

Universidade Técnica de Lisboa

Instituto Superior Técnico

Instituto de Telecomunicações – Pólo de Lisboa

Relatório Técnico

*“Estimação e verificação de parâmetros reais e inteiros sobre
modelo de observação linear”*

Março 2000

Artur Ferreira

ÍNDICE

| | |
|--|-----------|
| I. RESUMO | 3 |
| 1. DESCRIÇÃO GERAL DO PROBLEMA | 4 |
| 1.1 MODELO DE OBSERVAÇÃO..... | 4 |
| 1.2 COMPONENTES DO PROBLEMA | 5 |
| 1.2.1 <i>Estimação</i> | 5 |
| 1.2.2 <i>Verificação da estimação</i> | 9 |
| 1.3 CARACTERÍSTICAS DO PROBLEMA | 9 |
| 2. FORMULAÇÃO DO PROBLEMA DA ESTIMAÇÃO | 10 |
| 2.1 ANÁLISE DO CRITÉRIO DE MÁXIMA VEROSIMILHANÇA - <i>ML</i> | 10 |
| 2.2 EXPRESSÕES INDIVIDUAIS DE MÁXIMA VEROSIMILHANÇA..... | 12 |
| 2.3 ANÁLISE DOS PARÂMETROS A ESTIMAR | 13 |
| 2.3.1 <i>Mapeamento probabilístico</i> | 13 |
| 2.3.2 <i>Células de Voronoi</i> | 15 |
| 3. PROBLEMA DA VERIFICAÇÃO..... | 15 |
| 3.1 LIMITES INFERIOR E SUPERIOR DE <i>PC</i> | 16 |
| 3.2 PARÂMETROS REAIS..... | 18 |
| 4. PROBLEMA DA ESTIMAÇÃO..... | 18 |
| 4.1 FORMULAÇÃO A PARTIR DOS DADOS DO PROBLEMA DA VERIFICAÇÃO | 18 |
| 4.2 ALGORITMO SUB-ÓPTIMO PARA A ESTIMAÇÃO Z_{ML} | 19 |
| 4.3 ALGORITMO DE OPTIMIZAÇÃO GLOBAL | 21 |
| 5. EXTENSÕES AO PROBLEMA INICIAL | 23 |
| 5.1 CONHECIMENTO À <i>PRIORI</i> DA DISTRIBUIÇÃO DE <i>X</i> | 23 |
| 5.2 DADOS RECEBIDOS DE FORMA SEQUENCIAL (<i>STREAM-DATA</i>) | 23 |
| 6. ALGORITMO COMPLETO PARA A ESTIMAÇÃO DE PARÂMETROS INTEIROS | 25 |
| 7. CONCLUSÕES | 25 |
| 8. BIBLIOGRAFIA E REFERÊNCIAS | 26 |
| ANEXO A | 27 |
| ANEXO B | 30 |

I. Resumo

O presente trabalho tem como objectivos o estudo da estimação de parâmetros e respectiva verificação partindo de modelo linear de observação, que engloba parâmetros inteiros e reais. A estimação de parâmetros inteiros pode ser encarada como detecção, na medida em que se pode obter o valor exacto do parâmetro, isto é, sem aproximação.

Sobre o modelo de observação procura-se mostrar como o problema teórico da estimação por máxima verosimilhança (*maximum-likelihood*) que é classificado como *NP-hard*, pode ser resolvido de forma eficiente do ponto de vista computacional.

A verificação da estimativa dos parâmetros também se apresenta como um problema bastante difícil de resolver de forma eficiente em termos computacionais. Torna-se por isso necessário analisar o problema da verificação e desenvolver técnicas que permitam efectuar essa verificação em tempo polinomial (útil). Mostra-se como as medidas utilizadas para verificação são semelhantes ao processo da quantização vectorial, e explora-se a análise de funções densidade de probabilidade sobre células de *Voronoi*.

Apresentam-se os conceitos da teoria clássica da estimação de parâmetros e efectua-se o respectivo enquadramento com o problema em estudo, através da utilização dos critérios de estimação *maximum-likelihood* (*ML*) e *maximum à posteriori probability* (*MAP*).

Este problema de estimação (com parâmetros inteiros) tem aplicação prática, entre outras, no sistema *GPS* (*Global Positioning System*), quando se assume que o emissor e receptor estão em fase e se pretende estimar o número (inteiro) de ciclos da portadora entre ambos.

1. Descrição geral do problema

Nesta secção apresenta-se o modelo de observação e identificam-se os parâmetros desconhecidos (a estimar), a sua natureza em termos de formato dos dados (inteiros e reais)¹ e respectiva dimensionalidade. Em seguida caracterizam-se os problemas de estimação e verificação assentes sobre o modelo apresentado, após feito o enquadramento com a teoria clássica da estimação de parâmetros. Em seguida referem-se características e variantes do problema.

1.1 Modelo de observação

As observações são efectuadas na presença de ruído aditivo branco com distribuição de *Gauss* (*AWGN*²) de acordo com o modelo linear descrito pela expressão (1).

$$y = Ax + Bz + v \quad (1)$$

com :

$y \in \mathfrak{R}^N$, vector observação (real)

$x \in \mathfrak{R}^p$, vector de parâmetros (real)

$z \in Z^q$, vector de parâmetros (inteiro)

$v \in \mathfrak{R}^N$, vector com amostras de ruído (real).

Por sua vez A e B são matrizes reais conhecidas e bem condicionadas (*full column rank*) com as seguintes dimensões:

$$A \in \mathfrak{R}^{\lfloor N \times p \rfloor}$$

$$B \in \mathfrak{R}^{\lfloor N \times q \rfloor}$$

O ruído é modelado por uma variável aleatória com distribuição de *Gauss*, com média nula e variância unitária, representado por v (vector de dimensão N), constituído por concretizações dessa variável aleatória. Analisando a expressão (1) que traduz o modelo linear de observação, constata-se que na ausência de ruído ($v=0$) o vector observação y resulta da combinação linear dos vectores de parâmetros a estimar. Perante este modelo, o problema global da estimação e verificação pode ser enunciado da seguinte forma:

“Tem-se conhecimento à priori das matrizes A e B . Dada a observação y , na presença de ruído v , o objectivo é estimar os vectores x (real) e z (inteiro), verificando em seguida a qualidade dessa estimativa.”

¹ Ao longo do texto entende-se como real qualquer elemento pertencente ao conjunto dos números reais \mathfrak{R} .

² *Additive White Gaussian Noise*.

1.2 Componentes do problema

Identificado o modelo de observação e as características dos dados envolvidos torna-se necessário formular as duas componentes essenciais do problema:

- a forma como é efectuada a estimação;
- a verificação dessa estimação.

1.2.1 Estimação

Existem vários critérios para efectuar a estimação de parâmetros [1]. A diferença entre estes critérios consiste fundamentalmente na escolha da regra de estimação. A figura 1 mostra um modelo possível para a estimação de parâmetros, onde se pode verificar que o estimador é um sistema que tem como entrada a observação, efectuada na presença de ruído, e produz na saída o valor estimado, segundo uma regra de estimação que por sua vez se baseia numa função de custo.

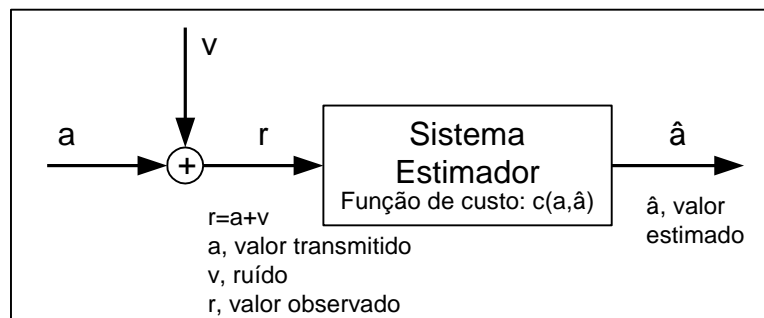


Figura 1 – Modelação do processo de estimação de parâmetros.

O problema consiste em determinar o valor de a , dada a observação r . De forma genérica, o erro existente entre o valor estimado e o valor produzido define-se como o valor absoluto da diferença entre o valor estimado e o valor efectivamente transmitido:

$$\varepsilon = |a - \hat{a}|$$

O valor estimado \hat{a} é aquele que minimiza a função de custo $c(a, \hat{a})$, intrínseca ao estimador. Esta função pode ser analisada de forma complementar, isto é, pode ser analisada como função a maximizar, sendo neste caso chamada de função de adequação. Em qualquer dos casos tem-se uma regra de estimação.

