Recent Developments on Multi-Channel Blind System Identification (BSI)

K. Abed-Meraim

ENST PARIS, Signal and Image Processing Dept.

abed@tsi.enst.fr

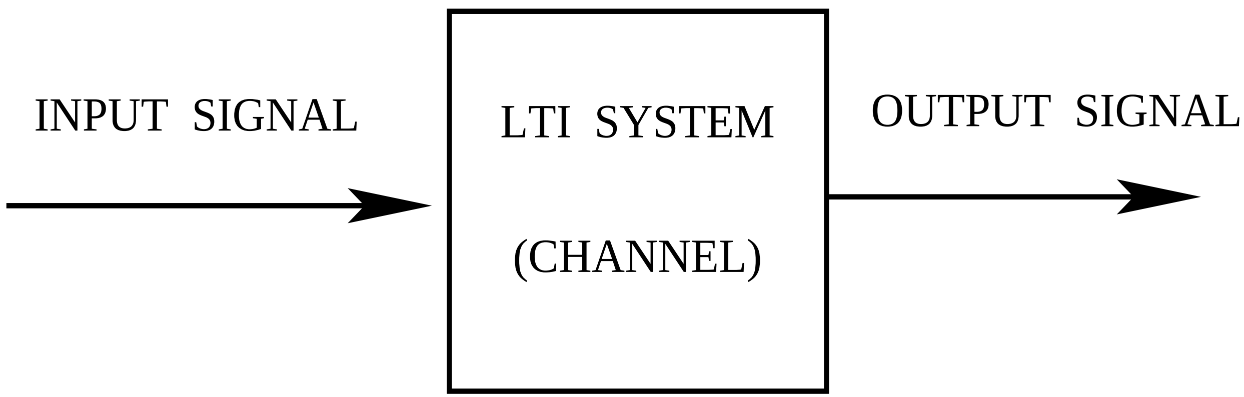
Presentation Outline

- Concepts and preliminaries
- BSI for SISO systems (mono-channel case)
- BSI for SIMO systems
- BSI for MIMO systems
- Concluding remarks

Blind System Identification:

Preliminaries

System identification



OBJECTIVE: Given the output signal and eventually certain side information (training sequence, physical or statistical information, partial channel knowledge, etc.), our objective is to estimate the channel (i.e., system transfer function) and restore the input signal.

Blind processing

We talk about *'BLIND PROCESSING'* in the situation where *'NO TRAINING SEQUENCE'* is available.

BSI $\Longleftrightarrow$ System identification using *only* the output data

Motivations:

- Increased channel throughput in communication systems.
- Robustness against channel modeling errors.
- Blind processing is necessary in certain applications (military applications, seismology, etc.)
- Flexibility and increased system autonomy.

Semi blind processing

Principle: Combining a data-aided (with training sequence) criterion J_{DA} with a blind criterion J_B, i.e:

$$J(h) = \alpha J_{DA}(h) + (1-\alpha) J_B(h)$$

Criterion choice: The blind criterion should be chosen according to the context. The data-aided criterion is usually chosen as the maximum likelihood (=MMSE) one.

The optimal value of α can be computed based on asymptotic performance analysis (Buchoux et al 1999).

Result: Improve the estimation accuracy and/or shorten the training sequence size and hence increase the 'useful' channel throughput.

Identification versus deconvolution

- **Blind identification**: Estimation of the channel state information using the observation data and certain 'statistical' information on the source signal.

- **Blind deconvolution**: Estimation of the input or the channel inverse (equalizer) using *only* the output (observation) signal. This is also known as the blind equalization problem in communication application.

Channel model

- *Parametric versus non-parametric*: Channel can be simply modeled by certain physical or statistical parameters, e.g. the specular channel model based on the paths delays, attenuations and angle of arrival.

- *Instantaneous versus convolutive*: In communication, convolutive model occurs when the channel delay spread is larger than the symbol duration.

- *Finite (FIR) versus infinite impulse response (IIR) channel*: For long memories channels (this is the case for example in echo cancellation), one model the channel and an IIR one using for example statistical ARMA representation.

Channel model

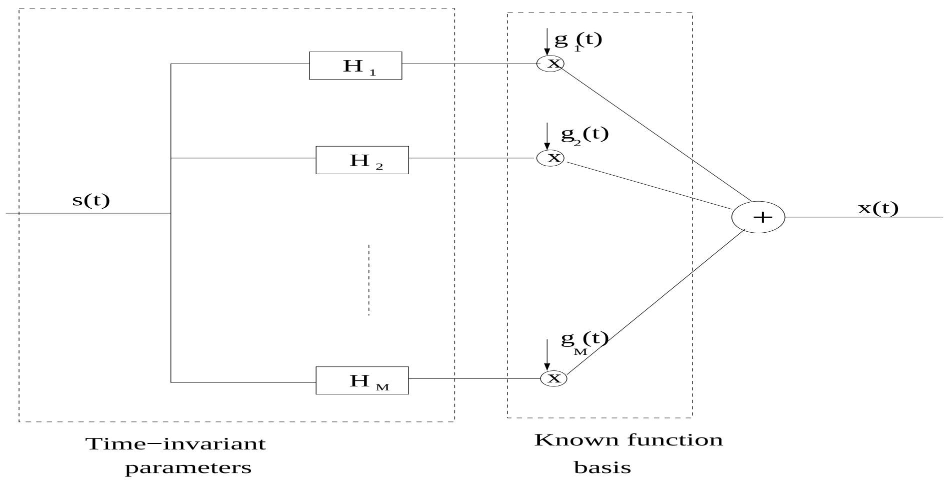
- *Linear versus non-linear*: Linear-quadratic or post-linear channel models have been considered in the literature. The non-linearity may be due, for example, to amplificator saturation (e.g. satellite communication).

- *Stationary versus non-stationary*: Stationarity is a ‘good’ approximation over a ‘large’ observation period in most real-life applications. Non-stationary model has been considered, for example, in the over-the-horizon channel deconvolution problem.

How to cope with non-stationarity

- By using adaptive and tracking algorithms, e.g. LMS, RLS, PAST, etc.

- By using channel representation with known basis functions.

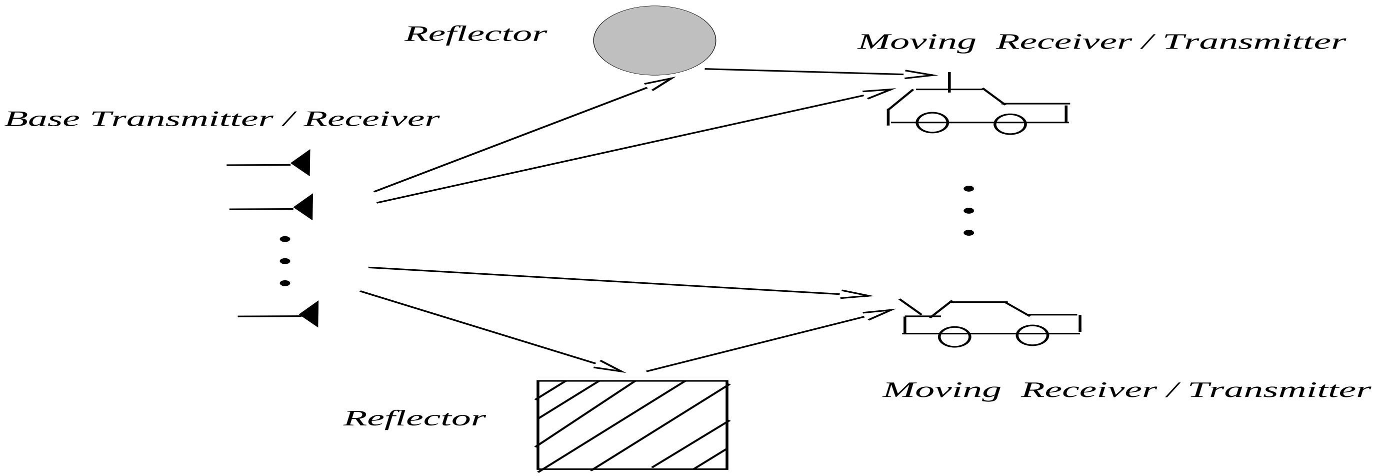


- By using time-frequency signal analysis.

Inherent ambiguities

- **Amplitude**: $y = h \star s = \lambda h \star \frac{1}{\lambda} s$.
- **Phase**: $y = h \star s = e^{j\theta} h \star e^{-j\theta} s$.
- **Delay**: In the stationary source case, $s(t)$ and $s(t + \tau)$ have the same statistical information.
- **Permutation**: This occurs in the multiple input case since the labeling of the source signals is arbitrary.

