEEE personal communications*, Volume 6, Número 4, Páginas 13-18, 1999.
- [9] TROPP, Joel Aaron: *Topics in sparse approximation*. 2004. University of Texas, Austin, 2004.

- [10] CAI, T Tony; WANG, Lie. *Orthogonal matching pursuit for sparse signal recovery with noise*. In: *IEEE Transactions on Information Theory*, Volume 57, Número 7, Páginas 4680-4688, 2011.
- [11] CHANET, Reinaldo and et al. *Análise de modelos de regressão linear com aplicações*: 1. Ed. Campinas: Editora UNICAMP, 1999.
- [12] ANATEL Agência Nacional de Telecomunicações. Disponível em: <http://www.anatel.gov.br/institucional/>. Acesso em: 20 de março de 2019.

# Apêndice

## Apêndice A - Teoremas e Definições Básicos

**Definição 7** Seja  $\mathbf{A} \in \mathbb{K}^{m \times n}$ . Os vetores linha da matriz  $\mathbf{A}$  estão em  $\mathbb{K}^n$  e são formados pelas linhas de  $\mathbf{A}$ . Os vetores coluna de  $\mathbf{A}$ , por sua vez, estão em  $\mathbb{K}^m$  e são formados pelas colunas de  $\mathbf{A}$ .

**Definição 8** O espaço linha de  $\mathbf{A}$  é o subespaço de  $\mathbb{R}^n$  gerado pelos vetores linha de  $\mathbf{A}$ .

**Definição 9** O espaço coluna de  $\mathbf{A}$  é o subespaço de  $\mathbb{R}^m$  gerado pelos vetores coluna de  $\mathbf{A}$ .

**Definição 10** O espaço nulo de  $\mathbf{A}$  é o subespaço de  $\mathbb{R}^n$  gerado pelas soluções do sistema de equações homogêneas  $\mathbf{A}\mathbf{x} = \mathbf{0}$ .

**Teorema 1** Os espaço linha e coluna de uma matriz  $\mathbf{A}$  possuem a mesma dimensão.

**Definição 11** A dimensão comum do espaço linha e do espaço coluna de uma matriz  $\mathbf{A}$  é denominada rank de  $\mathbf{A}$ , e é denotada por  $\text{rank}(\mathbf{A})$ . A dimensão do espaço nulo de  $\mathbf{A}$  é denominada nulidade de  $\mathbf{A}$  e é denotada por  $\text{nul}(\mathbf{A})$ .

**Teorema 2 (Teorema Espectral para matrizes reais)** Seja uma matriz  $\mathbf{A} \in \mathbb{R}^{n \times n}$ . As seguintes afirmações são equivalentes:

- A matriz  $\mathbf{A}$  é simétrica.
- Existe uma base ortonormal para o  $\mathbb{R}^n$  constituída exclusivamente de autovetores de  $\mathbf{A}$ .
- A matriz  $\mathbf{A}$  é uma matriz ortogonalmente diagonalizável, ou seja, existe uma matriz ortogonal  $\mathbf{V}$  tal que  $\mathbf{V}^T \mathbf{A} \mathbf{V} = \mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$ .
- Existe uma matriz diagonal  $\mathbf{D}$  e uma matriz ortogonal  $\mathbf{V}$  tal que  $\mathbf{A} = \mathbf{V} \mathbf{D} \mathbf{D}^T$ .

No livro *Matrix Computations*, Golub e Van Loan escrevem "a simetria garante que todos os autovalores de uma matriz simétrica  $\mathbf{A}$  são reais e que existe uma base de autovetores ortonormais para o espaço coluna da mesma."

**Teorema 3 (Teorema da Dimensão para matrizes)** Seja  $\mathbf{A}_{m \times n}$  tal que  $m \geq n$ . Então,

$$\text{rank}(\mathbf{A}) + \text{nul}(\mathbf{A}) = n,$$

onde  $\text{nul}(\mathbf{A})$  é a dimensão do espaço nulo de  $\mathbf{A}$  e  $\text{rank}(\mathbf{A})$  é o rank de  $\mathbf{A}$ .

**Teorema 4 (Teorema da Dimensão)** Seja  $\mathbb{K}^n$  um espaço vetorial de dimensão finita,  $T : \mathbb{K}^n \rightarrow \mathbb{K}^m$  uma transformação linear,  $N[T]$  o núcleo e  $\text{Im}(T)$ , a imagem de  $T$ . Então,  $\dim[N(T)] + \dim[\text{Im}(T)] = \dim(\mathbb{K}^n)$ .

## Apêndice B - Conceitos Matemáticos Fundamentais por trás do SVD

Toda a matriz  $\mathbf{A}_{m \times n}$  define uma transformação linear  $T_{\mathbf{A}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Sabemos, portanto, que o domínio  $D(T_{\mathbf{A}}) \subset \mathbb{R}^n$  e que a imagem  $Im(T_{\mathbf{A}}) \subset \mathbb{R}^m$ . Sabemos, também, da definição de vetores linha e vetores coluna de uma matriz, que as linhas de  $\mathbf{A}$  são vetores do  $\mathbb{R}^n$ , e que as colunas de  $\mathbf{A}$  são vetores do  $\mathbb{R}^m$ . Esses fatos sugerem que há uma relação entre os espaços matriciais fundamentais de  $\mathbf{A}$  e os espaços associados à transformação  $T_{\mathbf{A}}$ .

