Recuperação Cega de Sinais

Conceitos, Estratégias e Tendências

Charles Casimiro Cavalcante[1]

charles@gtel.ufc.br

João Marcos T. Romano[2]

romano@decom.fee.unicamp.br

[1] Grupo de Pesquisa em Telecomunicações Sem Fio – GTEL
Universidade Federal do Ceará – UFC
http://www.gtel.ufc.br

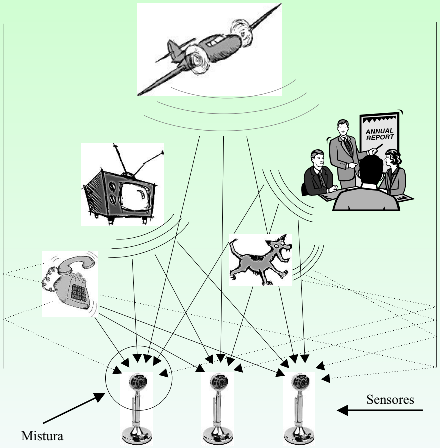
[2] Lab. de Processamento de Sinais para Comunicações – DSPCom
Universidade Estadual de Campinas – UNICAMP
http://www.dspcom.fee.unicamp.br

XXI Simpósio Brasileiro de Telecomunicações (SBT2004)
6-9 de Setembro, Belém-PA, Brasil

Motivação

- Modelagem abrange uma variedade muito grande de problemas
- Estratégias utilizam conhecimentos de várias disciplinas
  - Processamento de sinais
  - Estatística
  - Teoria da informação
- Aplicação em diversas áreas
  - Comunicações
  - Processamento de imagens
  - Reconhecimento de padrões
  - Econometria
  - ...

Contexto - *Cocktail party*



Objetivos

- Modelagem matemática do problema de recuperação cega de sinais
  - Multidimensional
  - Unidimensional
- Estratégias de recuperação cega de sinais
- Modelagem de sistemas de comunicação multi-usuário no contexto de recuperação de sinais
- Ilustrar casos particulares de interesse prático

Conteúdo

Part I

Separação Cega de Fontes

Considerações iniciais I

- Sistemas de comunicação digital: utilização de seqüências conhecidas para remover "dispersões" inseridas pelo meio transmissão
- Eficiência da estratégia é diminuída pela ocupação do espectro com sinais que não contém efetivamente mensagem a ser transmitida
- Estratégias que permitam a recuperação dos sinais envolvidos têm grande apelo prático

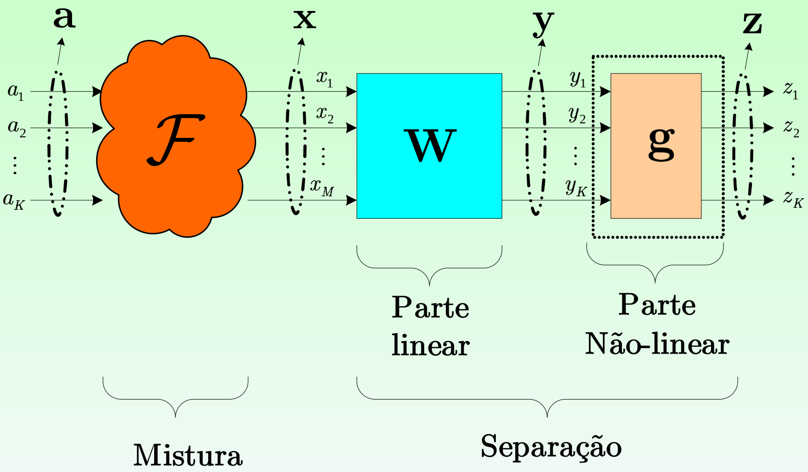
Denominações de tais estratégias

1. *Autodidata*
2. *Cega*
3. *Não-supervisionada*

Considerações iniciais II

- Estudos iniciais: anos 70 e 80, utilizando sistemas de comunicação digital com única fonte
- 1985: modelo genérico, múltiplas fontes, surgimento de uma nova linha de pesquisa denominada *separação cega de fontes*
- Confluência de várias áreas de conhecimento

Modelo geral



Modelagem matemática

Modelo

$$\mathbf{x}(n) = \mathcal{F}\left(\mathbf{a}(n), \mathbf{v}(n), n\right) \longleftarrow \text{Mapeamento}$$

$$\mathbf{a}(n) = \begin{bmatrix} a_1(n) & a_2(n) & \cdots & a_K(n) \end{bmatrix}^T \longleftarrow K \text{ fontes}$$

$$\mathbf{v}(n) = \begin{bmatrix} v_1(n) & v_2(n) & \cdots & v_V(n) \end{bmatrix}^T \longleftarrow V \text{ sinais de ruído}$$

$$\mathbf{x}(n) = \begin{bmatrix} x_1(n) & x_2(n) & \cdots & x_M(n) \end{bmatrix}^T \longleftarrow M \text{ sensores.}$$

Considerações usuais

- $\mathcal{F}$ é linear e invariante no tempo
- Fontes mutuamente independentes e independentes do ruído
- $V = M$
- $M \geq K$ (mais sensores que fontes no mínimo)

Resultando...

Mistura

$$\mathbf{x}(n) = \mathbf{H}\mathbf{a}(n) + \mathbf{v}(n)$$

Separação

$$\mathbf{y}(n) = \mathbf{W}^H\mathbf{x}(n) = \widehat{\mathbf{a}}(n)$$

Características

1. Indeterminação em relação a permutação e escalonamento

$$\mathbf{y}(n) = \mathbf{P}\mathbf{D}\mathbf{a}(n)$$

   $\mathbf{P}$ é uma matriz de permutação de ordem $K \times K$ e $\mathbf{D}$ é uma matriz diagonal e inversível de ordem $K \times K$.

2. Possível inserção de não-linearidade após $\mathbf{W}$

$$\boxed{\mathbf{x} = \mathbf{H}\mathbf{a}}$$

$$\begin{cases} x_1 = h_{11}a_1 + h_{12}a_2 \\ x_2 = h_{21}a_1 + h_{22}a_2 \end{cases}$$

Duas (K) variáveis e duas (M) equações!

Separando as fontes...

- Caso sem ruído: *separar é possível se* $\mathbf{W} = \mathbf{H}^{-1}$

Questão 1

Como identificar $\mathbf{H}$ para projetar $\mathbf{W}$?

Questão 2

Quais (e quantas) estatísticas são necessárias para prover a separação?

Resposta para questão 1: Como identificar $\mathbf{H}$?

- Observando a matriz de autocorrelação do vetor de sinais recebidos:

$$\mathbf{R}_x = \mathbb{E}\left\{\mathbf{x}(n)\mathbf{x}^T(n)\right\} = \mathbf{H}\mathbf{R}_a\mathbf{H}^T = \mathbf{H}\mathbf{H}^T,$$

- Observar que $\mathbf{H}\mathbf{Q}^T$, em que $\mathbf{Q}$ é uma matriz ortogonal também soluciona equação
- Conclui-se que: $\mathbf{H} = \mathbf{R}_x^{\frac{1}{2}}$
- Extração de raiz quadrada de matrizes: através de decomposição em valores singulares (SVD, Singular Value Decomposition)

Identificando $\mathbf{H}$: possibilidades

- Escrevendo
$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\mathcal{V}}^T,$$
em que $\mathbf{U}$ e $\boldsymbol{\mathcal{V}}$ são matrizes retangulares de ordem $K \times M$, tais que $\mathbf{U}\mathbf{U}^T = \boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{V}}^T = \mathbf{I}_M$ e $\mathbf{U}^T\mathbf{U} = \boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}} = \mathbf{I}_K$
- Então...
$$\begin{aligned}\mathbf{R}_x &= \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{V}}^T\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^T \\ &= \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T.\end{aligned}$$
- Resultado: através da matriz de autocorrelação (estatística de ordem 2) só é possível identificar as matrizes $\mathbf{U}$ e $\mathbf{\Lambda}$
- Estatísticas de ordem 2 (SOS, Second Order Statistics) não resolvem o problema por completo
- **Resposta parcial da questão 2: estatísticas necessárias para separação!**

Máximo possível com SOS

Processamento

Projeção dos dados na direção da inversa da matriz de autocorrelação: $\mathbf{T} = \mathbf{R}_x^{-\frac{1}{2}}$

- Assim, tem-se o seguinte conjunto de dados:

$$\bar{\mathbf{x}}(n) = \mathbf{T}\mathbf{x}(n)$$

de tal forma que

$$\boxed{\begin{aligned}\mathbf{R}_{\bar{x}} &= \mathbf{T}\mathbf{R}_x\mathbf{T}^T \\ &= \mathbf{U}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^T\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\mathbf{U}^T\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U} \\ &= \mathbf{I}_K\end{aligned}}$$

- **Branqueamento!** - em separação cega de fontes é denominado de *esferatização*

- Projeção dos dados na direção dos principais autovetores de $\mathbf{R}_x$
- Técnica de análise por componentes principais (PCA, Principal Component Analysis): projeção dos dados nas direções definidas pelos principais autovetores de $\mathbf{H}$
- Matriz de branquemento, ou *transformação de Mahalanobis*, é escrita como

$$\mathbf{T} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}_x^T.$$

- Redução do problema para uma matriz de mistura ortogonal

$$\bar{\mathbf{x}}(n) = \mathbf{THa}(n) = \mathbf{Qa}(n)$$

em que $\mathbf{QQ}^T = \mathbf{I}$.

Encontrando a matriz $\boldsymbol{\mathcal{V}}$

Considerar um atraso arbitrário ℓ para o qual, não haja duas fontes, i e j, com a mesma autocorrelação

$$\mathbb{E}\left\{a_i(n)a_i(n-\ell)\right\} \neq \mathbb{E}\left\{a_j(n)a_j(n-\ell)\right\} \qquad \forall i \neq j,$$

Diferentes densidades espectrais!

$$\boxed{\begin{aligned}
\mathbf{R}_{\bar{x}}(k) &= \mathbb{E}\left\{\bar{\mathbf{x}}(n)\bar{\mathbf{x}}^T(n-\ell)\right\} \\
&= \mathbf{T}\mathbf{H}\mathbf{R}_a(\ell)\mathbf{H}\mathbf{T}^T \\
&= \underbrace{\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^T}_{\mathbf{T}}\underbrace{\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\mathcal{V}}^T}_{\mathbf{H}}\mathbf{R}_a(\ell)\underbrace{\boldsymbol{\mathcal{V}}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}^T}_{\mathbf{H}^T}\underbrace{\mathbf{U}\boldsymbol{\Lambda}^{-\frac{1}{2}}}_{\mathbf{T}^T} \\
&= \boldsymbol{\mathcal{V}}^T\mathbf{R}_a(\ell)\boldsymbol{\mathcal{V}}.
\end{aligned}}$$

Informações restantes de $\mathbf{H}$ – II

- Fontes independentes $\Rightarrow \mathbf{R}_a(\ell)$ diagonal
- Autovalores distintos: fontes com diferentes autocorrelações
- Conclusão: *uso somente de SOS permite separar fontes com espectros diferentes!*
- Algoritmo clássico: *Algorithm for Multiple Unknown Signals Extraction* (AMUSE)
- Na prática, fontes com espectros similares (embora diferentes) não são separáveis
- Impossibilidade de separar fonte brancas ou i.i.d.

