Extrakce nezávislé komponenty

Zbyněk Koldovský

Acoustic Signal Analysis and Processing Group,
Faculty of Mechatronics, Informatics, and Interdisciplinary Studies,
Technical University in Liberec,
`https://asap.ite.tul.cz`
`zbynek.koldovsky@tul.cz`

Popis problematiky

- Cílem slepé separace signálů (BSS) je rozklad pozorované směsi signálů na její jednotlivé složky (komponenty) bez jejich znalosti. Známe tedy pouze pozorovanou směs, kterou jsme naměřili několika senzory.

- Příklad pro představu: Nahráváme několika mikrofony několik současně hovořících osob. Cílem BSS je separovat promluvy jednotlivých řečníků.

- Oblíbenou metodikou je Analýza nezávislých komponent (ICA), která separaci provádí na základě předpokladu, že původní signály (promluvy řečníků) jsou statisticky nezávislé procesy.

Nová tématika

- V mnoha praktických úlohách nám stačí separace (extrakce) pouze jediné komponenty, která nás zajímá. ICA provádí separaci všech komponent, což je v tomto případě zbytečné.
- Chceme se proto zabývat metodami, které separují pouze jednu komponentu (Extrakce nezávislé komponenty - Independent Component Extraction - ICE) a jejich aplikacemi např. ve zpracování zvukových nebo biologických signálů.

Nová tématika

- Zaměřit se chceme především na vývoj algoritmů, které extrakci provádí v reálném čase (on-line) a jsou schopné přizpůsobovat se změnám (pohyb zdroje v prostoru). Zde je nejdůležitějším problémem udržení konvergence algoritmů, tedy aby extrahovaly stále ten zdroj, který chceme.

- Problematika zahrnuje i teoretické a teoreticko-praktické úlohy jako je výpočet Rao-Cramérovým mezí pro maximální dosažitelnou přesnost ICE, stabilita algoritmů - zvětšení spádové oblasti konvergence ke správnému signálu, modifikace metod pro využití dodatečné informace (např. detekce aktivity zdroje cílového signálu) atd.

Epilog

- Zájemci o téma budou moci pracovat jako členové týmu A.S.A.P. Snažíme se spolupracovat a hledat řešení úloh společně.

- V případě dobrých výsledků je možné zapojit studentku/studenta do projektu.

- Tato prezentace pokračuje ukázkou jiné prezentace a to z mezinárodní konference, kde byl představen nový algoritmus ICE. Pro začátečníka jsou tyto informace nedostatečné a vytržené z kontextu. Prezentace je zde uvedena pouze pro doplnění představy, o jaký typ problematiky se jedná.

Part II

Prezentace z konference EUSIPCO 2017 v Řecku - Orthogonally Constrained Independent Component Extraction: Blind MPDR Beamforming

Motivation

- ICA/IVA are computationally expensive when

$$\#\ \text{microphones} \gg 2.$$

  (None of CHiME-4 solutions apply ICA or IVA although up to *six* microphones are available.)

- In many applications, only one signal of interest should be extracted. ICA/IVA aim to separate the same number of signals as that of microphones.

- Why not to estimate only one independent signal of interest?

Motivation

- ICA/IVA are computationally expensive when

$$\#\text{ microphones} \gg 2.$$

  (None of CHiME-4 solutions apply ICA or IVA although up to *six* microphones are available.)

- In many applications, only one signal of interest should be extracted. ICA/IVA aim to separate the same number of signals as that of microphones.

- Why not to estimate only one independent signal of interest?

Motivation

- ICA/IVA are computationally expensive when

$$\# \text{ microphones} \gg 2.$$

  (None of CHiME-4 solutions apply ICA or IVA although up to *six* microphones are available.)

- In many applications, only one signal of interest should be extracted. ICA/IVA aim to separate the same number of signals as that of microphones.

- Why not to estimate only one independent signal of interest?

Blind Signal Extraction methods

- The idea to extract only one "interesting" signal is not new: Projection Pursuit, nonlinear PCA, Maximum Kurtosis, ... Maximum Non-Gaussianity

- These methods extract the signal based on its "interestingness". They do not reflect its independence from the other (background) signals nor the properties of the background.