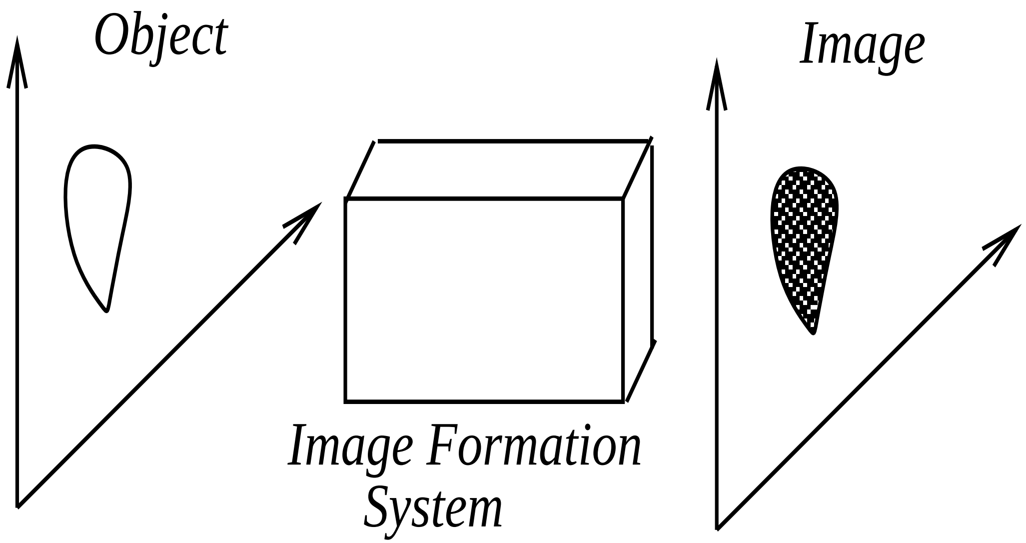
Application example: wireless communication



The objective here is to restore the transmitted signal that has been distorted by the propagation channel.

The blind processing helps in increasing the 'information data' throughput of the channel (for example, in GSM system the training data represents about 25% of transmitted data).

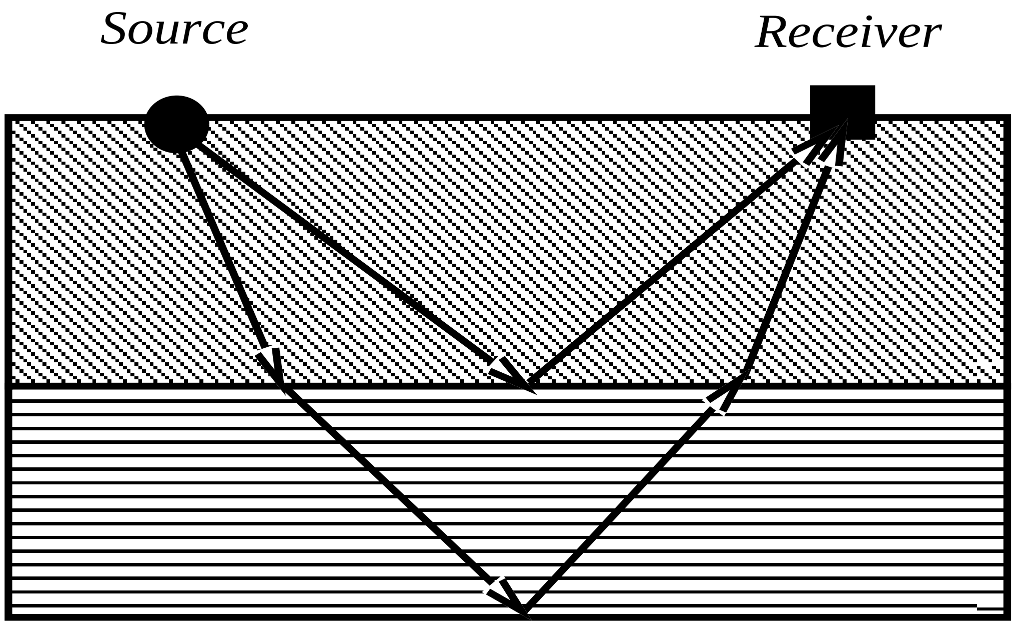
Application example: Image restoration



Objectives: From a blurred image retrieve the original one and/or the point spread function (channel).

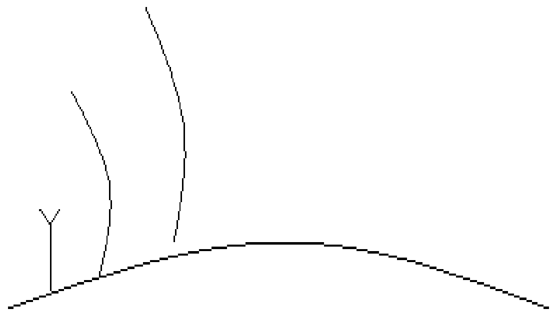
From several 'low quality' images form a 'high or improved quality' image.

Application example: Exploration seismology

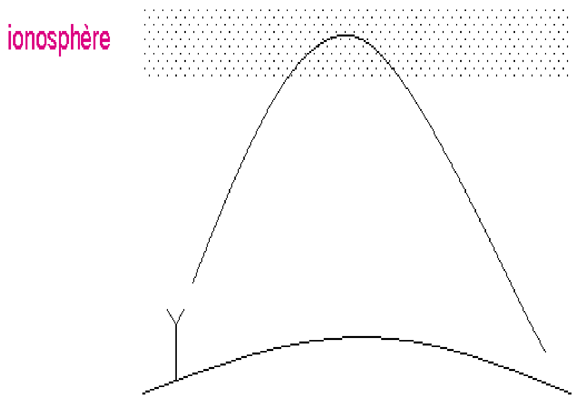


(Blind) channel estimation is used here to get information on the underground structure and the position of the reflectors.

Application example: Over-The-Horizon Radar (OTHR)

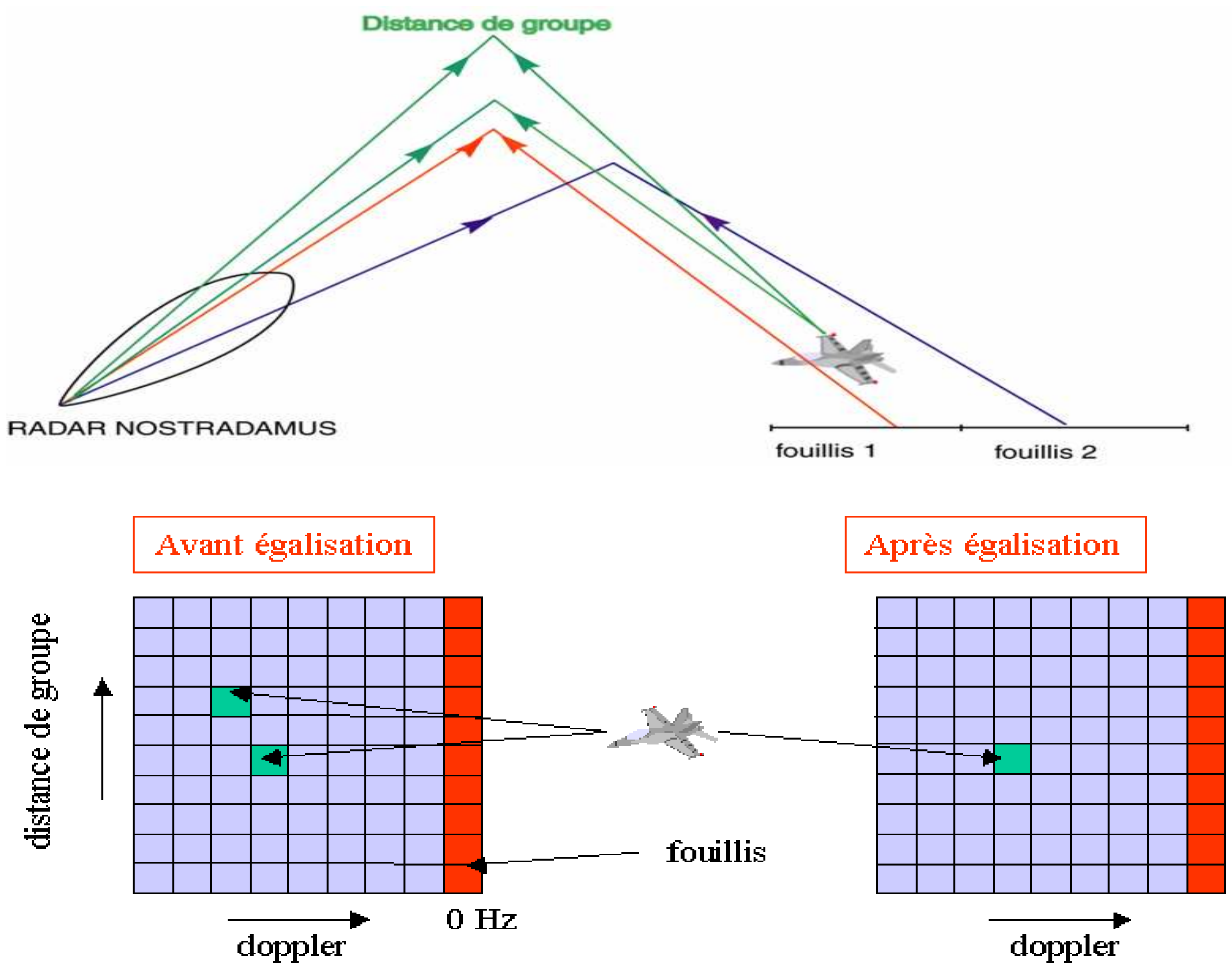


Classical radar: Limited horizon



OTHR: Early detection at all altitudes.