Seja  $\mathbf{v} \in N(T_{\mathbf{A}})$ . Assim,  $\mathbf{v}$  satisfaz a equação homogênea  $\mathbf{A}\mathbf{v} = 0$  e, consequentemente, pertence ao espaço nulo de  $\mathbf{A}$ . Dessa forma, o espaço nulo de  $\mathbf{A}$  é exatamente o núcleo de  $T_{\mathbf{A}}$  e, portanto,  $dim[N(T_{\mathbf{A}})] = nul(\mathbf{A})$ . Tendo em vista o Teorema da Dimensão, concluímos que  $dim[Im(T_{\mathbf{A}})] = rank(\mathbf{A})$ .

Seja  $k = rank(\mathbf{A})$ . As conclusões feitas nos permitem concluir que os  $k$  primeiros vetores coluna da matriz  $\mathbf{V}$  compõem uma base ortonormal para o espaço coluna de  $\mathbf{A}$  (onde se encontra  $D(T_{\mathbf{A}})$ ), enquanto os demais  $n - k$  vetores coluna de  $\mathbf{V}$  compõem uma base ortonormal para o espaço nulo de  $\mathbf{A}$ . Da mesma sorte, os  $k$  primeiros vetores coluna da matriz  $\mathbf{U}$  compõem uma base ortonormal para o espaço linha de  $\mathbf{A}$  (onde se encontra  $Im(T_{\mathbf{A}})$ ), enquanto os demais  $m - k$  vetores coluna de  $\mathbf{U}$  compõem uma base ortonormal para o restante do  $\mathbb{R}^m$ .

Por fim, lembremos que  $\mathbf{A}^T \mathbf{A}$  e  $\mathbf{A} \mathbf{A}^T$  são simétricas e, pelo Teorema Espectral, podem ser ortogonalizadas de forma *única* pela matriz dos seus autovetores. A unicidade da ortogonalização dessas duas matrizes implica na unicidade do SVD da matriz  $\mathbf{A}$ .

## Apêndice C - Cômputo do SVD

Seja  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , de forma que  $\mathbf{V}^* = \mathbf{V}^T$ . O cálculo do SVD de  $\mathbf{A}$  envolve a determinação das matrizes  $\mathbf{U}$ ,  $\mathbf{\Sigma}$  e  $\mathbf{V}$ , e está intimamente relacionado com os autovalores e autovetores das matrizes  $\mathbf{A}\mathbf{A}^T$  e  $\mathbf{A}^T\mathbf{A}$ . A fatora  o SVD de  $\mathbf{A}$   

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T,$$

de forma que

$$\mathbf{A}^T = (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T = (\mathbf{V}^T)^T\mathbf{\Sigma}^T\mathbf{U}^T.$$

Como  $(\mathbf{V}^T)^T = \mathbf{V}$ , podemos reescrever a transposta de  $\mathbf{A}$  como

$$\mathbf{A}^T = \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T.$$

Multiplicando  $\mathbf{A}$  pela esquerda por  $\mathbf{A}^T$ , obtemos

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T.$$

Como  $\mathbf{U}$    ortogonal, temos  $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ , onde  $\mathbf{I}$    a matriz identidade. Dessa forma,

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}(\mathbf{\Sigma}^T\mathbf{\Sigma})\mathbf{V}^T,$$

onde  $\mathbf{\Sigma}^T\mathbf{\Sigma}$    uma matriz diagonal de ordem  $n$ .

Sabemos que a matriz  $\mathbf{A}^T\mathbf{A}$    sim trica. Dessa forma, o Teorema Espectral garante que ela   ortogonalmente diagonaliz vel por uma matriz  $\mathbf{P}$  ortogonal cujos vetores coluna sejam autovetores de  $\mathbf{A}$ . Ou seja,  $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$ , onde  $\mathbf{D}$    uma matriz diagonal. Comparando essa equa o com a equa o  $\mathbf{V}(\mathbf{\Sigma}^T\mathbf{\Sigma})\mathbf{V}^T$  obtida, observa-se que  $\mathbf{V}$    a matriz dos autovetores de  $\mathbf{A}$ , e que as entradas da diagonal principal de  $\mathbf{\Sigma}^T\mathbf{\Sigma}$  s o autovalores de  $\mathbf{A}$ . Dessa forma, conclu mos que os vetores singulares direitos de  $\mathbf{A}$  (vetores coluna de  $\mathbf{V}$ ) s o os autovetores da matriz  $\mathbf{A}^T\mathbf{A}$ . Conclu mos, ainda, que as ra zes quadradas dos elementos da diagonal principal de  $\mathbf{\Sigma}^T\mathbf{\Sigma}$  s o os autovalores de  $\mathbf{A}$ , visto que aqueles s o o quadrado dos elementos da diagonal principal de  $\mathbf{\Sigma}$ . Repetindo o racioc nio desenvolvido, obtemos

