



**Célia Maria Gomes**

**Pereira da Costa**

**Representações esparsas de sinais e algoritmos gananciosos**



**Célia Maria Gomes  
Pereira da Costa**

**Representações esparsas de sinais e algoritmos  
gananciosos**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Matemática realizada sob a orientação científica do Prof. Dr. Uwe Kähler, Professor Auxiliar Convidado c./ Agregação do Departamento de Matemática da Universidade de Aveiro

Dedico este trabalho ao meu namorado, à minha mãe, ao meu pai, à minha irmã, à mãe e irmãos do meu namorado, pela esperança que sempre depositaram em mim e no meu futuro.

Ao meu orientador, Prof. Dr. Uwe Kähler, pela disponibilidade e paciência, mas principalmente pelo desafio que imprimiu a cada etapa deste estudo.

A todos os meus amigos, colegas e restantes familiares por todo o apoio incondicional que me deram.

## **O júri**

Presidente

**Prof. Dr. Helmuth Robert Malonek**  
Professor Catedrático do Departamento de Matemática da Universidade de Aveiro

**Prof. Dr. Gil Bernardes**  
Professor Auxiliar da Faculdade de Ciências e Tecnologia do Departamento de Matemática da Universidade de Coimbra

**Prof. Dr. Uwe Kähler**  
Professor Auxiliar Convidado com Agregação do Departamento de Matemática da Universidade de Aveiro

**palavras-chave**

Bases de Riesz, *Frames*, Dicionários, Aproximações Esparsas, Algoritmos Gananciosos, *Chirps*

**resumo**

Nesta dissertação apresentaremos novos resultados ligados ao uso de um algoritmo ganancioso, o dito *Orthogonal Matching Pursuit* (OMP), por forma a resolvermos problemas de aproximação esparsa em dicionários redundantes. Discutiremos, igualmente, uma modificação deste algoritmo, devida a Donoho, chamada *Basis Matching Pursuit* (BP). Apresentaremos uma condição (única) que assegura que ambos, OMP e BP, permitem recuperar sinais esparsos de forma exacta. Mais, mostraremos que ambos os algoritmos permitem efectuar a recuperação do sinal para uma vasta classe de dicionários. Com efeito, neste trabalho faremos um resumo de vários resultados recentes em BP, facilmente extensíveis ao caso de OMP. Adicionalmente, daremos também uma condição suficiente de garantia de que OMP pode recuperar átomos comuns a partir de todas as representações óptimas de um sinal não-esparso. Assim, OMP pode ser visto como um algoritmo de aproximação para o problema esparso num dicionário quasi-incoerente, isto é, para qualquer sinal dado, OMP permite calcular uma aproximação esparsa cujo erro não é muito pior do que o erro óptimo, e isto para o mesmo número de termos na aproximação.

**keywords**

Riesz basis, Frames, Dictionary, Sparse Approximation, Greedy methods, Chirp

**abstract**

This dissertation presents new results regarding the usage of a greedy algorithm, the so-called *Orthogonal Matching Pursuit* (OMP), in order to solve sparse approximation problems over redundant dictionaries. We also discuss a modification of this algorithm, due to Donoho, denoted *Basis Matching Pursuit* (BP). We present a single sufficient condition under which both OMP and BP can recover a sparse signal in an exact way. Moreover, it is shown that both algorithms allow such a recovering for a wide class of dictionaries. Indeed, in this work we give several recent results on BP which can easily be extended to OMP. Furthermore, we also give a sufficient condition under which OMP can retrieve the common atoms from all optimal representations of a non-sparse signal. Thus, OMP can be viewed as an approximation algorithm for the sparse problem over a quasi-incoherent dictionary, that is, for every input signal, OMP can calculate a sparse approximation whose error is only a small factor worse than the optimal error, and this done with same the same number of terms in the approximation.

# ÍNDICE

<b>Introdução</b>	<b>iii</b>
<b>1. Bases de Riesz e Frames</b>	<b>1</b>
1.1. Bases de Riesz	1
1.2. Frames	6
1.3. Bases de Riesz e Frames	8
1.4. Dicionário	9
1.4.1. Noção	9
1.4.2. Dicionários Ortonormais	10
1.4.3. Dicionários de Frames	11
<b>1.5 Aproximações</b>	<b>13</b>
1.5.1 Aproximações lineares	13
1.5.2. Aproximações não – lineares	14
1.5.3. Aproximações esparsas	15
<b>2. Algoritmos gananciosos</b>	<b>16</b>
<b>2.1. Algoritmos gananciosos</b>	<b>16</b>
2.1.1. Matching Pursuit	17
2.1.2. Orthogonal Matching Pursuit	18
2.1.3. Critérios de paragem	20
2.1.4. Um contra exemplo para MP	20
2.1.5. Breve história dos algoritmos gananciosos	21
<b>2.2. Comparação com as bases pursuit</b>	<b>24</b>

<b>3. Representação esparsa do sinal</b>	<b>26</b>
<b>3.1. Representação esparsa do sinal</b>	<b>27</b>
3.1.1. Selecção gananciosa dos átomos	27
3.1.2. Teorema da recuperação exacta	29
3.1.3. A avaliação da coerência	30
3.1.4. Necessidade de Recuperação Exacta	32
<b>3.2. Identificar os átomos para uma aproximação</b>	<b>32</b>
<b>3.3. O erro e a restrição da aproximação esparsa</b>	<b>35</b>
3.3.1. A avaliação da coerência	37
<b>4. Aplicação prática</b>	<b>39</b>
<b>4.1. Exemplos de chirps</b>	<b>39</b>
4.1.1. Chirp linear	39
4.1.2. Chirp geométrica	40
4.1.3. Chirp exponencial	40
4.1.4. Princípio de funcionamento dos radares	40
<b>4.2. Detecção de chirps</b>	<b>41</b>
4.2.1. A questão dos dicionários	41
4.2.2. Chirplet transform	43
<b>Conclusão</b>	<b>45</b>
<b>Bibliografia</b>	<b>46</b>



## INTRODUÇÃO

Em 2005, E. Candés, J. Romberg e T. Tao publicaram o artigo "*Stable signal recovery from incomplete and inaccurate measurements*", considerado agora um dos artigos fundamentais na área de processamento de sinais. Neste, os autores mostraram a possibilidade de reconstruir um sinal, de forma exacta, com baixos custos computacionais, tendo como ponto de partida o facto de apenas um número finito dos coeficientes da expansão serem diferentes de zero. Tais reconstruções têm várias aplicações em comunicação, processamento de imagens e em armazenamento de dados, entre outros, o que motivou um crescente interesse nos métodos expostos no mencionado artigo. A abordagem actual para a aproximação de funções usa a melhor aproximação em  $n$ -termos em dicionários. O uso e a escolha de dicionários revela-se necessário devido ao facto de bases ortonormais serem demasiado pequenas para lidar, de modo efectivo, com aproximações esparsas de sinais. Isto conduziu, nos últimos anos, á procura, e subsequente investigação, de novos métodos para a realização numérica, a baixo custo, da melhor aproximação em  $n$ -termos. O método que melhores resultados tem apresentado, até ao presente, são os chamados algoritmos gananciosos (greedy algorithms). Estes recebem o seu nome da estratégia que utilizam, isto é, para um pequeno número  $N$  de dados (aproximação esparsa), o algoritmo busca a melhor aproximação possível de entre uma vasta classe de funções. Obviamente, um  $N$  demasiado grande implica um crescimento incontrolável no custo numérico para obtenção da solução o que impede a aplicação deste algoritmo a aproximações não-esparsas.

O principal objectivo deste trabalho reside em mostrar as principais diferenças entre o *Matching Pursuit* (MP) e o *Orthogonal Matching Pursuit* (OMP). No primeiro capítulo realçaremos a importância das aproximações esparsas. Para esse efeito, introduzimos as noções de base de Riesz, de *Frames* e de Dicionários, bem como a sua importância para aproximações esparsas. O segundo capítulo é dedicado à descrição dos dois algoritmos previamente mencionados, *Matching Pursuit* e o *Orthogonal Matching Pursuit*, e sua comparação com o algoritmo inicial dito *Basis Pursuit* (Petr Vanicek, 1969). No terceiro capítulo faremos a análise de ambos os métodos em termos da representação esparsa do sinal. Finalizaremos este estudo com uma aplicação à detecção de chirps.

## CAPÍTULO I – Bases de Riesz e Frames

Este capítulo é dedicado aos conceitos necessários e fundamentais para os restantes capítulos da presente dissertação. Assim, introduziremos primeiro os conceitos de *base de Riesz* e de *frames*, os quais estão intimamente ligados à transformada de Fourier e às aproximações de resolução múltipla (ARM). Mencionaremos, de seguida, diferentes tipos de aproximação de sinais, como sejam, aproximações lineares, não-lineares e/ou esparsas,. Salientaremos a importância das aproximações esparsas na aplicação de algoritmos gananciosos (em Inglês, *greedy algorithms*, a abordar nos próximos capítulos).

### 1.1. Bases de Riesz

As bases ortonormais são indispensáveis em qualquer área do conhecimento científico (seja Matemática, Física, ou Electrónica) onde haja necessidade de expressar funções, ou sinais, em termos de certos *blocos elementares*, isto é, elementos de uma dada base. Por conveniência, as bases pretendem-se *ortonormadas*, mas esta pode revelar-se, como veremos à frente, uma exigência demasiado forte.

O seguinte lema caracteriza todas as bases ortonormais num dado espaço de Hilbert  $H$  (isto é, um espaço com produto interno, e completo relativamente à norma induzida por este - ver, e.g. [4,16,17]) em termos de uma única base ortonormal .

#### Definição 1.1. Base Ortonormal

Seja  $H$  um espaço de Hilbert com produto interno  $\langle \cdot, \cdot \rangle$ . Uma base  $(\mathbf{e}_k)_{k \in \mathbb{N}}$  diz-se

ortonormal se  $\langle \mathbf{e}_k, \mathbf{e}_j \rangle = \delta_{jk} = \begin{cases} 1, j = k \\ 0, j \neq k \end{cases}$ .

**Lema 1.1.**

Seja  $(e_k)_{k \in \mathbb{N}}$  uma base ortonormal para  $H$ . Qualquer outra base ortonormal de  $H$  pode ser escrita como  $(Ue_k)_{k \in \mathbb{N}}$ , onde  $U : H \rightarrow H$  é um operador unitário.

**Definição 1.2. Base de Riesz [17]**

Uma sequência de vetores  $(x_n)_{n \in A}$  num espaço de Hilbert  $H$  diz-se uma sequência de Riesz se existirem constantes  $0 < c < C$ , tais que

$$c \left( \sum_n |a_n|^2 \right)^{1/2} \leq \left\| \sum_n a_n x_n \right\| \leq C \left( \sum_n |a_n|^2 \right)^{1/2}, \quad (1.1)$$

para toda a sequência  $(a_n)_{n \in A}$ .

Se tivermos, adicionalmente,  $\text{span}(x_n)_{n \in A} = H$ , e a sequência for linearmente independente, então diremos que  $(x_n)_{n \in A}$  constitui uma base de Riesz.

**Observação 1.1.**

Uma vez que todo o sistema ortonormal satisfaz a condição (1.1) (basta tomar  $c=C=1$ ), uma base de Riesz é vista como uma generalização do conceito de base ortonormal.

**Lema 1.2.**

Para qualquer base de Riesz  $(x_n)_{n \in Z}$  em  $H$ , existe uma sequência biortogonal  $(x_n^*)_{n \in Z}$  de vetores em  $H$ , isto é, uma sequência de Riesz que satisfaz a condição de bi-ortogonalidade com a base inicialmente dada, isto é,

$$\langle x_n^*, x_m \rangle = \begin{cases} 1, n=m \\ 0, n \neq m \end{cases}.$$

A sequência  $(x_n^*)_{n \in Z}$  é de novo uma base de Riesz para  $H$ .

**Prova:**

Sejam  $e_n$ , definidos como  $(e_n)_k = \delta_{nk}$ , os elementos de uma base do espaço de Hilbert  $l^2(Z)$  e  $l : l^2(Z) \rightarrow H$  o operador definido por  $l e_n = x_n, \forall n \in Z$ .

Sendo  $(x_n)_{n \in \mathbb{Z}}$  uma base de Riesz, resulta que  $I$  é um isomorfismo entre  $l^2(\mathbb{Z})$  e  $H$  e, portanto, o adjunto do operador  $I^{-1}$ , denotado como  $(I^{-1})^*$  e que actua de  $l^2(\mathbb{Z})$  em  $H$ , é também um isomorfismo. Defina-se a sequência de elementos de  $H$  dada por  $x_n^* = (I^{-1})^* e_n$ .

Teremos assim que,  $\langle x_n^*, x_m \rangle = \langle (I^{-1})^* e_n, I e_m \rangle = \langle e_n, e_m \rangle = \delta_{nm}$ , e, portanto,  $(x_n^*)_{n \in \mathbb{Z}}$  constitui uma sequência biortogonal de  $(x_n)_{n \in \mathbb{Z}}$  em  $H$ . Para vermos que é uma base de Riesz, basta ter em atenção que  $(I^{-1})^*$  é um isomorfismo.

□

### Definição 1.3. Frame

Uma família de vectores  $(x_n)_{n \in A}$  num espaço de Hilbert  $H$  diz-se um frame para  $H$  se existirem constantes  $0 < c < C$ , tais que (1.1).

Se  $c=C$ , então o frame designa-se por *tight frame*.

### Observação 1.2.

Note-se que num *frame* não há necessidade da independência linear entre os seus elementos. Os *frames* são usados como substitutos para bases de Riesz, em situações onde a independência linear é uma condição demasiado restricta e inútil relativamente às aplicações em vista.

Um exemplo (trivial) de *tight frame* é a união disjunta de dois sistemas ortogonais em  $H$ . Por exemplo, para  $H = \mathbb{R}^2$ , então a união disjunta  $(e_1, e_2, Ae_1, Ae_2)$  é um *tight frame* para este, com  $c=C=2$ , onde  $e_1, e_2$  denotam os elementos da base canónica e  $A$  é a matriz  $2 \times 2$  que roda um dado vector de um ângulo de noventa graus (sentido positivo).

Vamos agora introduzir a transformada de Fourier, a qual tem a propriedade de preservar a norma. Para  $f \in L^1(\mathbb{R})$ , define-se a sua transformada de Fourier  $f^\wedge(w) = F(w)$  como sendo a função definida por (ver [16])

$$f^\wedge(w) = F(w) = \int_{-\infty}^{+\infty} f(t)e^{-itw} dt, w \in \mathbb{R},$$

e que denotaremos por transformada de Fourier (TF) da função  $f$ . É, por vezes, conveniente representar a transformada de Fourier a menos de um factor de escala  $2\pi$ , para efeitos de normalização.

