

## 2

### Estimação de máxima verossimilhança em arrays

Em um sistema de comunicações móveis, o processamento espacial realizado no *array* de antenas permite aumento no número de usuários e no ganho das antenas devido à diretividade. Além disso, possibilita redução na interferência total do sistema. Dois atributos fundamentais de um sistema móvel baseado em *arrays* são formação precisa do feixe e rastreamento em tempo real. O primeiro envolve qualidade de cobertura e o último, redução na probabilidade de terminação forçada de chamadas. Estimação de DOA é um ponto crítico no desempenho destas duas necessidades.

Um *array* de antenas é um conjunto de sensores para sinais eletromagnéticos dispostos no espaço obedecendo alguma geometria predeterminada. A Fig. 2.1 ilustra um instantâneo de um sinal propagado atingindo um *array* de cinco elementos. Considerando que o sinal-fonte possui banda estreita e

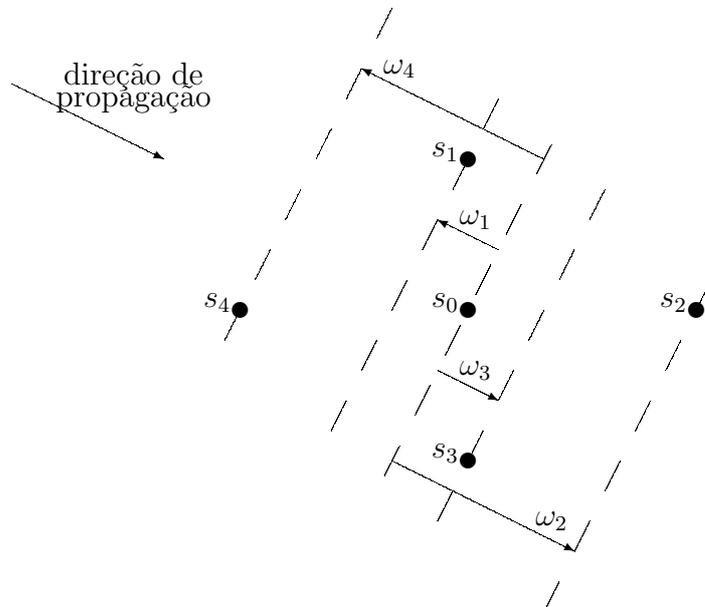


Figura 2.1: Diferenças de fase obtidas entre os elementos de um *array* em relação ao elemento de referência  $s_0$  para determinada direção de propagação.

que é originado a uma distância suficientemente grande do *array*, pode-se dizer que suas frentes de onda são planas ao atingirem os elementos do *array*.

Uma geometria de *array* de antenas amplamente utilizada na literatura é a do *array* linear uniforme (*uniform linear array*—ULA). Esta geometria, ilustrada na Fig. 2.2, determina que todos os elementos do *array* estão dispostos em uma mesma reta e são igualmente espaçados. Desta forma, as defasagens medidas em cada elemento com relação ao elemento de referência (que nesta geometria costuma se situar em uma das extremidades do *array*) serão sempre múltiplas da defasagem medida no elemento adjacente ao elemento de referência. Esta geometria, porém, é a mais passível de ambigüidades na estimação da DOA. Para reduzir ao máximo este problema, pode-se provar que a distância entre dois elementos adjacentes de um ULA não deve exceder a metade do comprimento de onda do sinal [27]. Um ULA cujos elementos estão espaçados de metade do comprimento de onda do sinal propagado é chamado de *array* linear padrão [27] (*standard linear array*—SLA). Esta configuração será adotada ao longo de toda esta tese. Ao mesmo tempo que esta configuração simplifica os cálculos posteriores, ela restringe a resolução das estimativas de DOA a um semiplano. Isto pode ser compreendido mais facilmente com o auxílio de um cone de revolução cujo eixo é a reta formada pelos sensores (linha do *array*). Todos os sinais que atingem o *array* em direções de propagação contidas na parede deste cone proporcionam as mesmas medidas e diferenças de fase nos elementos do *array*, o que impossibilita ao estimador precisar a direção de propagação do sinal em um contexto tridimensional. No entanto, um feixe formado por