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T \therefore \mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{U}^T,$$

que nos permite concluir que os vetores singulares esquerdos de  $\mathbf{A}$  (vetores coluna de  $\mathbf{U}$ ) são os autovetores da matriz  $\mathbf{A}\mathbf{A}^T$ . Concluimos, também, que as raízes quadradas dos elementos da diagonal principal de  $\mathbf{\Sigma}\mathbf{\Sigma}^T$  são os autovalores de  $\mathbf{A}$ . É importante frisar que, embora as matrizes  $\mathbf{A}^T\mathbf{A}$  e  $\mathbf{A}\mathbf{A}^T$  possuam dimensões distintas, seus autovalores são idênticos. Esse fato advém de um teorema da Álgebra Linear que afirma que, se  $\mathbf{A}\mathbf{A}^T$  possui  $rank(\mathbf{A})$  autovalores diferentes de 0,  $\mathbf{A}^T\mathbf{A}$  possui exatamente os mesmos  $rank(\mathbf{A})$  autovalores diferentes de 0. A demonstração desse teorema é feita como segue. Consideremos a equação

$$(\mathbf{A}^T\mathbf{A})\mathbf{x} = \lambda\mathbf{x} \therefore (\mathbf{A}\mathbf{A}^T)(\mathbf{A}\mathbf{x}) = \lambda(\mathbf{A}\mathbf{x}).$$

Se considerarmos  $\lambda \neq 0$  na equação acima, veremos que  $\mathbf{A}^T\mathbf{A}$  e  $\mathbf{A}\mathbf{A}^T$  possuem exatamente os mesmos  $rank(\mathbf{A})$  autovalores diferentes de 0. Dessa forma,  $\mathbf{\Sigma}^T\mathbf{\Sigma}$  e  $\mathbf{\Sigma}\mathbf{\Sigma}^T$  são matrizes diagonais cujas primeiras  $rank(\mathbf{A})$  entradas diagonais são iguais e correspondem aos quadrados dos valores singulares de  $\mathbf{A}$ ,  $\sigma_i^2$ .

Para permitir uma visualização dos procedimentos expostos, consideremos, sem perda de generalidade,  $\mathbf{A} \in \mathbb{R}^{4 \times 3}$  e  $rank(\mathbf{A}) = 3$ . Dessa forma, o SVD de  $\mathbf{A}$  será dado por  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ , onde

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{bmatrix} \quad \text{e} \quad \mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{bmatrix}.$$

Os produtos  $\mathbf{A}^T\mathbf{A}$  e  $\mathbf{A}\mathbf{A}^T$  serão escritos como

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix} \quad \text{e} \quad \mathbf{A}\mathbf{A}^T = \begin{bmatrix} g & k & l & m \\ k & h & n & o \\ l & n & i & p \\ m & o & p & j \end{bmatrix}.$$

A equação  $(\mathbf{A}^T \mathbf{A})\mathbf{V} = \mathbf{V}(\boldsymbol{\Sigma}^T \boldsymbol{\Sigma})$  pode ser reescrita como

$$\begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix} \cdot \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} \cdot \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} \therefore$$

$$\therefore \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix} \cdot \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 v_{11} & \sigma_2^2 v_{12} & \sigma_3^2 v_{13} \\ \sigma_1^2 v_{21} & \sigma_2^2 v_{22} & \sigma_3^2 v_{23} \\ \sigma_1^2 v_{31} & \sigma_2^2 v_{32} & \sigma_3^2 v_{33} \end{bmatrix}$$

Denotemos os vetores linha de  $\mathbf{A}^T \mathbf{A}$  por  $L(\mathbf{A}^T \mathbf{A})_j$  e os vetores coluna de  $\mathbf{V}$  por  $v_j$ . Assim, podemos escrever

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} L(\mathbf{A}^T \mathbf{A})_1 \\ L(\mathbf{A}^T \mathbf{A})_2 \\ L(\mathbf{A}^T \mathbf{A})_3 \end{bmatrix} \quad \text{e} \quad \mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}.$$

Assim, podemos reescrever  $(\mathbf{A}^T \mathbf{A})\mathbf{V} = \mathbf{V}(\boldsymbol{\Sigma}^T \boldsymbol{\Sigma})$  como

$$\begin{bmatrix} L(\mathbf{A}^T \mathbf{A})_1 \mathbf{v}_1 & L(\mathbf{A}^T \mathbf{A})_1 \mathbf{v}_2 & L(\mathbf{A}^T \mathbf{A})_1 \mathbf{v}_3 \\ L(\mathbf{A}^T \mathbf{A})_2 \mathbf{v}_1 & L(\mathbf{A}^T \mathbf{A})_2 \mathbf{v}_2 & L(\mathbf{A}^T \mathbf{A})_2 \mathbf{v}_3 \\ L(\mathbf{A}^T \mathbf{A})_3 \mathbf{v}_1 & L(\mathbf{A}^T \mathbf{A})_3 \mathbf{v}_2 & L(\mathbf{A}^T \mathbf{A})_3 \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 v_{11} & \sigma_2^2 v_{12} & \sigma_3^2 v_{13} \\ \sigma_1^2 v_{21} & \sigma_2^2 v_{22} & \sigma_3^2 v_{23} \\ \sigma_1^2 v_{31} & \sigma_2^2 v_{32} & \sigma_3^2 v_{33} \end{bmatrix} \therefore$$

$$\therefore \left[ (\mathbf{A}^T \mathbf{A})\mathbf{v}_1 \quad (\mathbf{A}^T \mathbf{A})\mathbf{v}_2 \quad (\mathbf{A}^T \mathbf{A})\mathbf{v}_3 \right] = \left[ \sigma_1^2 \mathbf{v}_1 \quad \sigma_2^2 \mathbf{v}_2 \quad \sigma_3^2 \mathbf{v}_3 \right].$$