Tanto a geração de valores por determinada fonte, como a respectiva estimação pode ser modelada de forma probabilística. A figura 2 mostra o modelo de estimação clássico, tal como apresentado em [1], onde se pode encontrar quatro componentes fundamentais:

- espaço de parâmetros;
- mapeamento probabilístico para o espaço de observação;

- espaço de observação;
- regra de estimação.

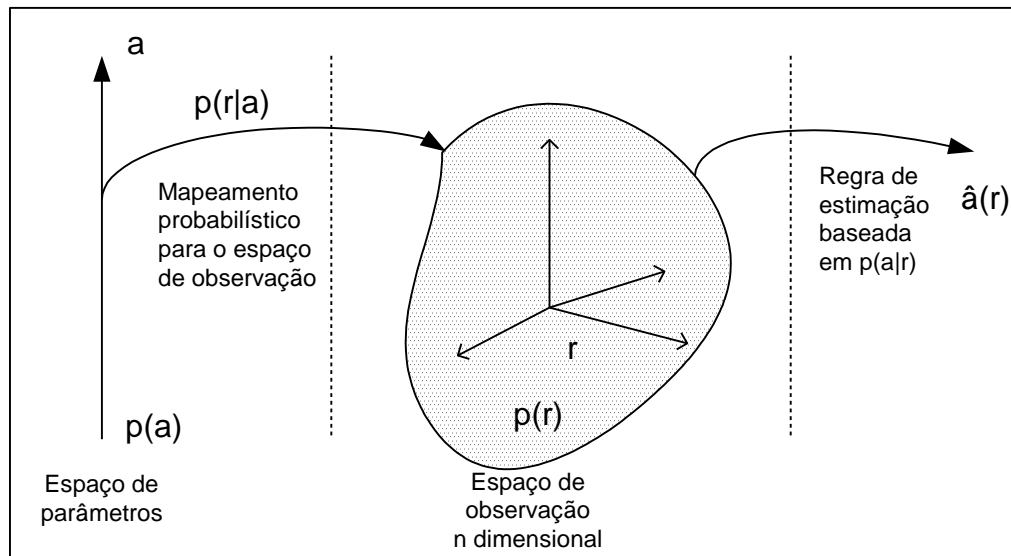


Figura 2 – Modelo clássico da estimação de parâmetros.

O **espaço de parâmetros** caracteriza-se por ser o conjunto dos símbolos (eventualmente infinito) produzidos por determinada fonte. Neste problema concreto existem dois espaços de parâmetros:

- a recta real (ou o conjunto \mathfrak{R}), para o vector x ;
- o conjunto dos inteiros relativos (Z) no caso do vector z .

O **mapeamento probabilístico para o espaço de observação**, traduz-se por uma lei de probabilidades que relaciona o parâmetro efectivamente transmitido com o observado. Neste problema assume-se ruído do tipo branco aditivo *gaussiano* pelo que esta lei de probabilidades leva em linha de conta a distribuição de *Gauss*, tal como explicitado pela expressão (1), no parâmetro v .

O **espaço de observação** tem n dimensões. A observação é constituída por um ponto neste espaço. No problema abordado, o vector observação y é constituído por N elementos.

A **regra de estimação** consiste na relação directa estabelecida entre a observação e o valor estimado. Dado que se tem a observação r , qual o valor do parâmetro a que originou essa observação. De forma geral a regra de estimação tem por base a análise da função densidade de probabilidade à *posteriori* $p_{A|R}(a/r)$, que sendo uma probabilidade condicionada pode ser relacionada com a probabilidade conjunta tal como se apresenta de seguida.

A probabilidade conjunta pode ser escrita em função de duas probabilidades condicionadas diferentes.

$$p_{A,R}(a,r) = p_{A|R}(a|r) \cdot p_R(r)$$

$$p_{A,R}(a,r) = p_{R|A}(r|a) \cdot p_A(a)$$

Igualando as duas expressões anteriores e resolvendo em ordem à expressão pretendida obtém-se uma forma equivalente da mesma, dada pela expressão (2).

$$p_R(r) \cdot p_{A|R}(a|r) = p_A(a) \cdot p_{R|A}(r|a)$$

$$p_{A|R}(a|r) = \frac{p_A(a) \cdot p_{R|A}(r|a)}{p_R(r)} \quad (2)$$

O critério de estimação *MAP* (*Maximum à Posteriori Probability*) consiste na maximização desta função densidade de probabilidade. A função de custo associada é uniforme tendo um aspecto gráfico semelhante ao mostrado na figura 3 a). Na figura 3 b) mostra-se uma possível função densidade de probabilidade à *posteriori* $p_{A|R}(a/r)$, notando que a mesma atinge o seu máximo no valor estimado.

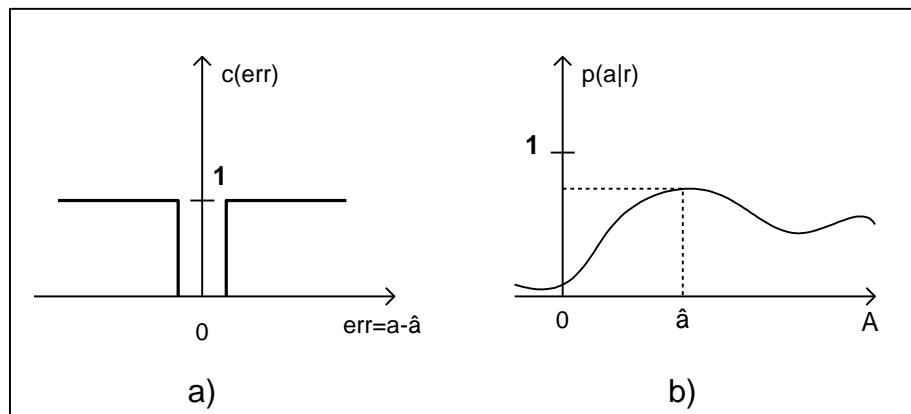


Figura 3 – a) Função de custo associada ao critério de estimação *MAP*.
 b) Exemplo de fdp^1 à *posteriori* exibindo um máximo e sua relação com o valor do parâmetro estimado.

A função densidade de probabilidade à *posteriori* $p_{A|R}(a/r)$, pode ser maximizada através do cálculo dos zeros da primeira derivada em ordem ao parâmetro a . A expressão do critério *MAP* pode ser obtida a partir da expressão (2) tomando logaritmos sobre ambos os termos e desenvolvendo de seguida.

¹ Função densidade de probabilidade

$$p_{A|R}(a | r) = \frac{p_A(a) \cdot p_{R|A}(r | a)}{p_R(r)} (=)$$

$$(=) \ln(p_{A|R}(a | r)) = \ln(p_A(a)) + \ln(p_{R|A}(r | a)) - \ln(p_R(r)) (=)$$

Dado que o termo $\ln(p_R(r))$ não é função do parâmetro a , pode ser ignorado no cálculo do máximo. Tomando $l(a) = \ln(p_{A|R}(a/r))$ a expressão anterior pode ser escrita na forma:

$$(=) l(a) = \ln(p_A(a)) + \ln(p_{R|A}(r | a))$$

O primeiro termo descreve o conhecimento *à priori* e o segundo define a dependência probabilística que o parâmetro a induz na observação r . Para encontrar o máximo deriva-se a expressão anterior e obtém-se a expressão (3) definida como **equação MAP** em [1].

$$\left. \frac{\partial l(a)}{\partial a} \right|_{a=\hat{a}(r)} = \left. \frac{\partial \ln(p_A(a))}{\partial a} \right|_{a=\hat{a}(r)} + \left. \frac{\ln(p_{R|A}(r | a))}{\partial a} \right|_{a=\hat{a}(r)} = 0 \quad (3)$$

Se assumirmos que não existe nenhum conhecimento *à priori* sobre a distribuição de probabilidades do parâmetro a estimar, a expressão reduz-se a apenas um termo.

$$\left. \frac{\partial l(a)}{\partial a} \right|_{a=\hat{a}(r)} = \left. \frac{\ln(p_{R|A}(r | a))}{\partial a} \right|_{a=\hat{a}(r)} = 0 \quad (4)$$

A expressão (4) define o critério de **máxima verosimilhança** (*maximum-likelihood*) abreviado por *ML*, constituindo um caso particular da estimação *MAP*.

1.2.1.1 Critério utilizado

No presente trabalho recorre-se ao critério *ML* dado que não se assume nenhum conhecimento *à priori* sobre a distribuição de probabilidades dos parâmetros a estimar. Dado que existem dois parâmetros a estimar, o critério *ML* para este problema é descrito analiticamente pela expressão (5).

$$(x_{ML}, z_{ML}) = \operatorname{argmax} P_{Y|X,Z}(y | x, z) \quad (5)$$

com $(x, z) \in \mathfrak{R}^p \cdot Z^q$

A estimação traduz-se pela maximização da probabilidade condicionada de observação (*à posteriori*), ou seja, obter o par de vectores (x, z) que maximiza essa probabilidade condicionada, para uma observação y . Note-se pela expressão (5) que a maximização da função não implica forçosamente o cálculo da sua derivada.