- We propose a new mixing model tailored for the signal extraction and apply the classical estimation theory: This is *Independent Component Extraction - ICE*

Blind Signal Extraction methods

- The idea to extract only one "interesting" signal is not new: Projection Pursuit, nonlinear PCA, Maximum Kurtosis, ... Maximum Non-Gaussianity
- These methods extract the signal based on its "interestingness". They do not reflect its independence from the other (background) signals nor the properties of the background.
- We propose a new mixing model tailored for the signal extraction and apply the classical estimation theory: This is *Independent Component Extraction - ICE*

Blind Signal Extraction methods

- The idea to extract only one "interesting" signal is not new: Projection Pursuit, nonlinear PCA, Maximum Kurtosis, . . . Maximum Non-Gaussianity
- These methods extract the signal based on its "interestingness". They do not reflect its independence from the other (background) signals nor the properties of the background.
- We propose a new mixing model tailored for the signal extraction and apply the classical estimation theory: This is *Independent Component Extraction - ICE*

Problem Formulation

- The problem we need to study (in the complex-valued domain) is

$$\mathbf{x} = \mathbf{a}s + \mathbf{y}$$

- ICA mixing model assumes

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \mathbf{a}s + \underbrace{\mathbf{A}_2\mathbf{u}}_{\mathbf{y}},$$

where $\mathbf{A} = \underbrace{[\mathbf{a} \quad \mathbf{A}_2]}_{\text{square}}$ and $\mathbf{s} = [s \quad \mathbf{u}]^T$.

- In ICE, it is *not* the goal to find $\mathbf{u}$, that is, the independent components of $\mathbf{y}$.

Problem Formulation

- The problem we need to study (in the complex-valued domain) is

$$\mathbf{x} = \mathbf{a}s + \mathbf{y}$$

- ICA mixing model assumes

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \mathbf{a}s + \underbrace{\mathbf{A}_2\mathbf{u}}_{\mathbf{y}},$$

where $\mathbf{A} = \underbrace{[\mathbf{a} \quad \mathbf{A}_2]}_{\text{square}}$ and $\mathbf{s} = [s \quad \mathbf{u}]^T$.

- In ICE, it is *not* the goal to find $\mathbf{u}$, that is, the independent components of $\mathbf{y}$.

Problem Formulation

- The problem we need to study (in the complex-valued domain) is

$$\mathbf{x} = \mathbf{a}s + \mathbf{y}$$

- ICA mixing model assumes

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \mathbf{a}s + \underbrace{\mathbf{A}_2\mathbf{u}}_{\mathbf{y}},$$

where $\mathbf{A} = \underbrace{[\mathbf{a} \quad \mathbf{A}_2]}_{\text{square}}$ and $\mathbf{s} = [s \quad \mathbf{u}]^T$.

- In ICE, it is *not* the goal to find $\mathbf{u}$, that is, the independent components of $\mathbf{y}$.

ICE Mixing and De-mixing Algebraic Model

Let $\mathbf{x}$ be written as

$$\mathbf{x} = \mathbf{a}s + \underbrace{\mathbf{y}}_{\mathbf{Qz}} = \underbrace{\begin{pmatrix}\mathbf{a} & \mathbf{Q}\end{pmatrix}}_{\mathbf{A}_{\mathrm{ICE}}} \begin{pmatrix} s \\ \mathbf{z} \end{pmatrix}.$$

How can we define $\mathbf{Q}$ and $\mathbf{z}$? Let

$$\mathbf{A}_{\mathrm{ICE}}^{-1} = \mathbf{W}_{\mathrm{ICE}}, \qquad \mathbf{W}_{\mathrm{ICE}} = \begin{pmatrix} \mathbf{w}^H \\ \mathbf{B} \end{pmatrix},$$

where

$$\mathbf{B}\mathbf{a} = \mathbf{0}.$$

$\mathbf{a}$	...	`steering vector`
$\mathbf{B}$	...	`blocking matrix`
$\mathbf{w}$	...	`separating vector`

ICE Mixing and De-mixing Algebraic Model

Defining

$$\mathbf{a} = \begin{pmatrix} \gamma \\ \mathbf{g} \end{pmatrix} \qquad \mathbf{w} = \begin{pmatrix} \beta \\ \mathbf{h} \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} \mathbf{g} & -\gamma \mathbf{I}_{d-1} \end{pmatrix},$$

we obtain

$$\mathbf{A}_{\mathrm{ICE}} = \begin{pmatrix} \mathbf{a} & \mathbf{Q} \end{pmatrix} = \begin{pmatrix} \gamma & \mathbf{h}^H \\ \mathbf{g} & \frac{1}{\gamma}\left(\mathbf{g}\mathbf{h}^H - \mathbf{I}_{d-1}\right) \end{pmatrix},$$

$$\mathbf{W}_{\mathrm{ICE}} = \begin{pmatrix} \mathbf{w}^H \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \overline{\beta} & \mathbf{h}^H \\ \mathbf{g} & -\gamma \mathbf{I}_{d-1} \end{pmatrix},$$

where

$$\overline{\beta}\gamma = 1 - \mathbf{h}^H\mathbf{g} \quad \Longleftrightarrow \quad \mathbf{w}^H\mathbf{a} = 1 \qquad \text{distrotionless const.}$$

It holds that $\mathbf{z} = \mathbf{B}\mathbf{x}$, and $\mathbf{y} = \mathbf{Q}\mathbf{z}$.