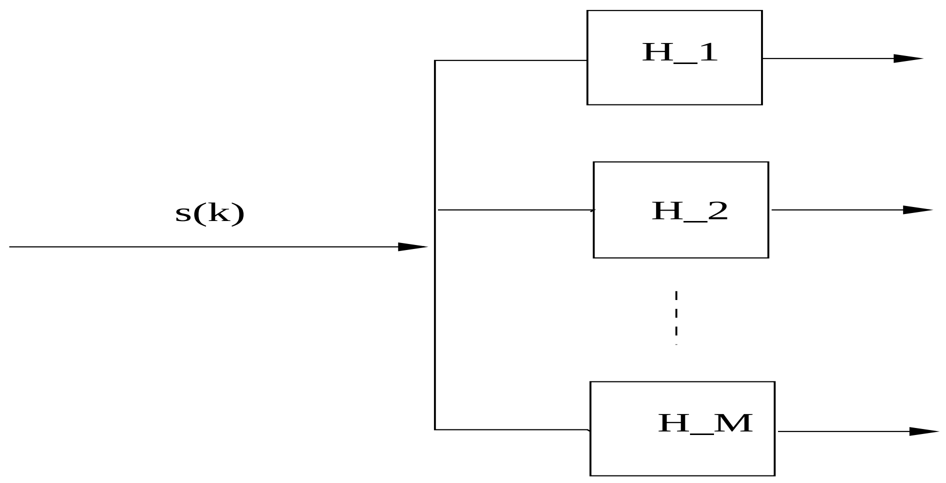
Application example: Over-The-Horizon Radar (OTHR)



Other potential applications

- Blind deconvolution for ultrasonic non-destructive testing (Nandi et al 1997, C.H. Chen et al 2002)
- ECG data processing (Sabry-Rizk et al., 1995): Fetal electro-cardiogram extraction,
- Acoustical and environmental robustness in automatic speech recognition (A. Acero et al 1993)
- Military applications, e.g. interference mitigation (M. Amin et al 1997), signal interception (Ph. Loubaton et al 2000), etc.

Multichannel processing: Diversity

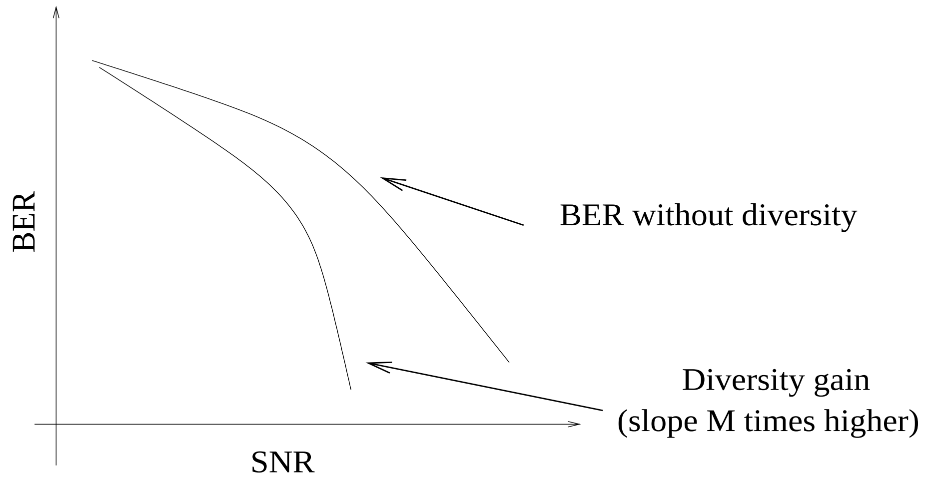


Multichannel processing is intimately linked to the concept of diversity:

Diversity: We would say that we have an order M diversity in the situation where we have several (M) replicas of the same input signal observed through M different and ‘independent’ channels.

Diversity gain

- **Improved restoration quality**: In communication, one can decrease the bit error rate (BER) by a factor of M (M being the diversity order)

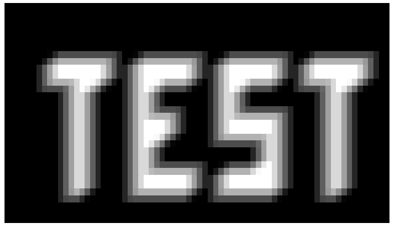


- **Increased transmission rate**: The diversity increases the channel capacity.

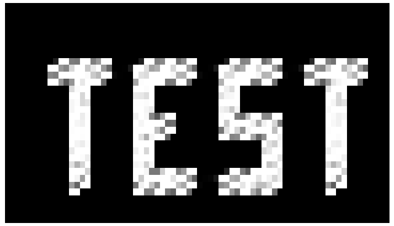
Example: Monochannel image restoration

Original Image

Blurred Image : motion filter

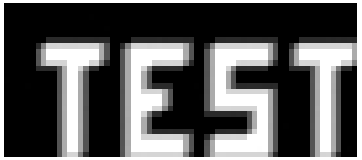


Deblurred image

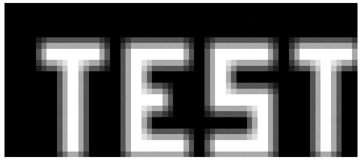


Example: Multichannel image restoration

Blurred noisy image: Motion filter

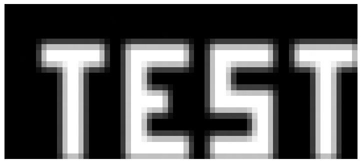


Degraded noisy image: Average filter

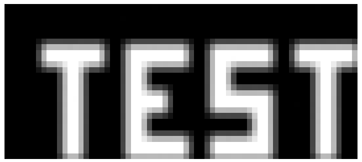


Original image

Degraded noisy image: Gaussian filter(v=0.8)



Blurred noisy image : Gaussian filter(v=1)



Deconvolved image

Multichannel system: processing strategy

- *Separate processing*: Perform the blind deconvolution for each channel followed by a maximum ratio combiner of the channel outputs (simplicity, SNR gain but loss of the multichannel diversity).
- *Selective approach*: Deconvolution based only on the ‘best’ channel (simplicity but difficulty to define the best channel in the convolutive case).
- *Joint processing*: Process the channel outputs jointly in order to restore the input signal (leads to the best performance gain).

Channel type (system dimension)

We consider a linear time-invariant finite impulse response channel in the following three cases:

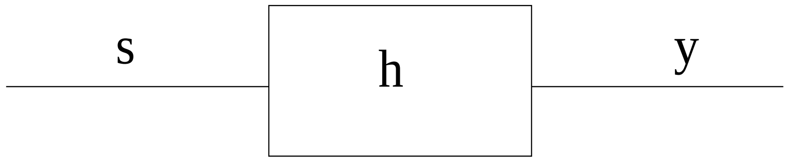
- *Single Input Single Output (SISO)* channel: This model is the most standard and the one considered first in the literature.

- *Single Input Multiple Output (SIMO)* channel: This is the situation for example when a multi-sensor antenna is used at the receiver.

- *Multiple Input Multiple Output (MIMO)* channel: This is an extension of the SIMO case when multiple users (signals) are considered. SIMO and MIMO cases have been studied extensively during the last decade.

Blind system identification for SISO channels

Data model



Noiseless observation:

$$y(k) = h(k) \star s(k)$$

$h(k)$ represents a LTI finite impulse response filter and $s(k)$ is a zero-mean stationary sequence of i.i.d (independent and identically distributed) *non-gaussian* random-variables of variance σ_s^2.

Need for higher order statistics (HOS)

- **Second order statistics (SOS) information**: The power spectral density of the output data is:

$$S_y(f) = |H(f)|^2 \sigma_s^2$$

$\Rightarrow$ No channel phase information from the observation SOS. BSI using the data SOS is only possible if the channel is of minimum phase.