1.2.2 Verificação da estimação

A estimação do par de vectores (x, z) , segundo o critério anterior, é feita de acordo com uma regra de verificação. Desta forma define-se o critério de verificação tendo em conta que o vector z é constituído por inteiros e o vector x contém reais. No caso do vector z , a estimação é exacta de forma que se passa a ter a detecção do seu valor. Analiticamente a probabilidade de estimar o valor correcto de z é dada pela expressão (6).

$$P_c = \text{Prob}(z_{ML} = z) \quad (6)$$

Para o vector x , constituído por reais, tem-se estimação adequada quando a norma *Euclideana* do erro entre o valor produzido e o valor estimado se situa abaixo do valor de erro pré-definido ε , o que se resulta na expressão (7).

$$\text{Prob}(\|x - x_{ML}\| < \varepsilon) \quad (7)$$

Note-se que esta expressão é uma forma de minimização de erro quadrático, porque a norma é uma medida quadrática. Dado que se estima o par (x, z) , a estimação apenas é adequada quando as duas últimas probabilidades tomam valores elevados (próximos de 1), de forma que se torna necessário conjugar as duas expressões numa única que traduza a qualidade da estimativa de ambos os parâmetros.

$$\text{Prob}(\|x - x_{ML}\| < \varepsilon, z_{ML} = z) \quad (8)$$

A expressão (8) representa a probabilidade conjunta de estimação correcta e pode ser escrita na forma de probabilidade condicionada.

$$\begin{aligned} \text{Prob}(\|x - x_{ML}\| < \varepsilon, z_{ML} = z) &= \\ \text{Prob}(\|x - x_{ML}\| < \varepsilon \mid z_{ML} = z) \cdot \text{Prob}(z_{ML} = z) & \quad (9) \end{aligned}$$

As expressões (6) e (9) constituem as medidas essenciais da qualidade da estimação, que serão usadas adiante. Verifica-se que o valor de x e x_{ML} são aproximados apenas quando se tem $z = z_{ML}$, o que intuitivamente constitui um critério aceitável, na medida em que os parâmetros reais só serão estimados correctamente quando os parâmetros inteiros também o forem. Nas secções 2.1 e 2.2 mostram-se as expressões que confirmam esta afirmação.

1.3 Características do problema

O modelo de observação pode ser simplificado de tal forma que ajuda a mostrar a natureza dos dois tipos de estimação aqui envolvidos e a sua relação com outros problemas de estimação conhecidos. Analisando a expressão (1) do modelo de

observação, e supondo que não existem parâmetros inteiros ($B=0$), obtém-se outro modelo linear simplificado dado pela expressão (10).

$$y = Ax + v \quad (10)$$

Sendo a matriz A conhecida e v o vector que modeliza o ruído, ficamos reduzidos a um problema clássico de estimação de x , que é resolvido através da minimização de erro quadrático $\|y-x\|^2$ (*least-squares problem*). Para o critério ML o vector estimado x_{ML} é dado pela matriz pseudo-inversa [3] de A tal como indicado pela expressão (11).

$$x_{ML} = (A^T A)^{-1} A^T y \quad (11)$$

Repare-se que nesta situação a minimização é feita apenas em ordem a x , após ter observado y . No problema global, considerando o modelo de observação dado pela expressão (1) é necessário efectuar conjuntamente duas minimizações, partindo de uma observação.

$$\begin{aligned} \min \|y - x\|^2 \\ \min \|y - z\|^2 \end{aligned} \quad (12)$$

Sendo z vector de inteiros, a minimização da segunda parcela da expressão (12) apresenta-se como a componente mais difícil de resolver (minimização de erro quadrático com parâmetros inteiros). Recordando a análise feita sobre a expressão (9) onde se verifica que o valor de x e x_{ML} são aproximados apenas quando se tem $z = z_{ML}$, conclui-se então que a estimação correcta de z é essencial. Torna-se pertinente desenvolver mecanismos de análise e formulação de tal forma que o problema global da estimação e verificação se possa efectuar da seguinte forma:

1. estimação correcta (com verificação) de z ;
2. estimar x , em função da estimação correcta de z .

Na próxima secção apresenta-se a análise do problema que permite abrir caminho para atingir os objectivos enunciados acima, no que respeita ao problema da estimação. Mais adiante formula-se o problema da verificação e respectiva resolução. Mostra-se de seguida como a estimação e verificação se relacionam.

2. Formulação do problema da estimação

2.1 Análise do critério de máxima verosimilhança - ML

Conjugando o modelo de observação dado pela expressão (1) e o critério ML dado pela expressão (2) é possível reescrever esse critério aplicado ao problema em estudo.

Considerando que v é um vector com concretizações de uma variável aleatória *gaussiana* com média nula e variância unitária, então pela expressão (1) o vector observação y (com dimensão N) segue uma distribuição de probabilidades *gaussiana* multivariável com média $E[y] = Ax + Bz$ e variância unitária, dada pela expressão (13).

$$P_{y|x,z}(y | x, z) = \frac{1}{\left(\sqrt{2\pi \cdot 1}\right)^N} \cdot e^{-\frac{\|y-(Ax+Bz)\|^2}{2}} = \frac{1}{\left(\sqrt{\pi}\right)^{\frac{N}{2}}} \cdot e^{-\frac{\|y-Ax-Bz\|^2}{2}} \quad (13)$$

Atendendo à expressão (2) a estimação do par (x_{ML}, z_{ML}) consiste na maximização desta função densidade de probabilidade condicionada. Maximizar uma função de *gauss* consiste em minimizar o respectivo expoente, de forma a tender para o ponto médio onde a função exhibe o seu único máximo, não sendo por isso necessário recorrer à sua derivada.

$$\begin{aligned} (x_{ML}, z_{ML}) &= \operatorname{argmax} P_{y|x,z}(y | x, z) \quad (=) \\ (x_{ML}, z_{ML}) &= \operatorname{argmin} \|y - Ax - Bz\|^2 \quad (14) \end{aligned}$$

Note-se que a desejada minimização conjunta de x e z , para uma observação y , indicada em (12), aparece naturalmente em (14). Após manipulação algébrica [2], o argumento a minimizar da expressão (14) pode ser escrito na forma da soma de três termos, que dependem de parâmetros diferentes, tal como se mostra de seguida.

$$(x_{ML}, z_{ML}) = \operatorname{argmin}(f(x, \hat{x}_z) + g(z, \hat{z}) + k) \quad (15)$$

$f(x, \hat{x}_z)$, representa o primeiro termo que depende apenas do valor de x e da sua estimativa após ter estimado z ;

$g(z, \hat{z})$, representa o segundo termo que depende apenas do valor de z e da sua estimativa;

k , é uma constante que depende apenas de A , B e y ;

Dado que k é constante não influencia o cálculo do mínimo do par (x_{ML}, z_{ML}) . Desta forma, torna-se necessário analisar os dois primeiros termos da expressão (15) que são variáveis e independentes entre si. Estes termos algo complexos, são apresentados nas expressões (16A) e (16B).

$$f(x, \hat{x}_z) = \left(-\hat{x}_z \right)^T \Xi^{-1} \left(-\hat{x}_z \right) \quad (16A)$$

$$g(z, \hat{z}) = \left(-\hat{z} \right)^T \Psi^{-1} \left(-\hat{z} \right) \quad (16B)$$

Com expressões auxiliares:

$$\begin{aligned}\Xi &= (A^T A)^{-1} \\ \Psi &= (B^T (I - A \Xi A^T) B)^{-1} \\ \hat{x}_z &= \Xi A^T (y - Bz) \\ \hat{z} &= \Psi B^T \begin{pmatrix} -A \Xi A^T \\ y \end{pmatrix}\end{aligned}$$

Este resultado é fundamental e permite tirar duas conclusões essenciais:

- o primeiro termo $f(x, \hat{x}_z)$ é sempre positivo ou nulo minimizando quando $x = \hat{x}_z$;
- o segundo termo $g(z, \hat{z})$ não depende de x , logo o seu mínimo pode ser encontrado apenas em ordem a z .

A partir destas duas conclusões é possível escrever expressões individuais de máxima verosimilhança para ambos os parâmetros, onde se verifica a ordem pela qual devem ser estimados e como a estimação de z afecta a estimação de x . Outro aspecto importante que importa referir é que o problema central está na estimação adequada de z , porque a estimação de x é obtida em sua consequência.