Statistical Model

Let s and $\mathbf{z}$ be independent, and

$$\begin{array}{lll} s & \text{be} & \text{non-Gaussian with pdf } p_s(\cdot) \\ \mathbf{z} & \text{be} & \text{circular Gaussian drawn from } \mathcal{CN}(\mathbf{0}, \mathbf{C}_\mathbf{z}) \end{array}$$

Hence, the log-likelihood function for N i.i.d. samples reads

$$\begin{aligned} \frac{1}{N}\mathcal{L}(\mathbf{a}, \mathbf{w}) &= \frac{1}{N} \log \prod_{n=1}^{N} p_\mathbf{x}(\mathbf{a}, \mathbf{w} | \mathbf{x}(n)) \\ &= \frac{1}{N} \sum_{n=1}^{N} \log p_s(\mathbf{w}^H \mathbf{x}(n)) + \frac{1}{N} \sum_{n=1}^{N} \log p_\mathbf{z}(\mathbf{B}\mathbf{x}(n)) + |\det \mathbf{W}_{\mathrm{ICE}}|^2 \end{aligned}$$

$\mathbf{C}_\mathbf{z}$ is a nuisance parameter.

Orthogonal Constraint

- The link between the steering vector **a** and the separating vector **w** is too "weak" $\Longrightarrow$ The contrast function has many spurious local extremes.

- The orthogonal constraint means that the sample correlation between $\widehat{s} = \mathbf{w}^H\mathbf{x}$ and $\widehat{\mathbf{z}} = \mathbf{B}\mathbf{x}$ is zero, i.e.,

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{w}^H\mathbf{x}(n)\mathbf{x}(n)^H\mathbf{B}^H = \mathbf{w}^H\widehat{\mathbf{C}}_{\mathbf{x}}\mathbf{B}^H = \mathbf{0}.$$

- By selecting **w** as the dependent variable on **a**, the orthogonal constraint is satisfied when

$$\mathbf{w} = \frac{\widehat{\mathbf{C}}_{\mathbf{x}}^{-1}\mathbf{a}}{\mathbf{a}^H\widehat{\mathbf{C}}_{\mathbf{x}}^{-1}\mathbf{a}} \qquad \text{(MPDR)}.$$

Orthogonal Constraint

- The link between the steering vector $\mathbf{a}$ and the separating vector $\mathbf{w}$ is too "weak" $\Longrightarrow$ The contrast function has many spurious local extremes.

- The orthogonal constraint means that the sample correlation between $\widehat{s} = \mathbf{w}^H\mathbf{x}$ and $\widehat{\mathbf{z}} = \mathbf{B}\mathbf{x}$ is zero, i.e.,

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{w}^H\mathbf{x}(n)\mathbf{x}(n)^H\mathbf{B}^H = \mathbf{w}^H\widehat{\mathbf{C}}_{\mathbf{x}}\mathbf{B}^H = \mathbf{0}.$$

- By selecting $\mathbf{w}$ as the dependent variable on $\mathbf{a}$, the orthogonal constraint is satisfied when

$$\mathbf{w} = \frac{\widehat{\mathbf{C}}_{\mathbf{x}}^{-1}\mathbf{a}}{\mathbf{a}^H\widehat{\mathbf{C}}_{\mathbf{x}}^{-1}\mathbf{a}} \qquad \text{(MPDR)}.$$

Orthogonal Constraint

- The link between the steering vector $\mathbf{a}$ and the separating vector $\mathbf{w}$ is too "weak" $\Longrightarrow$ The contrast function has many spurious local extremes.