Como separar fontes com mesmo espectro?

Proposição de uma nova técnica

- Não procura-se estimar $\mathbf{H}$ para projetar $\mathbf{W}$
- Fator decisivo: suposição de independência das fontes!
- Projetar $\mathbf{W}$ de maneira que sejam obtidas fontes *o mais independentes possível* na saída do dispositivo de separação
- *Independent Component Analysis (ICA)*
- Restrição: no máximo uma fonte pode ser gaussiana
  - Teorema Central do Limite: mistura de gaussianas é gaussiana!
- Interpretação: ponto chave é a não-gaussianidade das fontes.

Questão

Como medir a não-gaussianidade das fontes e utilizar o fato de que as mesmas são estatisticamente independentes para separá-las?

Medidas de independência/gaussianidade – I

Indepedência estatística

$$p_y(\mathbf{y}) \triangleq \prod_{i=1}^{K} p_{y_i}(y_i)$$

$$\mathbb{E}\{y_1 \cdot y_2 \cdots y_K\} = \mathbb{E}\{y_1\} \cdot \mathbb{E}\{y_2\} \cdots \mathbb{E}\{y_K\}$$

Descorrelação estatística

$$\mathbb{E}\{y_1 \cdot y_2 \cdots y_K\} - \mathbb{E}\{y_1\} \cdot \mathbb{E}\{y_2\} \cdots \mathbb{E}\{y_K\} = 0.$$

Independência $\underset{\not\Leftarrow}{\Rightarrow}$ Descorrelação

Entropia e informação mútua

$$\mathcal{H}(\mathbf{x}) \triangleq -\mathbb{E}\left\{\ln\left[p_x(\mathbf{x})\right]\right\} = -\int_{-\infty}^{\infty} p_x(\mathbf{x}) \cdot \ln\left[p_x(\mathbf{x})\right] d\mathbf{x}$$

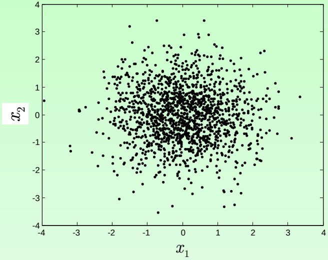
entropia diferencial

$$\mathcal{H}(\mathbf{x}|\mathbf{y}) = -\mathbb{E}\left\{\ln\left[p_{x|y}(\mathbf{x}|\mathbf{y})\right]\right\} = \int p_{x,y}(\mathbf{x},\mathbf{y}) \cdot \ln\left[p_{x|y}(\mathbf{x}|\mathbf{y})\right] d\mathbf{x}d\mathbf{y}$$

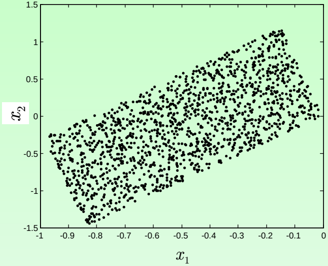
entropia condicional

$$I(\mathbf{x},\mathbf{y}) = \mathcal{H}(\mathbf{x}) - \mathcal{H}(\mathbf{x}|\mathbf{y})$$

informação mútua entre $\mathbf{x}$ e $\mathbf{y}$



Mistura de fontes gaussianas



Mistura de fontes uniformes

Não há direções preferenciais!

Divergência de Kullback-Leibler (KLD)

$$\boxed{D\left(p_x(\mathbf{x})||g_x(\mathbf{x})\right) \triangleq \int_{-\infty}^{\infty} p_x(\mathbf{x}) \cdot \ln\left[\frac{p_x(\mathbf{x})}{g_x(\mathbf{x})}\right] dx}$$

em que $p(x)$ e $g(x)$ são duas funções estritamente positivas

Propriedades:

1. é sempre de valor positivo ou zero; KLD é zero para o caso específico de $p_x(\mathbf{x}) = g_x(\mathbf{x})$.
2. é invariante com relação as seguintes mudanças nas componentes do vetor $\mathbf{x}$;
   - permutação de ordem
   - escalonamento de amplitude
   - transformação monotônica não-linear

Usando a KLD...

$$
\begin{aligned}
I(\mathbf{x}, \mathbf{y}) &= \int p_{x,y}(\mathbf{x}, \mathbf{y}) \cdot \ln \left[\frac{p_{x,y}(\mathbf{x}, \mathbf{y})}{p_x(\mathbf{x})\, p_y(\mathbf{y})} \right] d\mathbf{x} d\mathbf{y} \\
&= D\left(p_{x,y}(\mathbf{x}, \mathbf{y}) \| p_x(\mathbf{x})\, p_y(\mathbf{y})\right)
\end{aligned}
$$

ou ainda

$$
\begin{aligned}
I(\mathbf{y}) &= D\left(p_y(\mathbf{y}) \| \widetilde{p}_y(\mathbf{y})\right) \\
&= \int \cdots \int p_y(y_1, \ldots, y_K) \cdot \ln \left[\frac{p_y(y_1, \ldots, y_K)}{\prod_{i=1}^{K} p_{y_i}(y_i)} \right] dy_1 \cdots dy_K
\end{aligned}
$$

Finalmente...

$$I(\mathbf{y}) = -\mathcal{H}(\mathbf{y}) + \sum_{i=1}^{K} \mathcal{H}(y_i),$$

o que significa que minimizar a informação mútua entre os componentes do vetor $\mathbf{y}$ significa tornar a entropia de $\mathbf{y}$ o mais próximo possível da soma de suas entropias marginais.

Negentropia - I

- Medida de "não-gaussianidade" baseado na medida de entropia diferencial
- Diferença entre a entropia da v.a. $\mathbf{y}$ e a entropia de uma v.a. $\mathbf{y}^G$ de distribuição gaussiana e com os mesmos momentos de ordem um e dois (média e variância) de $\mathbf{y}$

$$N_G(\mathbf{y}) \triangleq \mathcal{H}(\mathbf{y}^G) - \mathcal{H}(\mathbf{y}).$$

ou

$$N_G(\mathbf{y}) \triangleq D\left(p_y(\mathbf{y}) \| p_{y^G}(\mathbf{y})\right)$$

- Problema: requer conhecimento ou estimativa da densidade de probabilidade de $\mathbf{y}$!

Medidas de independência/gaussianidade – VII

Negentropia - II

- Possibilidade de expressar estatísticas necessárias para separação
- Utilizando KLD e informação mútua pode-se escrever:

$$I\left(\mathbf{y}\right) = I\left(\mathbf{y}^{G}\right) + \left(N_G\left(\mathbf{y}\right) - \sum_{i=1}^{K} N_G\left(y_i\right)\right).$$

1. Primeiro termo utiliza SOS
2. Segundo termo mede não-gaussianidade do sinal

Kurtosis

- Cumulante de quarta ordem

$$\mathcal{K}\{y\} \triangleq \mathbb{E}\{y^4\} - 3 \cdot \left(\mathbb{E}\{y^2\}\right)^2 .$$

- Faixa de valores:
  - Distribuição gaussiana: $\mathcal{K}\{y\} = 0$
  - Distribuição sub-gaussiana: $\mathcal{K}\{y\} \leq 0$
  - Distribuição super-gaussiana: $\mathcal{K}\{y\} \geq 0$
- Propriedades

$$\mathcal{K}\{y_1 + y_2\} = \mathcal{K}\{y_1\} + \mathcal{K}\{y_2\}$$
$$\mathcal{K}\{\alpha \cdot y\} = \alpha^4 \cdot \mathcal{K}\{y\}$$

Funções de contraste

- Uma função $\Psi(\cdot)$, no espaço de K fdps (distintas ou não) é dita ser um *contraste* se respeita as seguintes condições:
  1. $\Psi(p_y)$ é invariante a permutações:

     $$\Psi(p_{Py}) = \Psi(p_y) \text{ para qualquer matriz de permutação } \mathbf{P}$$

  2. $\Psi(p_y)$ é invariante a mudanças de escala:

     $$\Psi(p_{Dy}) = \Psi(p_y) \text{ para qualquer matriz diagonal } \mathbf{D}$$

  3. Se $\mathbf{y}$ possui componentes independentes, então:

     $$\Psi(p_{Wy}) \leq \Psi(p_y) \text{ para qualquer matriz inversível } \mathbf{W}$$

Algumas funções de contraste

$$\Psi_{\mathsf{ICA}}\left(p_y\right) = -I\left(\mathbf{y}\right)$$

ou ainda em relação à matriz de separação

$$\Psi_{\mathsf{ICA}}\left(\mathbf{W}\right) = \ln\left[\left|\det\left(\mathbf{W}\right)\right|\right] - \mathbb{E}\left\{\ln\left[\prod_{i=1}^{K} p_{y_i}\left(y_i\right)\right]\right\}$$

Medidas de independência/gaussianidade – XI

Aproximações da negentropia

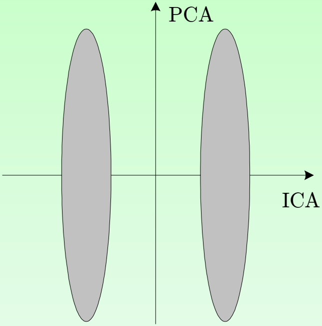
- Necessário para evitar estimações da fdp
- Momentos de ordem superior (clássica)

$$N_G(\mathbf{y}) \approx \frac{1}{12}\left[\mathbb{E}\left\{\mathbf{y}^3\right\}\right]^2 + \frac{1}{48}\left[\mathcal{K}\left\{\mathbf{y}\right\}\right]^2,$$

- Classe de aproximações (*FastICA*)

$$\boxed{\begin{gathered} N_G\left(\mathbf{y}\right) \approx \sum_{i=1}^{p} \varrho_i \cdot \left[\mathbb{E}\left\{g_i\left(\mathbf{y}\right)\right\} - \mathbb{E}\left\{g_i\left(\nu\right)\right\}\right]^2 \\ g_1(u) = \frac{1}{a_1}\log\left[\cosh\left(a_1 u\right)\right] \text{ e} \\ g_2(u) = -\exp\left(-\frac{u^2}{2}\right) \end{gathered}}$$

Diferença ICA × PCA



1. PCA: projeta numa dimensão menor preservando *máxima variância* dos dados
2. ICA: projeta numa dimensão menor preservando *estrutura* dos dados

MaxEnt/InfoMax - I

- Separação é obtida quando estimativas das fontes são independentes $\Rightarrow I(\mathbf{y}) = 0$

$$I(\mathbf{z}, \mathbf{x}) = \mathcal{H}(\mathbf{z}) - \mathcal{H}(\mathbf{z}|\mathbf{x})$$