- **HOS information**: Data HOS are needed to estimate the missing channel phase information.

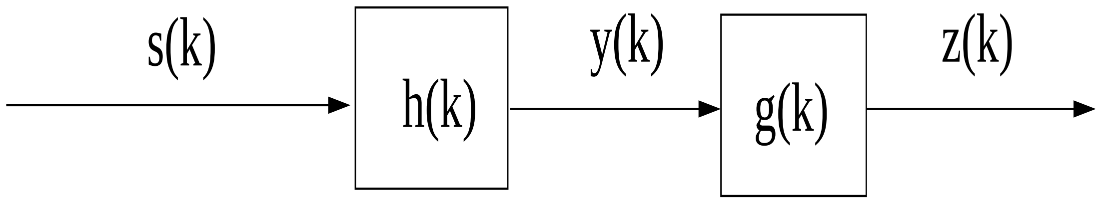
$\Rightarrow$ the source signal must be *non gaussian*.

HOS-based BSI

- **Explicit HOS methods**: Direct system identification through explicit use of the signal HOS, e.g. 4th order cumulant-based methods (J.A. Cadzow et al 1996), polyspectra based methods (C. Nikias et al 1993, D. Hatzinakos et al 1991), etc.

- **Implicit HOS methods**: Identification of the channel inverse filter (equalizer) through optimization of appropriate non-linear cost functions, Sato algorithm (Y. Sato 1975), CMA algorithm (D. Godard, Treichler et al 1980), Bussgang algorithms (A. Benveniste et al 1980), etc.

Example of an explicit HOS method



Shavi-Weinstein method: Estimate the channel inverse filter $g(k)$ in such a way that we maximize the (absolute value) of its output $z(k)$ fourth order cumulant (under constant power constraint).

Idea: maximize the nongaussianity of $z(k)$ by maximizing its 4th order cumulant.

Example of an implicit HOS method

Constant Modulus Algorithm (CMA): Introduced in communication (initially) for constant modulus constellation signals:

$$g = \arg\min E(|z(k)|^2 - R)^2$$

Idea: Restore the constant modulus property of the source signal at the equalizer output.

General features of HOS-based methods

- In general, HOS based methods require large sample sizes to achieve ‘good’ estimation performances.
- Non-linear optimization techniques are needed to estimate the channel (or the inverse channel) parameters. Often, stochastic gradient techniques are used for the optimization.
- The HOS based criteria suffer from the existence of local-minima.
- Convergence analysis is possible only in the noiseless case.

Blind system identification for SIMO channels

Motivation for multichannel processing

Blind deconvolution using SOS:

- *Single channel case*: Not possible unless the channel is minimum-phase. The minimum phase condition in the SISO case is a 'strong' condition that is, in general, not met in practice.

- *Multichannel case*: Almost always possible $\Rightarrow$ More robust and more accurate estimation. In fact, the minimum phase condition in the SIMO case is a 'mild' condition that is satisfied when the channels are sufficiently independent from one another.

Motivation for multichannel processing

More channel capacity in communication systems

- *Single channel case*:

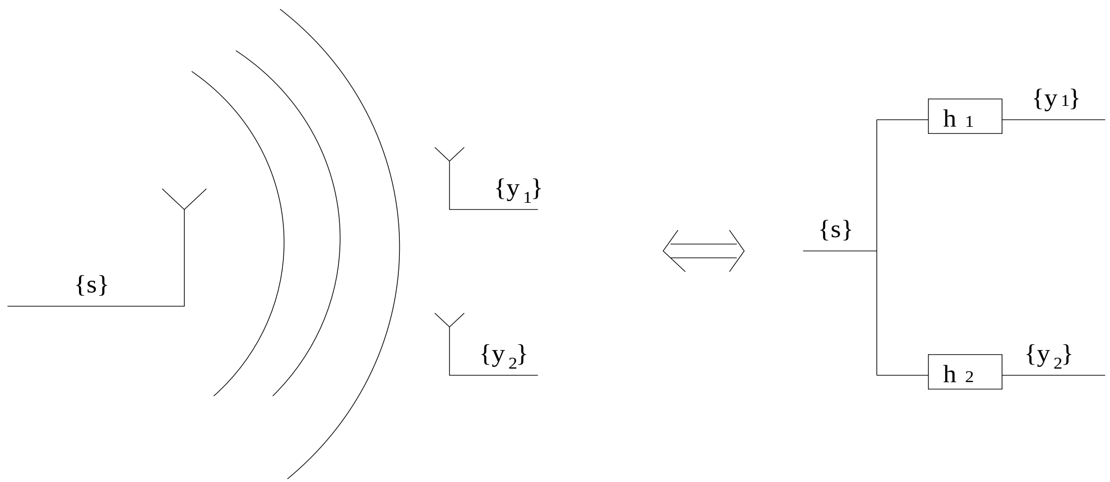
$$C = \log_2(1+\rho)$$

- *Multichannel case*: (M transmit and receive channels)

$$C = \log_2 \det(1 + \frac{\rho}{M}\mathbf{H}\mathbf{H}^H) \overset{M\to\infty}{\longrightarrow} M \log_2(1+\rho)$$

The capacity gain comes from the fact that having several replicas of the transmitted signal observed through independent channels reduces significantly the risk of information loss.

Space Diversity (Multiple receivers)

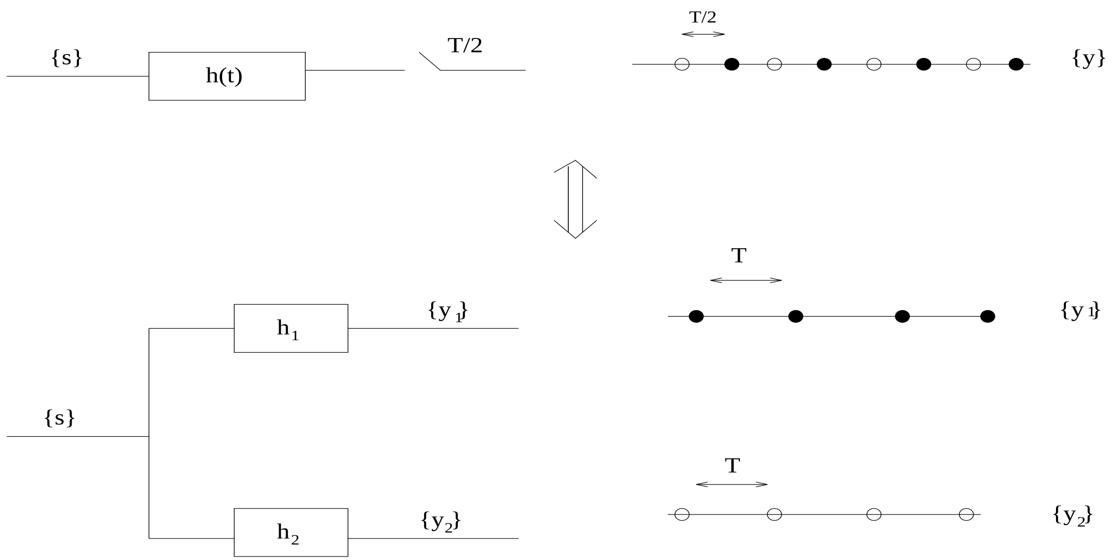


- $s(n)$: the source signal.
- $h_i(z)$: models the propagation between the emitting source and the i-th sensor.

$$h_i(z) = \sum_k h_i(k) z^{-k}$$

- $y_i(n)$: output at the i-th sensor.

Time diversity (oversampling)



$$y(t) = \sum_k h(t - kT)s(k) \quad \textbf{cyclostationary}$$

$\Rightarrow$ Exploit the cyclostationarity (time diversity) by oversampling wrt the symbol duration.