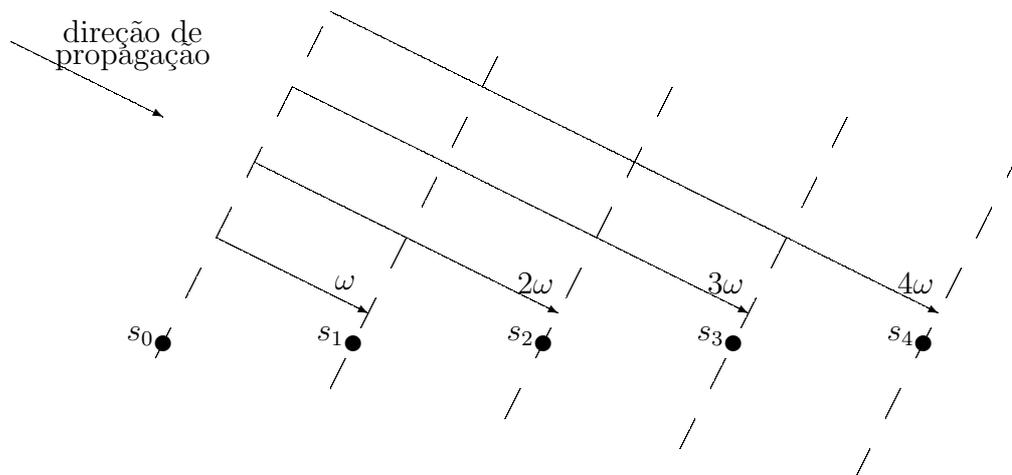


Figura 2.2: Ilustração de um *array* linear uniforme (ULA). As diferenças de fase obtidas são múltiplas do mesmo parâmetro  $\omega$ .

uma SLA cujos elementos são antenas isotrópicas apresenta diagrama de irradiação omnidirecional com relação a um plano perpendicular à linha do *array*, fornecendo simultaneamente cobertura a todas as direções de propagação contidas no mesmo cone. Assim, é possível reduzir o problema de estimação de DOA a um único semiplano sem perda de generalidade.

## 2.1 Modelo genérico

O modelo detalhado nesta seção é idêntico ao modelo genérico apresentado por Stoica e Nehorai em [21], salvo algumas mudanças de notação. Ele é baseado na transmissão simultânea de  $n$  sinais, que atingem um *array* com  $m$  sensores durante um intervalo de tempo suficiente para serem coletadas  $N$  amostras.

O  $u$ -ésimo sinal transmitido atinge o elemento de referência do *array* com envoltória complexa  $x_u(k)$ , onde  $k = 1, 2, \dots, N$ . De acordo com a Fig. 2.2, a leitura deste sinal no  $v$ -ésimo elemento do SLA é dada por

$$x_u(k) \cdot e^{j(v-1)\omega_u} \quad (2-1)$$

onde  $\omega_u$  representa a diferença de fase associada à DOA deste sinal. Como  $n$  sinais estão sendo transmitidos simultaneamente, a envoltória complexa da medida obtida no  $v$ -ésimo elemento do *array*,  $y_v(k)$ , pode ser escrita como

$$y_v(k) = \sum_{u=1}^n x_u(k) \cdot e^{j(v-1)\omega_u} + e_v(k) \quad (2-2)$$

onde  $e_v(k)$  é um ruído aditivo gaussiano branco (RAGB) complexo de média nula cujas componentes em fase e quadratura possuem variância desconhecida  $\sigma^2$ . Seja  $\mathbf{x}(k)$  um vetor-coluna  $n$ -dimensional descrito por

$$\mathbf{x}(k) = [ x_1(k) \quad x_2(k) \quad \cdots \quad x_n(k) ]^T, \quad k = 1, 2, \dots, N \quad (2-3)$$

e  $\mathbf{y}(k)$  um vetor-coluna  $m$ -dimensional contendo as envoltórias complexas das leituras em todos os elementos do *array*, ou seja,