2.2 Expressões individuais de máxima verosimilhança

Foram encontradas as expressões para a estimação dos dois parâmetros, que podem ser escritas na forma apresentada pelas expressões (17A) e (17B).

$$z_{ML} = \arg \min_{z \in Z^q, \hat{z} \notin Z^q} \begin{pmatrix} -A \Xi A^T \\ y \end{pmatrix}^T \Psi^{-1} \begin{pmatrix} -A \Xi A^T \\ y \end{pmatrix} \quad (17A)$$

$$x_{ML} = \hat{x}_{z_{ML}} = \Xi A^T (y - Bz_{ML}) \quad (17B)$$

Analisando a expressão (17A), verifica-se que a estimação Z_{ML} é dada pela solução de um problema de minimização de erro quadrático envolvendo variáveis do tipo inteiro. Na estimação acontece que o vector estimado \hat{z} não é constituído só por valores inteiros, tal como evidenciado na expressão (17A). A primeira aproximação para a obtenção de um vector só com valores do tipo inteiro, poderia passar pelo arredondamento de cada componente para o inteiro mais próximo, mas constata-se na prática que tal procedimento introduz erros elevados na estimação de z e conseqüentemente na estimação de x [2]. A estimação Z_{ML} nesta situação é classificada como um problema *NP-hard* [4], pelo que é necessário ser analisada de forma mais profunda para encontrar o processo de estimação adequado que possa ser realizada em tempo polinomial.

2.3 Análise dos parâmetros a estimar

Dada a necessidade de encontrar um processo de estimação de parâmetros inteiros que seja atractivo do ponto de vista computacional, efectua-se a análise da distribuição de probabilidades dos parâmetros a estimar, dada a presença do ruído *AWGN* caracterizado anteriormente. Esta análise é feita tendo em conta que não se assume nenhum conhecimento à priori sobre a distribuição de probabilidade dos parâmetros x e z .

Em termos do modelo da estimação clássica, analisam-se aqui o espaço de parâmetros e o mapeamento probabilístico para o espaço de observação, concluindo com a aproximação à regra de estimação.

2.3.1 Mapeamento probabilístico

Através das duas últimas expressões auxiliares em (16A) e (16B) podemos determinar tal como em [2]:

$$\begin{aligned} \hat{z}, \hat{x}_z &\text{ são combinações lineares da variável aleatória } y. \\ \text{cov}(\hat{z}) &= \Psi ; \text{cov}(\hat{x}_z) = \Xi \end{aligned}$$

Dado que y é uma variável aleatória *gaussiana*, qualquer combinação linear efectuada sobre a mesma é ainda uma variável aleatória *gaussiana*. É possível modelizar os parâmetros a estimar como variáveis aleatórias da seguinte forma:

$$\begin{aligned} \hat{z} &\text{ é uma variável aleatória gaussiana com média } z \text{ e variância } \Psi; \\ \hat{x}_z &\text{ é uma variável aleatória gaussiana com média } x \text{ e variância } \Xi. \end{aligned}$$

Verifica-se para qualquer um destes parâmetros que o desvio entre o valor estimado e o valor efectivamente transmitido depende apenas de uma variância. Esta constatação leva a que o valor estimado \hat{z} possa ser escrito em função do valor efectivamente transmitido tendo em conta essa variância, tal como se constata pelo modelo da figura 1.

$$\hat{z} = z + u \quad (18)$$

em que u é uma variável aleatória gaussiana com média nula e variância Ψ ;

A expressão (18) ilustra bem que os valores estimados para z , muito provavelmente não são inteiros, dado que as concretizações u_i tomam valores reais. A incerteza associada à estimação de z é resolvida estimando o valor da variável aleatória u . Dado que z (parâmetro efectivamente transmitido) tem apenas valores inteiros, torna-se bastante importante que w tenha variância unitária. A mesma expressão pode ser escrita de forma equivalente modificando a variável aleatória u para uma outra com variância unitária. A

forma de o conseguir (descorrelacionando as amostras) consiste em multiplicar ambos os termos da expressão (18) pela *whitening matrix* de u (descrita no anexo A), definida por:

$$G = \Psi^{-\frac{1}{2}}$$

Multiplicando ambos os termos de (18) por G , obtemos a expressão (19).

$$G\hat{z} = Gz + Gu \quad (=) \quad \tilde{y} = Gz + w \quad (=) \quad w = \tilde{y} - Gz \quad (19)$$

w , vector com concretizações de uma *gaussiana* de média nula e variância unitária.

$\tilde{y} = G\hat{z}$, por definição.

Tendo em conta a expressão (19) onde se apresenta nova variável aleatória w com variância unitária, verifica-se que a estimação z_{ML} pode ser escrita na forma de minimização de erro quadrático entre \tilde{y} e Gz .

$$z_{ML} = \operatorname{argmin} \|G\hat{z} - Gz\|^2 \quad (20A)$$

$$z_{ML} = \operatorname{argmin} \|\tilde{y} - Gz\|^2 \quad (20B)$$

$$z_{ML} = \operatorname{argmin} \|w\|^2 \quad (20C)$$

A última forma de escrita, expressão (20C), indica que a estimação exacta (detecção) $Z_{ML}=z$ acontece quando $w = [0 \ 0 \ \dots \ 0]$, ou seja, quando todas as concretizações da variável aleatória *gaussiana* tomam o valor central. Por outras palavras, esta situação é um caso particular que corresponde à ausência de ruído, conduzindo a um resultado intuitivamente aceitável. No entanto a expressão (20B) constitui a forma mais importante uma vez que representa **a minimização da distância Euclideana entre \tilde{y} e Gz** , num espaço com q dimensões. Dado que z é um vector constituído por q inteiros, o produto Gz define uma região em R^q . Desta forma tira-se a conclusão fundamental:

“A estimação Z_{ML} consiste em obter a região Gz mais próxima do vector \tilde{y} num espaço q -dimensional ”

Verifica-se que o vector y encontra-se afastado de Gz consoante o valor de w . Para valores pequenos de w significa que y está mais próximo de Gz do que de qualquer outro ponto, o que se traduz numa estimativa correcta. A matriz G é denominada por matriz geradora da região.

A secção seguinte trata da interpretação geométrica e analítica das referidas regiões em \mathcal{R}^q , e consequente utilização para efectuar a estimação Z_{ML} .

2.3.2 Células de Voronoi

Dado que z é um vector de inteiros, as regiões definidas pelo produto Gz são estruturas periódicas denominadas por células de *Voronoi*. A célula define-se como uma região com um ponto central. Um ponto está contido em determinada célula se a distância entre o mesmo e o ponto central dessa célula é menor do que todas as distâncias aos pontos centrais das restantes células. A figura 4 mostra a representação geométrica das células de *Voronoi*, onde se mostra V_0 como o ponto central da célula de origem.

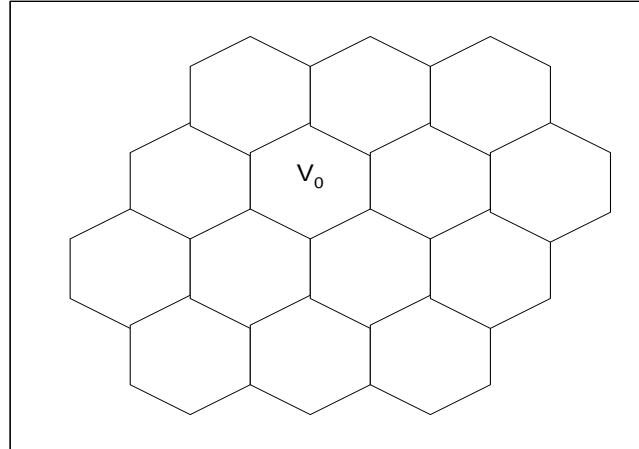


Figura 4 – Estrutura geométrica das células de *Voronoi*.

Dada a estrutura periódica das células de *Voronoi*, verifica-se que o produto Gz representa um deslocamento a partir da célula origem cujo centro é V_0 . Conjugando as expressões (6) e (20C) e recordando que w tem variância unitária, conclui-se que a probabilidade P_c pode ser escrita em função das células de *Voronoi*, tal como mostra a expressão (21).

$$P_c = \text{Prob}(z_{ML} = z) = \text{Prob}(w \in V_0) \quad (21)$$

A expressão (21) indica que o cálculo da probabilidade de estimação correcta Z_{ML} é dada pela probabilidade do vector w (q -dimensional) pertencer ao interior da célula de *Voronoi*. Este é um processo semelhante à quantização vectorial, em que determinado conjunto de pontos (vectors) é representado através de um ponto central, função da sua distância em relação a esse ponto. Na secção seguinte mostra-se como se pode efectuar o cálculo de probabilidades sobre células de *Voronoi*.