É importante atentar para uma peculiaridade da equação matricial acima: pela igualdade matricial, temos que

$$(\mathbf{A}^T \mathbf{A})\mathbf{v}_j = \sigma_j^2 \mathbf{v}_j,$$

que é compatível com os resultados que obtivemos anteriormente. O vetor  $\mathbf{v}_j$  é um autovetor da matriz  $\mathbf{A}^T \mathbf{A}$ , e  $\sigma_j^2$  é o autovalor de  $\mathbf{A}^T \mathbf{A}$  associado a  $\mathbf{v}_j$ . Ainda, os  $\sigma_j^2$  são todos reais, pelo Teorema Espectral (independente de  $\mathbf{A}$  ser real ou complexa). Denotemos tais autovalores por  $\lambda_1, \lambda_2$  e  $\lambda_3$ . Já foi visto que os autovalores de  $\mathbf{A}^T \mathbf{A}$  são  $\sigma_1^2, \sigma_2^2$  e  $\sigma_3^2$ .

Assim, concluímos que

$$\sigma_j^2 = \lambda_j \therefore \sigma_j = \sqrt{\lambda_j}, \quad (1 \leq j \leq 3).$$

Portanto, os valores singulares da matriz  $\mathbf{A}$  são as raízes quadradas dos autovalores de  $\mathbf{A}^T \mathbf{A}$ . Notamos, ainda, que  $\mathbf{A}^T \mathbf{A}$  possui 3 autovetores (associados aos seus 3 autovalores) localizados no  $\mathbb{R}^3$ , que é o espaço coluna de  $\mathbf{A}$ . Pelo Teorema Espectral, tais autovetores fazem parte de uma base ortonormal para esse espaço.

Agora, analisemos a equação

$$(\mathbf{A}\mathbf{A}^T)\mathbf{U} = \mathbf{U}(\mathbf{\Sigma}\mathbf{\Sigma}^T).$$

Essa equação pode ser reescrita como

$$\begin{bmatrix} g & k & l & m \\ k & h & n & o \\ l & n & i & p \\ m & o & p & j \end{bmatrix} \cdot \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{bmatrix} \cdot \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \therefore$$

$$\therefore \begin{bmatrix} g & k & l & m \\ k & h & n & o \\ l & n & i & p \\ m & o & p & j \end{bmatrix} \cdot \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 u_{11} & \sigma_2^2 u_{12} & \sigma_3^2 u_{13} & 0 \\ \sigma_1^2 u_{21} & \sigma_2^2 u_{22} & \sigma_3^2 u_{23} & 0 \\ \sigma_1^2 u_{31} & \sigma_2^2 u_{32} & \sigma_3^2 u_{33} & 0 \\ \sigma_1^2 u_{41} & \sigma_2^2 u_{42} & \sigma_3^2 u_{43} & 0 \end{bmatrix}.$$

Denotemos os vetores linha de  $\mathbf{A}\mathbf{A}^T$  por  $L(\mathbf{A}\mathbf{A}^T)_j$  e os vetores coluna de  $\mathbf{U}$  por  $\mathbf{u}_j$

$$\mathbf{A}\mathbf{A}^T = \begin{bmatrix} L(\mathbf{A}\mathbf{A}^T)_1 \\ L(\mathbf{A}\mathbf{A}^T)_2 \\ L(\mathbf{A}\mathbf{A}^T)_3 \\ L(\mathbf{A}\mathbf{A}^T)_4 \end{bmatrix} \quad \text{e} \quad \mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 \end{bmatrix}$$

Assim, podemos reescrever  $(\mathbf{A}\mathbf{A}^T)\mathbf{U} = \mathbf{U}(\mathbf{\Sigma}\mathbf{\Sigma}^T)$  como