Uma vez que a transformada de Fourier é uma isometria de  $(L^1 \cap L^2)(\mathbb{R})$  em  $L^2(\mathbb{R})$ , temos que, para cada  $f \in L^2(\mathbb{R})$  arbitrária, existe uma sequência de funções  $\{f_n\}_{n \in \mathbb{Z}}$  em  $(L^1 \cap L^2)(\mathbb{R})$  que converge para  $f$  em  $L^2(\mathbb{R})$ . Diremos então que  $L^2(\mathbb{R})$  constitui o fecho de  $(L^1 \cap L^2)(\mathbb{R})$  com respeito à norma  $L^2$ , e que a transformada de Fourier é extensível a um mapa unitário de  $L^2(\mathbb{R})$  em  $L^2(\mathbb{R})$ ; sem perda de generalidade, usaremos a mesma notação para representar essa extensão. Em particular, obtemos a equação de Parseval,

$$\langle f, g \rangle = 2\pi \langle f^\wedge, g^\wedge \rangle$$

para todos  $f, g \in L^2(\mathbb{R})$ , ou ainda, que  $\|f\|^2 = 2\pi \|f^\wedge\|^2$ . De salientar que a transformada de Fourier permite uma fácil visualização da frequência  $w$  do sinal original  $f(t)$ .

Dada uma função  $\phi \in L^2(\mathbb{R})$ , então a transformada de Fourier da expansão

$$\sum_n a_n \phi(x-n)$$

é dada por  $\sum_n a_n e^{-in\zeta} \phi^\wedge(\zeta)$ ; como a transformada de Fourier preserva a norma  $L^2$ , temos

$$\begin{aligned} \|\sum_n a_n \phi(\cdot-n)\|^2 &= 2\pi \|\sum_n a_n e^{-in\cdot} \phi^\wedge(\cdot)\|^2 = 2\pi \int_{-\infty}^{+\infty} |\sum_n a_n e^{-in\zeta}|^2 |\phi^\wedge(\zeta)|^2 d\zeta \\ &= 2\pi \int_0^{2\pi} |\sum_n a_n e^{-in\zeta}|^2 \sum_l |\phi^\wedge(\zeta+2\pi l)|^2 d\zeta \\ &\leq 2\pi (\sum_n |a_n|^2) \int_{-\infty}^{+\infty} \sum_l |\phi^\wedge(\zeta+2\pi l)|^2 d\zeta \end{aligned} \tag{1.2}$$

Como  $\int_0^{2\pi} \sum_l |\phi^\wedge(\zeta+2\pi l)|^2 d\zeta = \int_{-\infty}^{+\infty} |\phi^\wedge(\zeta)|^2 d\zeta$  é finito, a expansão dada,  $\sum_n a_n \phi(x-n)$ , converge em  $L^2(\mathbb{R})$ .

De seguida, caracterizaremos a função  $\phi \in L^2(\mathbb{R})$  por forma a que  $\{\phi(\cdot-m)\}_{m \in \mathbb{Z}}$  constitua uma base de Riesz.

**Proposição 1.1.**

Sejam  $\phi \in L^2(\mathbb{R})$  e  $0 < a < A$  constantes. As seguintes condições são equivalentes:

a) Para toda a sequência  $(a_n)_{n \in \mathbb{Z}}$  de escalares, tem-se

$$a \left( \sum_n |a_n|^2 \right)^{1/2} \leq \left\| \sum_n a_n \phi(x-n) \right\| = \left( \int_{-\infty}^{+\infty} \left| \sum_n a_n \phi(x-n) \right|^2 \right)^{1/2} \leq A \left( \sum_n |a_n|^2 \right)^{1/2}; \tag{1.3}$$

b) Para todo  $\zeta \in \mathbb{R}$ ,

$$\frac{a^2}{2\pi} \leq \sum_l |\hat{\phi}(\zeta + 2\pi l)|^2 \leq \frac{A^2}{2\pi}; \quad (1.4)$$

**Prova:**

b) implica a):

Dado que  $(a_n)_{n \in \mathbb{Z}} \in l^2(\mathbb{Z})$ , multiplique-se (1.4.) pelo termo finito  $|\sum_n a_n e^{-in\zeta}|^2$ ; integrando agora de 0 a  $2\pi$ , tendo em atenção (1.2) e a igualdade

$$\int_0^{2\pi} |\sum_n a_n e^{-in\zeta}|^2 d\zeta = 2\pi \sum_n |a_n|^2,$$

obtemos (1.3).

a) implica b):

Tome-se o conjunto  $\Gamma = \left\{ \zeta \in [0, 2\pi[ : \sum_l |\hat{\phi}(\zeta + 2\pi l)|^2 > \frac{A^2}{2\pi} \right\}$ . Suponha-se que  $\mu(\Gamma) > 0$ ,

onde  $\mu$  denota a medida de Lebesgue de  $\Gamma$ . Construa-se uma sequência  $(a_n)_{n \in \mathbb{Z}} \in l^2(\mathbb{Z})$  tal que  $\chi_\Gamma(\zeta) = \sum_n a_n e^{-in\zeta}$ , para todo  $\zeta \in [0, 2\pi[$ .

De (1.2), vem

$$\left\| \sum_n a_n \phi(x - n) \right\| = \left( \int_\Gamma \sum_l |\hat{\phi}(\zeta + 2\pi l)|^2 d\zeta \right)^{1/2} > \frac{A}{\sqrt{2\pi}} [\mu(\Gamma)]^{1/2}$$

Como

$$\frac{1}{\sqrt{2\pi}} [\mu(\Gamma)]^{1/2} = \frac{1}{\sqrt{2\pi}} \|\chi_\Gamma\|^{1/2} = \left( \sum_n |a_n|^2 \right)^{1/2}$$

temos

$$\left\| \sum_n a_n \phi(x - n) \right\| > A \left( \sum_n |a_n|^2 \right)^{1/2},$$

o que contraria a segunda desigualdade de (1.3). Então  $\mu(\Gamma) = 0$ , ou seja,  $\sum_l |\hat{\phi}(\zeta + 2\pi l)|^2 \leq \frac{A^2}{2\pi}$ .

Raciocínio análogo mostra o limite inferior de (1.4).

□

**Observação 1.3.**

a) Como consequência da Proposição 1.1, uma função  $\phi \in L^2(\mathbb{R})$  gera uma base de Riesz  $\{\phi(\cdot - m)\}_{m \in \mathbb{Z}}$  sse

$$\frac{a^2}{2\pi} \leq \sum_l |\hat{\phi}(\zeta + 2\pi l)|^2 \leq \frac{A^2}{2\pi}.$$

b) O sistema  $\{\phi(\cdot - m)\}_{m \in \mathbb{Z}}$  é ortonormal sse tivermos  $a=A=1$ , ou seja,

$$\sum_l |\hat{\phi}(\zeta + 2\pi l)|^2 = \frac{1}{2\pi}.$$

**Proposição 1.2**

Considere-se uma base de Riesz  $\{\phi(\cdot - m)\}_{m \in \mathbb{Z}}$  que satisfaz a propriedade (1.3).

Então,

i)  $g \in \text{span}\{\phi(\cdot - n)\}_{n \in \mathbb{Z}}$ , isto é,  $g(x) = \sum_n a_n \phi(x - n)$ , sse existir uma função  $f$ ,  $2\pi$ -

periódica e tal que  $\int_0^{2\pi} |f^\wedge(\zeta)|^2 d\zeta < +\infty$ , que verifique  $g(\zeta) = f(\zeta)\phi(\zeta)$ .

Mais,  $f(\zeta) = \sum_n a_n e^{-in\zeta}$ .

ii) As normas de  $g$  e  $f$  estão relacionadas por

$$(1/A)\|g\| \leq \|f^\wedge\| \leq (1/a)\|g\|,$$

onde  $a$  e  $A$  são as constantes da Proposição 1.1.

**Proposição 1.3**

Seja  $\phi \in L^2(\mathbb{R})$  tal que o sistema  $\{\phi(\cdot - m)\}_{m \in \mathbb{Z}}$  é uma base de Riesz.

Então existe  $\phi_1 \in \text{span}\{\phi(\cdot - m)\}_{m \in \mathbb{Z}}$ , tal que  $\{\phi_1(\cdot - m)\}_{m \in \mathbb{Z}}$  é também uma base ortonormal do espaço  $\text{span}\{\phi(\cdot - m)\}_{m \in \mathbb{Z}}$ .

**Observação 1.4**

Se  $\phi$  satisfaz a igualdade  $\sum_l |\hat{\phi}(\zeta + 2\pi l)|^2 = \frac{1}{2\pi}$ , então a função  $\varphi$ , dada por  $\varphi^\wedge(\zeta) = m(\zeta)\phi^\wedge(\zeta)$ , onde  $m$  é  $2\pi$ -periódica e  $|m(\zeta)|=1$ , também a satisfará.

## 1.2. Frames

*Frames* foram introduzidos em 1952, por Duffin e Schaeffer. Estas estruturas surgem de forma natural no estudo das séries de Fourier, isto é, em sequências do tipo  $\{e^{i\lambda_n x}\}_{n \in \mathbb{Z}}$ , onde  $\{\lambda_n\}_{n \in \mathbb{Z}}$  é uma família de números reais ou de números complexos.

Em 1985, Daubechies, Grossmann e Meyer observaram, no contexto de *wavelets*, que *frames* poderiam ser usados para determinar expansões em séries de funções em  $L^2(\mathbb{R})$ , expansões estas muito similares às obtidas em termos de bases ortonormais. Nesta dissertação, vamos abordar a teoria de *frames* encarada como uma análise de funções em espaços de Hilbert.

Dada a sua importância, vamos repetir a definição de *frame* dada na Secção 1.1.

### Definição 1.4

Uma família  $(f_n)_{n \in A}$  num espaço de Hilbert  $H$  diz-se um *frame* para  $H$  se existirem constantes  $0 < A < B < \infty$ , tais que

$$A\|f\|^2 \leq \sum_n |\langle f, f_n \rangle|^2 \leq B\|f\|^2 \quad (1.5)$$

ocorre para todo  $f \in H$ . Um *frame* diz-se *tight* se  $A=B$ .

Se, removendo um número arbitrário de elementos da família  $(f_n)_{n \in A}$ , a nova família já não constitui um *frame*, então  $(f_n)_{n \in A}$  diz-se um *frame exacto*.

Os números  $A$  e  $B$  são chamados os limites do *frame*. Eles não são únicos. O valor óptimo do limite  $B$  do *frame* é o ínfimo de todos os limites superiores do *frame*, e o valor óptimo para o limite  $A$  do *frame* é o supremo de todos os limites inferiores do *frame*. Da definição resulta que se  $(f_n)_{n \in A}$  é um *frame* para  $H$ , temos  $\overline{\text{span}}(f_n)_{n \in A} = H$ . Precisamos, por vezes, de considerar sequências em  $H$  que, não sendo *frames* para  $H$ , ainda assim são *frames* para um sub-conjunto fechado e linear de  $H$ .

Vamos agora mencionar alguns exemplos de *frames*.

### Exemplo 1.1

Seja  $(e_n)_{n \in \mathbb{N}}$  uma base ortonormal de  $H$ .

- i. Por repetição de cada elemento em  $(e_n)_{n \in \mathbb{N}}$  várias vezes, obtemos



$$(f_k)_{k \in \mathbb{N}} = \{e_1, e_1, e_2, e_2, \dots\}$$

que constitui um *tight frame* de limite  $A=2$ . Se apenas  $e_1$  estiver repetido, obtemos

$$(f_k)_{k \in \mathbb{N}} = \{e_1, e_1, e_2, e_3, \dots\}$$

que é agora um *frame* com limites  $A=1$  e  $B=2$ .

ii. Seja  $(f_k)_{k \in \mathbb{N}} = \{e_1, \frac{1}{\sqrt{2}}e_2, \frac{1}{\sqrt{2}}e_2, \frac{1}{\sqrt{3}}e_3, \frac{1}{\sqrt{3}}e_3, \frac{1}{\sqrt{3}}e_3, \dots\}$  uma sequência onde

cada vector  $\frac{1}{\sqrt{k}} e_k$  se repete  $k$  vezes. Para cada  $f \in H$  temos

$$\sum_k |\langle f, f_k \rangle|^2 = \sum_k k |\langle f, \frac{1}{\sqrt{k}} e_k \rangle|^2 = \|f\|^2$$

pelo que  $(f_k)_{k \in \mathbb{N}}$  é um *tight frame* para  $H$  com limite  $A=1$ .

### 1.3. Bases de Riesz e Frames

Um *frame*  $(f_k)_{k \in \mathbb{N}}$  combina muitas das propriedades das bases de *Riesz* em espaços de Hilbert. Dado um  $f \in H$ , existe uma sequência de coeficientes  $(c_k)_{k \in \mathbb{N}} \in \ell^2(\mathbb{N})$ , tais que

$$f = \sum_k c_k f_k.$$

Já vimos que bases de *Riesz* constituem frames. Vamos agora apresentar condições que garantem que um dado frame é uma base de *Riesz*.

#### Teorema 1.1

Seja  $(f_k)_{k \in \mathbb{N}}$  um frame para  $H$ . Logo as condições seguintes são equivalentes.

- i.  $(f_k)_{k \in \mathbb{N}}$  é uma base de *Riesz* para  $H$ .
- ii.  $(f_k)_{k \in \mathbb{N}}$  é um *frame exacto*.
- iii.  $(f_k)_{k \in \mathbb{N}}$  é *minimal*.
- iv.  $(f_k)_{k \in \mathbb{N}}$  tem uma sequência *biortogonal*.

- v.  $(f_k)_{k \in \mathbb{N}}$  e  $(S^{-1}f_k)_{k \in \mathbb{N}}$  são biortogonais.
- vi. Se  $\sum_k c_k f_k = 0$  para uma dada sequência  $(c_k)_{k \in \mathbb{N}} \in \ell^2(\mathbb{N})$ , então  $c_k = 0$  para todos os  $k \in \mathbb{N}$ .

### **Proposição 1.4**

Seja  $(f_k)_{k \in \mathbb{N}}$  um frame para  $H$ . Para  $n \in \mathbb{N}$ , seja  $A_n$  a representação do valor óptimo do limite do frame para a sequência  $(f_k)_{k \in \mathbb{N}}$ . As seguintes condições são equivalentes:

- i.  $(f_k)_{k \in \mathbb{N}}$  é uma Base de Riesz para  $H$ .
- ii.  $(f_k)_{k \in \mathbb{N}}$  é linearmente independente e  $\inf_n A_n > 0$ .
- iii.  $(f_k)_{k \in \mathbb{N}}$  é linearmente independente e o  $\lim_n A_n$  existe e é nulo.