3. Problema da verificação

A estimação Z_{ML} adequada e consequente verificação P_c são os passos essenciais para a resolução das restantes componentes do problema. Na secção anterior foi descrito que a medida de verificação é dada pela probabilidade do vector w (q -dimensional) pertencer ao interior da célula de *Voronoi* na origem designada por V_0 , o que significa integrar a

função densidade de probabilidade de w ao longo da célula. Do ponto de vista computacional, este problema é de difícil tratamento dada a forma geométrica das células, apresentada na figura 4. Mais uma vez torna-se necessário encontrar mecanismos que tornem este problema atractivo computacionalmente.

3.1 Limites inferior e superior de P_c

Dado que o valor exacto de P_c é bastante difícil de obter computacionalmente [2], surge a ideia de o calcular por enquadramento, ou seja, calculando os seus limites inferior e superior. Se existir garantia que ambos os limites tendem para o valor 1, então P_c está bastante próximo de 1, o que significa estimação correcta.

O limite inferior $P_{c_{low}}$, pode ser obtido de forma intuitiva através de uma hiper-esfera centrada na célula de origem com raio $d/2$, totalmente contida na célula, tal como ilustrado pela figura 5. Constata-se facilmente que qualquer ponto que pertencer à hiper-esfera pertence obrigatoriamente à célula de *Voronoi*. Como a hiper-esfera não ocupa toda a célula, a probabilidade de um valor estar contido no seu interior constitui um limite inferior para P_c .

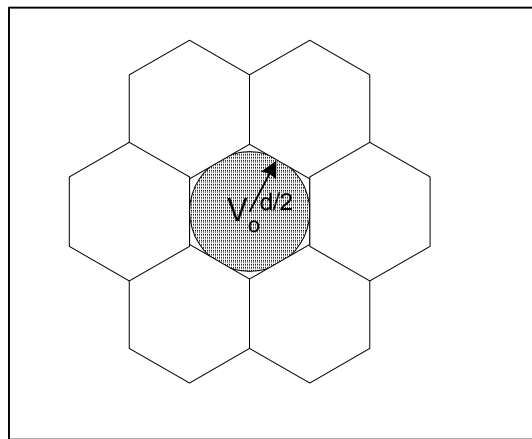


Figura 5 – Ilustração da obtenção do limite inferior de P_c .

Embora o raio da hiper-esfera seja intuitivo de obter como limite inferior, persiste a dificuldade de determinar o valor de d . Analisando a figura, conclui-se que a hiper-esfera desenhada é a maior hiper-esfera totalmente contida na célula. Verifica-se ainda que o valor $d/2$ é a distância entre a origem e os pontos vizinhos mais próximos (contidos em células vizinhas). A probabilidade de w estar contido nesta hiper-esfera é dada pela expressão (22).

$$P_{c_{low}} = \text{Prob}\left(\|w\| < \frac{d}{2}\right) \quad (22)$$

w tem média nula e variância unitária.

Dado que a norma do vector w consiste no somatório do quadrado das suas componentes, que têm distribuição normal com variância unitária, conclui-se que nestas condições $\|w\|^2$ tem distribuição de probabilidades χ^2 . Se a distribuição de probabilidades é conhecida, então pode ser representada pela sua função cumulativa, designada por $F_{\chi^2}(\cdot; q)$, onde se evidencia os q graus de liberdade. Então a expressão (22) escrita em termos de função cumulativa, toma a forma da expressão (23).

$$Pc_{low} = F_{\chi^2} \left(\frac{d^2}{4}; q \right) \quad (23)$$

Resta então verificar como se pode calcular o valor de d . Este é outro problema difícil de resolver computacionalmente, sendo classificado como *NP-hard*. No entanto o cálculo do limite inferior de d , pode ser resolvido de forma computacionalmente eficiente pelo processo de ortogonalização de *Gram-Schmidt*. Para tal efectua-se a ortogonalização das colunas da matriz G , obtendo nova matriz G^* .

$$G = \left[g_1; g_2; \dots; g_q \right]$$

Após o procedimento de Gram - Schmidt :

$$G^* = \left[g_1^*; g_2^*; \dots; g_q^* \right] \quad (24)$$

Após a ortogonalização, o limite inferior do valor de d é obtido pela menor das normas dos vectores de base resultantes.

$$d = \min \left(\|g_1^*\|, \|g_2^*\|, \dots, \|g_q^*\| \right) \quad (25)$$

Dado que se obtém um limite inferior de d está-se a minorar o valor de Pc_{low} , nunca existindo o perigo de obter valores que façam com que a hiper-esfera ultrapasse os limites da célula de *Voronoi*.

O limite superior Pc_{up} pode ser calculado de várias formas, sendo uma delas baseada no volume da célula de *Voronoi*. Dado que se tem uma distribuição normal (q -dimensional) a região que maximiza a probabilidade é a hiper-esfera Euclideana [2]. Tal como analisado para o limite inferior, é possível pensar agora numa hiper-esfera Euclideana que ocupa toda a célula de *Voronoi*. Dado que a matriz G é geradora da região, o volume da célula de *Voronoi* é dado por $|\det(G)|$ [2]. A probabilidade calculada sobre uma hiper-esfera com volume dado por $|\det(G)|$ é superior a Pc . O raio de uma hiper-esfera Euclideana q -dimensional de volume $|\det(G)|$ é dado pela expressão (26).

$$\ell = \sqrt[q]{|\det(G)| / \alpha_q} \quad , \quad \text{com} \quad \alpha_q = \frac{\pi^{q/2}}{(int(q/2))!} \quad (26)$$

Tendo estabelecido o valor do raio da hiper-esfera é possível estabelecer a expressão (27) que define o limite superior Pc_{up} .

$$Pc_{up} = Prob(\| w \| < \sqrt[q]{|\det(G)| / \alpha_q}) \quad (27)$$

w tem média nula e variância unitária.

Como se pode verificar pela expressão (27), o limite superior também consiste no cálculo de uma probabilidade onde o termo à esquerda é constituído pela norma de um vector q-dimensional. Tal como no caso da expressão do limite inferior, a expressão (27) pode ser escrita na forma da função cumulativa dando origem à expressão final do limite superior de Pc.

$$Pc_{up} = F\chi^2\left(\left(|\det(G)| / \alpha_q\right)^{\frac{2}{q}}; q\right) \quad (28)$$

Chega-se assim à conclusão que o limite superior de Pc é bastante eficiente de calcular computacionalmente. Basta calcular o valor do determinante de G, e em seguida aplicar esse valor na tabela da distribuição cumulativa de χ^2 . Resultados de simulações [2] provam que este limite superior é realmente eficiente de calcular, atingindo valores bastante adequados.

Nesta secção mostrou-se como é possível efectuar de forma computacionalmente eficiente o cálculo dos limites inferior e superior de Pc o que se traduz na correcta e eficiente verificação da estimação do parâmetro inteiro, ou de outra forma, a estimação adequada Z_{ML} . Apenas nestas condições é possível efectuar a estimação e verificação dos parâmetros reais.

3.2 Parâmetros reais

Após a estimação e respectiva verificação dos parâmetros inteiros, tem-se como dado adquirido a célula de *Voronoi* onde pesquisar pelo vector y, ou seja, a célula sobre a qual se efectua a análise para estimar o vector real x. Nesta altura, pela expressão (17B) tem-se a estimação X_{ML} , confirmando que os parâmetros reais apenas são estimados correctamente quando se assegura boa estimação para os parâmetros inteiros. A estimação dos parâmetros inteiros é agora o problema fundamental a considerar na secção seguinte.

4. Problema da estimação

4.1 Formulação a partir dos dados do problema da verificação

Foi apresentada anteriormente a forma eficiente de verificação da estimação do parâmetro inteiro z. Falta referir como é que se processa a estimação de z, o que constitui o sub-problema mais interessante do ponto de vista computacional, no problema global apresentado. Basicamente, a estimação de z é um problema de minimização de erro quadrático considerando parâmetros inteiros, em que se pretende obter o resultado por

máxima verosimilhança. Recordando a expressão (20B), onde está implícita a componente de verificação pela presença de G , tem-se que:

$$z_{ML} = \operatorname{argmin} \|\tilde{y} - Gz\|^2$$

O que se traduz pela minimização da distância *Euclideana* entre y e Gz , num espaço com q dimensões. Graficamente, sobre a célula de *Voronoi*, podemos representar a estimação tal como ilustra a figura 6, onde fica bem patente que a mesma pode ser encarada como a pesquisa do ponto central da célula que contém y .