$$\begin{aligned}
& \begin{bmatrix} L(\mathbf{A}\mathbf{A}^T)_1\mathbf{u}_1 & L(\mathbf{A}\mathbf{A}^T)_1\mathbf{u}_2 & L(\mathbf{A}\mathbf{A}^T)_1\mathbf{u}_3 & L(\mathbf{A}\mathbf{A}^T)_1\mathbf{u}_4 \\ L(\mathbf{A}\mathbf{A}^T)_2\mathbf{u}_1 & L(\mathbf{A}\mathbf{A}^T)_2\mathbf{u}_2 & L(\mathbf{A}\mathbf{A}^T)_2\mathbf{u}_3 & L(\mathbf{A}\mathbf{A}^T)_2\mathbf{u}_4 \\ L(\mathbf{A}\mathbf{A}^T)_3\mathbf{u}_1 & L(\mathbf{A}\mathbf{A}^T)_3\mathbf{u}_2 & L(\mathbf{A}\mathbf{A}^T)_3\mathbf{u}_3 & L(\mathbf{A}\mathbf{A}^T)_3\mathbf{u}_4 \\ L(\mathbf{A}\mathbf{A}^T)_4\mathbf{u}_1 & L(\mathbf{A}\mathbf{A}^T)_4\mathbf{u}_2 & L(\mathbf{A}\mathbf{A}^T)_4\mathbf{u}_3 & L(\mathbf{A}\mathbf{A}^T)_4\mathbf{u}_4 \end{bmatrix} = \\
& = \begin{bmatrix} \sigma_1^2 u_{11} & \sigma_2^2 u_{12} & \sigma_3^2 u_{13} & 0 \\ \sigma_1^2 u_{21} & \sigma_2^2 u_{22} & \sigma_3^2 u_{23} & 0 \\ \sigma_1^2 u_{31} & \sigma_2^2 u_{32} & \sigma_3^2 u_{33} & 0 \\ \sigma_1^2 u_{41} & \sigma_2^2 u_{42} & \sigma_3^2 u_{43} & 0 \end{bmatrix} \therefore \\
& \therefore \left[ (\mathbf{A}\mathbf{A}^T)\mathbf{u}_1 \quad (\mathbf{A}\mathbf{A}^T)\mathbf{u}_2 \quad (\mathbf{A}\mathbf{A}^T)\mathbf{u}_3 \quad (\mathbf{A}\mathbf{A}^T)\mathbf{u}_4 \right] = \\
& = \left[ \sigma_1^2 \mathbf{u}_1 \quad \sigma_2^2 \mathbf{u}_2 \quad \sigma_3^2 \mathbf{u}_3 \quad \mathbf{0} \right],
\end{aligned}$$

onde  $\mathbf{0}$  é a matriz nula. Novamente, observamos que o vetor  $u_j$  é um autovetor da matriz  $\mathbf{A}\mathbf{A}^T$ , e  $\sigma_j^2$  é um autovalor de  $\mathbf{A}\mathbf{A}^T$  associado a  $\mathbf{u}_j$ . É evidente, ainda, que os autovalores de  $\mathbf{A}\mathbf{A}^T$  são exatamente os mesmos de  $\mathbf{A}^T\mathbf{A}$ . Notamos, novamente, que  $\mathbf{A}\mathbf{A}^T$  possui 3 autovetores associados aos seus 3 autovalores. Tais autovetores estão no  $\mathbb{R}^4$ , que é o espaço linha de  $\mathbf{A}$ . Pelo Teorema Espectral, tais autovetores fazem parte de uma base ortonormal para esse espaço.

Resumidamente, observamos que os valores singulares de  $\mathbf{A}$  são iguais às raízes quadradas dos autovalores não nulos de  $\mathbf{A}^T\mathbf{A}$  (que são exatamente os mesmos autovalores de  $\mathbf{A}\mathbf{A}^T$ ), que os vetores singulares direitos de  $\mathbf{A}$  são os autovetores de  $\mathbf{A}^T\mathbf{A}$  e que os vetores singulares esquerdos de  $\mathbf{A}$  são os autovetores de  $\mathbf{A}\mathbf{A}^T$ . Assim, montamos as matrizes  $\mathbf{U}$ ,  $\mathbf{\Sigma}$  e  $\mathbf{V}$ .

## Apêndice D - Interpretação Geométrica do SVD

O SVD, geometricamente, é justificado pelo fato quando uma esfera unitária (ou seja, a esfera euclidiana usual em um espaço vetorial euclidiano  $n$ -dimensional) é transformada por meio de uma matriz  $\mathbf{A}_{m \times n}$ , sua imagem é uma hiperelipse (ou seja, uma elipse em um espaço vetorial euclidiano  $m$ -dimensional). Doravante, tal esfera será denotada por  $\mathbf{S}$  e a hiperelipse, por  $\mathbf{AS}$ .

Antes de prosseguir, é importante compreender a afirmação acima. Embora à primeira vista ela seja um tanto quanto "seca", ela é, de fato, consideravelmente simples. Quando nos referimos a uma esfera unitária, na realidade estamos nos definindo à superfície imaginária que envolve todos os vetores unitários e, em especial, aqueles que compõem bases ortonormais. Para exemplificar, consideremos os vetores da base canônica do  $\mathbb{R}^3$  (a saber,  $(0; 0; 1)$ ,  $(0; 1; 0)$  e  $(1; 0; 0)$ ). Esses três vetores são perpendiculares entre si e possuem o mesmo comprimento. Assim, podemos imaginar uma casca esférica imaginária envolvendo eles, como ilustrado na Figura 14. Nela, é possível visualizar a esfera imaginária à qual nos referimos. Enfatizamos que essa esfera envolve *todos* os vetores unitários do  $\mathbb{R}^3$ , dentre os quais destacamos aqueles que compõem bases ortonormais para o referido espaço vetorial.