## **1.4 Dicionário**

### **1.4.1 Noção**

Um dicionário  $D$  para um espaço de sinais é uma qualquer coleção finita de sinais elementares com norma unitária. Denotaremos o dicionário  $D$  como  $D = \{\varphi_w : w \in \Omega\}$ , onde  $\Omega$  representa o conjunto de índices do dicionário; os sinais elementares de  $D$  dizem-se átomos, e denotados por  $\varphi_w$ , onde o parâmetro  $w$  pode ser interpretados como tempo-frequência ou sendo uma mera etiqueta. Representaremos por  $N$  o número de átomos do dicionário. Temos  $N = \#D = \#\Omega$ , onde  $\#$  significa o cardinal do conjunto.

Seja  $M$  a dimensão do espaço de sinais. Se o espaço gerado pelo dicionário coincide com o espaço de sinais, diremos que o dicionário está completo ou é total. Neste caso, todo o sinal pode ser aproximado, sem erro, via uma combinação linear dos seus átomos. Se os átomos formarem um conjunto linearmente dependente, então o dicionário diz-se redundante. Neste caso, um sinal pode ter um número infinito de boas aproximações. Para um dicionário ser completo é necessário que  $N \geq M$  e para ser redundante é suficiente que  $N > M$ .

De agora em diante, consideraremos sempre o caso em que um dicionário é completo e redundante.

### 1.4.2 Dicionários Ortonormais

É frequentemente útil representar sinais como vectores, o que permite simplificar a notação recorrendo ao uso das ferramentas de Álgebra Linear. Para introduzir a expansão de sinais em bases ortogonais convém lembrar que, de forma análoga aos espaços vectoriais, nos quais podemos definir o produto vectorial e a norma, também para funções podemos escrever o produto interno entre duas funções (tomando valores reais ou complexos) definidas no intervalo  $[a, b]$  como sendo  $\langle x_1, x_2 \rangle = \int_a^b x_1(t)x_2^*(t)dt$ , onde  $x^*(t)$  denota a conjugação de  $x(t)$ . Se as funções forem ortogonais, temos que o seu produto interno é nulo. A partir da definição de produto interno obtemos a norma induzida

$$\|x\|^2 = \int_a^b |x(t)|^2 dt.$$

Um conjunto de sinais diz-se ortonormal se os seus elementos forem mutuamente ortogonais e tiverem norma unitária. Necessitamos, adicionalmente, da desigualdade triangular

$$\|x + y\| = \|x\| + \|y\|, \text{ para todos } x, y,$$

e a desigualdade de Schwartz

$$\left| \int_a^b x_1(t)x_2^*(t)dt \right|^2 \leq \left( \int_a^b |x_1(t)|^2 dt \right) \left( \int_a^b |x_2(t)|^2 dt \right),$$

na qual se obtem igualdade sse  $x_1 = c x_2$ , onde  $c$  é uma constante.

Considere-se o espaço de dimensão finita  $C^n$ , dito espaço dos sinais, munido do produto interno (complexo) acima indicado. Os elementos deste espaço serão referidos, de agora em diante, como sinais.

Seja  $D$  um dicionário ortonormal usado na reconstrucção de um sinal em  $C^n$ , e suponha-se  $\#D=N$ . Dado um sinal de entrada  $s$ , considere-se a série  $\sum_w \langle s, \varphi_w \rangle \varphi_w$ .

Se ordenarmos os termos da série por forma a que os produtos internos não incrementem, em magnitude, então podemos aproximar a série por uma soma finita de  $m$  termos; diremos que temos uma *melhor aproximação de  $m$  termos do sinal de entrada*, sendo os coeficientes da aproximação os produtos internos (ordenados) que aparecem na soma finita.

Este procedimento pode ser implementado com um baixo custo numérico, da ordem  $O(nN)$ , que representa o tempo que demora a calcular todos os produtos internos entre o sinal e os elementos do dicionário. Em certos casos, é possível implementar este método de uma forma ainda mais eficiente, como seja, por exemplo, no caso em que a matriz síntese do dicionário é a transformada de *Fourier*, então o tempo de cálculo de todos os produtos internos entre o sinal e os elementos do dicionário pode ser reduzido para  $O(n \ln n)$ .

Tal permite inferir que é mais fácil de lidar com aproximações esparsas, se imossermos que os átomos do dicionário sejam “quase” ortogonais.

### 1.4.3 Dicionários de *Frames*

#### **Dicionário discreto de Gabor**

As funções de Gabor (ou funções Gaussianas seno-moduladas) permitem uma localização optimal do tempo-frequência.

Uma função de *Gabor* de variável real pode ser expressada como

$$g_{\gamma}(t) = K(\gamma) \exp(-\pi(1-w)^2/t) \sin[2\pi w(t-u)/n + \phi],$$

onde  $n$  é o tamanho do sinal, que condicionará o dicionário a construir,  $K(\gamma)$  é tal que  $\|g_{\gamma}\|=1$  e  $\gamma=\{u, w, s, \phi\}$  representa o conjunto dos parâmetros das funções do dicionário.

O comprimento do sinal ( $n$  pontos) sugere a necessidade de construção de escalas para os parâmetros das funções de *Gabor*.

Todavia, para cada escala fixa seremos confrontados com um espaço contínuo para as funções do dicionário, isto é, o dicionário tem dimensão infinita. Para ultrapassar este problema, considerar-se-á apenas, na prática, sub-conjuntos do dicionário.

Na abordagem original de Mallat e Zhang, os parâmetros dos átomos do dicionário construído foram escolhidos como sequências diádicas de inteiros. Deste modo, a sua amostragem é governada por um parâmetro extra-oitava  $j$  (inteiro). A escala  $s$ , que corresponde ao suporte, em tempo, de um átomo, é obtida da sequência diádica  $s=2^j$ ,  $0 \leq j \leq L$  (tamanho do sinal  $n=2^L$ ). Os parâmetros  $u$  e  $w$ , que correspondem à posição de um átomo em tempo e frequência, são dados para cada oitava  $j$ , com intervalo  $s=2^j$ .

**Dicionário diádico discreto de Gabor**

Um átomo de um dicionário de tempo-frequência pode ser expresso em termos de uma translação ( $u$ ), uma dilatação ( $s$ ) e de uma modulação ( $w$ ) de uma certa função  $g(t) \in L^2(\mathbb{R})$ , como seja

$$g(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) e^{itw}.$$

A definição óptima do tempo-frequência é obtida para a função gaussiana  $g(t) = \exp(-t^2)$ , a qual, para uma análise de sinais discretos reais, gera o dicionário de funções de Gabor:

$$g_{\gamma,\phi}(t) = K(\gamma, \phi) \exp(-\pi(w-u)^2/s) \sin[2\pi w(t-u)/n + \phi].$$

De novo,  $K(\gamma, \phi)$  é tal que  $\|g_{\gamma,\phi}\|=1$ , com os parâmetros discretos  $u=1, \dots, n$ ,  $w=1, \dots, n/2$ , e  $s=1, \dots, n$ , onde  $n$  é o tamanho do sinal. Todavia, é visível que, mesmo para um relativamente pequeno valor de  $n$ , o dicionário gerado é enorme para efeitos práticos de cálculo.

Consequentemente, na execução clássica proposta por Mallat e Zhang, os parâmetros dos átomos do dicionário são escolhidos como sequências diádicas para um sinal discreto de comprimento  $n=2^L$ . A escala  $s$ , correspondente ao suporte do átomo, é escolhida como uma escala diádica  $s=2^j$ ,  $0 \leq j \leq L$ . Os parâmetros  $u$  e  $w$ , correspondem à posição do átomo em tempo e frequência.

O tamanho deste dicionário (e a definição da decomposição) pode ser aumentado para uma amostra  $2^j$  ( $0 \leq j$ ) via os parâmetros de tempo e frequência. O dicionário resultante tem  $O(2^{2j} n \ln(n))$  átomos. O tempo e a frequência aumentam proporcionalmente, isto é,

$$\Delta t = 2^{-j} \frac{2^j}{f_s}, \Delta f = 2^{-j} \frac{f_s}{2^j},$$

onde  $f_n$  está em frequência com a amostragem do sinal analisado. A

definição óptima de tempo-frequência é entendida assim como a distância entre os centros dos átomos dentro da mesma escala e depende da oitava  $j$  (escala  $s=2^j$ ). A escala corresponde, por sua vez, ao suporte de um átomo em tempo / frequência. Podemos definir assim um suporte em tempo de um átomo em tempo-frequência como metade do suporte da função  $g(t)$ :

$$T_{\frac{1}{2}} = 2 \frac{2^j}{f_s} \sqrt{\frac{\ln 2}{\pi}}.$$

Apesar do *oversampling*, o algoritmo procura ainda a expansão do sinal num

subconjunto relativamente pequeno das funções possíveis do dicionário, tornando o algoritmo aplicável em tempo útil.

## 1.5 Aproximações lineares

O desenvolvimento de bases de *wavelets* ortonormais abriu uma nova ponte entre a Teoria da Aproximação e Processamento de Sinais. Contudo, o estudo da Teoria da Aproximação tem vindo a mudar desde 1935.

Em particular, as propriedades das aproximações não-lineares motivam a escolha e preferência por certos algoritmos para o processamento de sinais não-lineares.

### 1.5.1. Aproximações lineares (ver [4])

A aproximação linear projecta o sinal  $f$  sobre o espaço gerado por  $M$  vectores, escolhidos à priori na base ortonormal  $B = (g_m)_{m \in \mathbb{N}}$ . Teremos então a expansão:

$$f_M = \sum_{m=0}^{M-1} \langle f, g_m \rangle g_m.$$

Desde que a base seja ortonormal, o erro da aproximação é dado pela seguinte soma de série

$$e[M] = \|f - f_M\|^2 = \sum_{m=M}^{\infty} |\langle f, g_m \rangle|^2$$

A exactidão desta aproximação depende, claramente, das propriedades de  $f$  relativamente à base  $B$ .

Uma base de Fourier produz uma aproximação linear uniforme e eficiente para os sinais periódicos, desde que projectados sobre  $M$  sinusoidais de baixa frequência. É igualmente óbvio que o erro  $e[M]$  está, muitas vezes, relacionado com a regularidade global do sinal  $f$ .

Numa base de *wavelets*, o sinal é projectado sobre várias escalas de *wavelets*, o que equivale a efectuar aproximações de sinais com resolução fixa.

Aproximações lineares com *aliasing* uniforme em bases de *wavelets* têm propriedades e comportamento similares às em bases de Fourier.

A aproximação linear efectuada a uma dada escala não depende das anteriores, o que não acontece nas aproximações não lineares.

### **Estimativa do erro (wavelets/Fourier)**

A representação em série de Fourier funciona como uma sobreposição de senos e co-senos, e é usada tanto na análise e solução numérica de equações diferenciais como na análise e tratamento de comunicação de sinais. A transformada de Fourier permite analisar um sinal no domínio temporal em termos da sua frequência.

Esta transformação transforma uma função temporal numa função no domínio das frequências. O sinal pode então ser analisado pelo seu conteúdo, sendo que a frequência dos coeficientes da transformada representa a contribuição de cada função seno e/ou co-seno nessa frequência. A transformada inversa de Fourier faz exactamente o que esperaríamos, ou seja, traduz os dados do domínio de frequência para o domínio de tempo.

#### **1.5.2. Aproximações não lineares**

A aproximação linear  $f_M = \sum_{m=0}^{M-1} \langle f, g_m \rangle g_m$  pode ser melhorada se escolhermos  $M$  à posteriori, isto é, retirando/introduzindo vectores  $g_m$ , de acordo a pior/melhor aproximação obtida para  $f$ . A aproximação assim obtida com  $M$  vectores indexados em  $I_M$  é dada por  $f_M = \sum_{m \in I_M} \langle f, g_m \rangle g_m$ .

Tal como anteriormente, o erro da aproximação é dado por:

$$e[M] = \|f - f_M\|^2 = \sum_{m \notin I_M} |\langle f, g_m \rangle|^2.$$

Para minimizar este erro, escolhemos  $I_M$  como o conjunto dos índices dos  $M$  vectores  $g_m$  que têm amplitude exactamente  $|\langle f, g_m \rangle|$ . Este condicionamento na escolha dos vectores  $g_m$  faz com que a aproximação seja não-linear, uma vez que esta escolha se altera com a escolha de  $I_M$ .

A amplitude do produto interno numa base de *wavelets* está relacionada com a regularidade do sinal. Uma aproximação não-linear que use uma vasta base de *wavelets* será equivalente à construção de uma aproximação gradual, cuja resolução é implementada localmente nos pontos onde o sinal é irregular.

Se o sinal está isolado, a aproximação não-linear é muito mais precisa do que um esquema que mantenha a mesma resolução sobre o suporte do sinal. Os espaços das funções que são bem aproximadas por esquemas não-lineares de *wavelets* são mais vastos do que os

obtidos via esquemas lineares, e incluem funções singulares. As variações ilimitadas de sinais são importantes, constituindo modelos úteis para processamento de imagens.

Se não tivermos uma base ortonormal, é usado, muitas das vezes, o processo de Gram-Schmidt. Todavia, na prática tal não é aconselhável devido à instabilidade numérica deste processo.

### 1.5.3 Aproximações esparsas (ver [1,3,9])

O problema básico de uma aproximação esparsa está ligado à escolha de um conjunto adequado (base ou frame) para o espaço de Hilbert  $H$  de sinais. Pretende-se aqui uma aproximação do sinal que possua apenas um número finito de coeficientes não nulos na sua representação em termos dos elementos da base ou frame escolhido.

Especificamente, procuramos a solução que minimiza o problema

$$\min_{\#\Lambda=m} \min_b \|f - \sum_{m \in \Lambda} b_m \varphi_m\|,$$

onde  $\Lambda \subset I_M$  e  $b \in \ell^2(\Lambda)$  representa a família de coeficientes complexos. Para um  $\Lambda$  fixo, o mínimo deste produto interno pode ser obtido por técnicas simples, mas pouco precisas. A dificuldade real está numa selecção óptima do conjunto  $\Lambda$ , uma vez que a estratégia de “esgotar” todas as hipóteses  $\binom{N}{m}$  é, na prática, irrealizável.