- Mapeamento é determinístico, logo $\mathcal{H}(\mathbf{z}|\mathbf{x}) = 0$

$$I(\mathbf{y}, \mathbf{x}) = \mathcal{H}(\mathbf{y})$$

- Entropia de $\mathbf{y}$ não é limitada, então

$$\mathbf{z} = \mathbf{g}(\mathbf{W}\mathbf{x})$$

Estratégias de separação cega de fontes - II

MaxEnt/InfoMax - II

- Funções $g_i(\cdot)$ monotonicamente crescentes, limitadas de tal forma que $g_i(-\infty) = 0$ e $g_i(\infty) = 1$
- Se $g_i(\cdot)$ for igual à função de distribuição cumulativa (fdc) da i-ésima fonte
$$p_z(\mathbf{z}) = U[0,1], \quad \text{para}$$
- Adaptação
$$\Delta\mathbf{W} \propto \left(\mathbf{W}^{-T}\right)^{-1} - 2 \cdot \tanh(\mathbf{W}\mathbf{x})\,\mathbf{x}^T,$$
- Função contraste
$$\max_{\mathbf{W}} \left(\Psi_{\mathsf{InfoMax}}(\mathbf{W}) \triangleq \ln[|\det(\mathbf{W})|] - \mathbb{E}\left\{ \ln\left[\prod_{i=1}^{K} g_i'(y_i)\right]\right\}\right).$$

Máxima verossimilhança - I

- Entropia

$$\mathcal{H}(\mathbf{z}) = -\int p_z(\mathbf{z}) \ln \left[\frac{p_z(\mathbf{z})}{\prod_{i=1}^{K} U(z_i)} \right] d\mathbf{z} \quad = -D\left(p_z(\mathbf{z}) \| U_N(\mathbf{z})\right)$$

- Modelo paramétrico $\boldsymbol{\Theta}$

$$\mathcal{L}_Q(\boldsymbol{\Theta}) = \frac{1}{Q} \ln \left[\prod_{q=1}^{Q} p_x\left(\mathbf{x}(q)|\boldsymbol{\Theta}\right) \right] = \frac{1}{Q} \sum_{q=1}^{Q} \ln \left[p_x\left(\mathbf{x}(q)|\boldsymbol{\Theta}\right) \right].$$

Máxima verossimilhança - II

- Lei dos grandes números

$$\mathcal{L}_Q\left(\boldsymbol{\Theta}\right) \xrightarrow{Q\to\infty} \mathcal{L}\left(\mathbf{W}\right) \triangleq \int p_y\left(\mathbf{y}|\mathbf{W}\right) \ln\left[p_{\widetilde{a}}(\mathbf{y})\right] d\mathbf{y} + \ln\left[\left|\det\left(\mathbf{W}\right)\right|\right].$$

- Usando KLD

$$\begin{aligned}\mathcal{L}\left(\mathbf{W}\right) &= -D\left(p_x\left(\mathbf{y}|\mathbf{W}\right)\|p_{\widetilde{a}}(\mathbf{y})\right) - \mathcal{H}\left(\mathbf{y}|\mathbf{W}\right) + \ln\left[\left|\det\left(\mathbf{W}\right)\right|\right] \\ &= -D\left(p_x\left(\mathbf{y}|\mathbf{W}\right)\|p_{\widetilde{a}}(\mathbf{y})\right) - \mathcal{H}\left(\mathbf{x}\right)\end{aligned}.$$

- Função contraste MV

$$\Psi_{\mathsf{MV}}\left(\mathbf{W}\right) = -D\left(\mathbf{WHa}\|\widetilde{\mathbf{a}}\right).$$

Máxima verossimilhança - III

- Interpretação da MV

$$-\Psi_{\mathsf{MV}}(\mathbf{W}) = -\Psi_{\mathsf{ICA}}(\mathbf{W}) + \sum_{i=1}^{K} D\left(\widetilde{z}_i \| \widetilde{a}_i\right)$$

ou seja

$$\begin{pmatrix} \text{Desvio} \\ \text{total} \end{pmatrix} = \begin{pmatrix} \text{Desvio da} \\ \text{independência} \end{pmatrix} + \begin{pmatrix} \text{Desvio} \\ \text{marginal} \end{pmatrix}.$$

e ainda

$$\Psi_{\mathsf{ICA}}(\mathbf{W}) = \max_{\widetilde{a}} \left(\Psi_{\mathsf{MV}}(\mathbf{W})\right).$$

Estratégias de separação cega de fontes - VI

Critério “universal”

$$J_{\mathsf{BSS}}(\mathbf{W}) = \ln\left[|\mathsf{det}|(\mathbf{W})\right] - \ln\left[\prod_{i=1}^{K}\phi_i(y_i)\right]$$

em que $\mathbf{y} = \mathbf{W}\mathbf{x}$ e, idealmente, as funções $\phi_i(\cdot)$ são as fdps das fontes.

Questão

Como encontrar ou estimar as fdps das fontes?

Resposta possível

Expansões polinomiais através dos cumulantes.

Critério geral

- Derivando J_{BSS} em relação a $\mathbf{W}$

$$\nabla J(\mathbf{W}) = \left(\mathbf{W}^T\right)^{-1} - \boldsymbol{f}\left(\mathbf{z}\right)\mathbf{x}^T$$

em que

$$\boldsymbol{f}(\mathbf{z}) = \begin{bmatrix} -\frac{\phi_1'(z_1)}{\phi_1(z_1)} & -\frac{\phi_2'(z_2)}{\phi_2(z_2)} & \cdots & -\frac{\phi_K'(z_K)}{\phi_K(z_K)} \end{bmatrix}^T.$$

e

$$\Delta\mathbf{W}(n) \propto \left(\mathbf{W}^T\right)^{-1} - \boldsymbol{f}\left(\mathbf{z}(n)\right)\mathbf{x}(n)^T.$$

- Se $\phi_i\left(z_i\right) = \frac{1}{1+\exp(-z_i)}$, o algoritmo torna-se *InfoMax/MaxEnt*

Algoritmo Jutten-Hérault - I

- Trabalho pioneiro em BSS: cancelamento das correlações cruzadas para obter sinais independentes
- Utilização de funções não-lineares
- Correlações do tipo $\mathbb{E}\left\{g_1(y_i)g_2(y_j)\right\}$, em que $g_1(\cdot)$ e $g_2(\cdot)$ são funções não-lineares adequadamente escolhidas
- Adaptação

$$\Delta \mathbf{W}_{ij} \propto g_1(y_i) \cdot g_2(y_j) \quad \forall i \neq j$$

$$\mathbf{y} = (\mathbf{I} + \mathbf{W})^{-1} \mathbf{x}$$

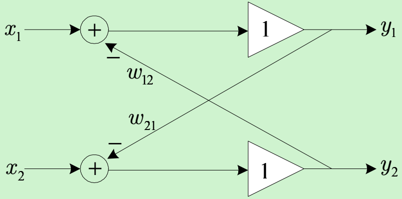
- Solução *ad hoc* - problemas de convergência

Algoritmo Jutten-Hérault - II

- Outra implementação

$$\Delta\mathbf{W} \propto \left(\mathbf{I} - g_1(\mathbf{y}) \cdot g_2\left(\mathbf{y}^T\right)\right) \cdot \mathbf{W},$$

em que $\mathbf{y} = \mathbf{W}\mathbf{x}$



Equivariância e gradiente relativo/natural - I

- Algoritmo Jutten-Hérault: necessidade de inversão de matriz a cada instante
- Condicionado ao tipo de matriz de mistura
- *Estimador equivariante*: $\widehat{\boldsymbol{E}}\left\langle \mathbf{W}\mathbf{x}\right\rangle = \mathbf{W}\widehat{\boldsymbol{E}}\left\langle \mathbf{x}\right\rangle$
- Logo

$$\begin{aligned}
\widehat{\mathbf{a}} &= \left(\widehat{\boldsymbol{E}}\langle\mathbf{x}\rangle\right)^{-1}\mathbf{x} \\
&= \left(\widehat{\boldsymbol{E}}\left\langle \mathbf{H}\mathbf{a}\right\rangle\right)^{-1}\mathbf{H}\mathbf{a} \\
&= \left(\mathbf{H}\widehat{\boldsymbol{E}}\left\langle \mathbf{a}\right\rangle\right)^{-1}\mathbf{H}\mathbf{a} \\
&= \left(\widehat{\boldsymbol{E}}\left\langle \mathbf{a}\right\rangle\right)^{-1}\mathbf{a}
\end{aligned}$$

Equivariância e gradiente relativo/natural - II

- Algoritmo EASI (*Equivariant Adaptive Separation via Independence*)

$$\Delta\mathbf{W}(n) \propto \left(\mathbf{I} - \boldsymbol{f}\left(\mathbf{z}(n)\right)\mathbf{z}^T(n)\right)\mathbf{W}(n)$$

- Na resposta combinada $\mathbf{G} = \mathbf{HW}$

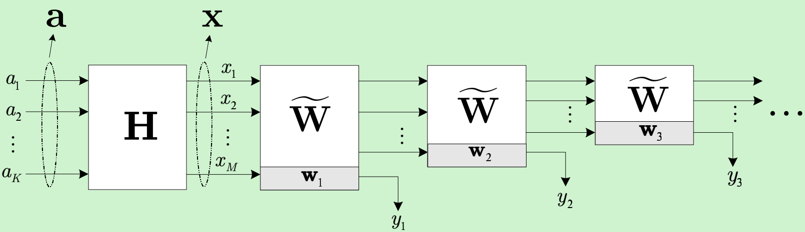
$$\Delta\mathbf{G}(n) \propto \left(\mathbf{I} - \boldsymbol{f}\left(\mathbf{z}(n)\right)\mathbf{z}^T(n)\right)\mathbf{G}(n)$$

- Gradiente relativo/natural

$$\widetilde{\nabla}\Lambda(\mathbf{W}) \triangleq \left.\frac{\partial\Lambda\left((\mathbf{I}+\boldsymbol{\Xi})\mathbf{W}\right)}{\partial\boldsymbol{\Xi}}\right|_{\boldsymbol{\Xi}=0}$$

Deflation

- Separação *serial* das fontes
- Interessante quando não se deseja todas as fontes da mistura
- Algum conhecimento é utilizado para identificar uma fonte e as outras são removidas por etapas



FastICA - I

- Tem por base a maximização da negentropia através de uma aproximação específica

$$\max_{\mathbf{G}^T\mathbf{G}=\mathbf{I}} \left(\Psi_{\text{FastICA}}(\mathbf{W}) \triangleq \left| \mathbb{E}\left\{ g(\mathbf{y}) - g\left(\mathbf{y}^{\mathbf{G}}\right) \right\} \right|^2 \right)$$

- Maximizar a divergência entre variável e sua versão gaussiana utilizando funções não-lineares que extraem informações das HOS
- Processamento com restrição

$$\begin{cases} \displaystyle\max_{\mathbf{w}_1,\cdots,\mathbf{w}_K} \sum_{i=1}^{K} \Psi_{\text{FastICA}}(\mathbf{w}_i) \\ \text{sujeito a} \qquad \mathbf{w}_i\mathbf{w}_j = \delta_{ij}, \end{cases}$$