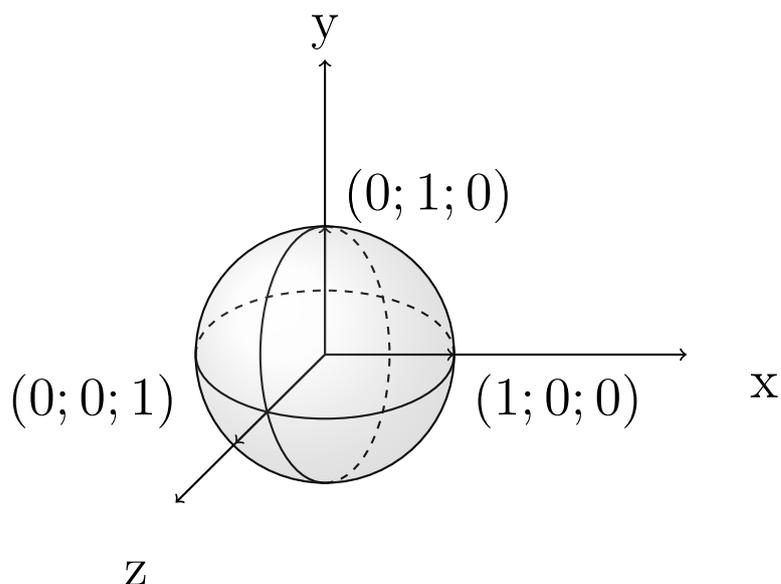


Figura 14: Esfera unitária no  $\mathbb{R}^3$

Quando foi dito que tal esfera é transformada por meio de uma matriz  $\mathbf{A}$  numa hiperelipse no  $\mathbb{R}^m$ , estamos simplesmente descrevendo geometricamente o efeito da transformação linear definida pela matriz  $\mathbf{A}$  sobre os vetores unitários do  $\mathbb{R}^n$ . Toda matriz  $A_{m \times n}$  sempre determina uma transformação linear  $T_{\mathbf{A}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Essa transformação reescala (i.e., modifica o módulo) e rotaciona os vetores unitários do  $\mathbb{R}^n$  no  $\mathbb{R}^m$ . Estes, por sua vez, podem ser imaginados no interior de uma superfície elíptica. Graficamente, a transformação de  $\mathbf{S}$  pode ser vista na Figura 15. Assim, vemos que a transformação  $\mathbf{A}$  deforma  $\mathbf{S}$ , fazendo com que os vetores  $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$  sejam rotacionados e reescalados no  $\mathbb{R}^m$ , transformando-os, assim, nos semieixos da hiperelipse  $\sigma_1 \mathbf{u}_1, \sigma_2 \mathbf{u}_2 \in \mathbb{R}^m$ .

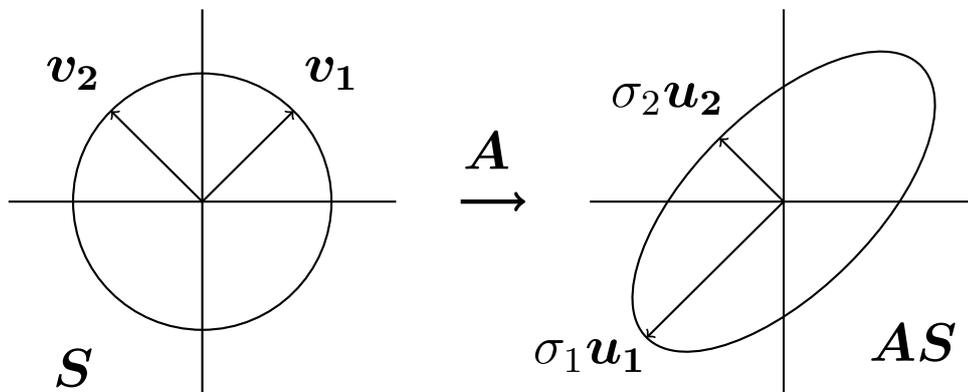


Figura 15: Deformação de  $\mathbf{S}$  por  $\mathbf{A}$

Agora, podemos dar uma interpretação geométrica para os valores e vetores singulares de  $\mathbf{A}$ . Os *vetores singulares direitos* da matriz  $\mathbf{A}$  (i.e.,  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ ) são os eixos principais da esfera euclidiana  $\mathbf{S}$  no  $\mathbb{R}^n$ , e constituem parte de uma base ortonormal para este. Os *vetores singulares esquerdos* da matriz  $\mathbf{A}$  (i.e.,  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ ) são os eixos principais da hiperelipse  $\mathbf{AS}$  no  $\mathbb{R}^m$ , e constituem parte de uma base ortonormal para este. Os *valores singulares* da matriz  $\mathbf{A}$  (i.e.,  $\sigma_1, \sigma_2, \dots, \sigma_n$ ) são os comprimentos dos semieixos de  $\mathbf{AS}$ .

Agora que vimos a interpretação geométrica da deformação de  $\mathbf{S}$  pela transformação  $\mathbf{A}$ , analisemos a interpretação geométrica da deformação de  $\mathbf{S}$  pelas matrizes  $\mathbf{U}$ ,  $\mathbf{\Sigma}$  e  $\mathbf{V}$ , que juntas formam o SVD de  $\mathbf{A}$ . Como vimos, a deformação de  $\mathbf{S}$  nada mais é do que a rotação e o reescalonamento (no espaço linha de  $\mathbf{A}$ ) dos vetores que compõem  $\mathbf{S}$  (pertencentes ao espaço coluna de  $\mathbf{A}$ ). Como  $\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \mathbf{V}^T$ , ao multiplicarmos um  $\mathbf{x} \in \mathbb{R}^n$  por  $\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$ , o resultado deve ser o mesmo de que obteríamos multiplicando-o por  $\mathbf{A}$ .