Time diversity (oversampling)

By oversampling we have multiple 'virtual' channels:

$$\begin{cases} h_1(z) & = & \sum_k h(kT)z^{-k} \\ h_2(z) & = & \sum_k h(kT+T/2)z^{-k} \end{cases}$$

The cyclostationary oversampled signal can be represented as a *stationary multivariate* signal as:

$$\begin{cases} y_1(k) & = & x(kT) \\ y_2(k) & = & x(kT+T/2) \end{cases} \quad \textbf{stationary}$$

Multichannel model

$$\left\{\begin{array}{lcl} y_1(k) & = & s(k) * h_1(k) \\ y_2(k) & = & s(k) * h_2(k) \\ \vdots & & \\ y_M(k) & = & s(k) * h_M(k) \end{array}\right. \qquad k = 0, \cdots, N-1$$

- $s(n)$: single unknown source signal.
- To each output i corresponds the FIR transfer function $h_i(z)$

$$h_i(z) = \sum_{k=0}^{L} h_i(k) z^{-k}$$

Multichannel model

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_M \end{bmatrix} \begin{bmatrix} s(-L) \\ \vdots \\ s(N-1) \end{bmatrix} = \mathbf{H}\mathbf{s}$$

$\mathbf{s}$ is the input vector, $\mathbf{y}_i$ is the observation vector at sensor i and $\mathbf{H}_i$ is the $N \times (N+L)$ Sylvester matrix

$$\mathbf{H}_i = \begin{bmatrix} h_i(L) & \cdots & h_i(0) & \cdots & 0 \\ \vdots & \ddots & & \ddots & \vdots \\ 0 & \cdots & h_i(L) & \cdots & h_i(0) \end{bmatrix}$$

Some properties of SIMO systems

- Weak *minimum phase condition*

$$\mathbf{h}(z) \stackrel{\text{def}}{=} \begin{bmatrix} h_1(z) \\ \vdots \\ h_M(z) \end{bmatrix} \neq 0 \ \text{ for } |z| > 1$$

Satisfied as soon as $h_i(z),\ 1 \leq i \leq M$ do not share common zeros.

- **Left invertible system**: as soon as the $MN \times (N+L)$ matrix $\mathbf{H}$ is full column rank, i.e. when we have more equations than unknowns.

Some properties of SIMO systems

- **Finite zero-forcing inverse filters**: if $\mathbf{h}(z) \neq 0, \ \forall \ z$, $\exists \ \mathbf{g}(z) = [g_1(z), \cdots, g_M(z)]$ a polynomial vector such that:

$$\underbrace{\mathbf{g}(z)\mathbf{h}(z) = \sum_{k=1}^{M} g_i(z)h_i(z) = 1}_{\textbf{Bezout equality}}$$

- **Exact identification** in the noiseless case from a finite sample size observation vector.

SIMO versus SISO

- FIR equalizer for SIMO versus IIR equalizer for SISO.
- Causal equalizer for SIMO versus non-causal equalizer for SISO.
- Exact estimation using finite sample size for SIMO (not possible for SISO).
- Equalizer delay plays an important role in SIMO case and not in the SISO case.
- SOS-based BSI for SIMO versus HOS-based BSI for SISO.

Strict identifiability

Definition

The system is *strictly identifiable* if a given output $\mathbf{y}$ implies a unique input $\mathbf{s}$ and a unique system matrix $\mathbf{H}$ up to an unknown scalar, i.e.,

$$\mathbf{H}'\mathbf{s}' = \mathbf{H}\mathbf{s} \Longrightarrow \mathbf{s}' = \alpha\mathbf{s} \ \text{ and } \ \mathbf{h}'(z) = \frac{1}{\alpha}\mathbf{h}(z)$$

where α is a given non-zero scalar.

Strict identifiability

Necessary condition: The system is identifiable only if the followings are true:

$$\left\{ \begin{array}{c} \mathbf{h}(z) \neq 0, \ \forall z \\ p \geq L+2 \\ N \geq L+2 \end{array} \right.$$

where p is the number of modes in the input sequence.

Strict identifiability

Sufficient condition : The system is identifiable if the followings are true:

$$\begin{cases} \mathbf{h}(z) \neq 0, \ \forall z \\ p \geq 2L + 1 \\ N \geq 3L + 1 \end{cases}$$

Strict identifiability

The identifiability conditions shown above essentially ensure the following intuitive requirements:

- *All channels in the system must be different enough from each other.* They can not be identical, for example.

- *The input sequence must be complex enough.* It can not be zero, a constant or a single sinusoid, for example.

- *There must be enough number of output samples available.* A set of available data can not yield sufficient information on a larger set of unknown parameters, for example.

Estimation techniques

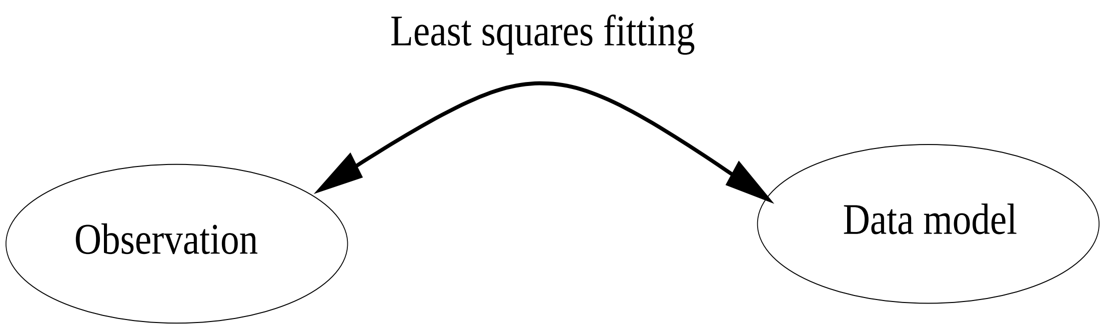
- Direct estimation of **system function**:
  - Maximum likelihood (ML) method.
  - Cross-relations (CR) method.
  - Channel subspace (CS) method.
- Direct estimation of **system input**:
  - Signal subspace (SS) method.
  - Mutually referenced equalizers (MRE) method.
  - Linear prediction (LP) method.

Maximum likelihood method

Principle: Assuming a circular white Gaussian noise vector $\mathbf{w}$

$$p(\mathbf{y}) = \frac{1}{\pi^N \sigma^{2N}} \exp(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2)$$

Thus the ML estimate is given by $\boxed{(\mathbf{H}, \mathbf{s})_{ML} = \arg\min_{\mathbf{H},\mathbf{s}} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2}$



Maximum likelihood method

Separable problem: Minimize over $\mathbf{s}$:

$$\mathbf{s}_{ML} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}$$

Then over $\mathbf{H}$:

$$\mathbf{H}_{ML} = \arg\min_{\mathbf{H}} \|\mathbf{P}_H^{\perp}\mathbf{y}\|^2$$

$\mathbf{P}_H^{\perp}$ = orthogonal projection matrix onto Range$(\mathbf{H})^{\perp}$.

Orthogonal Complement Matrix (OCM)

Idea: One can obtain noise vectors by observing that

$$[0, \cdots, \underset{\text{i-th}}{-h_j(z)}, 0, \cdots, 0, \underset{\text{j-th}}{h_i(z)}, \cdots, 0] \begin{bmatrix} h_1(z) \\ \vdots \\ h_M(z) \end{bmatrix} = 0$$

Result (Y. Hua 1995) : One can form an OCM $\mathbf{G}$ that is a linear function of the channel parameters such that its column vectors form a basis of the noise subspace, i.e.

$$\mathbf{P}_G = \mathbf{P}_H^{\perp}$$

Two step estimation technique

ML criterion:

$$\mathbf{h}_{ML} = \arg \min_{\|\mathbf{h}\|=1} \mathbf{y}^H \mathbf{G}(\mathbf{G}^H \mathbf{G})^{\#} \mathbf{G}^H \mathbf{y}$$

where $\mathbf{h}$ is the vector of all channels' impulse responses. From the *commutativity property* of linear convolution:

$$\mathbf{G}^H \mathbf{y} = \mathbf{Y}\mathbf{h}$$

we obtain

$$\mathbf{h}_{ML} = \arg \min_{\mathbf{h}} \mathbf{h}^H \mathbf{Y}^H (\mathbf{G}^H \mathbf{G})^{\#} \mathbf{Y}\mathbf{h}$$

Two step estimation technique

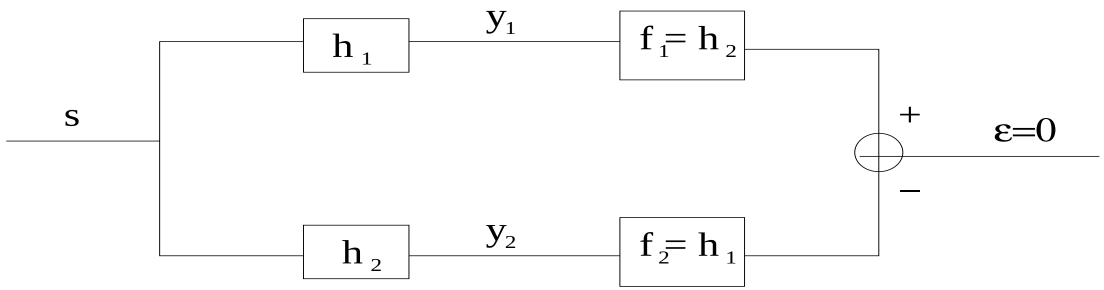
Two Step Maximum Likelihood (TSML):

1. $\mathbf{h}_c = \arg\min \; \mathbf{h}^H \mathbf{Y}^H \mathbf{Y} \mathbf{h}$

2. $\mathbf{h}_e = \arg\min \; \mathbf{h}^H \mathbf{Y}^H (\mathbf{G}_c^H \mathbf{G}_c)^{\#} \mathbf{Y} \mathbf{h}$, where $\mathbf{G}_c$ is $\mathbf{G}$ constructed from $\mathbf{h}_c$.