$$\mathbf{y}(k) = [ y_1(k) \quad y_2(k) \quad \cdots \quad y_m(k) ]^T, \quad k = 1, 2, \dots, N \quad (2-4)$$

pode-se então generalizar (2-2) para uma forma matricial, como em

$$\mathbf{y}(k) = \mathbf{S}(\boldsymbol{\omega}) \cdot \mathbf{x}(k) + \mathbf{e}(k), \quad k = 1, 2, \dots, N \quad (2-5)$$

onde  $\mathbf{S}(\boldsymbol{\omega})$  é a matriz de direcionamento (*steering matrix*), função apenas da geometria do *array* e das DOAs dos sinais, e  $\mathbf{e}(k)$  é um vetor complexo gaussiano de média nula cujas componentes em fase e quadratura são independentes, possuindo ambas matriz correlação igual a  $\sigma^2 \mathbf{I}$  (ruído branco). A matriz  $\mathbf{S}(\boldsymbol{\omega})$ , de dimensão  $m \times n$ , é constituída por  $n$  vetores-coluna  $\mathbf{s}(\omega_u)$  chamados vetores de direcionamento, isto é,

$$\mathbf{S}(\boldsymbol{\omega}) = [ \mathbf{s}(\omega_1) \quad \mathbf{s}(\omega_2) \quad \cdots \quad \mathbf{s}(\omega_n) ] \quad (2-6)$$

onde cada vetor de direcionamento é função de uma única DOA e é definido de acordo com a geometria do *array*. Para SLA, tem-se que

$$\mathbf{s}(\omega) = [ 1 \quad e^{j\omega} \quad e^{j2\omega} \quad \dots \quad e^{j(m-1)\omega} ]^T \quad (2-7)$$

é definido em um espaço  $m$ -dimensional complexo. Variando  $\omega$  pelo intervalo  $[-\pi, \pi]$ , obtém-se todos os possíveis valores do vetor  $\mathbf{s}(\omega)$ . Este conjunto de valores é conhecido como *array manifold* e compreende a solução do problema de estimação de DOA em formato de vetor de direcionamento [27]. Alguns métodos baseados em subespaço de sinais, como o ESPRIT [17], podem apresentar uma solução geométrica para a estimação de DOA através da interseção entre o *array manifold* e o subespaço de sinais [14]. Uma representação geométrica similar também pode ser desenvolvida para os métodos baseados em subespaço de ruído, como o MUSIC [19]. Métodos baseados em subespaço não serão estudados neste trabalho.

## 2.2

### Estimador de máxima verossimilhança determinístico

Nesta seção, o estimador ML desenvolvido em [21] para o modelo de sinais detalhado na seção anterior é apresentado. Estimadores ML utilizam o conhecimento probabilístico do modelo para descobrir quais valores seriam mais adequados aos parâmetros dado um determinado conjunto de observações.

A análise apresentada em [21] considera que os sinais  $\mathbf{x}(k)$  são parâmetros (funções) complexos desconhecidos a estimar. Como o modelo desenvolvido tem como objetivo principal abranger o maior número possível de aplicações dentro do mesmo cenário (sinais de banda estreita, RAGB, etc.), a variância do ruído  $\sigma^2$  também é um parâmetro desconhecido. Como o vetor de ruído é gaussiano de média nula, a função densidade de probabi-

lidade (fdp) condicional do vetor  $\mathbf{y}(k)$ ,  $p_{\mathbf{y}(k)|\boldsymbol{\omega}, \mathbf{x}(k), \sigma^2}(\mathbf{Y}(k))$ , é dada por

$$\begin{aligned} p_{\mathbf{y}(k)|\boldsymbol{\omega}, \mathbf{x}(k), \sigma^2}(\mathbf{Y}(k)) &= \\ &= \frac{1}{(2\pi\sigma^2)^m} \exp \left\{ -\frac{1}{2\sigma^2} [\mathbf{Y}(k) - \mathbf{S}(\boldsymbol{\omega}) \cdot \mathbf{x}(k)]^H [\mathbf{Y}(k) - \mathbf{S}(\boldsymbol{\omega}) \cdot \mathbf{x}(k)] \right\} \end{aligned} \quad (2-8)$$