- The orthogonal constraint means that the sample correlation between $\widehat{s} = \mathbf{w}^H\mathbf{x}$ and $\widehat{\mathbf{z}} = \mathbf{B}\mathbf{x}$ is zero, i.e.,

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{w}^H\mathbf{x}(n)\mathbf{x}(n)^H\mathbf{B}^H = \mathbf{w}^H\widehat{\mathbf{C}}_{\mathbf{x}}\mathbf{B}^H = \mathbf{0}.$$

- By selecting $\mathbf{w}$ as the dependent variable on $\mathbf{a}$, the orthogonal constraint is satisfied when

$$\mathbf{w} = \frac{\widehat{\mathbf{C}}_{\mathbf{x}}^{-1}\mathbf{a}}{\mathbf{a}^H\widehat{\mathbf{C}}_{\mathbf{x}}^{-1}\mathbf{a}} \qquad \text{(MPDR)}.$$

Gradient Ascent Algorithm: Blind MPDR

Let the unknown density $p_s(\cdot)$ be replaced by some $f(\cdot)$, and the unknown $\mathbf{C}_\mathbf{z}$ be replaced by $\widehat{\mathbf{C}}_{\widehat{\mathbf{z}}}$. Then, the gradient of the contrast is

$$\frac{\partial \mathcal{C}}{\partial \mathbf{a}^H} = \mathbf{w} - \lambda_\mathbf{a} \widehat{\mathbf{C}}_\mathbf{x}^{-1} \mathbf{X} \phi(\widehat{\mathbf{s}})^H / N,$$

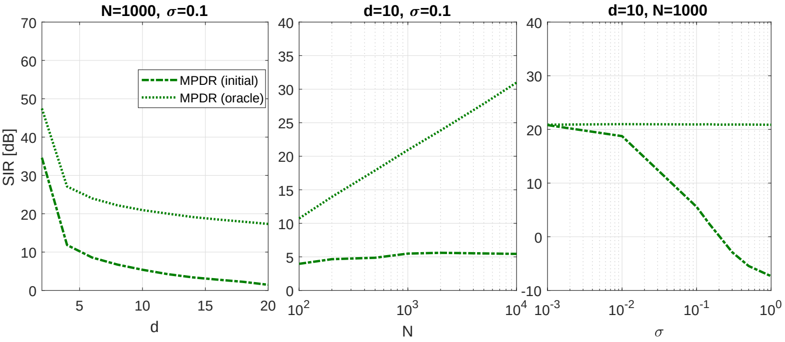
where $\phi = -\partial/\partial \overline{z}(\log f)$, $\lambda_\mathbf{a} = (\mathbf{a}^H \widehat{\mathbf{C}}_\mathbf{x}^{-1} \mathbf{a})^{-1}$ and $\phi(\cdot)$ must be normalized so that

$$\frac{1}{N} \widehat{\mathbf{s}} \phi(\widehat{\mathbf{s}})^H = 1$$

Main steps of the algorithm (repeated until convergence):

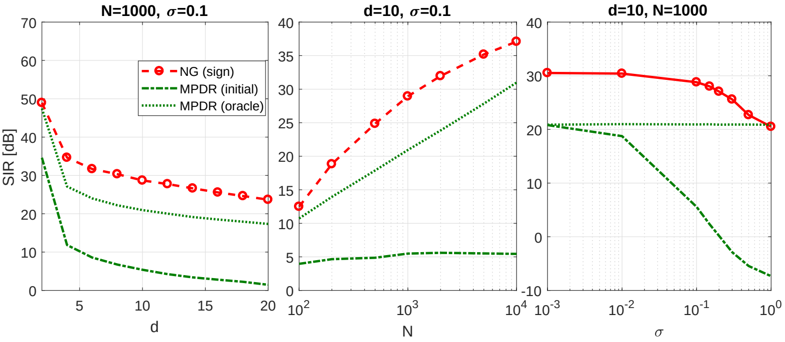
1. $\widehat{\mathbf{s}} = \mathbf{w}^H \mathbf{X}$
2. $\nu = \widehat{\mathbf{s}} \phi(\widehat{\mathbf{s}})^H / N$
3. $\mathbf{a} \leftarrow \mathbf{a} + \mu \nu^{-1} \frac{\partial \mathcal{C}}{\partial \mathbf{a}^H}$
4. $\mathbf{w} = \lambda_\mathbf{a} \widehat{\mathbf{C}}_\mathbf{x}^{-1} \mathbf{a}$ `% MPDR`