Podemos distinguir quatro problemas básicos na aproximação esparsa, cada um partindo de pressupostos diferentes relativamente ao erro de aproximação e custos da representação dessa mesma aproximação. São estes:

- 1) Encontrar a representação para o sinal que seja esparsa;
- 2) Dado um sinal, encontrar o coeficiente do vector esparso  $\varphi_m$  que forneça uma aproximação do sinal com um erro aceitável;
- 3) De entre os coeficientes dos vectores com um aceitável nível de esparsidade, encontrar os que produzem a melhor aproximação para o sinal;
- 4) Dado um sinal, encontrar o coeficiente do vector  $\varphi_m$  que constitui um compromisso entre o desejo de esparsidade e a qualidade do erro.



## CAPÍTULO II - Algoritmos gananciosos

Nos últimos quinze anos têm sido propostos diferentes métodos para resolver os problemas de aproximação esparsa. Entre os métodos mais utilizados, contam-se os algoritmos gananciosos. Um algoritmo ganancioso para efeitos de aproximação esparsa constrói a aproximação do sinal passo por passo, seleccionando em cada passo o átomo com maior correlação com a parte residual do sinal.

### 2.1 Algoritmos gananciosos (ver [14])

Se um dicionário é ortonormal, é possível escolher esparsamente os átomos, maximizando ao mesmo tempo o produto interno entre estes e o sinal. Um procedimento habitual é de se escolher o átomo com maior correlação com o sinal, subtrair a contribuição do sinal (cálculo do erro da aproximação), e iterar. Os algoritmos gananciosos para aproximações esparsas refinam este procedimento de modo a torná-lo aplicável a classes de dicionários.

Primeiro, introduzimos o *Matching Pursuit*, que é uma extensão simples do algoritmo básico acima descrito. Em seguida, descreveremos o *Orthogonal Matching Pursuit*, o qual inclui uma condição de minimização por forma a aumentar a eficiência do método. Resta analisarmos o critério de paragem a adoptar. Sempre que em presença de um algoritmo interactivo, o critério de paragem mais simples é o do número de iterações, e dependente do problema de aproximação esparsa que estivermos a tentar resolver. Concluiremos esta secção com um pequeno comentário relativo à história dos algoritmos gananciosos.

### 2.1.1. *Matching Pursuit* (ver [2,5,6,10,11])

Vamos começar com o procedimento formal para o *Matching Pursuit* (MP). Fixemos um dicionário  $D$  e escolha-se um critério de paragem adequado ao problema em tratamento.

#### Algoritmo (*Matching Pursuit*)

Entrada:

Dimensão  $d$  do sinal  $s$

Saída:

Vector  $c$  dos coeficientes em  $C^d$

Procedimento:

- 1) Inicialmente, tomar o vector  $c=c(\gamma_0)=0$ , o resíduo  $r_0=s$ , e o índice do laço  $k=1$ ;
- 2) Determinar um índice  $\gamma_k$  tal que  $\gamma_k \in \arg \max_w |\langle r_{k-1}, \varphi_w \rangle|$ ;
- 3) Actualizar o coeficiente do vector:  $c(\gamma_k) \leftarrow c(\gamma_k) + \langle r_{k-1}, \varphi_{\gamma_k} \rangle$ ;
- 4) Calcular o novo resíduo  $r_k \leftarrow r_{k-1} - \langle r_{k-1}, \varphi_{\gamma_k} \rangle \varphi_{\gamma_k}$ .
- 5) Aumentar  $k$  para  $k+1$ ;
- 6) Se o critério de paragem não for verificado, então re-iniciamos o passo 2).

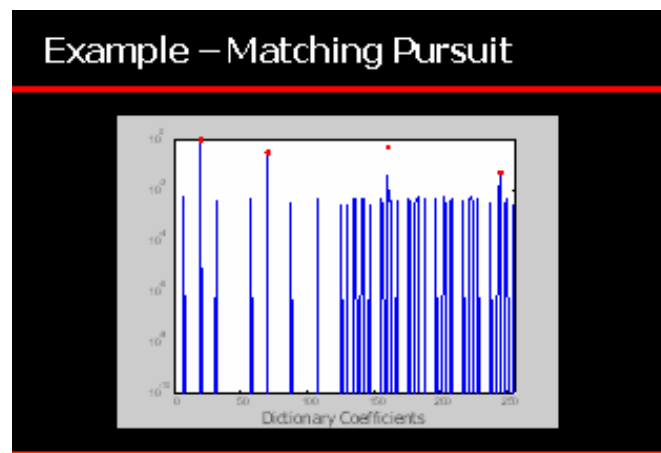
No passo 2) é realizada uma selecção gananciosa, onde as escolhas dos átomos são condicionadas pela existência de uma forte correlação com o resíduo do sinal. É de notar que, nos casos em que o dicionário não é ortogonal, este método pode seleccionar várias vezes o átomo com o mesmo índice. Essa repetição ocorre devido ao facto de o produto interno entre o átomo e o resíduo não registar as contribuições dos outros átomos para o resíduo.

O passo 3) procede à actualização do corrente coeficiente do vector, registando o átomo correspondente a  $\gamma_k$ .

No passo 4) calcula-se o novo resíduo, subtraindo a componente na direcção do átomo  $\varphi_{j_k}$ . Se o dicionário fôr completo, a norma do correspondente resíduo converge para zero quando  $k$  tende para infinito.

O laço é repetido até o critério de paragem ser satisfeito.

Vamos estimar o custo computacional do procedimento. A selecção gananciosa no passo 2) envolve o cálculo de todos os possíveis produtos internos entre o resíduo e os elementos do dicionário, pelo que requiere  $O(dN)$  operações. Nos passos 3) e 4) são necessárias apenas  $O(d)$  operações. Finalmente, se o procedimento fôr repetido  $T$  vezes, o custo do algoritmo é da ordem  $O(dTN)$ .



### 2.1.2 Orthogonal Matching Pursuit (ver [2,5,6,10,11])

Vamos concentrar-nos no algoritmo *Orthogonal Matching Pursuit* (OMP). Este introduz uma condição de minimização dos quadrados mínimos (minimização da energia) por forma a aumentar a eficiência da aproximação relativamente ao algoritmo MP. O objectivo é obter uma melhor aproximação relativamente aos átomos previamente escolhidos.

Considere-se, de novo, um dicionário fixo  $D$  e um certo critério de paragem.

#### **Algoritmo (Orthogonal Matching Pursuit)**

Entrada:

Dimensão  $d$  do sinal  $s$

Saída:

Vector  $c$  dos coeficientes em  $C^d$

Procedimento:

- 1) Inicialmente, o índice do conjunto  $\lambda_0=0$ , o resíduo  $r_0=s$ , e o índice do laço  $t=1$ ;
- 2) Determinar um índice  $\lambda_y$  tal que  $\lambda_y \in \arg\max_w |\langle r_{t-1}, \varphi_w \rangle|$ ;
- 3) Actualizar o coeficiente do conjunto:  $\Lambda_t \rightarrow \Lambda_t \cup \{\lambda_t\}$ ;
- 4) Encontrar a solução  $c$  para o problema do quadrado mínimo
 
$$\min_{c \in C^{\Lambda_t}} \left\| s - \sum_{j=1}^t c(\lambda_j) \varphi_{\lambda_j} \right\|.$$
- 5) Calcular o novo resíduo o novo  $c$ :  $r_t \rightarrow s - \sum_{j=1}^t c(\lambda_j) \varphi_{\lambda_j}$ ;
- 6) Incrementar o laço:  $t \rightarrow t+1$ .
- 7) Se o critério de paragem não for verificado, então retomamos o passo 2).

A anatomia do OMP é idêntica à do MP. O passo 2) selecciona um novo átomo usando o mesmo critério ganancioso que no PM, e no passo 3) adiciona um conjunto de listas dos átomos que têm de ser escolhidos. Os coeficientes actuais são calculados no passo 4) via a resolução do problema de mínimos quadrados. A solução deste determina a aproximação  $\alpha_t$  para o sinal:  $\alpha_t = \sum_{j=1}^t c(\lambda_j) \varphi_{\lambda_j}$ , enquanto que no passo 5) determinamos o novo resíduo. O laço repete-se até o critério de paragem ser satisfeito.

O comportamento de OMP difere significativamente do de MP. Observe-se que no algoritmo OMP, o passo 2) selecciona sempre átomos linearmente independentes relativamente aos átomos previamente escolhidos. Em consequência, o resíduo pode anular-se ao fim de um número finito de passos. Note-se, além disso, que a solução do problema de mínimos quadrados é única.

Estime-se agora o custo computacional do algoritmo.

O passo dois pode ser sempre implementado com  $O(dN)$  operações, envolvendo o cálculo de todos os produtos internos entre o resíduo e os elementos do dicionário. O problema de mínimos quadrados é calculável em  $t$  iterações; tendo em conta que o passo 4) requiere

apenas  $O(td)$  operações, a solução do problema de mínimos quadrados pode ser reduzida para  $(t-1)$  iterações. Não havendo possibilidade de aumentar a eficiência, o custo total do algoritmo é  $O(dT(T+N))$ , onde  $T$  é o número total de iterações.

### 2.1.3 Critérios de paragem

Para terminar a discussão dos dois algoritmos, resta estabelecer um critério para determinar quando devemos parar a iteração. Variando o critério de paragem, determinar-se-ão diferentes aproximações esparsas para os problemas em estudo. Existem três possibilidades:

➤ Critério EXACT

Parar quando a norma do resíduo  $r_t$  for nula. Este critério é apropriado para o problema da reconversão do sinal de entrada (*EXACT*).

➤ Critério do ERROR

Parar quando a norma do resíduo  $r_t$  diminuir abaixo de uma especificação limite. Este critério é apropriado para o erro do problema da aproximação (*ERROR*).

➤ Critério SPARSE

Parar após a selecção de  $m$  átomos distintos. Este método é apropriado para resolver o problema da aproximação esparsa (*SPARSE*).

Suponha-se que o dicionário é ortonormal. Então não é difícil verificar que MP e OMP, equipados com essas regras de paragem, são ambos algoritmos correctos para os respectivos problemas da aproximação esparsa, isto é, ambos fornecem uma solução ao fim de um número finito de passos.

### 2.1.4 Um contra exemplo para MP

Contudo, em geral não é de esperar que os métodos gananciosos tenham sucesso na obtenção de uma aproximação esparsa. Com efeito, existem exemplos na literatura onde ambos, MP e OMP, são incapazes de fornecer a representação esparsa que se sabe o sinal possuir sobre o dicionário.

Suponhamos que  $D$  é um dicionário ortonormal para  $C^d$ , onde  $d \geq 3$ . Se juntarmos o átomo  $\psi = \alpha \left[ \varphi_1 + \varphi_2 + \sum_{n=3}^d \frac{1}{n-2} \varphi_n \right]$ , onde  $\varphi$  é escolhido tal que  $\psi$  tem norma unitária, e executarmos o algoritmo *MP* relativamente ao sinal de entrada  $s = \varphi_1 + \varphi_2$ , então o algoritmo poderá continuar indefinidamente com um erro de aproximação  $\|s - \alpha_t\|_2 \approx t^{-\frac{1}{2}}$ .

No entanto, o sinal de entrada tem, claramente, dois termos na sua representação. Este sinal também pode ser usado para a construção de exemplos onde o OMP precisa de todos os  $d$  passos por forma a construir o sinal que tenha somente dois termos na representação.

### 2.1.5 Breve história dos algoritmos gananciosos

Os algoritmos gananciosos para aproximações esparsas apareceram pela primeira vez na literatura estatística, durante o ano de 1950, onde foram usados para resolver os problemas de subconjuntos seleccionados. A técnica base, chamada *forward selection* é uma versão mais elaborada do MP.

O algoritmo MP foi inventado em 1981 sobre o nome *Projection Pursuit Regression*. Este algoritmo foi introduzido na literatura relativa a processamento de sinais por Mallat e Zhang, que o conotaram *Matching Pursuit*.

O OMP foi estudado independentemente por vários matemáticos e físicos. A primeira referência a este surge em 1989, com Chen, Billings e Luo. O primeiro processamento de sinal usando OMP aparece por volta de 1993.

#### **Método dos Frames (ver [6,7])**

O método dos *Frames* (MOF) selecciona, de entre todas as soluções do sistema  $\Phi \alpha = s$ , os coeficientes que possuem norma  $l^2$  mínima:  $\min \|\alpha\|$  para  $\Phi \alpha = s$ .

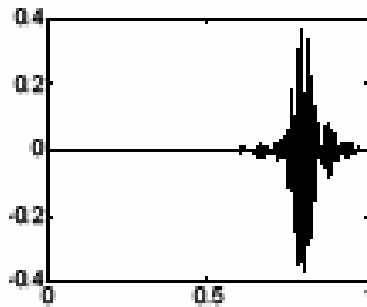
A solução deste problema é única, e representada por  $\alpha^\dagger$ . A colecção de todas as soluções de  $\Phi \alpha = s$  é um sub-espço fechado de  $IR^p$ . O MOF selecciona o elemento desse sub espaço com norma  $l^2$  mínima. Algumas vezes este elemento é também chamado *solução de duração mínima*. Há uma matriz  $\Phi^\dagger$ , inversa generalizada de  $\Phi$ , que calcula a solução de duração mínima para um sistema de equações lineares:  $\alpha^\dagger = \Phi^\dagger s = \Phi^\dagger (\Phi \Phi^\dagger)^{-1} s$ .

No caso de dicionário de *frames*, o MOF é avaliado em forma fechada.

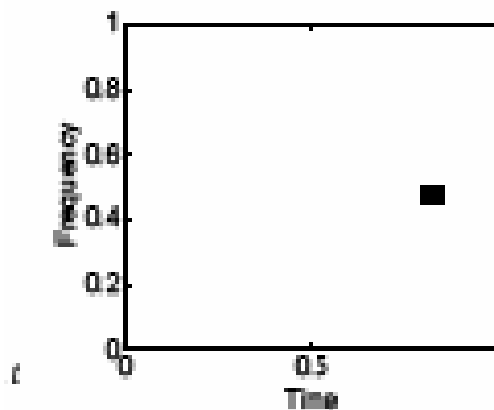
Bom exemplo: o dicionário standard de caixa de *wavelets*. Podemos calcular todos os vectores  $v$ , e teremos  $\|\Phi^T v\|^2 = L_n; L_n = \log_2(n)$ . Em resumo,  $\Phi^t = L_n^{-1} \Phi^T$ . Repare-se que  $\Phi^T$  é um operador.

Há dois problemas chaves com o Método dos *Frames*. Primeiro, o MOF não preserva a esparsidade: se o sinal tiver uma representação esparsa em termos do dicionário, os coeficientes determinados via MOF serão, muito provavelmente, menos esparsos, uma vez que cada átomo do dicionário para o qual o produto interno com o sinal dado é não nulo constitui, a menos de um potencial, um membro da solução procurada.