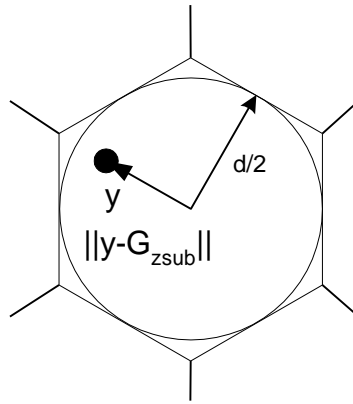


Figura 6 – Vector estimado y sobre a célula de *Voronoi* cujo centro é Gz .

Também foi referido que a estimação Z_{ML} pode ser escrita de outra forma, nomeadamente a expressão (17A) pode tomar outra forma em função de G , apresentada na expressão (29).

$$z_{ML} = \operatorname{argmin} \left\langle -\hat{z}^T, \Psi^{-1} \left\langle -\hat{z} \right\rangle \right\rangle \quad z \in Z^q, \hat{z} \notin Z^q \quad (29)$$

com

$$\hat{z} = G^{-1}y$$

$$\Psi = (G^T G)^{-1}$$

4.2 Algoritmo sub-óptimo para a estimação Z_{ML}

Identificadas as formas gráfica e analítica do problema, resta então encontrar algoritmos eficientes que consigam efectuar a estimação em tempo polinomial. Consideram-se agora dois casos particulares e de seguida analisa-se a situação genérica.

- No caso particular da inexistência de ruído, analisando a expressão (29), verifica-se que se \hat{z} é constituído apenas por valores inteiros tem-se imediatamente $Z_{ML} = \hat{z}$.

- Outra situação particular verifica-se quando a matriz Ψ é diagonal. Nesta situação Z_{ML} pode ser estimado através do arredondamento de todos os elementos de z para o inteiro mais próximo. A diagonalização de Ψ apenas existe quando G é uma matriz ortogonal (colunas de G ortogonais entre si). No caso das colunas de G serem “quase ortogonais” entre si, tem-se que Ψ é uma matriz “quase diagonal”.

Passa-se agora à apresentação do algoritmo sub-óptimo para determinar o mínimo de $\|y - Gz\|$. Designa-se por sub-óptimo, dado que não garante que se obtenha o mínimo global da expressão pretendida, embora na prática seja uma situação que ocorre com alguma frequência [2]. Existe interesse na análise deste tipo de algoritmo por duas razões:

- mesmo que não se consiga atingir o máximo global, obtém-se uma boa aproximação inicial a qual pode ser utilizada como ponto de partida por um algoritmo de optimização global (tal como descrito adiante);
- este tipo de algoritmo é eficiente e tem baixa complexidade computacional no pior caso possível.

A condição de optimalidade consiste em a distância Euclideana entre o vector y e o centro da célula em causa ser igual ou inferior a $d/2$, ou seja, ser um ponto contido na hiper-esfera de raio $d/2$, que tal como visto anteriormente define o limite inferior Pc_{low} . Nesta situação Gz toma a designação de Gz_{sub} (*sub-optimo*), tal como referido pela figura 6. A expressão (30) traduz a condição de optimalidade.

$$\|y - Gz_{sub}\| \leq \frac{d}{2} \Rightarrow Z_{sub} = Z_{ML} \quad (30)$$

O algoritmo deve então determinar Gz_{sub} a partir do qual se consegue extrair directamente o valor de z_{sub} . O algoritmo tem como parâmetros de entrada:

- G , matriz geradora da região (*whitening matrix* de u em (18));
- y , vector observado;
- d , dobro do raio da hiper-esfera totalmente contida na célula de *Voronoi*, que pode ser calculado pela expressão (25).

Como parâmetro de saída tem-se o vector de inteiros z , que respeita a condição enunciada na expressão (30) se for o óptimo global, ou o que mais se aproxima se for apenas óptimo local. A figura 7 descreve o algoritmo.

$Z_{sub} = Sub_Optimal_Z(G, \tilde{y}, d)$

1. $[F, G_{\#}]$ = aplicação do algoritmo LLL (anexo B) sobre as colunas da matriz G ;
2. $Z_{sub} = F \cdot \text{Int}(G_{\#}^{-1} \cdot \tilde{y})$, produto da matriz F pelo arredondamento de cada elemento, para o próximo inteiro, do vector resultante de $G_{\#}^{-1} \cdot \tilde{y}$
3. Devolver Z_{sub} . (Se Z_{sub} respeita a condição da expressão (30), então é um ótimo global.)

Fim

Figura 7 – Algoritmo polinomial sub-ótimo para a estimação z_{ML} .

Sobre a descrição acima importa descrever o algoritmo LLL [2], e os resultados que o mesmo produz. Começando por analisar os parâmetros de entrada e de saída do mesmo constatamos que recebe como parâmetro de entrada a matriz geradora das regiões (células) G , e que produz duas outras matrizes F e $G_{\#}$. Estas três matrizes relacionam-se da forma indicada na expressão (31).

$$G_{\#} = GF \quad (31)$$

A matriz $G_{\#}$ caracteriza-se por ser quase “quase ortogonal” e também é geradora da região. As colunas de $G_{\#}$ (vectors de base da região) caracterizam-se por ter norma aproximadamente unitária. A “quase ortogonalidade” da matriz $G_{\#}$ conjuntamente com a matriz unimodular¹ F permitem o arredondamento de cada componente de z_{sub} para o próximo valor inteiro. O algoritmo LLL [5] encontra-se descrito em pormenor no anexo B. Note-se que existem outros algoritmos sub-ótimos [6], para além de existirem variantes do algoritmo aqui apresentado [2].

4.3 Algoritmo de optimização global

Partindo do raciocínio anterior utilizado para o algoritmo sub-ótimo, é possível acrescentar mais funcionalidades de forma a obter optimização global. A mais valia da optimização global consiste em garantir que o ponto encontrado é realmente o mínimo global e para tal tem que se adicionar uma restrição à estimação anterior. A restrição consiste em verificar se existe algum ponto mais próximo de Gz_{sub} , que tenha como coordenadas valores inteiros, o que significa efectuar a pesquisa no interior de um elipsóide centrado em Gz_{sub} e encontrar (se existir) o ponto mais próximo com (q) coordenadas todas com valor inteiro. Se existir um ponto nessas condições então esse

¹ Uma matriz F é unimodular quando todos os seus elementos e da sua inversa F^{-1} são inteiros e verifica ainda $|\det(F)| = 1$.

ponto é o óptimo, se não existir então acabámos de provar que não existe nenhum ponto com coordenadas com valor inteiro dentro do elipsóide e Gz_{sub} constitui a solução óptima procurada.

Supondo que foi encontrado um ponto sub-óptimo Z_{sub} , esse ponto define o raio r (ponto 3 do algoritmo da figura 8) do elipsóide sobre o qual vamos pesquisar. Atendendo à expressão (29), analiticamente tem-se que o elipsóide a pesquisar é dado pela expressão (32).

$$\left(\mathbf{z} - \hat{\mathbf{z}} \right)^T \Psi^{-1} \left(\mathbf{z} - \hat{\mathbf{z}} \right) < r^2 \quad (32)$$

O método de pesquisa de pontos no interior do elipsóide não é trivial, e merece um tratamento mais rigoroso que não se enquadra no contexto do presente trabalho. O tratamento adequado sobre esta questão pode ser encontrado em [2].

Identificada a constrição, o algoritmo de optimização global deriva-se com relativa facilidade a partir do anterior. É um algoritmo que converge em tempo polinomial.

$Z_{ML} = \text{Optimal_Z}(G, \tilde{y}, d)$

- 1 $Z_{\text{sub}} = \text{Sub_Optimal_Z}(G, \tilde{y}, d)$
- 2 Se Z_{sub} respeita a condição da expressão (30), então
 $Z_{ML} = Z_{\text{sub}}$, é um óptimo global.
Parar.
- 3 Faça-se $r = \|y - Gz_{\text{sub}}\|$ e $Z^* = Z_{\text{sub}}$
- 4 Repetir indefinidamente:
 - a) Pesquisar um ponto inteiro no interior do elipsóide $\|y - Gz\| < r$.
 Se não for encontrado nenhum ponto nestas condições, $Z_{ML} = Z^*$, é um óptimo global. **Parar.**
 - b) Seja Z^* o ponto inteiro encontrado no passo anterior.
 Faça-se $r = \|y - Gz^*\|$
 - c) Se $r \leq \frac{d}{2}$ então $Z_{ML} = Z^*$. **Parar.**

Fim

Figura 8 – Algoritmo polinomial óptimo para a estimação z_{ML} .