Sabemos que a multiplicação de um vetor por uma matriz ortonormal apenas faz com que ele seja rotacionado, sem que haja alteração do valor do seu módulo. Sabemos, também, que a multiplicação de um vetor por uma matriz diagonal implica no seu reescalonamento, sem que ele seja rotacionado.

Dessa forma, o SVD é exatamente o conjunto dessas operações de rotação e reescalonamento. As matrizes  $\mathbf{U}$  e  $\mathbf{V}$ , ortonormais, realizam a rotação dos vetores que compõem a esfera  $\mathbf{S}$ , e a matriz  $\mathbf{\Sigma}$  realiza o reescalonamento dos mesmos. Para melhor ilustrar essa afirmação, consideremos a Figura 16.

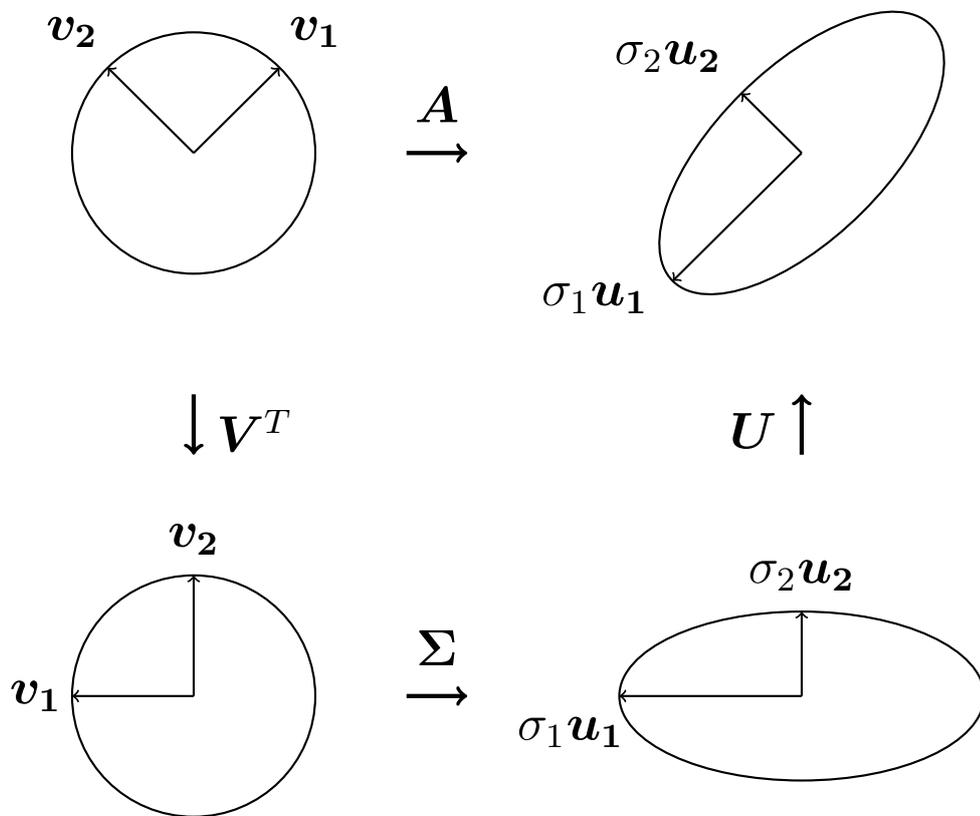


Figura 16: Deformação de  $\mathbf{S}$  por  $\mathbf{U}$ ,  $\mathbf{V}$  e  $\mathbf{\Sigma}$

Na Figura 16, vemos perfeitamente o resultado da atuação de cada um dos vetores que compõem o SVD de  $\mathbf{A}$  sobre a esfera  $\mathbf{S}$ . Novamente, sem perda de generalidade, consideremos  $\mathbf{A} \in \mathbb{R}^{4 \times 3}$  com  $\text{rank}(\mathbf{A}) = 3$ . Assim, temos

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{bmatrix} \quad \text{e} \quad \mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{bmatrix}.$$

Consideremos, também, o vetor  $\mathbf{x} \in \mathbb{R}^3$ ,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad |\mathbf{x}| = 1.$$

Fazendo uso da associatividade do produto matricial, fazamos  $\mathbf{Ax} = (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)\mathbf{x} = \mathbf{U}\mathbf{\Sigma}(\mathbf{V}^T\mathbf{x})$ . Podemos escrever  $\mathbf{V}^T\mathbf{x}$  como

$$\begin{aligned} \mathbf{V}^T\mathbf{x} &= \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \\ &= \begin{bmatrix} v_{11}x_1 + v_{21}x_2 + v_{31}x_3 \\ v_{12}x_1 + v_{22}x_2 + v_{32}x_3 \\ v_{13}x_1 + v_{23}x_2 + v_{33}x_3 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1^T x_1 + \mathbf{v}_2^T x_2 + \mathbf{v}_3^T x_3 \\ \mathbf{v}_2^T x_1 + \mathbf{v}_1^T x_2 + \mathbf{v}_3^T x_3 \\ \mathbf{v}_3^T x_1 + \mathbf{v}_2^T x_2 + \mathbf{v}_1^T x_3 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1\mathbf{x} \\ \mathbf{v}_2\mathbf{x} \\ \mathbf{v}_3\mathbf{x} \end{bmatrix}, \end{aligned}$$