Simulations



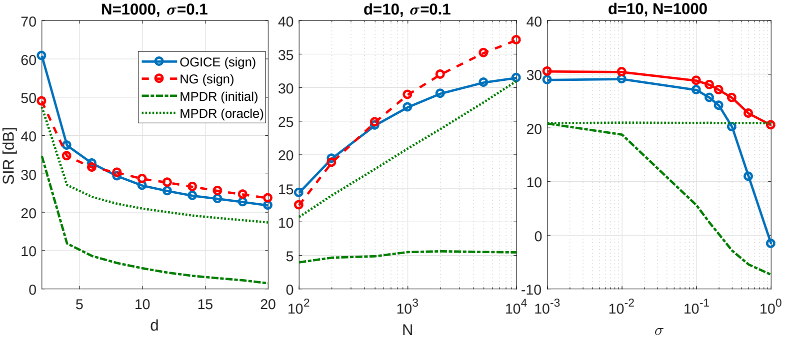
- Target signal: Circular Laplacean, Background: Gaussian
- Algorithms randomly initialized in the σ-vicinity of correct $\mathbf{a}$
- d is the number of sensors
- Averages taken from 1000 trials

Simulations



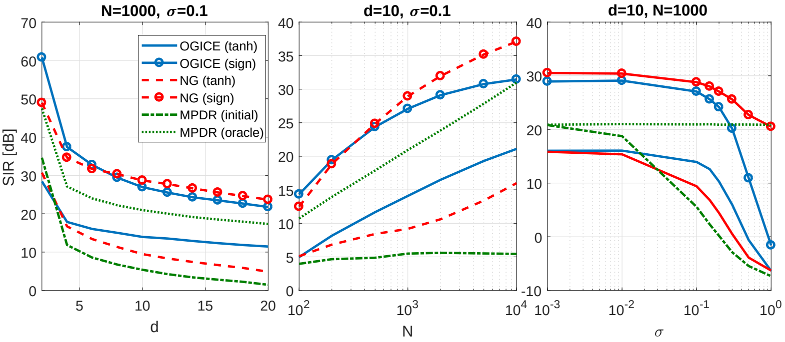
- Target signal: Circular Laplacean, Background: Gaussian
- Algorithms randomly initialized in the σ-vicinity of correct $\mathbf{a}$
- d is the number of sensors
- Averages taken from 1000 trials

Simulations



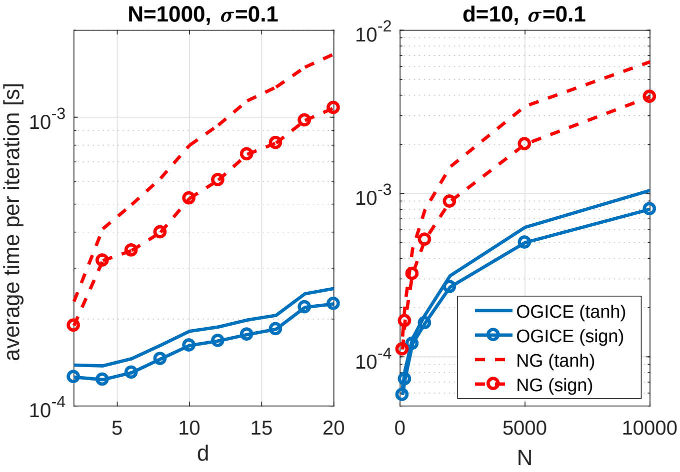
- Target signal: Circular Laplacean, Background: Gaussian
- Algorithms randomly initialized in the σ-vicinity of correct $\mathbf{a}$
- d is the number of sensors
- Averages taken from 1000 trials

Simulations



- Target signal: Circular Laplacean, Background: Gaussian
- Algorithms randomly initialized in the σ-vicinity of correct $\mathbf{a}$
- d is the number of sensors
- Averages taken from 1000 trials

Computational Time



Perspectives

- The idea can be extended for extracting vector components.
- A priori information (e.g. in the form of piloting signal) can be used to extend the area of convergence to the target signal.
- Faster optimization strategies can replace the gradient-based method.
- The background need not be modeled as Gaussian. Can we approach the accuracy that is attainable through ICA?

Perspectives

- The idea can be extended for extracting vector components.
- A priori information (e.g. in the form of piloting signal) can be used to extend the area of convergence to the target signal.
- Faster optimization strategies can replace the gradient-based method.
- The background need not be modeled as Gaussian. Can we approach the accuracy that is attainable through ICA?

Perspectives

- The idea can be extended for extracting vector components.
- A priori information (e.g. in the form of piloting signal) can be used to extend the area of convergence to the target signal.
- Faster optimization strategies can replace the gradient-based method.
- The background need not be modeled as Gaussian. Can we approach the accuracy that is attainable through ICA?

Perspectives

- The idea can be extended for extracting vector components.
- A priori information (e.g. in the form of piloting signal) can be used to extend the area of convergence to the target signal.
- Faster optimization strategies can replace the gradient-based method.
- The background need not be modeled as Gaussian. Can we approach the accuracy that is attainable through ICA?

Thank You!