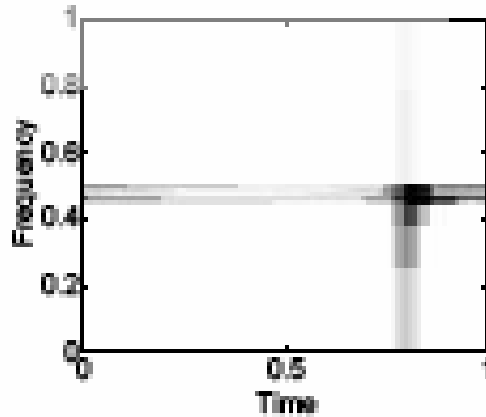
A figura seguinte



mostra-nos o sinal de um átomo de hidrogénio num dicionário de *wavelets*. A imagem seguinte mostra um exemplo da janela tempo-frequência de um certo átomo do dicionário de *wavelets*,



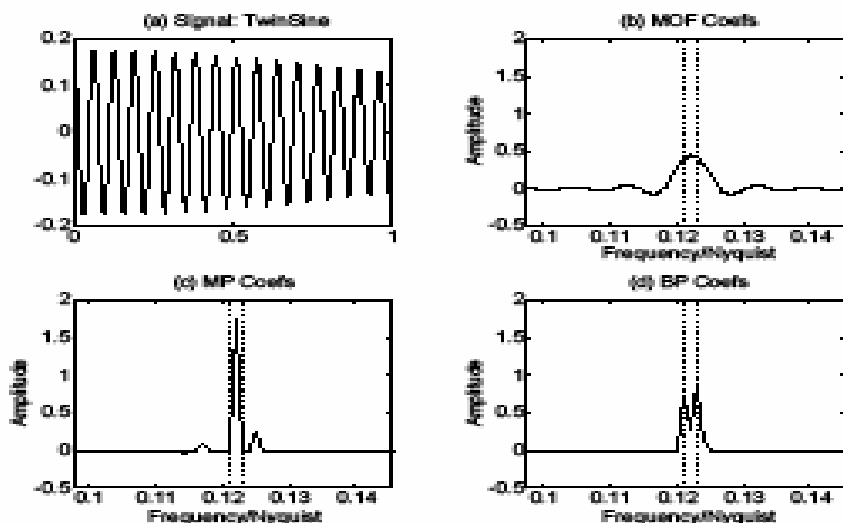
sendo que o resultado da decomposição do sinal relativamente a esse átomo do dicionário de *wavelets*, visível na figura seguinte:



Sumariando, enquanto que o sinal pode ser sintetizado por um único átomo do dicionário, a sua decomposição em termos do *frame* pode envolver vários outros átomos.

Em segundo lugar, o MOF é um algoritmo complexo, de resolução limitada. Nenhum sinal pode ser reconstruído com mais características do que as admitidas na construção do operador  $\Phi^\dagger \Phi$ . Para vermos isto, suponha-se que o sinal se encontra está claramente localizado em  $\chi\{\gamma=\gamma_0\}$ . Todavia, a reconstrução pode não estar em  $\chi\{\gamma=\gamma_0\}$ , mas estar, em vez disso, no intervalo  $\Phi^\dagger \Phi \chi\{\gamma=\gamma_0\}$  o qual, em casos de dicionários completos, pode ser programado por forma a “cair fora” do intervalo desejado.

As figuras seguintes representam um sinal que consiste na sobreposição de duas sinusóides, separadas de fase  $2\pi/n$ .



Podemos analisar a quarta imagem em termos do dicionário discreto completo. Neste caso, a reconstrução por MOF, na figura superior direita, constitui uma simples convolução com o núcleo de Dirichlet. O resultado é que os coeficientes obtidos possuem uma larga oscilação



aparente, consistindo não em duas, mas em várias frequências, mas dando indícios visuais de que o objecto pode ser sintetizado com apenas duas frequências.

Vamos agora descrever, e comparar de seguida, o método de *Basis Pursuit* (BP) com os métodos anteriormente mencionados.

## 2.2 O Método de *Basis Pursuit*: Descrição e Comparação

Vamos aqui descrever o método de *Basis Pursuit* (BP) e fazer a sua comparação com o método de *frames*.

O método BP determina as representações completas do sinal, em termos dos dicionários, via optimizações convexas: isso processa-se através da decomposição que minimiza a norma  $l^1$  do vector-coeficientes que ocorre na representação. Devido à não diferenciabilidade da norma  $l^1$ , esta optimização (essencial ao método) precisa de uma decomposição, que se revela muito diferente da efectuada pelo método de *frames* (MOF) – em particular esta última deve ser muito esparsa.

O BP pode ser usado para a resolução da optimização de problemas, utilizando uma combinação quadrática com uma norma  $l^p$  do vector-coeficientes. Recentemente, tem sido aplicada também uma programação de ponto interior, onde, com recurso a certos dicionários, é possível resolver o problema da optimização muito rapidamente.

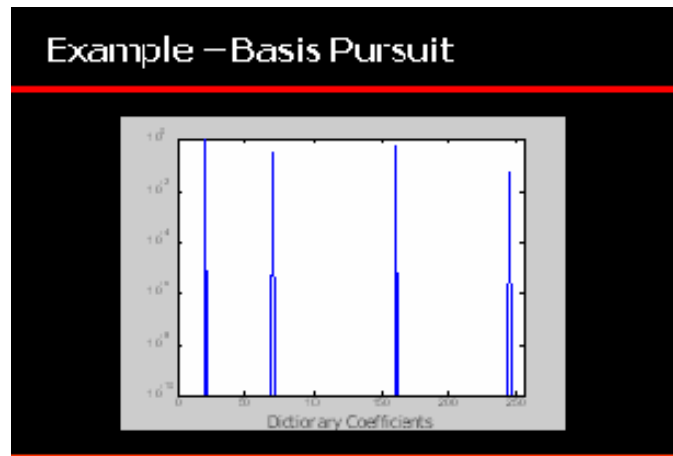
Vamos agora focar os problemas das representações completas. Assuma-se que o dicionário é completo, isto é, o sinal  $s$  tem representações  $s = \sum_{\gamma} \alpha_{\gamma} \phi_{\gamma}$ .

O princípio do BP é encontrar uma representação do sinal tal que minimize a norma  $l^1$  dos coeficientes. Formalmente,

$$\min \|\alpha\|_1 \quad \text{sujeito a} \quad \Phi\alpha = s.$$

De um ponto de vista formal, este problema é idêntico ao MOF (ver mais à frente), na medida em apenas se substitui a norma  $l^2$  por a norma  $l^1$ . O MOF é usado como método base nos problemas de optimização quadrática com igualdades lineares de comparação, e envolve apenas a solução de um sistema de equações lineares. Já o BP, em contrapartida, requiere

uma solução convexa para o problema da otimização não quadrática, o que envolve consideravelmente mais esforço e sofisticação.



### ***Basis Pursuit versus Orthogonal Matching Pursuit (ver [5, 7, 10])***

À primeira vista, os métodos OMP e BP parecem diferentes. O OMP é um algoritmo iterativo, cujo esquema não busca uma meta concreta, consistindo numa regra simples, passível de aplicar repetidamente. Em contraste, o BP assenta num princípio global de otimização, sem ligação a um algoritmo específico. Analisar o contraste entre estes dois algoritmos pode ser instrutivo. O OMP parte de um “modelo vazio” para construir uma representação do sinal, adicionando passo a passo um átomo ao modelo anterior. Já o BP inicia-se com um modelo prévio (isto é, uma representação preliminar do sinal na base) e, iterativamente, progride para o modelo completo, retirando termos que não sejam usados no modelo, e trocando por termos novos e mais adequados à representação. Assim, OMP pode ser visto como uma construção aproximada, enquanto que o BP é uma aproximação por ordenação/troca.

### ***Bases Pursuit versus Método dos Frames***

Como já dizemos atrás, o MOF e o BP diferem apenas na norma usada no processo de otimização (norma  $l^2$ , no primeiro caso, norma  $l^1$  no segundo). Assim, o BP pode ser entendido como um refinamento em termos de norma do MOF.

## CAPÍTULO III - Representação Esparsa do Sinal

É sabido que a aplicação dos algoritmos gananciosos para aproximações esparsas produz, em geral, uma aproximação convergente para o sinal de entrada (assumindo que o dicionário é completo). Todavia, vimos anteriormente um contra-exemplo em que, dado um sinal representável por dois átomos do dicionário, o MP construía para este uma representação contendo vários outros átomos, sem alteração significativa no erro de aproximação.

Um recente resultado de Gilbert, Muthurishnan e Strauss [18] para o OMP sugere um procedimento para ultrapassar esta dificuldade, permitindo assim a consideração dos algoritmos gananciosos para cálculo efectivo de aproximações esparsas de sinais. Os autores provaram que OMP permite identificar  $m$  termos da representação de um sinal arbitrário com um erro a menos de um factor constante do erro óptimo, na condição de  $m$  ser suficientemente pequeno. Em particular, OMP resolve este problema correctamente, com o critério de paragem EXACT, para todo o sinal que possua uma representação suficiente esparsa. O mesmo acontece para o critério de paragem SPARSE, e usando dicionários incoerentes.

Iremos mostrar que o OMP resolve correctamente o problema da aproximação esparsa para os diferentes critérios EXACT, ERROR e SPARSE, se bem que os resultados dependam significativamente das propriedades geométricas do dicionário.

A primeira secção considera o comportamento do OMP com paragem EXACT, no problema da determinação de representações esparsas do sinal de entrada. Daremos uma condição que assegura que uma selecção gananciosa identifica todos os átomos essenciais à representação esparsa óptima do sinal e mostraremos que essa condição não pode ser melhorada.

A segunda secção é dedicada a uma análise semelhante, agora para sinais que não possuam uma representação esparsa. Nas terceira e quarta secções, aplicaremos estes resultados para os critérios de paragem ERROR e SPARSE. Em particular, provaremos que

OMP, com ERROR, pode produzir uma representação que usa apenas os átomos óptimos e cujo erro de aproximação é igual, a menos de uma constante, ao erro óptimo.

### 3.1. Representação esparsa do sinal (ver [9])

Começamos com o problema da determinação da representação esparsa do sinal de entrada  $s$ . Formalmente, o problema é expresso como

$$\min_{c \in \mathbb{C}^d} \|c\|_0 \quad \text{para} \quad \phi c = s \quad \text{EXACT}$$

Suponhamos que o sinal de entrada  $s$  pode ser escrito como uma combinação linear de  $m$  átomos e não menos do que. Portanto,

$$s = \sum_{\lambda \in \Lambda_{opt}} c_\lambda \varphi_\lambda$$

onde  $\Lambda_{opt}$  é um subconjunto de  $\Omega$  com cardinalidade  $m$ . Repare-se que os átomos associados aos índices de  $\Lambda_{opt}$  são linearmente independentes entre si e os coeficientes  $c_\lambda$  são não-nulos. Se existir uma representação para o sinal, não pode usar menos do que  $m$  átomos.

Seja  $\phi_{opt}$  a representação da matriz síntese associada ao sub-dicionário  $\{\varphi_\lambda, \lambda \in \Lambda_{opt}\}$ .

Então, o sinal pode ser expresso como  $s = \phi_{opt} c_{opt}$ , onde  $c_{opt} \in \mathbb{C}^{\Lambda_{opt}}$ . Uma vez que os átomos óptimos são linearmente independentes,  $\phi_{opt}$  tem ainda demasiadas colunas. Defina-se agora uma segunda matriz  $\Psi_{opt}$  cujas colunas correspondem aos  $N-m$  átomos indexados por  $\Omega \setminus \Lambda_{opt}$ . Deste modo, asseguramos que  $\Psi_{opt}$  contem os átomos que não participam na representação óptima.

#### 3.1.1. Selecção gananciosa dos átomos

O fundamental num algoritmo ganancioso é escolha (dita, gananciosa) dos átomos que “melhor” aproximam o sinal. O nosso objectivo é desenvolver uma condição que assegure que essa escolha gananciosa coincide com a escolha dos átomos do conjunto óptimo indexado de  $\Lambda_{opt}$ .

Recorde-se que a selecção gananciosa determina um conjunto  $\gamma$  satisfazendo a

$$\gamma \in \arg \max_{w \in \Omega} |\langle r, \varphi_w \rangle|,$$

sendo  $r$  o resíduo do sinal. Observe-se que o vector  $\phi_{opt}^* r$  corresponde ao produto interno entre o resíduo do sinal e os átomos óptimos para a representação deste. A norma  $\|\phi_{opt}^* r\|_\infty$  dá-nos o valor máximo que estes produtos podem tomar. De forma idêntica,  $\|\Psi_{opt}^* r\|_\infty$  expressa o

máximo que o produto interno entre resíduo e um átomo não-óptimo pode tomar. Em consequência, o quociente entre estas quantidades expressa o equilíbrio da selecção gananciosa,

$$\rho(r) = \|\Psi_{opt}^* r\|_{\infty} / \|\phi_{opt}^* r\|_{\infty} \quad (3.1)$$

Vemos que a escolha gananciosa ocorrer num átomo óptimo se e só se  $\rho(r) < 1$ . O seguinte lema fornece uma condição suficiente para a confirmar a escolha de um átomo óptimo.

**Lema 3.1**

Uma condição suficiente para que  $\rho(r) < 1$  ocorra é que  $ERC(\Lambda_{opt}) > 0$ , sendo

$$ERC(\Lambda_{opt}) = 1 - \max_{\psi} \|\phi_{opt}^T \Psi_{opt}\|_1.$$

**Demonstração:**

A expressão (3.1) é semelhante à usada para o cálculo de norma de operadores, facto que iremos utilizar de seguida. Relembro que  $\phi_{opt} \phi_{opt}^T$  é um projector ortogonal no espaço das colunas de  $\phi_{opt}$ . Uma vez que a projecção é conjugada simétrica de si própria, temos

$$(\phi_{opt}^T)^* \phi_{opt}^* r = \phi_{opt} \phi_{opt}^T r = r.$$

Portanto, podemos calcular que

$$\rho(r) = \|\Psi_{opt}^* r\|_{\infty} / \|\phi_{opt}^* r\|_{\infty} = \|\Psi_{opt}^* (\phi_{opt}^T)^* \phi_{opt}^* r\|_{\infty} / \|\phi_{opt}^* r\|_{\infty} \leq \|\Psi_{opt}^* (\phi_{opt}^T)^*\|_{\infty, \infty}.$$

Desde que  $\|\cdot\|_{\infty, \infty}$  seja igual à soma máxima e absoluta da linha desse argumento e a norma  $\|\cdot\|_{1,1}$  igual à soma máxima e absoluta da coluna desse argumento, tomaremos uma transposta conjugada com as normas trocadas. Assim, temos

$$\rho(r) \leq \|\phi_{opt}^T \Psi_{opt}\|_{1,1} = \max_{\psi} \|\phi_{opt}^T \Psi_{opt}\|_1$$

onde a maximização ocorre agora sobre as colunas de  $\Psi_{opt}$  correspondentes aos átomos não óptimos. □

Este lema é suficiente simples de aplicar e usar, mas mostra que OMP pode determinar os átomos óptimos para uma representação esparsa de sinais.