FastICA - II

- Se cada vetor $\mathbf{w}_i$ possui norma unitária e $g\left(\mathbf{y}^{\mathbf{G}}\right)$ é constante então

$$\max_{\|\mathbf{g}_i\|=1} \mathbb{E}\left\{F\left(\mathbf{w}_i^T\mathbf{x}\right)\right\}$$

- A solução da equação é dada por $\mathbb{E}\left\{\mathbf{x}f\left(\mathbf{w}_i^T\mathbf{x}\right)\right\} - \beta\mathbf{w}_i = \mathbf{0}$
- Aplicação da aproximação de Newton no gradiente da solução fornece

$$\mathbf{w}_i^{\ddagger}(n+1) = \mathbb{E}\left\{\mathbf{x}f\left(\mathbf{w}_i^T(n)\mathbf{x}\right)\right\} - \mathbb{E}\left\{f'\left(\mathbf{w}_i^T(n)\mathbf{x}\right)\right\}\mathbf{w}_i(n)$$

$$\mathbf{w}_i(n+1) = \frac{\mathbf{w}_i^{\ddagger}(n+1)}{\left\|\mathbf{w}_i^{\ddagger}(n+1)\right\|}$$

FastICA

Nonlinear PCA - I

- Generalização de PCA através de uma projeção não-linear
- Função custo

$$\Psi_{\mathsf{nPCA}}(\mathbf{W}) = \mathbb{E}\left\{\|\mathbf{x} - \mathbf{W}^T\mathbf{g}(\mathbf{W}\mathbf{x})\|^2\right\}$$

- Para $\mathbf{W}$ ortogonal (branqueamento)

$$\begin{aligned}\Psi_{\mathsf{nPCA}}(\mathbf{W}) &= \mathbb{E}\left\{\|\mathbf{W}\mathbf{x} - \mathbf{g}(\mathbf{W}\mathbf{x})\|^2\right\} \\ &= \mathbb{E}\left\{\|\mathbf{y} - \mathbf{g}(\mathbf{y})\|^2\right\} \\ &= \sum_{i=1}^{K} \mathbb{E}\left\{|y_i - g_i(y_i)|^2\right\}\end{aligned}$$

- Forte ligação com critério de Bussgang (equalização cega)!

Nonlinear PCA - II

- Adaptação possível

$$\Delta \mathbf{w}_i \propto g\left(y_i\right)\mathbf{x} - g\left(y_i\right)\sum_{j=1}^{i} g\left(y_i\right)\mathbf{w}_j$$

- Simplificação: *algoritmo bigradiente*

$$\Delta \mathbf{W}(n) \propto \mu(n) g\left[\mathbf{W}(n)\mathbf{x}(n)\right]\mathbf{x}^T(n) + \alpha\left[\mathbf{I} - \mathbf{W}(n)\mathbf{W}^T(n)\right]\mathbf{W}(n)$$

- Ou de forma equivalente

$$\Delta \mathbf{W}(n) \propto \left[\mathbf{x}^T(n) - \mathbf{W}(n)\boldsymbol{f}(\mathbf{z}^T(n))\right]\boldsymbol{f}(\mathbf{z}(n))$$

Part II

Processamento multi-usuário

Características de processamento multi-usuário

- Múltiplos usuários dividindo recursos do canala
- Diversidade espacial no receptor é considerada para fornecer remoção de interferência (múltiplo acesso e inter-simbólica): geralmente é empregado um arranjo de antenas
- Modelo espacial ou espaço-temporal: depende da faixa considerada do canal comparada com a faixa dos sinais
- **Processamento multi-usuário**: separar *todos* os usuários de uma combinação dos sinais

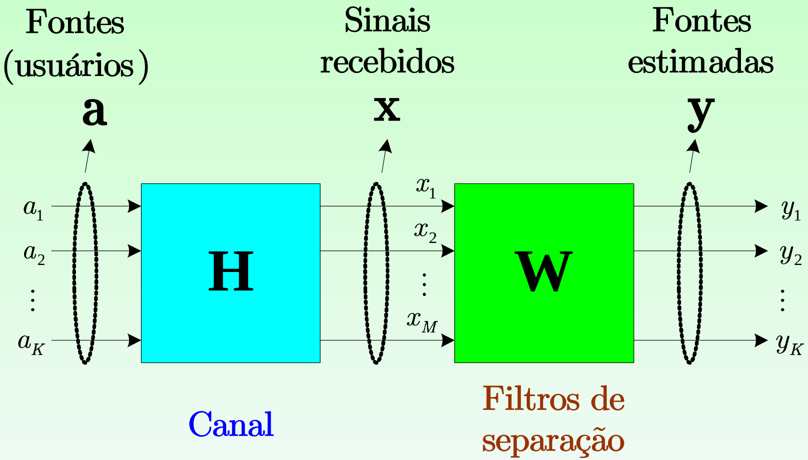
Caso particular de BSS

Características da fonte

- Fontes discretas
- Alfabeto finito
- Mesma distribuição de probabilidade

Processamento é possível com estruturas lineares!

Esquema de processamento multi-usuário: notação



Modelando o canal... I

- Resposta do canal entre o k-ésimo usuário e o elemento de referência do arranjo

$$h_{0,k}(t) = \sum_{i=0}^{L_k-1} \alpha_{k,i} \cdot \delta\left(t - \tau_{k,i}\right),$$

onde L_k é o número de multipercursos, $\alpha_{k,i}$ é uma v.a. que modela o desvanecimento e rotação de fase e $\tau_{k,i}$ são os atrasos relativos de cada multipercurso do k-ésimo usuário

- Resposta do canal entre k-ésimo usuário e m-ésimo elemento do arranjo

$$h_{m,k}(t) = f_m(\theta_k) \cdot \sum_{i=0}^{L_k-1} \alpha_{k,i} \cdot \delta\left(t - \tau_{k,i}\right),$$

Modelando o canal... II

- Vetor de resposta impulsiva

$$\mathbf{h}_{m,k}(t) = \sum_{i=0}^{L_k-1} \mathbf{f}\left(\theta_{k,i}\right) \cdot \alpha_{k,i} \cdot \delta\left(t - \tau_{k,i}\right),$$

Interferência de múltiplo acesso: modelo I

- Faixa do canal é mais estreita que a da portadora
- Todos os delays relativos são aproximadamente iguais: $\tau_{k,i} = \tau_k$

$$\mathbf{h}_{m,k}(t) = \delta\left(t - \tau_k\right) \sum_{i=0}^{L_k - 1} \mathbf{f}\left(\theta_{k,i}\right) \cdot \alpha_{k,i} = \delta\left(t - \tau_k\right) \cdot \mathbf{s}_k,$$

em que $\mathbf{s}_k$ é a *assinatura espacial* do k-ésimo usuário, dada por

$$\mathbf{s}_k = \sum_{i=0}^{L_k - 1} \mathbf{f}\left(\theta_{k,i}\right) \cdot \alpha_{k,i}.$$

Interferência de múltiplo acesso: modelo II

- No m-ésimo elemento do arranjo o sinal é dado por

$$x_m(t) = \sum_{k=1}^{K} \sqrt{P_k} a_k \left(t - \tau_k\right) \cdot \alpha_{k,m} \cdot f_m \left(\theta_k\right) + v(t),$$

  em que P_k é a potência transmitida do k-ésimo usuário e $v(t)$ é uma v.a. de um processo estocástico gaussiano

- Considerando
  1. τ_k múltiplos do período de símbolo (T)
  2. perfeito sincronismo

$$\boxed{\mathbf{x}(n) = \sum_{k=1}^{K} \sqrt{P_k} a_k(n) \cdot \alpha_{k,m} \cdot \mathbf{f}\left(\theta_k\right) + \mathbf{v}(n)}$$

Notação matricial

- Utilizando notação matricial

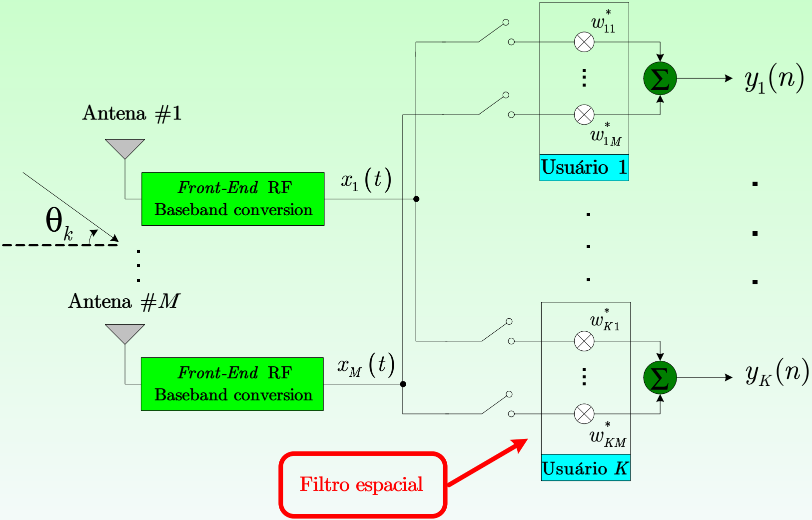
$$\mathbf{x}(n) = \mathbf{F}\mathbf{a}(n) + \mathbf{v}(n),$$

em que

$$\mathbf{x}(n) = \begin{bmatrix} x_1(n) \\ \vdots \\ x_M(n) \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{f}(\theta_1) | & \cdots & | \mathbf{f}(\theta_K) \end{bmatrix}_{M \times K} \quad \text{e}$$

$$\mathbf{a}(n) = \begin{bmatrix} a_1(n) \\ \vdots \\ a_K(n) \end{bmatrix}$$

Esquema de processamento espacial com arranjos de antenas



Modelo do receptor

- Saída do k-ésimo filtro

$$y_k(n) = \mathbf{w}_k^H(n)\mathbf{x}(n),$$

- Na forma matricial

$$\mathbf{y}(n) = \mathbf{W}^H(n)\mathbf{x}(n),$$

em que

$$\mathbf{y}\left(n\right) = \begin{bmatrix} y_1(n) \\ \vdots \\ y_K(n) \end{bmatrix} \quad \text{e} \quad \mathbf{W} = \begin{bmatrix} \mathbf{w}_1(n) | & \cdots & | \mathbf{w}_K(n) \end{bmatrix}_{M \times K}.$$

Interferências de múltiplo acesso e inter-simbólica: modelo I

- Faixa do canal é da ordem da portadora
- $\tau_{k,i} \neq \tau_k$
- Vetor de resposta impulsica do usuário k para o elemento m é

$$\mathbf{h}_{m,k}(t) = \sum_{i=0}^{L_k-1} \mathbf{f}\left(\theta_{k,i}\right) \cdot \alpha_{k,i} \cdot \delta\left(t - \tau_{k,i}\right)$$

Interferências de múltiplo acesso e inter-simbólica: modelo II

- Representação do sinal recebido

$$\begin{aligned}\mathbf{x}(t) &= \sum_{k=1}^{K} a_k(t) \star \mathbf{h}_k(t) + \mathbf{v}(t)\\ &= \sum_{k=1}^{K} \sum_{i=0}^{L_k-1} \mathbf{f}\left(\theta_{k,i}\right) \cdot \alpha_{k,i} \cdot a_k\left(t-\tau_k\right) + \mathbf{v}(t),\end{aligned}$$