At each step the solution is given by the least eigenvector associated to the least eigenvalue of the considered quadratic form.

Cross-relations (CR) method



- **Principle**: For every pair of channels, we have

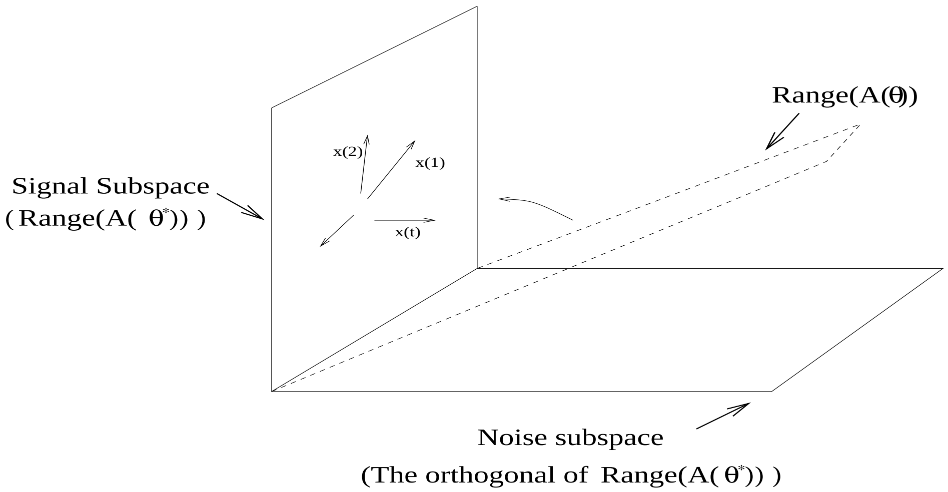
$$y_i(k) * h_j(k) = y_j(k) * h_i(k)$$

- **Algorithm**: By collecting all possible pairs of M channels, one can easily establish a set of linear equations:

$$\mathbf{Y}\mathbf{h} = 0$$

This yields to $\mathbf{h}_{CR} = \arg\min \mathbf{h}^H \mathbf{Y}^H \mathbf{Y}\mathbf{h}$ (first step of TSML).

Subspace method



Principle: Assume the following model: $\mathbf{x}(n) = \mathbf{A}(\theta)\mathbf{s}(n)$ with

$$\mathrm{Range}(\mathbf{A}(\theta)) = \mathrm{Range}(\mathbf{A}(\theta')) \Longleftrightarrow \theta = \theta'$$

Thus, θ can be estimated as:

$$\hat{\theta} = \arg\min_{\theta} \mathrm{d}(\mathrm{Range}\{\mathbf{x}(n)\},\ \mathrm{Range}(\mathbf{A}(\theta)))$$

Channel subspace (CS) method

- **Model**:

$$
\begin{aligned}
\mathbf{y}(n) &= \mathbf{H}\mathbf{s}(n)\ n = 0, ..., N - W \\
\mathbf{y}(n) &= [\mathbf{y}_1^T(n), \cdots, \mathbf{y}_M^T(n)]^T \\
\mathbf{y}_i(n) &= [y_i(n), \cdots, y_i(n + W - 1)]^T
\end{aligned}
$$

In our case: $\mathbf{A} \longleftrightarrow \mathbf{H}$ and $\theta \longleftrightarrow \mathbf{h}$.

- **Main result**: If $W \geq L + 1$ and the M channels do not share a common zero, then

$$
\text{Range}(\mathbf{H}) = \text{Range}(\mathbf{H}') \Longleftrightarrow \mathbf{h}' = \alpha \mathbf{h}
$$

where α is a scalar constant.

CS algorithm

- Estimate the signal (resp. noise) subspace as the principal (resp. minor) eigen-subspace of the data covariance matrix $\mathbf{R}_y$:

$$\mathbf{R}_y = \sum_n \mathbf{y}(n)\mathbf{y}^H(n) = [\mathcal{E}_s \ \mathcal{E}_n] \begin{bmatrix} \mathbf{\Lambda}_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathcal{E}_s^H \\ \mathcal{E}_n^H \end{bmatrix}$$

where $\quad \text{Range}(\mathcal{E}_s) = \text{Range}(\mathbf{H}) \perp \text{Range}(\mathcal{E}_n).$

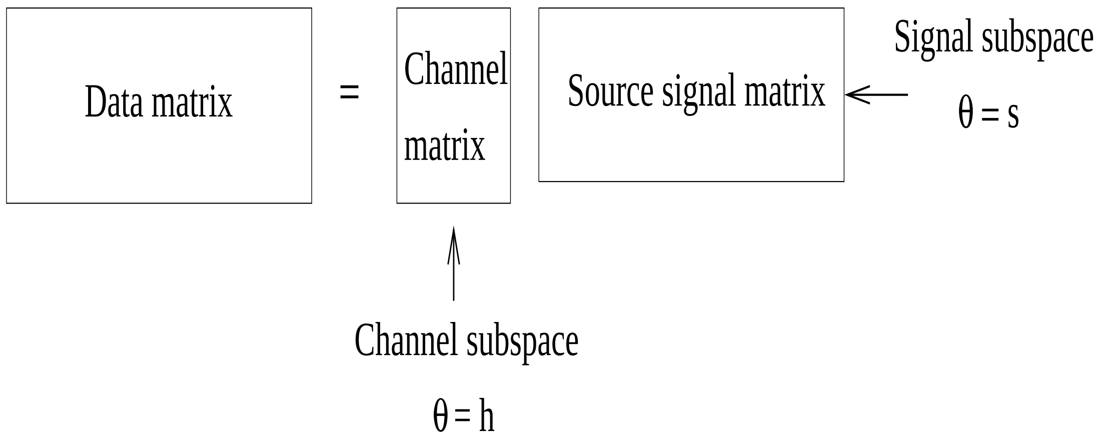
- Compute the least square error solution to

$$\mathbf{h}_{CS} = \arg \min_{\|\mathbf{h}\|=1} \|\mathcal{E}_n^H \mathbf{H}\|^2.$$

Comparison of the ML, CR, and CS methods

- **ML method**
  - Large computational cost
  - Very good estimation accuracy
- **CR method**
  - Low computational cost
  - Moderate estimation accuracy
- **CS method**
  - Moderate computational cost
  - Good estimation accuracy

Signal Subspace (SS) method



- **Model**: $\mathbf{Y} = [\mathbf{y}(0), \cdots, \mathbf{y}(N-W)] = \mathbf{HS}$

  $\mathbf{S}$ being a the source signal matrix of Hankel structure.

- **Principle**: $\mathbf{A} \longleftrightarrow \mathbf{S}$ and $\theta \longleftrightarrow \mathbf{s}$

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} = \mathrm{d}(\mathrm{Row}(\mathbf{Y}),\ \mathrm{Row}(\mathbf{S}))$$

Signal Subspace (SS) method

- **Result** (Xu et al 1995): Assume that $\mathbf{H}$ is full column rank and that the input sequence $\{s(n)\}_{-L \leq n \leq N-1}$ contains more than $W + L + 1$ modes, then

$$\mathrm{Row}(\mathbf{S}) = \mathrm{Row}(\mathbf{S}') \Longleftrightarrow \mathbf{s}' = \alpha \mathbf{s}$$

where α is a scalar constant.