### 3.1.2. Teorema da Recuperação Exacta

Usando o Lema 3.1 é fácil de provar uma condição suficiente para a recuperação da representação esparsa do sinal de entrada.

#### Teorema 3.1 (Recuperação exacta para OMP)

Suponha-se que  $c_{opt}$  é a representação esparsa do sinal de entrada  $s$ , com o conjunto de índices  $\Lambda_{opt}$ . A condição

$$ERC(\Lambda_{opt}) > 0$$

é suficiente para que o OMP recupere  $c_{opt}$  ao fim de  $\#\Lambda_{opt}$  passos.

#### Demonstração:

Suponha-se que, depois de  $t$  iterações, o OMP seleccionou exactamente  $t$  átomos óptimos. Esta hipótese é trivialmente satisfeita à primeira iteração. Vamos agora desenvolver uma condição para garantir que o átomo seleccionado de seguida (iteração  $t+1$ ) é também átomo óptimo.

Seja  $r_t$  o resíduo do sinal ao fim de  $t$  iterações. O OMP selecciona agora um átomo óptimo se a razão da selecção gananciosa  $\rho(r_t)$  for inferior a 1. Já o algoritmo mostra que o resíduo  $r_t$  é dado pela diferença entre o sinal  $s$  e a combinação linear dos  $t$  átomos previamente escolhidos. Sendo o sinal uma combinação linear de átomos indexados por  $\Lambda_{opt}$ , a hipótese indutiva mostra que o resíduo  $r_t$  corresponde à respectiva coluna de  $\phi_{opt}$ . O Lema 3.1 mostra que  $\rho(r_t) < 1$  para  $ERC(\Lambda_{opt}) > 0$ . Portanto, esta última condição garante a que OMP selecciona um átomo óptimo ao passo  $t+1$ . □

Desde que esteja assegurado que o OMP escolhe um novo átomo, o método identificará todos os  $m$  átomos óptimos que constituem a representação esparsa do sinal dado. Assim, dizemos que o vector-resposta obtido via o OMP é um sinal perfeito.

### Corolário 3.1

Assuma-se que o sinal pode ser representado usando uma lista de átomos de  $\Lambda_{opt}$ . Se  $ERC(\Lambda_{opt}) > 0$ , então o MP selecciona um átomo de  $\Lambda_{opt}$  durante várias iterações.

#### 3.1.3 Avaliação da Coerência

Uma vez que o Teorema 3.1 pressupõe o conhecimento, à priori, dos átomos óptimos, o seu resultado parece inútil. Note-se, todavia, que para alguns dicionários, a condição  $ERC(\Lambda_{opt}) > 0$  é satisfeita por conjuntos de índices  $\Lambda$  cuja cardinalidade é suficientemente pequena para propósitos práticos.

Defina-se o parâmetro coerência  $\mu$  como sendo o máximo do produto interno entre dois átomos distintos:

$$\mu = \max_{j \neq k} |\langle \varphi_{w_j}, \varphi_{w_k} \rangle|$$

e estabeleça-se a nova função

$$\mu_1(m) = \max_{|\Lambda|=m} \max_{w \notin \Lambda} \sum_{\lambda \in \Lambda} |\langle \varphi_w, \varphi_\lambda \rangle|.$$

Por convenção, estabelecemos  $\mu_1(0) = 0$ .

#### Proposição 3.1

Suponha-se que  $\#\Lambda \leq m$ . Temos para o coeficiente de recuperação exacta que

$$ERC(\Lambda) > \frac{1 - \mu_1(m-1) - \mu_1(m)}{1 - \mu_1(m-1)}.$$

Adicionalmente, se  $\mu_1(m-1) + \mu_1(m) < 1$ , ter-se-á  $ERC(\Lambda) > 0$ .

Note-se que para qualquer sub-dicionário de bases ortonormais, o coeficiente da recuperação exacta  $ERC$  é 1. Combinando agora o resultado do Teorema 3.1 com esta proposição, obtemos o seguinte corolário.

**Corolário 3.2**

Se  $\mu_1^{(m-1)} + \mu_1^{(m)} < 1$ , o OMP determina o sinal perfeito, usando o critério de paragem EXACT, desde que o sinal de entrada seja constituído por  $m$  termos do dicionário. Teremos assim que os  $m$  termos da representação são únicos.

**Demonstração**

De acordo com a Proposição 3.1, a condição  $\mu_1^{(m-1)} + \mu_1^{(m)} < 1$  garante que se tem  $ERC(\mathcal{A}) > 0$  para todo o conjunto indexado  $\mathcal{A}$  de cardinalidade  $m$ . Se a representação esparsa do sinal de entrada tiver uma representação usando os  $m$  átomos indexados em  $\mathcal{A}$ , o Teorema 3.1 garante que o uso do critério EXACT no OMP estabelece a recuperação óptima do sinal de entrada. Mais, essa representação será única.

□

Uma outra consequência é que OMP pode também recuperar sinais que possuam representações esparsas sobre dicionários incoerentes desde que  $\mu_1(m) < m\mu$ , isto é, desde que seja satisfeita uma condição em termos do parâmetro coerência  $\mu$ .

**Corolário 3.3**

Para  $m < \frac{1}{2}(\mu^{-1} + 1)$  o OMP, com o critério EXACT, recupera de forma perfeita e única o sinal de entrada que tenha uma representação em termos de  $m$  átomos do dicionário.

No caso dos dicionários possuírem mais estrutura, as condições suficientes acima descritas podem ser melhoradas. Como exemplo, tome-se o seguinte corolário.

**Corolário 3.4**

Suponha-se que o dicionário é formado por  $J$  bases ortonormais, com parâmetro de coerência  $\mu$  e suponha-se que  $m < \left(\sqrt{2} - 1 + \frac{1}{2(J-1)}\right)\mu^{-1}$ .

Então o OMP, com o critério EXACT, recupera de forma perfeita e única o sinal de entrada que tenha uma representação com  $m$  termos sobre o dicionário.



### 3.1.4 Necessidade de Recuperação Exacta

Há necessidade de se exigir uma recuperação exacta? A resposta é afirmativa, como iremos provar de seguida.

#### **Teorema 3.2 (Recuperação exacta)**

*Considere-se um sub-dicionário indexado por  $\Lambda_{opt}$  para o qual  $ERC(\Lambda_{opt}) \leq 0$ , e assumam-se que as representações sobre  $\Lambda_{opt}$  são únicas.*

*Então existirá um sinal de entrada, com representação sobre  $\Lambda_{opt}$ , que não pode ser obtido pelo OMP com o critério EXACT.*

#### **Demonstração**

A prova segue linhas idênticas à do Teorema 3.1. Pelas propriedades das representações esparsas sobre  $\Lambda_{opt}$ , todo o sinal que tenha uma representação em termos de átomos indexados em  $\Lambda_{opt}$ , induzirá as mesmas matrizes  $\phi_{opt}$  e  $\Psi_{opt}$ .

Uma vez que  $ERC(\Lambda_{opt}) \leq 0$ , teremos agora  $\|\phi_{opt}^T \Psi_{opt}\|_{1,1} \geq 1$ . Escolhendo  $b_{bad}$  em  $\mathbb{C}^{\#\Lambda_{opt}}$  tal que

$$\|\Psi_{opt}^* (\phi_{opt}^T)^* b_{bad}\|_{\infty} / \|b_{bad}\|_{\infty} = \|\Psi_{opt}^* (\phi_{opt}^T)^*\|_{\infty, \infty},$$

donde resulta  $\|\Psi_{opt}^* (\phi_{opt}^T)^* b_{bad}\|_{\infty} / \|b_{bad}\|_{\infty} \geq 1$ .

A matriz  $\phi_{opt}$  tem colunas e linhas cheias, pelo que esta contém um sinal  $s_{bad}$  para o qual que  $\phi_{opt}^* s_{bad} = b_{bad}$ . Concluimos assim que  $\rho(s_{bad}) = \|\Psi_{opt}^* s_{bad}\|_{\infty} / \|\phi_{opt}^* s_{bad}\|_{\infty} \geq 1$ , e o OMP com  $s_{bad}$  como entrada, escolherá um átomo não óptimo no primeiro passo. Sendo a representação de  $s_{bad}$  sobre  $\Lambda_{opt}$  única, a representação obtida por OMP será incorrecta.

## 3.2 Identificar os Átomos para uma Aproximação

Nesta última secção vamos mostrar que o comportamento dos algoritmos gananciosos não difere quando o sinal de entrada não possui uma representação esparsa. O argumento usado na demonstração do Lema 3.1 requer que o vector-resíduo  $r$  se situe no espaço gerado

pelos átomos óptimos. Em consequência, não é aplicável ao tratamento de sinais com representações “fora” do espaço gerado pelos átomos óptimos.

Vamos agora mostrar uma condição que permite o tratamento sinais de entrada arbitrários. Para o efeito, temos de estudar o desempenho da selecção gananciosa quando aplicada a sinais gerados por átomos arbitrários, quando a selecção se refere a um sub-dicionário com índices em  $\Lambda$ .

Fixemos um sinal arbitrário  $s$  com representação  $a$  num dicionário indexado por  $\Omega$ , considere-se um sub-dicionário indexado por  $\Lambda$  e suponha-se  $ERC(\Lambda) > 0$ . Defina-se, de novo,  $\phi_\Lambda$  como sendo a matriz cujas colunas são átomos indexados por  $\Lambda$ , e seja  $\psi_\Lambda$  a representação da matriz cujas colunas são os átomos indexados por  $\Omega \setminus \Lambda$ . A melhor aproximação para o sinal de entrada sobre os átomos indexado por  $\Lambda$  pode ser escrita como  $a_\Lambda = P_\Lambda a$ , onde  $P_\Lambda$  é a projecção ortogonal para o espaço gerado pelos átomos indexados por  $\Lambda$ , ou seja, pelas colunas de  $\phi_\Lambda$ .

**Lema 3.2 (Recuperação geral)**

Recordemos que, a razão da selecção gananciosa é dada por:

$$\rho(r) = \frac{\|\psi_{opt}^* r\|_\infty}{\|\phi_{opt}^* r\|_\infty}$$

e que o máximo da correlação entre o sinal  $r$  e o dicionário é definido como

$$\max cor(r) = \frac{\|\phi^* r\|_\infty}{\|r\|_2}$$

Suponhamos que  $a$  é um vector gerado pelas colunas de  $\phi_\Lambda$ . A condição

$$\|s - a\|_2 < \left[ 1 + \left( \frac{\max cor(s - a_\Lambda) \|\phi_\Lambda^T\|_{2,1}}{ERC(\Lambda)} \right)^2 \right]^{\frac{1}{2}} \|s - a_\Lambda\|_2 \tag{3.2}$$

é suficiente para garantir que  $\rho(s - a) < 1$ .

Por outras palavras, o lema estabelece que a selecção gananciosa escolherá átomos de  $\Lambda$  sempre que o vector  $a$  não possuir uma comparação favorável com a melhor aproximação do sinal sobre o sub-dicionário.

### Demonstração

Podemos dividir  $\rho(s-a)$  em duas partes, que calcularemos separadamente.

$$\begin{aligned} \rho(s-a) &= \frac{\|\psi_{\Lambda}^*(s-a)\|_{\infty}}{\|\phi_{\Lambda}^*(s-a)\|_{\infty}} = \frac{\|\psi_{\Lambda}^*(s-a_{\Lambda}) + \psi_{\Lambda}^*(s-a)\|_{\infty}}{\|\phi_{\Lambda}^*(s-a_{\Lambda}) + \phi_{\Lambda}^*(a_{\Lambda}-a)\|_{\infty}} \leq \frac{\|\psi_{\Lambda}^*(s-a_{\Lambda})\|_{\infty}}{\|\phi_{\Lambda}^*(a_{\Lambda}-a)\|_{\infty}} + \frac{\|\psi_{\Lambda}^*(a_{\Lambda}-a)\|_{\infty}}{\|\phi_{\Lambda}^*(a_{\Lambda}-a)\|_{\infty}} \\ &= \rho_{err} + \rho_{opt}. \end{aligned} \quad (3.3)$$

O termo  $\phi_{\Lambda}^*(s-a_{\Lambda})$  desaparece do denominador desde que  $(s-a_{\Lambda})$  seja ortogonal ao sub-espço das colunas de  $\phi_{\Lambda}$ .