- Modelo discreto

$$\mathbf{x}(n) = \sum_{k=1}^{K} \sum_{i=0}^{L_k-1} h_k\left(i\right) \cdot a_k\left(n-i\right) \cdot \mathbf{f}\left(\theta_{k,i}\right) + \mathbf{v}\left(n\right),$$

em que $h_k(i)$ representa o i-ésimo coeficiente da resposta impulsiva do canal do k-ésimo usuário, representando $\alpha_{k,i}$

Notação matricial

- Matriz do canal do k-ésimo usuário é

$$\mathbf{H}_k = \mathbf{F}\left(\theta_k\right) \cdot \boldsymbol{\alpha}_k,$$

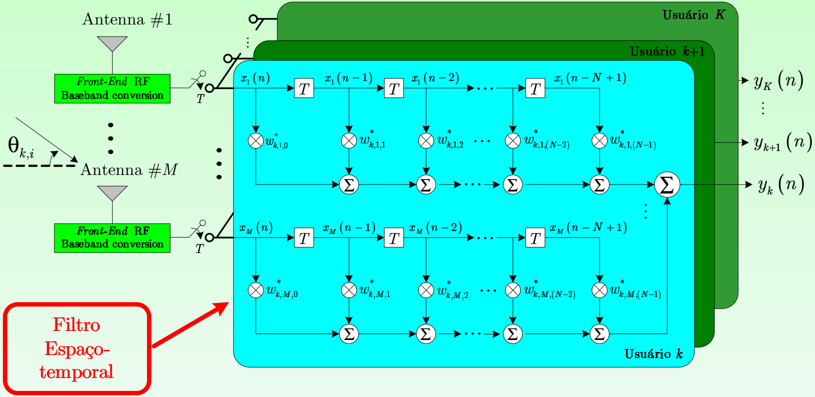
em que

$$\mathbf{F}\left(\theta_k\right) = \left[\begin{array}{c|c|c|c} \mathbf{f}\left(\theta_{k,0}\right) & \mathbf{f}\left(\theta_{k,1}\right) & \cdots & \mathbf{f}\left(\theta_{k,L_k-1}\right) \end{array}\right]_{M \times L}$$

e

$$\boldsymbol{\alpha}_k = \text{diag}\left(\left[\begin{array}{cccc} h_k\left(0\right) & h_k\left(1\right) & \cdots & h_k\left(L_k - 1\right) \end{array}\right]\right)$$

Estrutura de processamento espaço-temporal



Representação de sinais I

- Representação **tensorial**
- Notação mais compacta

$$\mathcal{X}(n) = \sum_{k=1}^{K} \boldsymbol{\mathcal{H}}_k \mathcal{A}_k(n) + \mathcal{V}(n)$$

em que

$$\mathcal{X}(n) = \begin{bmatrix} \mathbf{x}(n) \\ \vdots \\ \mathbf{x}(n-N+1) \end{bmatrix}_{MN\times 1} \qquad \mathbf{x}(n) = \begin{bmatrix} x_1(n) \\ \vdots \\ x_M(n) \end{bmatrix},$$

$$\mathcal{A}_k(n) = \begin{bmatrix} a_k\,(n) \\ \vdots \\ a_k\,(n-L_k-N+2) \end{bmatrix}_{(L_k+N-1)\times 1} \qquad \mathcal{V}(n) = \begin{bmatrix} \mathbf{v}(n) \\ \vdots \\ \mathbf{v}(n-N+1) \end{bmatrix}_{MN\times 1}$$

Matriz de convolução espaço-temporal

$$\boldsymbol{\mathcal{H}}_k = \left[\begin{array}{ccccc} \mathbf{H}_k(n) & & \mathbf{0}_{M\times(N-1)} & & \\ \mathbf{0}_{M\times 1} & & \mathbf{H}_k(n-1) & & \mathbf{0}_{M\times(N-2)} \\ \mathbf{0}_{M\times 1} & \mathbf{0}_{M\times 1} & \mathbf{H}_k(n-2) & & \mathbf{0}_{M\times(N-3)} \\ \vdots & \cdots & & \ddots & \\ & \mathbf{0}_{M\times(N-1)} & & & \mathbf{H}_k(n-N+1) \end{array}\right]_{MN\times(N+L-1)},$$

em que $\mathbf{0}_{k\times l}$ é uma matriz de dimensão$k \times l$ de zeros

- $M = 4$, $L = 3$ and $N = 3$

$$\boldsymbol{\mathcal{H}}_k = \begin{bmatrix} h_{k,1,0}(n) & h_{k,1,1}(n) & h_{k,1,2}(n) & 0 & 0 \\ h_{k,2,0}(n) & h_{k,2,1}(n) & h_{k,2,2}(n) & 0 & 0 \\ h_{k,3,0}(n) & h_{k,3,1}(n) & h_{k,3,2}(n) & 0 & 0 \\ h_{k,4,0}(n) & h_{k,4,1}(n) & h_{k,4,2}(n) & 0 & 0 \\ 0 & h_{k,1,0}(n-1) & h_{k,1,1}(n-1) & h_{k,1,2}(n-1) & 0 \\ 0 & h_{k,2,0}(n-1) & h_{k,2,1}(n-1) & h_{k,2,2}(n-1) & 0 \\ 0 & h_{k,3,0}(n-1) & h_{k,3,1}(n-1) & h_{k,3,2}(n-1) & 0 \\ 0 & h_{k,4,0}(n-1) & h_{k,4,1}(n-1) & h_{k,4,2}(n-1) & 0 \\ 0 & 0 & h_{k,1,0}(n-2) & h_{k,1,1}(n-2) & h_{k,1,2}(n-2) \\ 0 & 0 & h_{k,2,0}(n-2) & h_{k,2,1}(n-2) & h_{k,2,2}(n-2) \\ 0 & 0 & h_{k,3,0}(n-2) & h_{k,3,1}(n-2) & h_{k,3,2}(n-2) \\ 0 & 0 & h_{k,4,0}(n-2) & h_{k,4,1}(n-2) & h_{k,4,2}(n-2) \end{bmatrix}_{12\times 5}$$

Modelo do sinal do receptor

$$y_k(n) = \mathcal{W}_k^H(n)\mathcal{X}(n),$$

em que

$$\mathcal{W}_k = \begin{bmatrix} \mathbf{w}_{k0} \\ \mathbf{w}_{k1} \\ \vdots \\ \mathbf{w}_{k(N-1)} \end{bmatrix}_{MN\times 1} \quad , \quad \mathbf{w}_{k\widehat{n}} = \begin{bmatrix} w_{k,1,\widehat{n}} \\ w_{k,2,\widehat{n}} \\ \vdots \\ w_{k,M,\widehat{n}} \end{bmatrix}_{M\times 1}$$

Algoritmos cegos

Sinal desejado não disponível: métodos através de HOS

- Sinal desejado é aproximado por alguma medida estatística
- Complexidade de estimação de momentos de ordem superior
- Análise mais elaborada e difícil
- Diferentes abordagens
    1. Critério **Constant modulus** (CM) ou de Godard
    2. **Directed-decision** (DD)
    3. Critério de **Sato**
    4. Critério de **Shalvi-Weinstein**
    5. Algoritmo de **Benveniste-Goursat**
    6. Critérios baseados em **teoria da informação**

Estratégias de processamento multi-usuário cegas

- Estratégias lineares baseadas em técnicas do domínio temporal podem ser aplicadas
- Critérios adicionais são utilizados para recuperação de todas as fontes
- Caso contrário: a recuperação das fontes pode ser errônea devido ao efeito "near-far" ou desbalanceamento de potência

Multiuser Constant Modulus Algorithm (MU-CMA)

- Penaliza sinais fora de um raio de módulo constante: $J_{\mathsf{CMA}}\left(\mathbf{w}_k\right) = \mathbb{E}\left\{\left(\left|y(n)\right|^2 - \rho_2\right)^2\right\}$, em que $\rho_2 = \frac{\mathbb{E}\left\{a_k^4(n)\right\}}{\mathbb{E}\left\{a_k^2(n)\right\}}$
- Critério adicional: descorrelação explícita

$$\boxed{J_{\mathsf{MU\text{-}CMA}}\left(\mathbf{w}_k\right) = J_{\mathsf{CMA}}\left(\mathbf{w}_k\right) + \gamma \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \left|r_{ij}\right|^2}$$

$r_{ij} = \mathbb{E}\left\{y_i(n) y_j^*(n)\right\}$ e γ é um fator de regularização

MU-CMA: adaptação (caso espacial)

$$\boxed{\mathbf{w}_k(n+1) = \mathbf{w}_k(n) + \mu \cdot \left(1 - |y(n)|^2\right) \cdot y_k^* \cdot \mathbf{x}(n) - \gamma \sum_{\substack{i=1 \\ i \neq k}}^{K} \widehat{r}_{ik}(n) \widehat{\mathbf{p}}_i(n)}$$

$\widehat{r}_{ik}(n)$ é o (i,k)-ésimo elemento de $\widehat{\mathbf{R}}_y(n)$ e $\widehat{\mathbf{p}}_i(n)$ é a i-ésima coluna de $\widehat{\mathbf{P}}(n)$.

$$\widehat{\mathbf{R}}_y(n+1) = \varsigma \widehat{\mathbf{R}}_y(n) + (1-\varsigma)\mathbf{y}(n)\mathbf{y}^H(n)$$
$$\widehat{\mathbf{P}}(n+1) = \varsigma \widehat{\mathbf{P}}(n) + (1-\varsigma)\mathbf{x}(n)\mathbf{y}^H(n),$$

MU-CMA: adaptação (caso espaço-temporal)

$$\boxed{\mathcal{W}_k(n+1) = \mathcal{W}_k(n) + \mu \cdot \left(1 - |y_k(n)|^2\right) \cdot y(n)_k^* \cdot \mathcal{X}(n) - \gamma \sum_{\substack{i=1 \\ i \neq k}}^{K} \sum_{\ell=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \widehat{r}_{ik,\ell}(n) \widehat{\mathbf{p}}_{i,\ell}(n)}$$

em que

$$\mathbf{R}_{y,\ell}(n+1) = \varsigma \mathbf{R}_{y,\ell}(n) + (1-\varsigma)\mathbf{y}(n)\mathbf{y}^H(n-\ell)$$

$$\mathbf{P}_{\ell}(n+1) = \varsigma \mathbf{P}_{\ell}(n) + (1-\varsigma)\mathcal{X}(n)\mathbf{y}^H(n-\ell)$$