Note-se que o primeiro passo deste algoritmo consiste na aplicação do algoritmo sub-óptimo para encontrar um ponto de partida. É possível que este ponto de partida já constitua o óptimo global e nesta situação já não é necessário continuar a pesquisa. Caso contrário efectua-se a referida pesquisa sobre o elipsóide, para encontrar o óptimo global ou para provar que o mesmo já foi encontrado. Nesta situação, o valor de r obtido no

passo 3 do algoritmo é geralmente pequeno, o que conduz a uma redução no espaço de pesquisa sobre o elipsóide.

Após a obtenção de Z_{ML} pelo processo de optimização global referido acima, para efectuar a estimação X_{ML} basta seguir a expressão (17B). Recorde-se que a estimação Z_{ML} pelo algoritmo de optimização global leva em conta os critérios de verificação (medida P_c) enunciados nas secções anteriores, garantido que o valor estimado é correcto.

5. Extensões ao problema inicial

5.1 Conhecimento à priori da distribuição de x

Até este ponto do texto não se assumiu nenhum conhecimento sobre a distribuição de probabilidades dos parâmetros a estimar. Nesta secção analisam-se as alterações ao processo de estimação no caso de se admitir algum conhecimento sobre a estatística de x .

Neste cenário, a estimação *MAP* descrita pela expressão (3), é uma alternativa à estimação *ML*. Recordando o modelo de observação (1), e o critério *ML* (5) para o problema em causa, tem-se a equação *MAP* correspondente dada pela expressão (33).

$$\begin{aligned} (x_{Map}, z_{Map}) = \operatorname{argmax}_{(x, z) \in \mathfrak{R}^p \cdot Z^q} P_{x, z|Y}(x, z | y) \end{aligned} \quad (33)$$

A introdução do conhecimento à priori tem como consequência a utilização doutro critério de estimação. Verifica-se que a aplicação do critério *MAP* permite que o problema seja abordado da mesma forma usada no critério *ML*, conduzindo à alteração de algumas expressões, mantendo o raciocínio idêntico.

5.2 Dados recebidos de forma sequencial (*stream-data*)

Na análise feita até este ponto, admite-se que os dados estão totalmente disponíveis quando se parte para a estimação e verificação. No entanto quando se trata de aplicações reais, é típico os dados serem recebidos de forma sequencial ao longo do tempo, sem efectuar armazenamento. Em termos analíticos significa que o vector y (parâmetro observado) vai sendo construído à medida que se vão recebendo as suas componentes.

A expressão (34) mostra a forma genérica de evolução do vector de observações. Mais concretamente considerando uma observação com dimensão N , a expressão mostra a que corresponde a observação i .

$$r_i = Ca + w_i, \quad i = 1, 2, \dots, N \quad (34)$$

Sobre a expressão (34), r é o vector de observações, C é uma matriz conhecida com dimensões $[1 \times k]$, w_i são amostras de ruído, independentes e identicamente distribuídas,

geralmente com distribuição *gaussiana*. O parâmetro (k -dimensional) a estimar designa-se por a .

No caso do problema em estudo, é possível calcular \tilde{y} e G de forma recursiva, com baixo custo computacional. Assumindo que o vector x tem uma estrutura crescente x_k tal que $k=0,1,2,\dots,N$ é possível escrever o vector x na forma recursiva dada pela expressão (35).

$$x_{k+1} = F_k x_k + G_k u_k \quad (35)$$

De acordo com a expressão (1) que define o modelo de observação tem-se uma observação que avança ao ritmo de k , permitindo estabelecer a relação (36).

$$y_k = H_k x_k + J_k z + v_k \quad (36)$$

Com F_k , G_k , H_k , J_k , são matrizes conhecidas e $x(0)$, u_k , v_k , são variáveis aleatórias *gaussianas* com média nula. Nesta situação, o objectivo é efectuar a estimação *MAP* de z e x_k dado y_0, \dots, y_N . Dado que este é um problema típico de **estimação sequencial** com conhecimento *a priori* sobre a distribuição de probabilidades dos parâmetros a estimar, a forma de efectuar a estimação *MAP* consiste na utilização do filtro de *Kalman* [1]. Constata-se que a expressão (35) é referida em [1] como a discretização do filtro de *Kalman*, estabelecendo a relação entre o mesmo e o problema da minimização de erro quadrático com parâmetros inteiros.

A partir da relação (36) e do critério *MAP* apresentado anteriormente deriva-se a expressão (37) que mostra o cálculo recursivo de Z_{MAP} .

$$z_{MAP}^{(k)} = \operatorname{argmin} \left\| y^{(k)} - G^{(k)} z \right\|^2 \quad (37)$$

As restantes fórmulas derivadas a partir desta expressão e das duas anteriores podem ser encontradas em [2]. Como nota conclusiva nesta secção resta acrescentar que a probabilidade de estimação correcta de z nesta situação (na observação de índice k) é dada pelo integral de uma variável *gaussiana* (q -dimensional) sobre a célula de *Voronoi* da origem, significando que toda a análise efectuada para o critério *ML* mantém-se válida quando se aplica o critério *MAP*. Após o cálculo de Z_{MAP} , a estimação X_{MAP} pode ser obtida também através da filtragem de *Kalman*.

6. Algoritmo completo para a estimação de parâmetros inteiros

Nesta secção reúnem-se todos os conceitos explorados ao longo do trabalho para apresentar, em forma de conclusão, o algoritmo completo para a estimação de parâmetros inteiros, sobre um modelo linear de observação.

Integer And Real Parameter Estimation

1. Actualizar G e y_{est} por cada medição, tal como referido pela expressão (35);
2. Calcular o limite superior de P_c , designado por $P_{c_{up}}$, pela expressão (28);
3. Enquanto $P_{c_{up}}$ for baixo ($P_{c_{up}} < 0.99$), repetir o ponto 2;
4. Calcular o limite inferior de P_c , designado por $P_{c_{low}}$, pela expressão (23);
5. Enquanto $P_{c_{low}}$ for baixo ($P_{c_{low}} < 0.99$), repetir o ponto 4;
6. Estimar Z_{ML} , pelo algoritmo sub-óptimo (figura 7) ou pelo algoritmo óptimo (figura 8);
7. Nas condições anteriores, a estimação Z_{ML} é correcta e a obtenção do parâmetro real X_{ML} é imediata, pela expressão correspondente à (17B), aplicada ao critério *MAP*;

Fim

Figura 9 – Algoritmo de estimação de parâmetros inteiros e reais.

Pela análise do algoritmo, conclui-se que este acaba por se revelar bastante simples, no que toca ao seu aspecto global. A complexidade da estimação de parâmetros inteiros aparece dividida pelo cálculo dos limites inferior e superior da probabilidade de estimação correcta, juntamente com a aplicação de algoritmos sub-óptimos ou óptimos realizáveis em tempo polinomial, descritos anteriormente.

7. Conclusões

No presente trabalho foi abordado o problema da estimação de parâmetros inteiros e reais a partir de um modelo de observação linear, na presença de ruído aditivo gaussiano. Dado que o modelo contempla a existência de parâmetros inteiros, o problema de estimação pode ser encarado como detecção, introduzindo uma complexidade computacional acrescida, que deriva directamente da minimização de erro quadrático com o uso de parâmetros inteiros.

Apresentou-se o conceito clássico de estimação, tal como referido na literatura e estabeleceu-se a relação com o problema em causa, através da aplicação de dois critérios de estimação: *ML (Maximum Likelihood)* e *MAP (Maximum à Posteriori Probability)*.

Para além da estimação de parâmetros foi descrita a componente de verificação, a qual também constitui um desafio do ponto de vista computacional. Para encontrar uma solução viável foi necessário efectuar a análise exaustiva do problema de verificação estabelecendo medidas de adequação bem como os seus limites inferior e superior. Sobre estas medidas foi derivado o seu tratamento sobre as células de *Voronoi* e abordou-se a semelhança com a quantização vectorial.

No final refere-se como o filtro de *Kalman* pode ser utilizado para a estimação recursiva dos parâmetros, através da aplicação do critério *MAP*.

8. Bibliografia e Referências

[1] H. Van-Trees, **Detection, Estimation, and Modulation Theory – Part One**, John Wiley and Sons, 1968.

[2] A. Hassibi, A. Boyd, “*Integer Parameter Estimation in Linear Models with Applications to GPS*”, IEEE Transactions on Signal Processing, Vol. 46, Nº 11, November 1998.

[3] G. Strang, K. Borre, **Linear Algebra, Geodesy, and GPS**, Wellesley-Cambridge Press, 1997.

[4] P. van Emde Boas, “*Another NP-complete partition and the complexity of computing short vectors in a lattice*”, Technical Report 81-04, Math. Inst., University of Amsterdam, Amsterdam, The Netherlands, 1981.

[5] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, “*Factoring polynomials with rational coefficients*”, Math. Ann., vol. 261, pp. 515-534, 1982.