onde  $\mathbf{v}_j$  são os vetores coluna de  $\mathbf{V}$ . Como os vetores  $\mathbf{v}_j$  são unitários, o resultado do produto  $\mathbf{V}^T\mathbf{x}$  é a rotação do vetor  $\mathbf{x}$  no  $\mathbb{R}^3$ , o espaço coluna de  $\mathbf{A}$ . Agora, escrevemos  $\mathbf{\Sigma}(\mathbf{V}^T\mathbf{x})$  como

$$\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_{11}x_1 + v_{21}x_2 + v_{31}x_3 \\ v_{12}x_1 + v_{22}x_2 + v_{32}x_3 \\ v_{13}x_1 + v_{23}x_2 + v_{33}x_3 \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma_1(v_{11}x_1 + v_{21}x_2 + v_{31}x_3) \\ \sigma_2(v_{12}x_1 + v_{22}x_2 + v_{32}x_3) \\ \sigma_3(v_{13}x_1 + v_{23}x_2 + v_{33}x_3) \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 \mathbf{v}_1^T x_1 + \sigma_2 \mathbf{v}_2^T x_2 + \sigma_3 \mathbf{v}_3^T x_3 \end{bmatrix}.$$

Na equação acima, vemos que o resultado do produto  $\Sigma(\mathbf{V}^T \mathbf{x})$  é o reescalonamento do vetor  $\mathbf{V}^T \mathbf{x}$  no  $\mathbb{R}^4$ , o espaço linha de  $\mathbf{A}$ . Dessa forma, é a matriz  $\Sigma$  que associa o vetor  $\mathbf{V}^T \mathbf{x}$  à sua imagem no  $\mathbb{R}^4$ . Ou seja, *a matriz  $\Sigma$  é uma forma simplificada da matriz  $\mathbf{A}$* . De fato, esse é o objetivo da diagonalização de uma matriz: simplificá-la por meio da sua transformação em uma simples matriz diagonal. Por fim, analisamos  $\mathbf{U}(\Sigma \mathbf{V}^T \mathbf{x})$ . Tal qual  $\mathbf{V}^T \mathbf{x}$ , ela resulta na rotação do vetor  $\Sigma \mathbf{V}^T \mathbf{x}$  no  $\mathbb{R}^4$ . De forma similar, podemos escrever  $\mathbf{U}(\Sigma \mathbf{V}^T \mathbf{x})$  como

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 \mathbf{v}_1^T \mathbf{x} \\ \sigma_2 \mathbf{v}_2^T \mathbf{x} \\ \sigma_3 \mathbf{v}_3^T \mathbf{x} \\ 0 \end{bmatrix} = \begin{bmatrix} u_{11}\sigma_1 \mathbf{v}_1^T \mathbf{x} + u_{12}\sigma_2 \mathbf{v}_2^T \mathbf{x} + u_{13}\sigma_3 \mathbf{v}_3^T \mathbf{x} + u_{14}0 \\ u_{21}\sigma_1 \mathbf{v}_1^T \mathbf{x} + u_{22}\sigma_2 \mathbf{v}_2^T \mathbf{x} + u_{23}\sigma_3 \mathbf{v}_3^T \mathbf{x} + u_{24}0 \\ u_{31}\sigma_1 \mathbf{v}_1^T \mathbf{x} + u_{32}\sigma_2 \mathbf{v}_2^T \mathbf{x} + u_{33}\sigma_3 \mathbf{v}_3^T \mathbf{x} + u_{34}0 \\ u_{41}\sigma_1 \mathbf{v}_1^T \mathbf{x} + u_{42}\sigma_2 \mathbf{v}_2^T \mathbf{x} + u_{43}\sigma_3 \mathbf{v}_3^T \mathbf{x} + u_{44}0 \end{bmatrix} =$$

$$\left[ (\mathbf{u}_1 \sigma_1 \mathbf{v}_1^T) \mathbf{x} + (\mathbf{u}_2 \sigma_2 \mathbf{v}_2^T) \mathbf{x} + (\mathbf{u}_3 \sigma_3 \mathbf{v}_3^T) \mathbf{x} + \mathbf{u}_4 0 \right].$$

onde  $\mathbf{u}_j$  são os vetores coluna de  $\mathbf{U}$ . Como os vetores  $\mathbf{u}_j$  são unitários, o resultado do produto  $\mathbf{U}(\Sigma \mathbf{V}^T \mathbf{x})$  é a rotação dos vetores coluna de  $\Sigma \mathbf{V}^T \mathbf{x}$  no  $\mathbb{R}^4$ , o espaço linha de  $\mathbf{A}$ . Dessa forma, é latente o significado geométrico do SVD: ele tão somente é a "quebra" da transformação  $\mathbf{A}$  em três etapas mais simples (uma rotação, um reescalonamento e uma segunda rotação). Ao fazermos  $\mathbf{A}\mathbf{x}$ , realizamos todas essas três operações simultaneamente. Ao fazermos  $(\mathbf{U}\Sigma \mathbf{V}^T)\mathbf{x}$ , realizamos essas três operações separadamente.