SS algorithm

- Perform the SVD of the data matrix $\mathbf{Y} = [\mathbf{y}(0), \cdots, \mathbf{y}(N-W)]$

$$\mathbf{Y} = \mathbf{U} \begin{bmatrix} \mathbf{\Sigma}_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_s^H \\ \mathbf{V}_n^H \end{bmatrix}$$

$\mathbf{V}_n$ is the orthogonal matrix to the row space of $\mathbf{S}$

$$\mathbf{V}_n \mathbf{S}^H = 0$$

- Estimate s by minimizing the quadratic criterion

$$\hat{\mathbf{s}} = \arg \min_{\|\mathbf{s}\|=1} \|\mathbf{V}_n \mathbf{S}^H\|^2$$

Blind equalization

- **Definition**: $\mathbf{g}(z)$ is a blind equalizer iff:

$$\mathbf{g}(n) * \mathbf{y}(n) = \alpha s(n-m) \iff \mathbf{g}(z)\mathbf{h}(z) = \alpha z^{-m}$$

- **Characterization**:
  - *Statistical criterion*: If $s(n)$ is i.i.d.

$$\mathbf{g}(z) \longrightarrow \hat{s}(n) = \mathbf{g}(n) * \mathbf{y}(n) \text{ is i.i.d.}$$

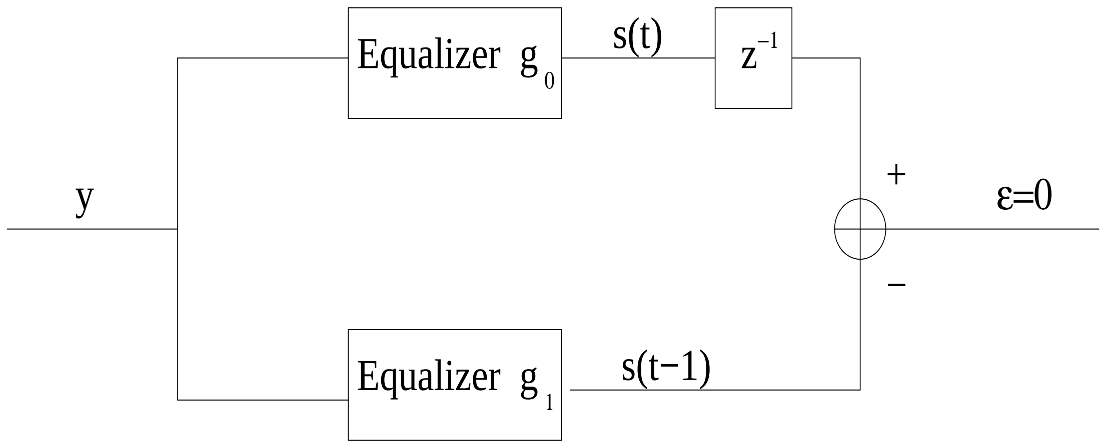
e.g., Linear prediction , Bussgang , etc.

  - *Geometrical criterion*: If $s(n) \in \mathcal{A}$

$$\mathbf{g}(z) \longrightarrow \hat{s}(n) = \mathbf{g}(n) * \mathbf{y}(n) \in \mathcal{A}$$

e.g., CMA algorithms.

Mutually referenced equalizers (MRE) method



MRE relations: Let $\mathbf{g}_i(n)\ i = 0, \cdots, W + L - 1$ be equalizer filters satisfying

$$\mathbf{g}_i(n) \star \mathbf{y}(n) = \alpha s(n-i),\ i = 0, 1, \cdots$$

Then, filters $\mathbf{g}_i$ should satisfy (MRE relations):

$$\mathbf{g}_i \star \mathbf{y}(n) = \mathbf{g}_{i+1} \star \mathbf{y}(n+1)$$

Mutually referenced equalizers (MRE) method

- **Result** (D. Gesbert et al 1994) : Vice versa, the previous relations characterize uniquely the equalizer filters, i.e. if $\mathbf{g}_0, \cdots, \mathbf{g}_{d-1}$ $(d = W + L)$ satisfy the MRE relations, then

$$\mathbf{g}_i(n) \star \mathbf{y}(n) = \alpha s(n-i), \quad \forall\, i$$

- **Algorithm**: $\{\mathbf{g}_i\}$ are estimated by minimizing (under a suitable constraint) the quadratic criterion

$$J = \sum_{n,i} \|\mathbf{g}_i \star \mathbf{y}(n) - \mathbf{g}_{i+1} \star \mathbf{y}(n+1)\|^2$$

Linear prediction (LP) method

- **Model**:

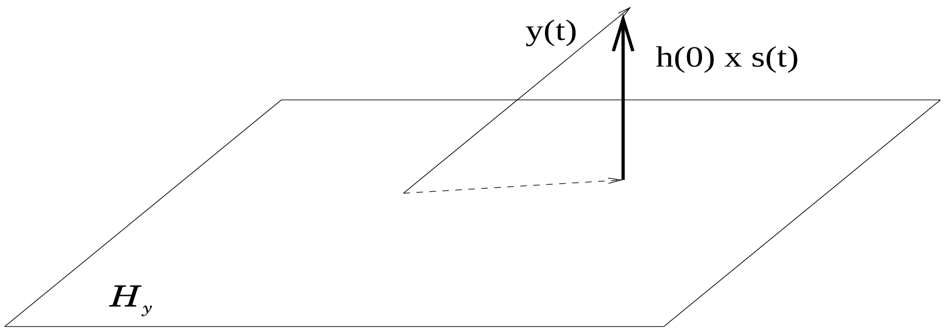
$$\begin{cases} \mathbf{y}(n) = [y_1(n), \cdots, y_M(n)]^T = [\mathbf{h}(z)]s(n), \quad s(n): \ i.i.d \\ \mathbf{h}(z) = [h_1(z), \cdots, h_M(z)]^T \neq 0, \forall\, z \end{cases}$$

- **Principle**: **Bezout equality**: $\exists\, \mathbf{g}(z) = [g_1(z), \cdots, g_M(z)]$ such that $\mathbf{g}(z)\mathbf{h}(z) = 1$

$$\Longrightarrow [\mathbf{g}(z)]\mathbf{y}(n) = s(n)$$

- **Result**: $\mathbf{y}(n)$ is an AR process of order L. Its innovation process is $\mathbf{i}(n) = \mathbf{h}(0)s(n)$.

LP algorithm



Prediction (projection) subspace

- Estimate the prediction coefficients of $\mathbf{y}(n)$ by solving the Yule-Walker equations:

$$\mathbf{y}(n) + \sum_{k=1}^{P} \mathbf{A}(k)\mathbf{y}(n-k) = \mathbf{i}(n) = \mathbf{h}(0)s(n)$$

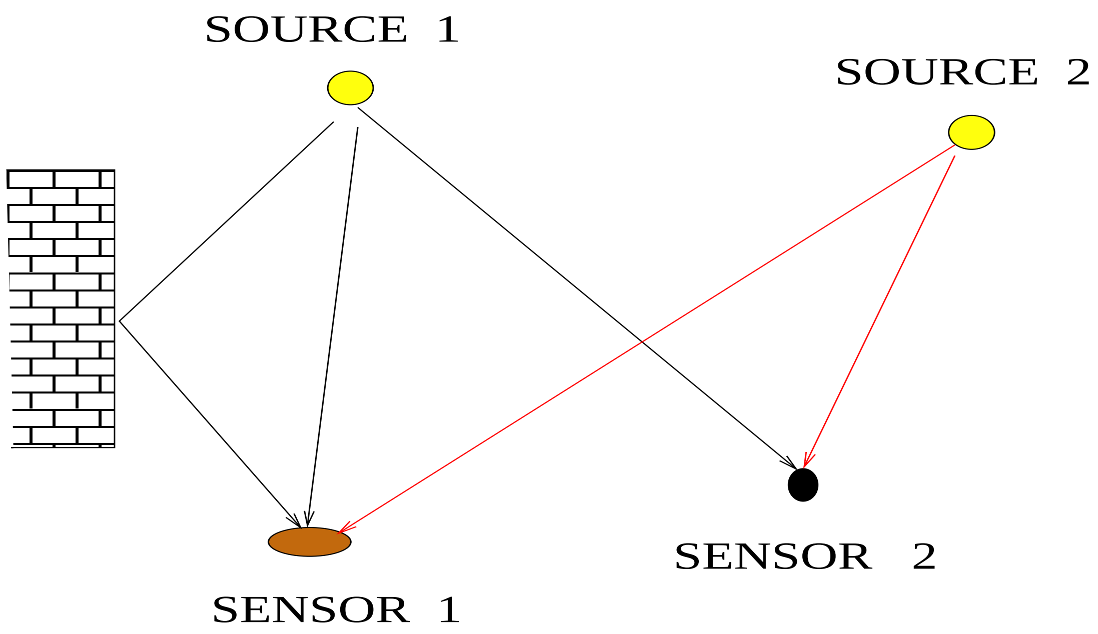
- Estimate vector $\mathbf{h}(0)$ (up to a constant) as the principal eigenvector of the innovation covariance matrix $\mathbf{D} = E(\mathbf{i}(n)\mathbf{i}(n)^H) = \mathbf{h}(0)\mathbf{h}(0)^H$.