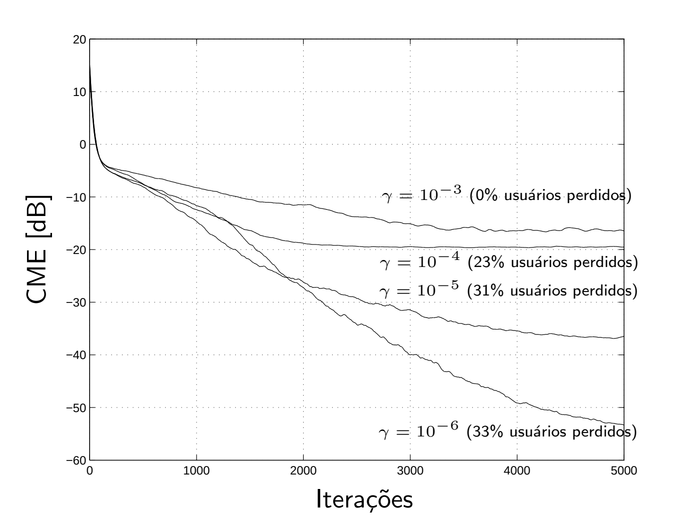
$$\mathbf{y}(n-\ell) = \begin{bmatrix} y_1(n-\ell) & \cdots & y_K(n-\ell) \end{bmatrix}^T$$

$$\ell = -\frac{\Delta}{2}, \ldots, \frac{\Delta}{2}$$

MU-CMA: características

1. Robustez e simplicidade
2. Funciona mesmo para sinais que não apresentam módulo constante
3. Convergência lenta
4. Problema de inicialização
5. Compromisso entre termo de regularização e desempenho em regime

Desempenho: usuários perdidos × erro do módulo constante (CME)



Fast Multiuser Constant Modulus (FMU-CMA)

Metas

- Melhorar a taxa de convergência por meio de um procedimento recursivo e um algoritmo tipo DMI utilizando o critério do módulo constante para cálculo de $\mathbf{R}_{yy}$ e $\mathbf{p}_y$
- Melhorar o desempenho de erro em regime sem perda de usuários utilizando um termo de regularização adaptativo que mede a correlação dos sinais de saída dos filtros e penaliza-os quando esta é alta

$$\mathbf{w}_k(n) = \mathbf{R}_{xy,k}^{-1}(n)\mathbf{d}_{xy,k}(n)$$

$$\mathbf{R}_{xy,k}(n+1) = \zeta\mathbf{R}_{xy,k}(n) + (1-\zeta)\left|y_k(n)\right|^2\mathbf{x}(n)\mathbf{x}^H(n)$$

$$\mathbf{d}_{xy,k}(k+1) = \zeta\mathbf{d}_{xy,k}(n) + (1-\zeta)\cdot\rho_2\cdot y_k^*(n)\mathbf{x}(n) - \gamma(n)\sum_{\substack{i=1\\ i\neq k}}^{K}\widehat{r}_{ik}(n)\widehat{\mathbf{p}}_i($$

$$\widehat{\mathbf{R}}_y(n+1) = \varsigma\widehat{\mathbf{R}}_y(n) + (1-\varsigma)\mathbf{y}(n)\mathbf{y}^H(n)$$

$$\widehat{\mathbf{P}}(n+1) = \varsigma\widehat{\mathbf{P}}(n) + (1-\varsigma)\mathbf{x}(n)\mathbf{y}^H(n),$$

FMU-CMA: termo de regularização adaptativo

$$\bar{r}_k(n) = \frac{1}{K-1} \sum_{\substack{i=1 \\ i \neq k}}^{K} |\widehat{r}_{ik}(n)|^2$$

$$\boxed{\gamma_k(n) = \tanh\left[\bar{r}_k(n)\right]}$$

FMU-CMA: características

Propriedades

- Maior taxa de convergência
- Menor erro em regime sem perda de usuários
- Maior complexidade computacional (inversão de matrizes)

Multiuser Kurtosis (MUK) Criterion

- Baseado na generalização das condições para recuperação cega de sinais
- Generalização do teorema de Shalvi-Weinstein
  1. $a_l(n)$ é i.i.d. e de média nula $(l = 1, \ \dots \ , K)$;
  2. $a_l(n)$ e $a_q(n)$ são estatisticamente independentes para $l \neq q$ e têm a mesma fdp;
  3. $|\mathcal{K}\left[y_l(n)\right]| = |\mathcal{K}_a| \quad (l = 1, \ \dots \ , K)$;
  4. $\mathbb{E}\left\{|y_l(n)|^2\right\} = \sigma_a^2 \quad (l = 1, \ \dots \ , K)$;
  5. $\mathbb{E}\left\{y_l(n)y_q^*(n)\right\} = 0, \quad l \neq q$,

  em que $\mathcal{K}_a$ é a kurtosis dos sinais transmitidos e $\mathcal{K}[\cdot]$ é o operador de kurtosis

MUK: descrição

- Maximizar a kurtosis da saída e igualar as potências da entrada e saída: remoção da interferência
- Necessidade de diferentes (descorrelação) fontes
- Evitar termo de regularização
- Alternativa: forçar a resposta global (canal combinado + filtros de separação) a ser uma matriz identidade
- **Problema com restrições!**

$$\boxed{\begin{cases} \max\limits_{\mathbf{G}} \quad J_{\mathsf{MUK}}(\mathbf{G}) = \sum_{k=1}^{K} |\mathcal{K}[y_k]| \\ \text{subject to : } \mathbf{G}^H\mathbf{G} = \mathbf{I} \end{cases}}$$

MUK: adaptação

Etapas

1. Remoção da interferência: $\mathbf{W}^e$

$$\nabla J_{\mathsf{MUK}}\left(\mathbf{G}\right) = 4 \cdot \operatorname{sign}\left(\mathcal{K}_a\right) \cdot \sum_{k=1}^{K} \mathbb{E}\left\{ \left|y_k(n)\right|^2 \cdot y_k(n) \cdot \mathbf{y}^*(n) \right\}$$

2. Ortogonalização: $\mathbf{W}$

$$\mathbf{w}_k(n+1) = \frac{\mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1}\left[\mathbf{w}_l^H(n+1)\mathbf{w}_k^e(n+1)\right]\mathbf{w}_l(n+1)}{\left\|\mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1}\left[\mathbf{w}_l^H(n+1)\mathbf{w}_k^e(n+1)\right]\mathbf{w}_l(n+1)\right\|}$$

Propriedades

- Ortogonalização de Gram-Schmidt
- Descarta termo de regularização
- Baixa complexidade (entretanto deve ter dados esferatizados para garantir separação!)
- Baixa taxa de convergência

Observação sobre as estratégias de BSS para processamento multi-usuário

Questão

Há um método cego com as desejáveis características de alta velocidade de convergência, erro em regime baixo e computacionalmente simples?

Proposta: aspectos gerais

- Proposição inicial para sistemas SISO (equalização mono-usuário)
- Caracterização da função de densidade de probabilidade (fdp) da saída do *equalizador ideal*
- Estimação paramétrica
- Critério: estimação/entropia
  - Divergência de Kullback-Leibler (KLD)
  - Função contraste

Matematicamente...

Sinal na saída do equalizador ideal

$$\begin{aligned} y(n) &= \left(\mathcal{H}^H \mathbf{a}(n) + \mathbf{v}(n)\right)^H \mathbf{w}_{\text{ideal}} \\ &= \mathbf{a}^H(n)\mathcal{H}\mathbf{w}_{\text{ideal}} + \mathbf{v}^H(n)\mathbf{w}_{\text{ideal}} \\ &= \mathbf{a}^T(n) \underbrace{\mathcal{H}\mathbf{w}_{\text{ideal}}}_{\mathbf{g}_{\text{ideal}}} + \mathbf{v}^H(n)\mathbf{w}_{\text{ideal}} \\ &= \mathbf{a}^H(n)\mathbf{g}_{\text{ideal}} + \vartheta(n) \\ &= a(n-\ell) + \vartheta(n), \end{aligned}$$

fdp de $y(n)$

$$p_Y(y) = \frac{1}{\sqrt{2\pi\sigma_\vartheta^2}} \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n) - \mathfrak{a}_i|^2}{2\sigma_\vartheta^2}\right) \Pr(\mathfrak{a}_i),$$

Minimizar a divergência

Modelo paramétrico

$$\Phi(y) = \underbrace{\frac{1}{\sqrt{2\pi\sigma_r^2}}}_{A} \cdot \frac{1}{\mathfrak{C}} \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n) - \mathfrak{a}_i|^2}{2\sigma_r^2}\right), \qquad (1)$$

Como medir a similaridade?

Divergência de Kullback-Leibler (KLD)

$$D\left(p_Y(y) || \Phi(y)\right) = \int_{-\infty}^{\infty} p_Y(y) \cdot \ln\left[\frac{p_Y(y)}{\Phi(y)}\right] dy$$

Fitting pdf Criterion

Termo da KLD dependente de $\Phi(y)$

$$
\begin{aligned}
J_{\mathsf{FP}}(\mathbf{w}) &= \int_{-\infty}^{\infty} p_Y(y) \cdot \ln\left(\frac{1}{\Phi(y)}\right) dy \\
&= -\mathbb{E}\left\{\ln\left[\Phi(y)\right]\right\} \\
&= -\mathbb{E}\left\{\ln\left[A \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y-\mathfrak{a}_i|^2}{2\sigma_r^2}\right)\right]\right\}
\end{aligned}
$$

Fitting pdf Algorithm

$$\nabla J_{\mathsf{FPC}}(\mathbf{w}) = \frac{\sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n)-\mathfrak{a}_i|^2}{2\sigma_r^2}\right)\left[y(n) - \mathfrak{a}_i^*\right]}{\sigma_r^2 \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n)-\mathfrak{a}_i|^2}{2\sigma_r^2}\right)} \mathbf{x}(n),$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla J_{\mathsf{FPC}}(\mathbf{w}).$$

Equivalências

- Solução do FPA quando se dispõe dos dados coincide com a solução MMSE
- Pode ser visto como um caso generalizado dos critérios CM e de Sato
- Equivalência do comportamento da função custo com algoritmo de decisão dirigida
- Relação importante com critério de minimização da probabilidade de erro (SBT2004)

FP $\times$ MAP

$$J_{\mathrm{MAP}} = J_{\mathrm{FPC}} - J_{\mathrm{MMSE}}$$

Opção de descorrelação: descorrelação explícita

- Critério adicional para penalizar a adaptação quando há réplicas do mesmo sinal

Critério MU-FPC

$$J_{\text{MU-FPC}}\left(\mathbf{w}_k\right) = J_{\text{FPC}}\left(\mathbf{w}_k\right) + \gamma \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \left|r_{ij}\right|^2$$

Algoritmo - Processamento espacial

$$\boxed{\begin{aligned}\mathbf{R}_{yy}(n+1) &= \xi\mathbf{R}_{yy}(n) + (1-\xi)\mathbf{y}(n)\mathbf{y}^H(n)\\ \mathbf{P}(n+1) &= \xi\mathbf{P}(n) + (1-\xi)\mathbf{x}(n)\mathbf{y}^H(n),\end{aligned}}$$

$$\boxed{\nabla J_{\mathsf{FPC}}(\mathbf{w}) = \frac{\sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n)-\mathfrak{a}_i|^2}{2\sigma_r^2}\right)[y(n) - \mathfrak{a}_i^*]}{\sigma_r^2 \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y(n)-\mathfrak{a}_i|^2}{2\sigma_r^2}\right)}\mathbf{x}(n)}$$

$$\boxed{\mathbf{w}_k(n+1) = \mathbf{w}_k(n) - \mu\nabla J_{\mathsf{FPC}}(\mathbf{w}_k) - \gamma\sum_{\substack{i=1\\ i\neq k}}^{K}\widehat{r}_{ik}(n)\mathbf{p}_i(n)}$$