[6] L. Babai, “*On Lovász’ lattice reduction and the nearest lattice point problem*”, Combinatorica, vol. 6, pp. 1-13, 1986.

[7] A. K. Jain, **Fundamentals of Digital Image Processing**, Prentice Hall, 1989.

[8] Hyperlink:

<http://www-ee.eng.hawaii.edu/~neng/Class/NeuralNet/ICAimage/sld007.htm>

Anexo A

“Whitening Matrix”

Neste anexo descreve-se a “Whitening Matrix” G , referida na subsecção 2.3.1 de forma a ilustrar a sua obtenção e o porquê da sua utilização.

Descrição

Efectuar *whitening* (branqueamento) a um conjunto de dados significa tornar os mesmos incorrelacionados, com covariância unitária.

Este processo é semelhante à transformada de *Karhunen-Loève* [1] [7] (ou análise em componentes principais), em que a matriz de transformação é constituída pelos vectores próprios da matriz de auto-correlação dos dados a transformar, tal como se descreve de seguida.

- Seja X uma matriz real $[N \times N]$, com média nula.
- A matriz de autocorrelação de X é dada por $R_X = X.X^T$;
- R_X é uma matriz real simétrica, pelo que os seus vectores próprios são ortogonais entre si, e os seus valores próprios são reais;
- Calculando os valores e vectores próprios de R_X tem-se:
 - $[E, D] = \text{eig}(R_X)$;
 - $E = [e_1 \ e_2 \ \dots \ e_n]$, matriz ortogonal com os vectores próprios;
 - $D = \text{diag}(\lambda_1 \ \lambda_2 \ \dots \ \lambda_n)$, matriz diagonal com os valores próprios;
- A matriz transformada X_{KL} é dada por $X_{KL} = E^T \cdot X$;
- A matriz de autocorrelação de X_{KL} é dada por $R_{X_{KL}} = X_{KL} \cdot X_{KL}^T = D$, é uma matriz diagonal;

A transformada para efectuar *whitening* dos dados consiste em introduzir mais um factor, relacionado com a matriz dos valores próprios, na expressão de transformação, tal que se obtém:

$$X_{\text{White}} = D^{(-0.5)} \cdot E^T \cdot X;$$

Demonstra-se nesta situação [8] que a matriz de covariância (que corresponde à matriz de auto-correlação quando se tem média nula) de X_{White} é a matriz identidade I_N .

De seguida mostra-se um *script* exemplo em MATLAB, que ilustra a aplicação da transformada *KL* sobre uma matriz de dados (com média nula). Sobre a mesma matriz de dados é aplicada a transformação que efectua o referido *whitening*. Mostram-se os resultados obtidos em ambas as situações e verifica-se o funcionamento da transformada referida.

```

% IST - Instituto Superior Técnico.
% MEEC - Mestrado em Engenharia Electrotécnica e de Computadores.
% TC - Teoria da Comunicação.
%
% Artur Ferreira N° 4243
%
% Exemplificação da descorrelação de dados, resultando
% em variância unitária.

% X é uma matriz de dados [3x3].
% Tem média nula
X = [
    1 5 3 ;
   -1 -7 7 ;
   -5 -6 3 ]

% Matriz de autocorrelação de X.
% Dado que tem média nula é igual à
% matriz de covariância.
RX = X * X'

% Calcular os valores e vectores próprios da matriz
% de auto-correlacao. Dado que é uma matriz real simétrica
% os valores próprios são reais e os vectores próprios
% ortogonais entre si.
[E, D] = eig( RX )

% Calcular a transformada KL, que descorrelaciona
% os valores de X.
XKL = E' * X

% Mostrar a matriz de autocorrelação da
% matriz transformada e verificar a descorrelação.
RXKL = XKL * XKL'

% Efectuar a potenciação da matriz dos valores
% próprios.
D = power(D, -0.5 );
[NRows, NCols] = size(D);
for i=1:NRows
    for j=1:NCols
        if i~=j
            D(i,j) = 0;
        end
    end
end

% Calcular a transformada, que descorrelaciona
% os valores de X, com variância unitária.
X_WHITE = D * E' * X

% Mostrar a matriz de autocorrelação da
% matriz transformada e verificar a descorrelação,
% e variância unitária.
RX_WHITE = X_WHITE * X_WHITE'

```

X =

| | | |
|----|----|---|
| 1 | 5 | 3 |
| -1 | -7 | 7 |
| -5 | -6 | 3 |

RX =

| | | |
|-----|-----|-----|
| 35 | -15 | -26 |
| -15 | 99 | 68 |
| -26 | 68 | 70 |

E =

| | | |
|---------|---------|---------|
| -0.8568 | -0.4666 | -0.2195 |
| -0.4548 | 0.4833 | 0.7481 |
| 0.2429 | -0.7408 | 0.6263 |

D =

| | | |
|---------|--------|----------|
| 34.4108 | 0 | 0 |
| 0 | 9.2577 | 0 |
| 0 | 0 | 160.3315 |

XKL =

| | | |
|---------|----------|---------|
| -1.6168 | -2.5585 | -5.0250 |
| 2.7540 | -1.2714 | -0.2388 |
| -4.0990 | -10.0915 | 6.4569 |

RXKL =

| | | |
|---------|--------|----------|
| 34.4108 | 0.0000 | 0.0000 |
| 0.0000 | 9.2577 | 0.0000 |
| 0.0000 | 0.0000 | 160.3315 |

X_WHITE =

| | | |
|---------|---------|---------|
| -0.2756 | -0.4362 | -0.8566 |
| 0.9051 | -0.4179 | -0.0785 |
| -0.3237 | -0.7970 | 0.5099 |

RX_WHITE =

| | | |
|--------|--------|--------|
| 1.0000 | 0.0000 | 0.0000 |
| 0.0000 | 1.0000 | 0.0000 |
| 0.0000 | 0.0000 | 1.0000 |

Anexo B

Algoritmo LLL¹

O nome deste algoritmo deriva dos nomes dos seus autores: **Lenstra, Lenstra e Lovász**.

Trata-se de um algoritmo de complexidade polinomial que permite obter uma base reduzida para uma região partindo de uma matriz geradora dessa região.

O algoritmo retorna uma matriz “quase ortogonal” $G_{\#}$, que também é geradora da mesma região e outra unimodular² F , tal que $G = G_{\#}F$

[$G_{\#}, F$] = LLL(G)

Considere-se uma região L , gerada por uma matriz G , tal que $G = [g_1 g_2 \dots g_q]$.

1. Aplicar o procedimento de *Gram-Schmidt* sobre as colunas de G : $g_1 g_2 \dots g_q$ e obter os vectores $g_1^* g_2^* \dots g_q^*$, que constituem uma base ortogonal, para o espaço gerado por G , através da recursão dada pela expressão (B1).

$$g_j^* = g_j - \sum_{i=1}^{j-1} \frac{g_j^T g_i^*}{\|g_i^*\|^2} g_i^*, \text{ com } j = 2, 3, \dots, q \quad (\text{B1})$$

Obtém-se uma matriz quase ortogonal $G_{\#} = [g_1^* g_2^* \dots g_q^*]$;

As bases do espaço verificam a seguinte relação:

Qualquer vector g_j pode ser escrito na forma de combinação linear dos vectores $g_1^* g_2^* \dots g_q^*$, tal como refere a expressão (B2).

$$g_j = \sum_{i=1}^j \mu_{ji} g_i^*, \quad \mu_{ji} = g_j^T g_i^* / \|g_i^*\|^2, \text{ com } i = 1, 2, \dots, j-1 \text{ e } \mu_{jj} = 1 \quad (\text{B2})$$

2. Para $j=1, 2, \dots, q$, e dado j para $i=1, 2, \dots, j-1$, substituir g_j por $g_j - \text{int}(\mu_{ji} \cdot g_i)$. Actualizar os g_j^* s e μ_{ji} s com os novos g_j^* s de acordo com as expressões (B1) e (B2).

3. Se existir um índice j que viole a condição da expressão (B3),

$$\|g_{j+1}^* - \mu_{(j+1)j} g_j^*\|^2 \geq \frac{3}{4} \|g_j^*\|^2 \quad (\text{B3})$$

então trocar g_j com g_{j+1} e voltar ao primeiro passo do algoritmo.

Caso contrário **parar**.

$G_{\#} = [g_1^* g_2^* \dots g_q^*]$ é a matriz reduzida geradora da região L .

A matriz unimodular F obtém-se através de $F = \text{int}(G^{-1} \cdot G_{\#})$;

¹ Para maior detalhe consultar referência [5].

² Uma matriz F é unimodular quando todos os seus elementos e da sua inversa F^{-1} são inteiros e verifica ainda $|\det(F)| = 1$.