onde  $^H$  representa o operador hermitiano. Como as amostras de ruído para instantes de tempo distintos são independentes, tem-se que a função de verossimilhança  $\ell$  dos dados observados, definida como a fdp condicional conjunta para todos os instantes é

$$\begin{aligned} \ell(\boldsymbol{\omega}, \mathbf{x}(k), \sigma^2) &= \\ &= \frac{1}{(2\pi\sigma^2)^{mN}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^N [\mathbf{Y}(k) - \mathbf{S} \cdot \mathbf{x}(k)]^H [\mathbf{Y}(k) - \mathbf{S} \cdot \mathbf{x}(k)] \right\} \end{aligned} \quad (2-9)$$

sendo o vetor de parâmetros  $\boldsymbol{\omega}$  omitido por simplicidade de notação. A função de menos-log-verossimilhança  $\mathcal{L}$  é então

$$\mathcal{L}(\boldsymbol{\omega}, \mathbf{x}(k), \sigma^2) = \text{cte} + mN \ln \sigma^2 + \frac{1}{2\sigma^2} \sum_{k=1}^N [\mathbf{Y}(k) - \mathbf{S} \cdot \mathbf{x}(k)]^H [\mathbf{Y}(k) - \mathbf{S} \cdot \mathbf{x}(k)] \quad (2-10)$$

Derivando  $\mathcal{L}$  com relação a  $\sigma^2$  e às partes real e imaginária de  $\mathbf{x}(k)$ , obtém-se

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = \frac{mN}{\sigma^2} - \frac{1}{2\sigma^4} \sum_{k=1}^N [\mathbf{Y}(k) - \mathbf{S} \cdot \mathbf{x}(k)]^H [\mathbf{Y}(k) - \mathbf{S} \cdot \mathbf{x}(k)] \quad (2-11)$$

$$\frac{\partial \mathcal{L}}{\partial \Re\{\mathbf{x}(k)\}} = -\frac{1}{\sigma^2} \Re \{ \mathbf{S}^H [\mathbf{Y}(k) - \mathbf{S} \cdot \mathbf{x}(k)] \} \quad (2-12)$$

$$\frac{\partial \mathcal{L}}{\partial \Im\{\mathbf{x}(k)\}} = -\frac{1}{\sigma^2} \Im \{ \mathbf{S}^H [\mathbf{Y}(k) - \mathbf{S} \cdot \mathbf{x}(k)] \} \quad (2-13)$$

Igualando estas derivadas a zero, tem-se as seguintes estimativas para  $\sigma^2$  e  $\mathbf{x}(k)$

$$\hat{\sigma}^2 = \frac{1}{2mN} \sum_{k=1}^N [\mathbf{Y}(k) - \mathbf{S} \cdot \mathbf{x}(k)]^H [\mathbf{Y}(k) - \mathbf{S} \cdot \mathbf{x}(k)] \quad (2-14)$$

$$\hat{\mathbf{x}}(k) = (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{Y}(k) = \mathbf{S}^\dagger \mathbf{Y}(k) \quad (2-15)$$

onde  $\mathbf{S}^\dagger$  é a pseudo-inversa da matriz  $\mathbf{S}$ . Substituindo primeiramente a estimativa de  $\sigma^2$  e, finalmente, a estimativa de  $\mathbf{x}(k)$  em (2-10), obtém-se

que a estimativa de  $\boldsymbol{\omega}$  é dada por

$$\begin{aligned}\hat{\boldsymbol{\omega}} &= \min^{-1} \frac{1}{N} \sum_{k=1}^N \mathbf{Y}^H(k) [\mathbf{I} - \mathbf{S}\mathbf{S}^\dagger]^H [\mathbf{I} - \mathbf{S}\mathbf{S}^\dagger] \mathbf{Y}(k) = \\ &= \min^{-1} \text{tr} \left\{ [\mathbf{I} - \mathbf{S}\mathbf{S}^\dagger]^2 \frac{1}{N} \sum_{k=1}^N \mathbf{Y}(k) \mathbf{Y}^H(k) \right\} = \min^{-1} \text{tr} \left\{ [\mathbf{I} - \mathbf{S}\mathbf{S}^\dagger] \hat{\mathbf{R}}_{\mathbf{Y}} \right\}\end{aligned}\quad (2-16)$$