Processamento espaço-temporal

$$J_{\text{MU-FPC}}(\mathcal{W}_k) = J_{\text{FPC}}(\mathcal{W}_k) + \gamma \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \sum_{\ell=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} |r_{ij}(\ell)|^2$$

$$\mathcal{W}_k(n+1) = \mathcal{W}_k(n) - \mu \nabla J_{\text{FPC}}(\mathcal{W}_k) - \gamma \sum_{\substack{i=1 \\ i \neq k}}^{K} \sum_{\ell=-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \widehat{r}_{ik,\ell}(n) \widehat{\mathbf{p}}_{i,\ell}(n)$$

$$\begin{aligned}
\mathbf{R}_{y,\ell}(n+1) &= \xi \mathbf{R}_{y,\ell}(n) + (1-\xi)\mathbf{y}(n)\mathbf{y}^H(n-\ell) \\
\mathbf{P}_\ell(n+1) &= \xi \mathbf{P}_\ell(n) + (1-\xi)\mathcal{X}(n)\mathbf{y}^H(n-\ell) \\
\mathbf{y}(n-\ell) &= \begin{bmatrix} y_1(n-\ell) & \cdots & y_K(n-\ell) \end{bmatrix}^T \\
\ell &= -\frac{\Delta}{2}, \ldots, \frac{\Delta}{2}
\end{aligned}$$

Características MU-FPA

- Recuperação da fase do sinal transmitido
- Baixa complexidade: algoritmo tipo LMS
- Grau de liberdade: σ_r^2
- Limitações:
  - Dependência com o fator γ
  - *Trade-off* – erro em regime $\times$ usuários perdidos

Condições - Teorema de Shalvi-Weinstein

C1. $a_l(n)$ é i.i.d. e de média zero $(l = 1, \ \ldots \ , K)$;

C2. $a_l(n)$ e $a_q(n)$ são estatisticamente independentes para $l \neq q$ e têm a mesma fdp;

C3. $|\mathcal{K}\left[y_l(n)\right]| = |\mathcal{K}_a| \quad (l = 1, \ \ldots \ , K)$;

C4. $\mathbb{E}\left\{|y_l(n)|^2\right\} = \sigma_a^2 \quad (l = 1, \ \ldots \ , K)$;

C5. $\mathbb{E}\left\{y_l(n)y_q^*(n)\right\} = 0, \quad l \neq q$.

Critério com restrição - II

- Condições necessárias para recuperação
- Maximização da *kurtosis* – MUK
- Garantia da obtenção de solução
  - Ortogonalização da matriz de resposta global

Critério MU-CFP

$$\begin{cases} \min\limits_{\mathbf{W}} J_{\mathsf{FP}}(\mathbf{W}) = \sum\limits_{i=1}^{K} D\left(p_Y(y_i) \,\|\, \Phi(y_i)\right) \\ \text{sujeito a: } \mathbf{G}^H\mathbf{G} = \mathbf{I} \end{cases}$$

MU-CFPA

$$\boxed{\mathbf{w}_k^e(n+1) = \mathbf{w}_k(n) - \mu \left(\frac{\sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y_k(n)-\mathfrak{a}_i|^2}{2\sigma_r^2}\right)[y_k(n) - \mathfrak{a}_i^*]}{\sigma_r^2 \cdot \sum_{i=1}^{\mathfrak{C}} \exp\left(-\frac{|y_k(n)-\mathfrak{a}_i|^2}{2\sigma_r^2}\right)} \mathbf{x}(n) \right)}$$

$$\mathbf{w}_1(n+1) = \frac{\mathbf{w}_1^e(n+1)}{\|\mathbf{w}_1^e(n+1)\|}$$

$$\mathbf{w}_k(n+1) = \frac{\mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1} \left[\mathbf{w}_l^H(n+1)\mathbf{w}_k^e(n+1)\right] \mathbf{w}_l(n+1)}{\left\| \mathbf{w}_k^e(n+1) - \sum_{l=1}^{k-1} \left[\mathbf{w}_l^H(n+1)\mathbf{w}_k^e(n+1)\right] \mathbf{w}_l(n+1) \right\|},$$

- Comparar desempenho de dois algoritmos com a mesma estrutura
- Diferença: número de estatísticas de ordem superior (HOS) consideradas

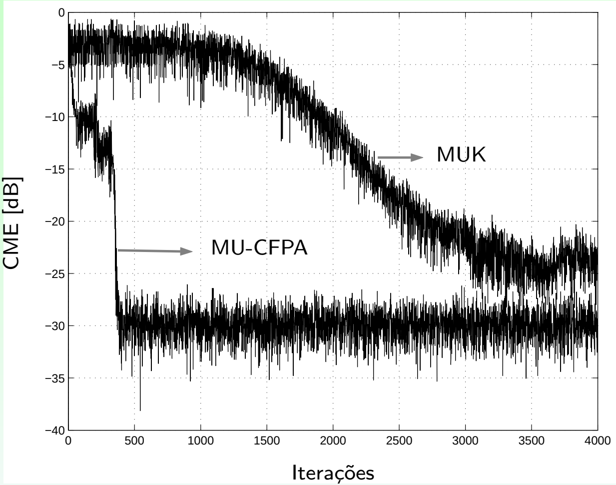
MUK

$$\begin{cases} \max\limits_{\mathbf{G}} \quad J_{\mathsf{MUK}}(\mathbf{G}) = \sum_{k=1}^{K} |\mathcal{K}[y_k]| \\ \text{sujeito a : } \mathbf{G}^H\mathbf{G} = \mathbf{I} \end{cases}$$

MU-CFPA

$$\begin{cases} \min\limits_{\mathbf{W}} J_{\mathsf{FP}}(\mathbf{W}) = \sum_{i=1}^{K} D\left(p(y_i) \,\|\, \Phi(y_i)\right) \\ \text{sujeito a: } \mathbf{G}^H\mathbf{G} = \mathbf{I} \end{cases}$$

MUK × MU-CFPA



CME [dB]
Iterações
MUK
MU-CFPA
0
−5
−10
−15
−20
−25
−30
−35
−40
0
500
1000
1500
2000
2500
3000
3500
4000

Impacto das HOS em BSS

Quais HOS são importantes para BSS?

Pelo menos uma. (Teoremas BGR e SW)

Há algum ganho em usar-se mais?

Sim, em versões adaptativas!

Como verificar?

Estimação da fdp do sinal desejado – Critério “universal” de BSS.

Aproximações da fdp: expansões em série

- Aproximação da fdp em séries ortonormais
- Expansões em torno de distribuições conhecidas
- Coeficientes da série dados pelas estatísticas do sinal

Aproximação em série

$$p_Y(y) = p_0(y) \sum_{k=1}^{\infty} C_k \cdot \mathfrak{b}_k(y)$$

Expansão em série (cont.)

Coeficientes - cumulantes

$$
\begin{aligned}
C_1 &= 0 \\
C_2 &= 0 \\
C_3 &= \frac{c_3}{6} \\
C_4 &= \frac{c_4}{24} \\
C_5 &= \frac{c_5}{120} \\
C_6 &= \frac{1}{720}\left(c_6 + 10c_3^2\right) \\
C_7 &= \frac{1}{5040}\left(c_7 + 35c_4c_3\right) \\
C_8 &= \frac{1}{40320}\left(c_8 + 56c_5c_3 + 35c_4^2\right)
\end{aligned}
$$

Coeficientes - momentos

$$
\begin{aligned}
C_1 &= \kappa_1 \\
C_2 &= \frac{1}{2}\left(\kappa_2 - 1\right) \\
C_3 &= \frac{1}{6}\left(\kappa_3 - 3\kappa_2\right) \\
C_4 &= \frac{1}{24}\left(\kappa_4 - 6\kappa_2 + 3\right) \\
C_5 &= \frac{1}{120}\left(\kappa_5 - 10\kappa_3 + 15\kappa_1\right) \\
C_6 &= \frac{1}{720}\left(\kappa_6 - 15\kappa_4 + 45\kappa_2 - 15\right) \\
C_7 &= \frac{1}{5040}\left(\kappa_7 - 21\kappa_5 + 105\kappa_3 - 105\kappa_1\right) \\
C_8 &= \frac{1}{40320}\left(\kappa_8 - 28\kappa_6 + 210\kappa_4 - 420\kappa_2 + 105\right)
\end{aligned}
$$

Expansão de Gram-Charlier

Densidade de referência

$$p_0(y) = p_G(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)$$

Base ortonormal – Polinômios de Hermite

$$\mathfrak{h}_k(y) = (-1)^k \cdot \frac{1}{p_0(y)} \cdot \frac{d^k p_0(y)}{dy^k},$$

Gram-Charlier

$$p_Y(y) = p_G(y)\left(1 + \sum_{k=3}^{\infty} C_k \cdot \mathfrak{h}_i(y)\right)$$

Expansão de Edgeworth

- Basicamente consiste na aproximação de Gram-Charlier ordenada em ordem decrescente dos coeficientes da série

$$p_y(\mathbf{y}) = p_G(\mathbf{y}) \Big[1 + \frac{1}{3!} c_3(\mathbf{y}) \mathfrak{h}_3(\mathbf{y}) + \frac{1}{4!} c_4(\mathbf{y}) \mathfrak{h}_4(\mathbf{y}) + \frac{10}{6!} c_3^2(\mathbf{y}) \mathfrak{h}_6(\mathbf{y}) + \frac{1}{5!} c_5(\mathbf{y}) \mathfrak{h}_5(\mathbf{y}) + \frac{35}{7!} c_3(\mathbf{y}) c_4(\mathbf{y}) \mathfrak{h}_7(\mathbf{y}) + \cdots \Big]$$

Expansões existentes - Gram-Charlier e Edgeworth

- Não aproximam todos os tipos de fpds
- Somente aquelas que são "próximas" de uma forma gaussiana (concentradas em torno da média)
- Sinais de comunicação digital não são contemplados
- Necessidade de uma expansão que permita avaliar sinais de modulações digitais

Expansão em torno de misturas gaussianas

- Necessidade de expansão em torno de distribuições típicas em comunicações digitais
- Somatório de gaussianas (estimador de Parzen)

Distribuição de referência

$$p_{SG}(y) = \frac{1}{\mathfrak{C}} \sum_{i=1}^{\mathfrak{C}} \frac{1}{\sqrt{2\pi\sigma_{\vartheta}^2}} \exp\left[-\frac{(y-\mathfrak{a}_i)^2}{2\sigma_{\vartheta}^2}\right],$$

Recursão da base ortonormal

$$\mathfrak{c}_{k+1}(y) = \mathfrak{C} \cdot y \cdot \mathfrak{c}_k(y) - k\mathfrak{c}_{k-1}(y),$$