Desde que  $(a_{\Lambda}-a)$  pertença ao sub-espço das colunas de  $\phi_{\Lambda}$ , o Lema 3.1 garantirá

$$\rho_{opt} \leq 1-ERC(\Lambda). \quad (3.4)$$

Para o segundo termo, sabemos que a seguinte desigualdade é válida toda a matriz  $A$

$$\min_{\substack{z \in R \\ z \neq 0}} \frac{\|Az\|_q}{\|z\|_q} \geq \|A^T\|_{q,p}^{-1}.$$

Assim, temos

$$\rho_{err} \leq \frac{\|\psi_{\Lambda}^*(a_{\Lambda}-a)\|_{\infty}}{\|\phi_{\Lambda}^T\|_{2,1}^{-1} \|a_{\Lambda}-a\|_2}.$$

Desde que  $(s-a_{\Lambda})$  seja ortogonal ao sub-espço gerado pelas colunas de  $\phi_{\Lambda}$ , podemos escrever o numerador dessa fracção como

$$\rho_{err} \leq \frac{\|\phi_{\Lambda}^*(a_{\Lambda}-a)\|_{\infty} \|\phi_{\Lambda}^T\|_{2,1}}{\|a_{\Lambda}-a\|_2}. \quad (3.5)$$

Combinando as equações (3.3), (3.4) e (3.5), resulta que  $\rho(s-a) < 1$  para

$$\frac{\|\phi_{\Lambda}^*(a_{\Lambda}-a)\|_{\infty} \|\phi_{\Lambda}^T\|_{2,1}}{\|a_{\Lambda}-a\|_2} < ERC(\Lambda)$$

ou seja,

$$\left\| \phi^*(s - a_\Lambda) \right\|_{\infty} \frac{\left\| \phi_\Lambda^T \right\|_{2,1}}{ERC(\Lambda)} < \|a_\Lambda - a\|_2.$$

Elevando ambos os termos da desigualdade ao quadrado, adicionando  $\|s - a_\Lambda\|_2^2$  em ambos os membros, e aplicando o Teorema de Pitágoras ao lado direito, obtemos finalmente

$$\|s - a_\Lambda\|_2^2 + \left( \left\| \phi^*(s - a_\Lambda) \right\|_{\infty} \frac{\left\| \phi_\Lambda^T \right\|_{2,1}}{ERC(\Lambda)} \right)^2 < \|s - a\|_2^2.$$

Pondo em evidência o factor  $\|s - a_\Lambda\|_2^2$ , e usando a identificação da correlação máxima de  $(s - a_\Lambda)$  com o dicionário, obtemos

$$\left[ 1 + \frac{\left( \max_{\Lambda} \text{cor}(s - a_\Lambda) \left\| \phi_\Lambda^T \right\|_{2,1} \right)^2}{ERC(\Lambda)} \right] \|s - a_\Lambda\|_2^2 < \|s - a\|_2^2.$$

□

### 3.3 O erro e a restrição da aproximação esparsa

O Lema 3.2 tem aplicação imediata para o controlo do erro no problema da aproximação esparsa. Para um sinal de entrada  $s$  e um erro de tolerância  $\varepsilon$ , pretende-se resolver

$$\min_{c \in \mathcal{C}^a} \|c\|_0 + \frac{1}{2} \|s - \phi_c\|_2 \varepsilon^{-1} \quad \text{para} \quad \|s - \phi_c\|_2 \leq \varepsilon. \quad \text{ERROR}$$

Suponha-se que o vector  $c_{opt}$  resulta do algoritmo aplicado, e seja  $a_{opt} = \phi c_{opt}$  a correspondente aproximação do sinal no dicionário. Fixe-se  $\Lambda_{opt} = \text{supp}(c_{opt})$  e suponha-se que  $ERC(\Lambda_{opt}) > 0$ .

### Teorema 3.3

Suponha-se que o OMP pára quando a norma do resíduo  $r_t$  satisfaz a desigualdade

$$\|r_t\|_2 \leq \left[ 1 + \left( \frac{\max_{\Lambda_{opt}} \text{cor}(s - a_{opt}) \|\phi_{opt}^T\|_{2,1}}{\text{ERC}(\Lambda_{opt})} \right)^2 \right]^{\frac{1}{2}} \varepsilon.$$

Logo, todo o átomo que o OMP escolheu pertence ao conjunto de índices  $\Lambda_{opt}$

Por outras palavras, o OMP com o critério ERROR calcula sempre uma representação do sinal que usa apenas os átomos óptimos e o resíduo realiza só um factor que tenha o erro pior. Podemos discutir essa constante detalhadamente depois de demonstrar o teorema.

### Demonstração

Suponha-se que, após  $t$  iterações, o OMP seleccionou  $t$  átomos distintos de  $\Lambda_{opt}$ . A hipótese é satisfeita trivialmente à primeira iteração.

Represente-se por  $r_t$  o resíduo ao fim de  $t$  iterações. Na iteração  $t+1$  o OMP seleccionará um outro átomo óptimo se e só se  $\rho(r_t) < 1$ .

A descrição do algoritmo mostra que o resíduo  $r_t$  é igual à diferença entre o sinal  $s$  e a combinação linear dos  $t$  átomos previamente escolhidos. O Lema 3.2 estabelece que OMP seleccionará outro átomo de  $\Lambda_{opt}$  na condição de a norma do resíduo satisfazer

$$\|r_t\|_2 > \left[ 1 + \left( \frac{\max_{\Lambda_{opt}} \text{cor}(s - a_{opt}) \|\phi_{opt}^T\|_{2,1}}{\text{ERC}(\Lambda_{opt})} \right)^2 \right]^{\frac{1}{2}} \|s - a_{opt}\|_2.$$

Enquanto esta condição for satisfeita, o algoritmo continua, assegurando assim que todo o átomo escolhido pertence à lista  $\Lambda_{opt}$ , e dando, em simultâneo, um limite superior para a norma do resíduo, ou um erro para a aproximação. Desde que a aproximação  $a_{opt}$  represente uma solução do OMP com critério de paragem ERROR e parâmetro  $\varepsilon$ , mostra-se que

$$\|s - a_{opt}\|_2 \leq \varepsilon.$$

□

A condição no Teorema 3.3 depende de vários factores. Em particular, o coeficiente da recuperação exacta de  $\Lambda_{opt}$  pode ser positivo, e o limite do erro melhora significativamente quando  $ERC(\Lambda_{opt})$  é limitado. Por causa do termo  $\|\phi_{opt}^T\|_{2,1}$ , o limite do erro aumenta quando o sub-dicionário aumenta. O erro também aumenta quando o resíduo óptimo está mal correlacionado com o dicionário.

Uma questão prática é estimar o tamanho do sub-dicionário. Suponha-se que o sinal tem uma boa aproximação para um pequeno sub-dicionário, indexado por  $\Lambda_{opt}$ , com um aceitável coeficiente de recuperação exacto. Se o dicionário é incoerente, é de esperar que o máximo da correlação entre o resíduo e o dicionário tenha ordem  $d^{1/2}$  e que o termo  $\|\phi_{opt}^T\|_{2,1}$  seja de ordem  $\sqrt{\#\Lambda_{opt}}$ . Nestas condições, a constante será menor do que  $\sqrt{2}$ .

Uma aplicação interessante deste teorema é quando usado com dicionários ortonormais. Neste caso, o coeficiente de recuperação exacta e o termo  $\|\phi_{opt}^T\|_{2,1}$  são ambos iguais a um. A correlação máxima entre o resíduo e o dicionário nunca será superior a um. Portanto, o Teorema 3.3 implica que devemos parar o algoritmo quando o resíduo da aproximação satisfizer  $\|r\| \leq \varepsilon\sqrt{2}$ . Isso conduz também a que não seja possível melhorar este limite.

Considere-se o sinal  $s = \varepsilon(\varphi_\lambda + \varphi_\xi)$ , com norma do resíduo da aproximação a satisfazer  $\varepsilon\sqrt{2}$  (porque o dicionário é ortonormal). O Teorema 3.3 recomenda que se páre o algoritmo e se re-inicie com aproximação nula. Este conselho é, em parte, correcto: note-se que a representação  $\varepsilon\varphi_\lambda$  nos dá uma  $\varepsilon$ -representação óptima para  $s$  no sub-dicionário indexado por  $\Lambda_{opt}$  (ao qual  $\lambda$  pertence). Mas se aplicarmos de novo o OMP ao sinal  $s$ , a selecção gananciosa pode “hesitar” entre escolher de novo o átomo associado a  $\lambda$ , ou um associado a  $\xi$ . Para evitarmos a última escolha, que pode não pertencer a  $\Lambda_{opt}$ , não usamos um número par de passos no algoritmo.

### 3.3.1 A avaliação da coerência

Usando o Teorema 3.3, descrevemos agora alguns métodos para escolha do critério de paragem à priori.

**Corolário 3.5**

Suponha-se que  $\#\Lambda_{opt} = m$ , e a norma do resíduo  $r_t$  resultante do OMP satisfaz

$$\|r_t\|_2 \leq \left[ 1 + \frac{m[1 - \mu_1(m-1)]}{[1 - \mu_1(m-1) - \mu_1(m)]^2} \max cor(s - a_{opt})^2 \right]^{\frac{1}{2}} \varepsilon.$$

Então, todo o átomo escolhido pelo OPM está na lista de  $\Lambda_{opt}$ .

Usando as propriedades da coerência da função acumulativa, podemos desenvolver uma versão simples para este corolário.

**Corolário 3.6**

Suponha-se que  $\#\Lambda_{opt} = m$ , e que  $\mu_1(m) < 1/3$ . Se a norma do resíduo  $r_t$  resultante do OMP satisfaz

$$\|r_t\|_2 \leq \varepsilon \sqrt{1 + 6m \max cor(s - a_{opt})^2},$$

então todo o átomo escolhido pelo OPM estará na lista de  $\Lambda_{opt}$ .

## CAPÍTULO IV- Aplicação prática

Neste capítulo descrevemos um possíveis dicionários para detecção de *chirps*, isto é, sinais analíticos com frequência instantânea. Estes sinais ocorrem, por exemplo, se quisermos detectar o som emitido por um submarino. Um *chirp* pode ser visto como uma sinusóide em que a frequência aumenta ou diminui com o tempo, e estão frequentemente associados sinais emitidos por sonares e radares.

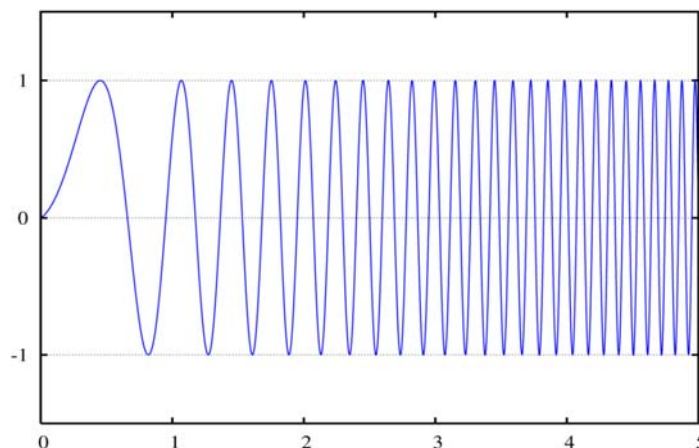
### 4.1 Exemplos de *chirps*

#### 4.1.1. *Chirp* linear

Um *chirp* em que a frequência  $f(t)$  varia linearmente com tempo  $f(t) = f_0 + kt$ , onde  $f_0$  é a frequência inicial e  $k$  o valor do aumento da frequência, diz-se um *chirp* linear. A função correspondente será assim

$$x(t) = \sin\left(2\pi \int_0^t f(s) ds\right) = \sin\left(2\pi (f_0 + (k/2)t) t\right)$$

com gráfico





### 4.1.2. Chirp geométrica

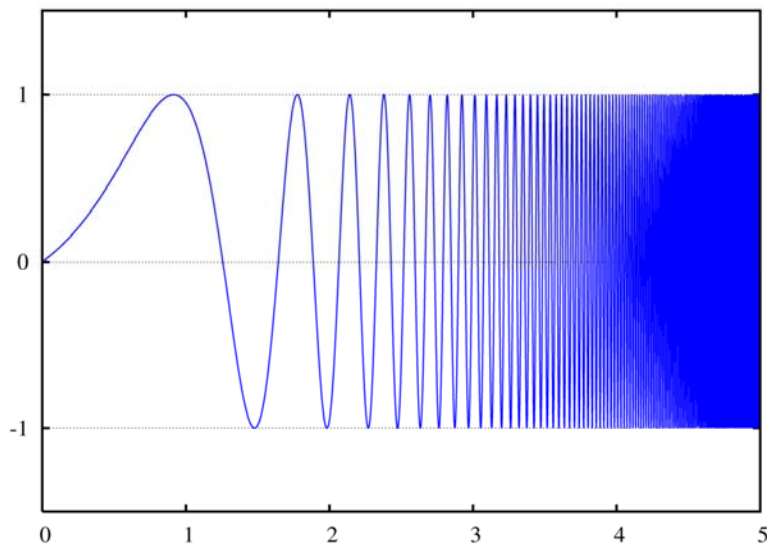
Num *chirp* geométrico, a frequência do sinal varia em relação geométrica com o tempo. Por outras palavras, dados dois pontos de uma onda,  $t_1$  e  $t_2$ , se o intervalo de tempo  $t_2-t_1$  é constante, o quociente das frequências  $f(t_2)/f(t_1)$  também é constante.

### 4.1.3. Chirp exponencial

Um *chirp* diz-se exponencial se a frequência do sinal varia exponencialmente em função do tempo, ou seja, se  $f(t) = f_0 k^t$ , onde, de novo,  $f_0$  é a frequência inicial e  $k$  é o aumento da frequência exponencial. O *chirp* correspondente assim a

$$x(t) = \sin\left(2\pi f_0 \int_0^t k^s ds\right) = \sin\left(\frac{2\pi f_0}{\ln k}(k^t - 1)\right)$$

com gráfico



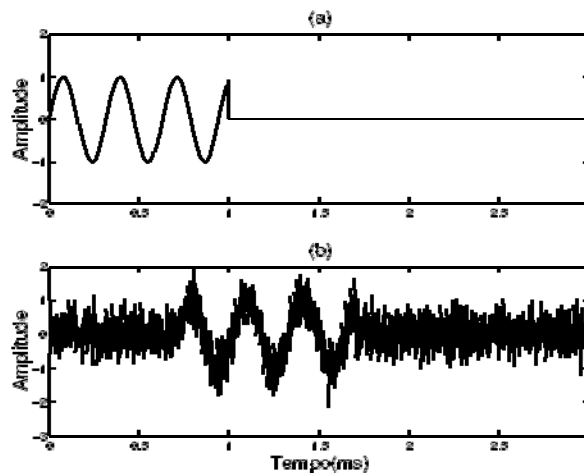
### 4.1.4. Princípio de funcionamento dos radares

O radar é um sistema que permite detectar objectos, determinando também a sua posição, direcção de deslocamento e velocidade. Consiste na emissão de um impulso electromagnético que, reflectindo-se no objecto, é retransmitido de volta como um eco.

Conhecendo o intervalo de tempo  $T$  entre o impulso emitido e o eco recebido, a velocidade de propagação  $c$  das ondas electromagnéticas na atmosfera, é possível determinar a distância  $R = cT/2$  à qual o objecto se encontra da antena emissora.

Na prática, este cálculo não é fácil. Primeiro, se o objecto estiver distante, o eco recebido tem amplitude muito baixa; por outro lado, o sinal é também deformado devido à reflexão no objecto-alvo. Pode também acontecer que não exista um só objecto, mas vários objectos sobrepostos, pelo que o impulso reflectido, corresponderá a uma multitude de ecos distintos, tornando difícil a tarefa de isolar o eco desejado.

A figura seguinte mostra os gráficos para amplitude-tempo de um impulso emitido (imagem (a)) e do eco recebido (imagem (b)). Pode ver-se em (b) que não é claro quando se inicia o eco, o que leva a uma incerteza para  $T$  e, conseqüentemente, para o valor de  $R$ .