onde ‘tr’ representa o operador traço da matriz e  $\hat{\mathbf{R}}_{\mathbf{Y}} = \frac{1}{N} \sum_{k=1}^N \mathbf{Y}(k) \mathbf{Y}^H(k)$  é a matriz de correlação assintótica do vetor de observações. Ela tende para a matriz de correlação real quando  $N$  tende para infinito. Como a matriz  $\mathbf{S}$  é uma função não-linear de  $\boldsymbol{\omega}$ , minimizar (2-16) é uma tarefa que demanda certa carga computacional. Mais ainda, a função obtida em (2-16) apresenta vários mínimos locais, fazendo com que a obtenção da estimativa por força bruta (varredura em *grid*  $n$ -dimensional) seja mais recomendada que através de um algoritmo baseado em gradiente descendente. Uma boa alternativa seria a utilização de um algoritmo combinado, obtendo uma estimativa inicial próxima da verdadeira através de varredura em escala grosseira (força bruta) e refinando esta estimativa com um algoritmo de gradiente descendente (convergência iterativa).

### 2.3

#### Estimador de máxima verossimilhança estocástico

Nesta segunda abordagem, os sinais  $\mathbf{x}(k)$  não são mais considerados como parâmetros desconhecidos, e sim como amostras de um processo estocástico de média nula, decorrelatado no tempo e com matriz de correlação espacial  $\mathbf{R}_{\mathbf{x}}$  desconhecida. Desta forma, a função de verossimilhança não será mais função dos parâmetros  $\boldsymbol{\omega}$ ,  $\mathbf{x}(k)$  e  $\sigma^2$ , mas dos parâmetros  $\boldsymbol{\omega}$ ,  $\mathbf{R}_{\mathbf{x}}$  e  $\sigma^2$ . A matriz de correlação espacial do vetor de observações,  $\mathbf{R}_{\mathbf{y}}$ , é dada por

$$\mathbf{R}_{\mathbf{y}} = E[\mathbf{y}(k) \cdot \mathbf{y}^H(k)] = \mathbf{S}(\boldsymbol{\omega}) \mathbf{R}_{\mathbf{x}} \mathbf{S}^H(\boldsymbol{\omega}) + 2\sigma^2 \mathbf{I} \quad (2-17)$$

já que  $\mathbf{x}(k)$  e  $\mathbf{e}(k)$  são independentes. Enunciando o lema de inversão de matrizes para a expressão definida em (2-17), tem-se

$$\mathbf{R}_{\mathbf{y}}^{-1} = \frac{1}{2\sigma^2} \left[ \mathbf{I} - \mathbf{S} (\mathbf{R}_{\mathbf{x}} \mathbf{S}^H \mathbf{S} + 2\sigma^2 \mathbf{I})^{-1} \mathbf{R}_{\mathbf{x}} \mathbf{S}^H \right] \quad (2-18)$$

Se o processo que origina o vetor  $\mathbf{x}(k)$  for gaussiano ou se o número  $n$  de sinais for suficientemente grande, pode-se adotar a fdp condicional de  $\mathbf{y}(k)$

como sendo gaussiana. Logo, a função de menos-log-verossimilhança é

$$\mathcal{L}(\boldsymbol{\omega}, \mathbf{R}_x, \sigma^2) = \text{cte} + \frac{N}{2} \ln \det [\mathbf{R}_y] + \frac{N}{2} \text{tr} \left\{ \mathbf{R}_y^{-1} \hat{\mathbf{R}}_Y \right\} \quad (2-19)$$

onde  $\hat{\mathbf{R}}_Y$  já foi definido na seção anterior. Considerando que [9]