Base ortonormal

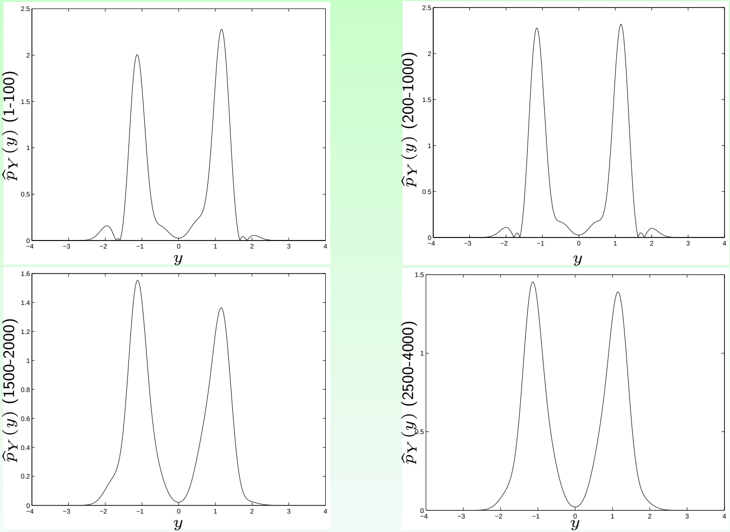
$$\mathfrak{c}_0(y) = 1$$
$$\mathfrak{c}_1(y) = \mathfrak{C} \cdot y$$
$$\mathfrak{c}_2(y) = (\mathfrak{C} \cdot y)^2 - 1$$
$$\mathfrak{c}_3(y) = (\mathfrak{C} \cdot y)^3 - 3(\mathfrak{C} \cdot y)$$
$$\mathfrak{c}_4(y) = (\mathfrak{C} \cdot y)^4 - 6((\mathfrak{C} \cdot y)^2 + 3$$
$$\mathfrak{c}_5(y) = (\mathfrak{C} \cdot y)^5 - 10(\mathfrak{C} \cdot y)^3 + 15(\mathfrak{C} \cdot y)$$
$$\mathfrak{c}_6(y) = (\mathfrak{C} \cdot y)^6 - 15(\mathfrak{C} \cdot y)^4 + 45(\mathfrak{C} \cdot y)^2 - 15$$
$$\mathfrak{c}_7(y) = (\mathfrak{C} \cdot y)^7 - 21(\mathfrak{C} \cdot y)^5 + 105(\mathfrak{C} \cdot y)^3 - 105(\mathfrak{C} \cdot y)$$
$$\mathfrak{c}_8(y) = (\mathfrak{C} \cdot y)^8 - 28(\mathfrak{C} \cdot y)^6 + 210(\mathfrak{C} \cdot y)^4 - 420(\mathfrak{C} \cdot y)^2 + 105$$

Algoritmos adaptativos e HOS

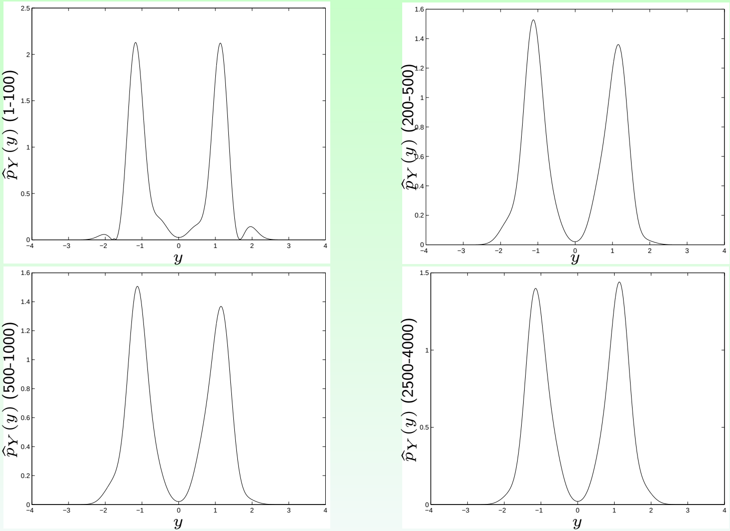
Destaques

- Estimativa da fdp a partir dos dados disponíveis
- HOS envolvidas no critério compõem a expansão da fdp
- Dependência dos cumulantes (momentos) e da base ortonormal
- Mais HOS aproveitam a "amplificação" fornecida pela base ortonormal
- Maior convergência

Dinâmica da estimativa da fdp - Algoritmo MUK



Dinâmica da estimativa da fdp - Algoritmo MU-CFPA



HOS: observações

- Possibilidade de análise da importância das HOS na dinâmica de algoritmos de recuperação cega de sinais
- Nova expansão da densidade de probabilidade permite aproximar densidades de sinais típicos em sistemas de comunicação digital
- Possibilidade de avaliar compromisso velocidade-desempenho em algoritmos baseados em HOS

Part III

Casos particulares

Características

- Redução do número de fontes/sensores: $K = M = 1$
- Canal dispersivo: multipercursos (amplitude e fase)
- Filtros para correção dos efeitos do canal
- Diversidade temporal no receptor
- Medidas estatísticas para estimativa do sinal desejado

Equalização cega

Critérios

- Teorema de Benveniste-Goursat-Rouget: igualar as fdps é necessário para equalização cega
- Teorema de Shalvi-Weinstein: igualar dois momentos estatísticos (potência e o de quarta ordem)
- Técnicas de Bussgang
  - *Decision-directed*
  - Sato
  - *Constant modulus* ou Godard
  - Benveniste-Goursat
  - *Stop-and-Go*
- Técnicas baseadas em teoria da informação

Equalização cega

Algumas equivalências "conhecidas"

1. Shalvi-Weinstein e *Constant Modulus*
2. *Constant Modulus* e solução de Wiener
3. *Fitting PDF* $\rightarrow$ Wiener, Sato, CMA

Equivalências com critérios de BSS

- *FastICA* e Algoritmo super-exponencial
- *InfoMax*/*MaxEnt* e Benveniste-Goursat-Rouget (BGR)
- Funções de Contraste e Shalvi-Weinstein

Equivalência *InfoMax* e critério BGR

- Maximizar a informação mútua entre a entrada e saída de um sistema de separação
- BGR: igualar a densidade de probabilidade da entrada e saída
- Informação mútua:

$$\begin{aligned} I(\mathbf{y}, \mathbf{x}) &= \mathcal{H}(\mathbf{y}) \\ &= \mathcal{H}(\mathbf{W}\mathbf{x}) \\ &= \mathcal{H}(\mathbf{W}\mathbf{H}\mathbf{a}) \end{aligned}$$

- Máximo da entropia/informação mútua ocorre quando $p_a(\mathbf{a}) = p_y(\mathbf{y})$

Equivalência *FastICA* e SEA

- *FastICA*

$$\mathbf{w}_i^{\ddagger}(n+1) = \mathbb{E}\left\{\mathbf{x} f\left(\mathbf{w}_i^T(n)\mathbf{x}\right)\right\} - \mathbb{E}\left\{f'\left(\mathbf{w}_i^T(n)\mathbf{x}\right)\right\}\mathbf{w}_i(n)$$

$$\mathbf{w}_i(n+1) = \frac{\mathbf{w}_i^{\ddagger}(n+1)}{\left\|\mathbf{w}_i^{\ddagger}(n+1)\right\|}$$

- Escolhendo fator de passo $\mu(n) = \frac{1}{\beta(n)}$

$$\mathbf{w}_i^{\ddagger}(n+1) = \mathbb{E}\left\{\mathbf{x} f\left(\mathbf{w}_i^T(n)\mathbf{x}\right)\right\}$$

$$\mathbf{w}_i(n+1) = \frac{\mathbf{w}_i^{\ddagger}(n+1)}{\left\|\mathbf{w}_i^{\ddagger}(n+1)\right\|}$$

- Sendo $F(y) = y^{2p}$, e daí $f(y) = 2py^{2p-1}$: **SEA**!

Equivalência função de contraste e SW

- SW: estatísticas de ordem superior
- Função contraste: medida de não-gaussianidade sujeito a respeitar propriedades
- Função contraste pode ser definida como função da kurtosis e do momento de segunda ordem

Características

- Caso específico: $K = 1$ e M sensores
- Subcanais são gerados por diversidade (temporal ou espacial)
- Modelo análogo ao de BSS: simplificação pois todas as variáveis são do mesmo sinal
- Redução de complexidade

Critérios

- Cicloestacionaridade: construção de técnicas que usam tal fato
- Diversidade espacial: múltiplas cópias do sinal (modelos equivalentes)
- Técnicas de BSS se aplicam: $M \geq K$
- Possibilidade de mais técnicas adaptativas
- Subespaços (PCA), predição linear (função contraste)!

Part IV

Tendências e perspectivas

Separação de fontes

- Misturas convolutivas: maioria das técnicas derivada para misturas instantâneas
- Misturas não-lineares: maior generalização de problemas passíveis de serem abordados
- Novas estratégias: simples, rápidas e eficientes
- Independência em misturas não-lineares não é preservada: como fazer?
- Não-estacionariedade
- Assincronismo
- Problemas sub-determinados $(M \leq K)$
- álgebra multi-linear
- Geometria diferencial
- ...

Processamento multi-usuário

- Critério e algoritmos simples: aplicações em tempo real
- Misturas convolutivas necessitam de estratégias mais elaboradas
- Modelagem do sistema (canal, usuários e interferência) apresenta ganhos quando modelada com tensores
- Critérios com melhor relação de compromisso velocidade/desempenho podem ser derivados
- Assincronismo
- ...

Equalização cega mono e multi-canal

- Equalizar é encontrar amostras temporais independentes
- Novos critérios baseados em ICA
- Possibilidade de equalização cega com redes neurais
- Critérios mais elaborados
- Novas estratégias de aprendizado
- ...

Part V

Sugestões bibliográficas

- A. Hyvärinen, E. Oja, and J. Karhunen, *Independent Component Analysis*. John Wiley & Sons, 2001.
- A. Cichocki and S. ichi Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, 2002.
- S. Haykin, Ed., *Unsupervised Adaptive Filtering*, ser. (Series on Adaptive and Learning Systems for Signal Processing, Communications and Control). John Wiley & Sons, 2000, vol. I: Source Separation.

Livros e teses - II

- S. Haykin, Ed., *Unsupervised Adaptive Filtering*, ser. (Series on Adaptive and Learning Systems for Signal Processing, Communications and Control). John Wiley & Sons, 2000, vol. II: Blind Deconvolution.
- C. C. Cavalcante, "Sobre Separação Cega de Fontes: Proposições e Análise de Estratégias para Processamento Multi-Usuário," Tese de Doutorado, Universidade Estadual de Campinas (UNICAMP) – DECOM, Campinas, SP - Brasil, Abril 2004.

- A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.
- A. Hyvärinen, "Survey on Independent Component Analysis," *Neural Computing Surveys*, vol. 2, pp. 94–128, 1999.
- P. Comon, "Independent Component Analysis: A New Concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, April 1994.
- J.-F. Cardoso, "Blind Signal Separation: Statistical Principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, October 1998.