Visivelmente  $T$  não pode ser determinado com exactidão e, na prática, teremos apenas  $r$  uma estimativa para a transformada de Fourier  $T^{\wedge}$  de  $T$ .

O caso de sonares é muito semelhante ao do dos radares. Uma das principais diferenças é a natureza das ondas usadas: ondas sonoras. Na água, por exemplo, estas propagam-se muito melhor do que as ondas electromagnéticas dos radares (inúteis neste meio). Outra diferença essencial entre as ondas sonoras e as electromagnéticas consiste na velocidade de propagação  $c$  das ondas, que para as electromagnéticas é sensivelmente constante, mas para as acústicas varia com a distância e o meio.

## 4.2 Detecção de chirps (ver [14,17,18])

### 4.2.1. A questão dos dicionários

Os dicionários a usar podem variar entre a típica base de Fourier (senos e co-senos), ou serem mais refinados, por exemplo, os de *wavelets*, *frames* de Gabor, ou do tipo

$$f(t) = A(t) \cos(\lambda \phi(t))$$

onde a amplitude  $A(t)$  e a fase  $\varphi(t)$  são funções suaves em tempo, cuja oscilação  $\lambda$  é grande. Um dicionário em termos de base de Fourier está particularmente concebido para descrever sinais periódicos, ou seja, com uma certa regularidade. Sinais cuja amplitude e frequência variam rapidamente em tempo não são recuperados por um dicionário deste tipo.

Consideremos o átomo  $f(t) = \cos(n\pi t^2/2)$ , com início em  $t=1/N$ . Este oscila rapidamente pelo que é particularmente adequado para descrever sinais que oscilem violentamente, como sejam os *chirps*. O mesmo pode ser dito dos *frames* de Gabor.

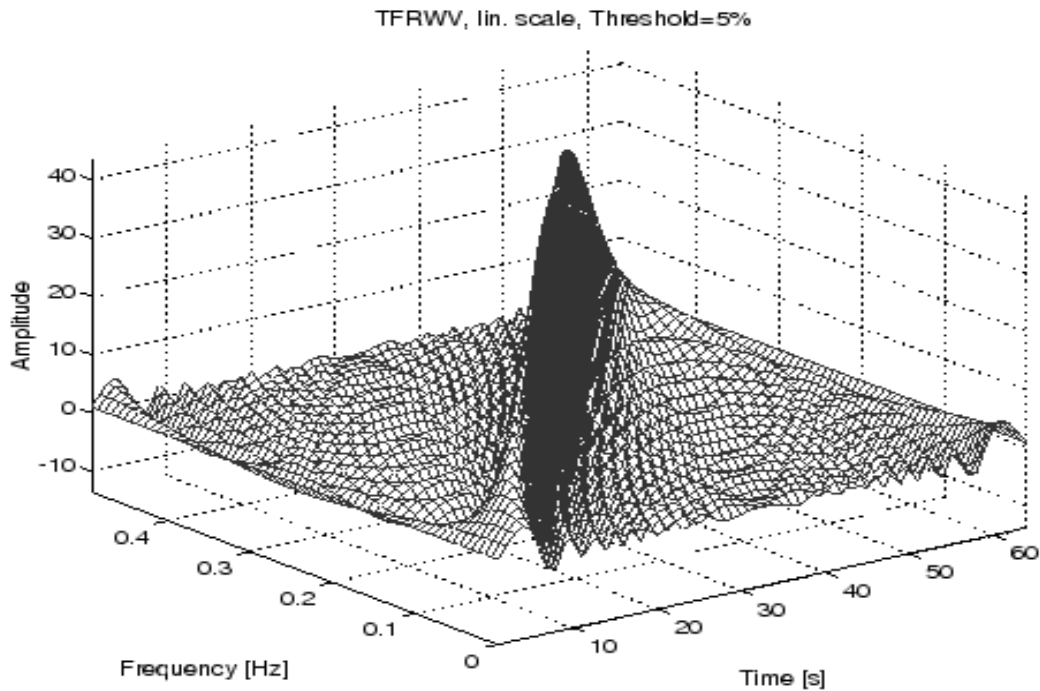
Já as *wavelets* têm muitas aplicações em processamento de sinais cuja oscilação, não sendo periódica, não é demasiado violenta. As *wavelets* de Daubechies, pela sua ortogonalidade e suporte compacto, são frequentemente empregues nestes casos. As *wavelets* de Gabor revelam-se úteis para aplicações em segmentação de imagens. As ferramentas de implementação variam de aplicação para aplicação, mas nota-se um recurso a técnicas de filtragem para implementação da transformada *wavelet*. O trabalho pioneiro de Mallat (1989) usava as *wavelets* para determinação de textura e análise fractal de imagens. Já Stollnitz, entre outros, mostraram as capacidades selectivas das *wavelets* no realce de imagens, alterando aspectos globais das linhas de contorno sem afectar os detalhes na imagem. As *wavelets* têm sido aplicadas em remoção de ruído, reconhecimento e extracção de padrões, e realce de características de sinais dados. Têm sido empregues também na interpretação de imagens e sinais biomédicos, na detecção de sinais estelares e em análise sísmica.

### **Distribuição Wigner-Ville**

A transformada de Wigner-Ville (WVD) foi introduzida por Eugene Wigner, em 1932, e é definida como

$$W_z(t, \nu) = \int_{-\infty}^{+\infty} x(t+t/2) x^*(t-t/2) e^{-j2\pi\nu t} dt.$$

Esta transformada satisfaz um largo número de propriedades matemáticas. Em particular, a WVD preserva o tempo e frequência do sinal  $x$ , como pode ser visto na figura seguinte, que representa a transformada WVD de um *chirp* linear.



#### 4.2.1. Transformada *Chirplet*

O termo *transformada chirplet* foi introduzido por Steve Mann, em 1991. Por outras palavras, se uma *wavelet* é entendida como parte de um sinal, já uma *chirplet* é uma parte de um sinal sonoro. Mais precisamente, uma transformada *chirplet* constitui uma “janela de enquadramento” para o sinal *chirp*, e que permite assim localizar certas propriedades. Tal como as *wavelets*, as *chirplets* são transformações no tempo-frequência que dependem de rotações e de outras acções no domínio temporal.

Considere-se um sinal de tamanho  $N$ . Introduza-se o intervalo  $M=[0,1]$ , e para cada  $j \geq 0$ , seja  $I_{jk}$  o intervalo diádico  $I_{jk} = [k2^j, (k+1) 2^j]$ , onde  $k=0,1, \dots, 2^{j-1}$ .

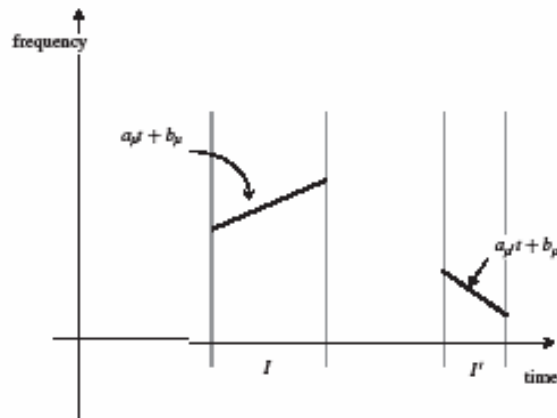
Definimos o dicionário *chirplet* como sendo a família de funções definidas por

$$f_{j,k,\mu}(t) = 2^{j/2} \exp[ i(a_{\mu}t^2 / (2+b_{\mu}t) ) \chi_M(t)],$$

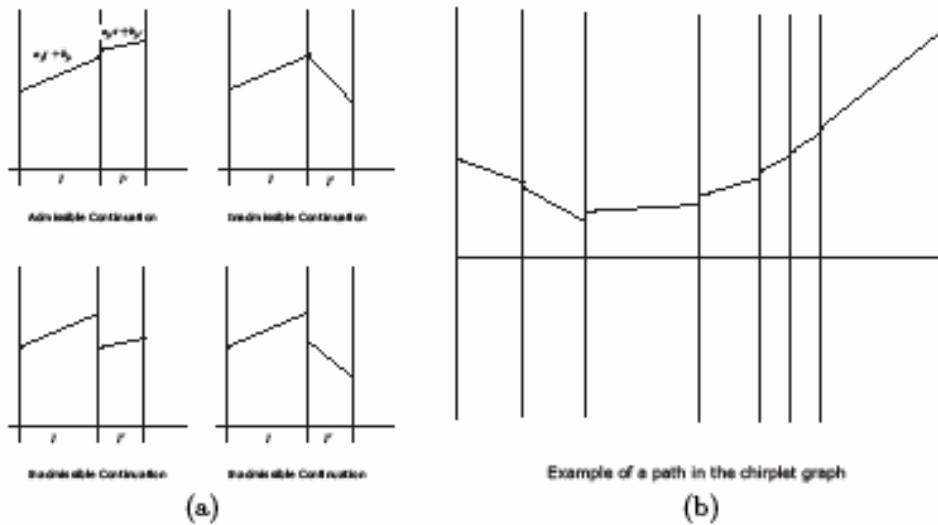
onde  $a_{\mu}, b_{\mu} \in M$ . Os parâmetros  $j, k, \mu$  podem depender da escala e conter informação de interesse sobre os objectos em análise. Efectue-se a normalização dos *chirplets*:  $\| f_{j,k,\mu} \| = 1$ .

Este sistema tem a vantagem de, após aplicada a WVD, fornecer indicação no tempo-frequência sobre as direcções possíveis, sua localização, frequência média, e, mais importante, retratar *chirplets* com mudanças lineares da frequência instantânea durante o intervalo de

geração  $I_{jk}$ . Isto permite pensar, inicialmente, na frequência da *chirplet* como linear e dada por  $a_{\mu}t+b_{\mu}$  (ver imagem abaixo).



As *chirplets* são “coladas” impondo uma certa regularidade à sua frequência instantânea.



Assim, dois *chirplets* podem ser ligados se forem adjacentes no mesmo intervalo de tempo, se a frequência da ligação é pequena e se a diferença entre os seus declives não muito grande. A ideia é de “moldar” um *chirp* com frequência instantânea  $\lambda\varphi(t)$  como uma sequência de segmentos lineares de tamanho infinitesimal.

## CONCLUSÃO

Esta tese menciona algumas das contribuições actuais para a teoria da aproximação esparsa de sinais, nomeadamente na aplicação de algoritmos gananciosos para representação e recuperação de sinais, por forma a termos um erro mínimo e uma aproximação óptima.

Estes métodos foram desenvolvidos tendo em vista aplicações práticas envolvendo detecção e subsequente representação de sinais, por vezes de fraca intensidade (sinais estelares), outras vezes demasiado irregulares para serem expressos em séries de Fourier. Como exemplos, ficam o problema da transmissão de sinais a partir de um avião, ou a captação dos sons emitidos por golfinhos. Todos estes procedimentos podem ser efectuados através na aproximação esparsa de sinais, mais precisamente, através de *chirps*.

## BIBLIOGRAFIA

[1] A. Rahmoune, P. Vandergheynst, P. Frossard - *Sparse Approximation Using M-Term Pursuits with Applications to Image and Video Compression*, Signal Processing institute Technical Report TR-ITS-2005.003, January 2005

[2] C. Rozell, D. Johnson, R. Baraniuk, *Locally competitive algorithms for sparse approximation*, **em:** *Proceedings of the International Conference on Image Processing (ICIP)*, San Antonio, TX, September 2007.

[http://www.ece.rice.edu/~dhj/rozell\\_icip2007.pdf](http://www.ece.rice.edu/~dhj/rozell_icip2007.pdf)

[3] C. Kotsakis, *Multiresolution aspects of linear approximation methods in Hilbert Spaces using Gridded data*, PHD Thesis, University of Calgary June 2000

[4] S. S. Chen, D. L. Donoho, M. A. Saunders, *Atomic Decomposition by Basis Pursuit*, SIAM Journal on Scientific Computing, 20 (1998) 1, 33 – 61

[5] N. A. Ziyad, E. T. Gilmore, M. F. Chouikha, *Dictionary approaches to image compression and reconstruction*, NASA Technical Report 19990009121, 1998.

<http://citeseer.ist.psu.edu/417232.html>

[6] L. Lovisolo, E. A. B. da Silva, P. S. R. Diniz, *Construção de frames exponenciais decrescentes para algoritmos de decomposição adaptativos de sinais*, **em:** XXI Simpósio Brasileiro de Telecomunicações - 2004, Belém, 2004

[7] A. Ferreira, *O procedimento de Gram-Schmidt: definição, exemplos e aplicações*, IST - Lisboa, 15/Out/2001,

<http://www.deetc.isel.ipl.pt/analisedesinai/tss/Bibliografia/Tutorials/GramSchmidt.pdf>

- [8] M. F. Duarte, M. A. Davenport, M. B. Wakin, R. G. Baraniuk, *Sparse signal detection from incoherent projections*, Proceedings of the International Conference on Acoustics Speech and Signal Processing (ICASSP), 2006, Toulouse, France, 305 – 308.
- [9] V. Guigue, A. Rakotomamonjy, S. Canu, *Kernel Basis Pursuit*, em: Actes de CAP 05, Conference francophone sur l'apprentissage automatique 2005, François Denis (Ed.), (2005) 93 – 106.
- [10] S. G. Mallat, Z. Zhang, *Matching Pursuits with time-frequency dictionaries*, IEEE Transactions on Signal Processing 41(42), December 1993.
- [11] J. Tropp, *Greed is good: Algorithmic results for sparse approximation*, IEEE Trans, Info. Theory, 50, (2004), 2231 – 2242.
- [12] St. Hengstler, D. P. Kasilingam, A. H. Costa, *A novel chirp modulation spread for multiple access Spread Spectrum Techniques and Applications*, 2002, IEEE Seventh International Symposium on volume 1, (2002) 73 - 77
- [13] J. Tropp, *Topics in Sparse Approximation*, PhD dissertation Computational and Applied Mathematics Univ. Texas at Austin Aug, 2004
- [14] U. Kähler, Apontamentos da Disciplina de Análise Complexa Aplicada (manuscripto), 2005.
- [15] O. Christensen, *An introduction to frames and Base de Riesz*, Birkhäuser Boston, (2002).
- [16] E. J. Candès, P. Charlton and H. Helgason, *Detecting highly oscillatory signals by chirplet path pursuit*, Appl. Comput. Hammon Anal, 24 (2006), 14 - 40
- [17] E. J. Candès, *Multiscale chirplets and near-optimal recovery of chirps*, Technical Report, Stanford University, 2002.
- [18] A. C. Gilbert, S. Muthukrishnan, M. Strauss, *Approximation of functions over redundant dictionaries using coherence*, SODA, 2003, 243-252.