$$\frac{\partial f(\mathbf{R}_y)}{\partial \theta} = \text{tr} \left\{ \left[ \frac{\partial f(\mathbf{R}_y)}{\partial \mathbf{R}_y} \right]^T \frac{\partial \mathbf{R}_y}{\partial \theta} \right\} \quad (2-20)$$

onde  $\theta$  é um parâmetro escalar e  $f$  é uma função qualquer, e considerando também que [9]

$$\frac{\partial \ln \det [\mathbf{R}_y]}{\partial \mathbf{R}_y} = [\mathbf{R}_y^{-1}]^T \quad (2-21)$$

$$\frac{\partial \text{tr} [\mathbf{R}_y^{-1} \hat{\mathbf{R}}_Y]}{\partial \mathbf{R}_y} = -\mathbf{R}_y^{-1} \hat{\mathbf{R}}_Y \mathbf{R}_y^{-1} \quad (2-22)$$

tem-se que as derivadas com relação a  $\mathbf{R}_x$  e  $\sigma^2$  são as seguintes [9],[30]

$$\frac{\partial \mathcal{L}}{\partial \mathbf{R}_x} = \frac{N}{2} (\mathbf{S}^H \boldsymbol{\Phi} \mathbf{S})^T \quad (2-23)$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = N \text{tr}(\boldsymbol{\Phi}) \quad (2-24)$$

onde  $\boldsymbol{\Phi} = \mathbf{R}_y^{-1} (\mathbf{R}_y - \hat{\mathbf{R}}_Y) \mathbf{R}_y^{-1}$ . Igualando a zero e utilizando (2-18), obtém-se as estimativas [9],[30]

$$\hat{\mathbf{R}}_x = \mathbf{S}^\dagger \left( \hat{\mathbf{R}}_Y - 2\hat{\sigma}^2 \mathbf{I} \right) \mathbf{S}^{\dagger H} \quad (2-25)$$

$$\hat{\sigma}^2 = \frac{1}{2(m-n)} \text{tr} \left\{ [\mathbf{I} - \mathbf{S}\mathbf{S}^\dagger] \hat{\mathbf{R}}_Y \right\} \quad (2-26)$$

Substituindo estas estimativas em (2-19), tem-se que

$$\hat{\boldsymbol{\omega}} = \min^{-1} \det \left[ \mathbf{S}\mathbf{S}^\dagger \hat{\mathbf{R}}_Y \mathbf{S}\mathbf{S}^\dagger + \frac{1}{m-n} \text{tr} \left\{ [\mathbf{I} - \mathbf{S}\mathbf{S}^\dagger] \hat{\mathbf{R}}_Y \right\} (\mathbf{I} - \mathbf{S}\mathbf{S}^\dagger) \right] \quad (2-27)$$

é a estimativa ML das diferenças de fase ocasionadas pela DOA. Percebe-se que, tanto neste estimador quanto no estimador determinístico, a estimativa obtida é a do vetor de parâmetros  $\boldsymbol{\omega}$ , ao invés das DOAs propriamente ditas. Pelo fato da DOA ter uma relação biunívoca com a diferença de fase  $\omega$ , pode-se provar [26] que as estimativas ótimas comutam.

Mais uma vez, encontra-se o mesmo problema de não-linearidade para a obtenção do mínimo desta função. Os mesmos procedimentos de-

talhados na seção anterior também são aplicáveis nesta situação. Conforme foi apresentado, a principal diferença entre as abordagens determinística e estocástica se refere à função de verossimilhança. Enquanto que na abordagem determinística a fdp condicional de  $\mathbf{y}(k)$  era caracterizada como  $N(\mathbf{S}x, 2\sigma^2\mathbf{I})$ , na abordagem estocástica esta função é dada por  $N(\mathbf{0}, \mathbf{S}R_x\mathbf{S}^H + 2\sigma^2\mathbf{I})$ . Conforme será detalhado nos próximos capítulos, o modelo estatístico desenvolvido utiliza a abordagem determinística para obtenção das expressões que regem os estimadores propostos nesta tese.