Comparison of the SS, MRE, and LP methods

- **SS method**
  - Large computational cost
  - Good estimation accuracy
  - Deterministic input
- **MRE method**
  - Moderate computational cost
  - Good estimation accuracy
  - Deterministic input
- **LP method**
  - Low computational cost
  - Moderate estimation accuracy
  - Stochastic decorrelated input

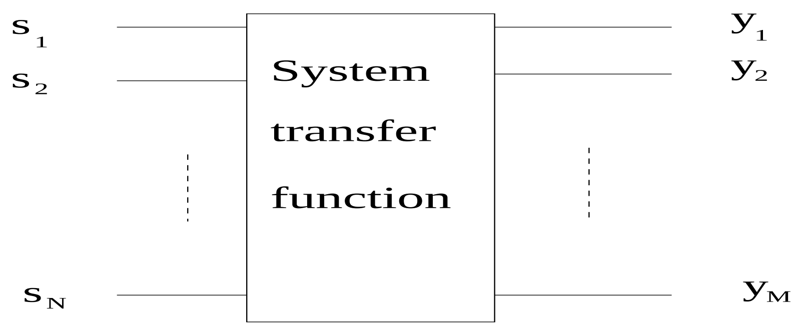
Blind system identification for MIMO channels

The situation of interest



M different (possibly noisy) linear combinations of N independent source signals are observed at the sensors.

Convolutive linear mixture model



$$\mathbf{y}(n) = \mathbf{H}(n) \star \mathbf{s}(n)$$

- $\mathbf{y}(n)$: $M \times 1$ observation vector (array output),
- $\mathbf{s}(n)$: $N \times 1$ *unknown* source vector,
- $\mathbf{H}(z) = \sum_n \mathbf{H}(n) z^{-n}$: $M \times N$ *unknown* transfer function matrix assumed, in general, of finite impulse response, i.e. $\deg(\mathbf{H}(z)) = L$.

Basic assumptions

- **System dimension**: We assume here strictly *more sensors than sources*, i.e. $M > N$.
- **Source signals**: They are assumed to be *mutually independent* stationary random processes.
- **System matrix**: The transfer function $\mathbf{H}(z)$ is assumed to be *irreducible* (rank$(\mathbf{H}(z)) = N$ for all z) and *column reduced* (rank$(\mathbf{H}(L)) = N$).

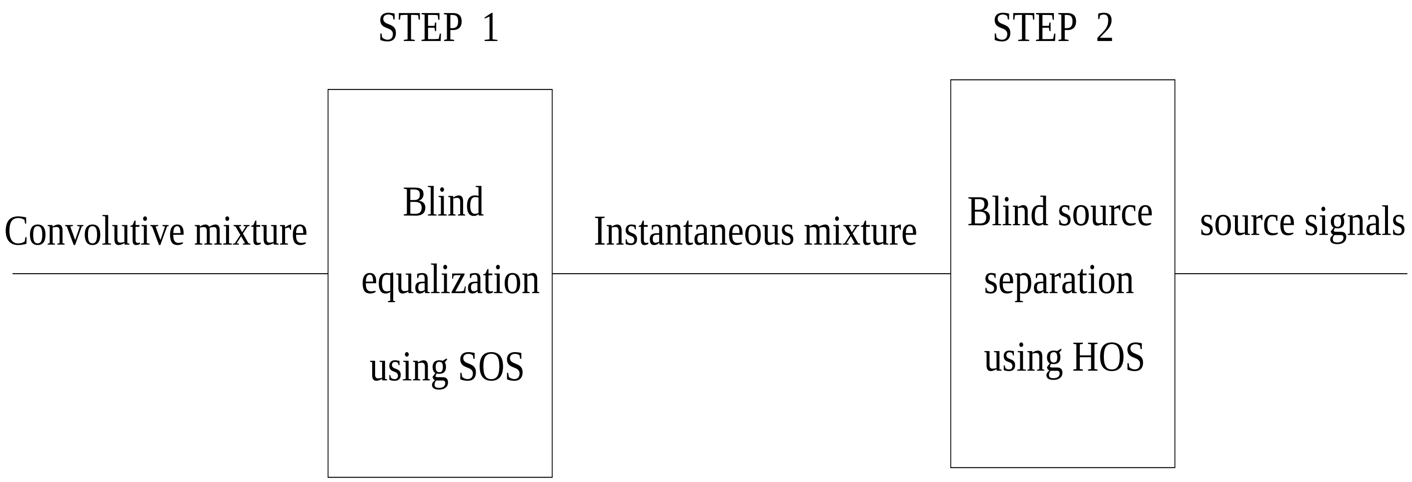
Objectives

- **SIMO case**: In the SIMO case we have to get rid of the inter-symbol interference (ISI) only *(blind equalization problem)*.

- **MIMO case**: In the MIMO case we have to get rid of the ISI *(blind equalization problem)* and to get rid also of the inter-user interferences *(blind source separation (BSS))* .

MIMO blind deconvolution $\Longleftrightarrow$ Blind equalization + BSS.

Deconvolution approach



- **Step 1**: Blind equalization using second order statistics. This step transforms the convolutive mixture into an instantaneous mixture.
- **Step 2**: Application of a BSS algorithm (using, in general, the data HOS) to the instantaneous mixture obtained at the previous step.

Other possible deconvolution approaches

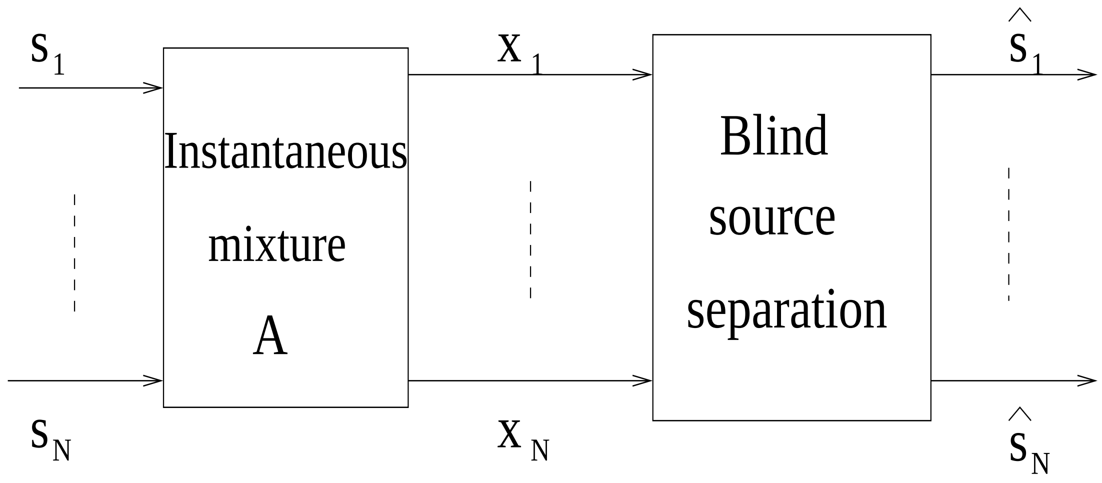
- Blind identification and deconvolution using HOS (Nikias et al 1993, Liu et al 2002, etc.).
- Blind separation followed by M parallel SIMO blind equalization (Bousbia-Salah et al 2000).
- Joint blind equalization and source separation by decorrelation (Y. Hua et al 2000).
- Iterative blind deconvolution with interference cancellation (Delfosse et al 1996).

First step: Blind equalization

The same algorithms (except for certain details) for SIMO blind identification can be applied to MIMO identification.

However, in the SIMO case we estimate the channel transfer function (resp. the source signal) up to a 1×1 constant factor α, i.e. $\hat{\mathbf{h}}(z) = \mathbf{h}(z)\alpha$, while in the MIMO case we estimate the channel transfer function (resp. the source vector) up to a $N \times N$ constant matrix $\mathbf{A}$, i.e. $\hat{\mathbf{H}}(z) = \mathbf{H}(z)\mathbf{A}$.

Second step: Blind source separation



Instantaneous linear mixture model:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$

BSS versus ICA (Independent Component Analysis)

1. *BSS = signal synthesis*: Identify the mixture matrix and/or recover the input signals from the observed signal by exploiting the *statistical independence* or other features of the sources.

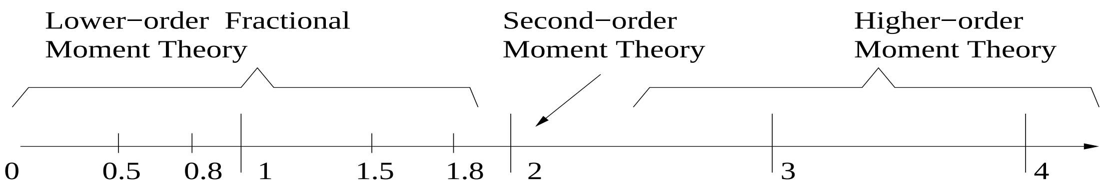
2. *ICA = signal analysis*: Analyse a multi-variate signal by decomposing it into a set of independent components (independent component analysis ICA).

ICA versus PCA

- **Principal component analysis**: seeks directions in feature space that best represent the data in *least squares* sense.

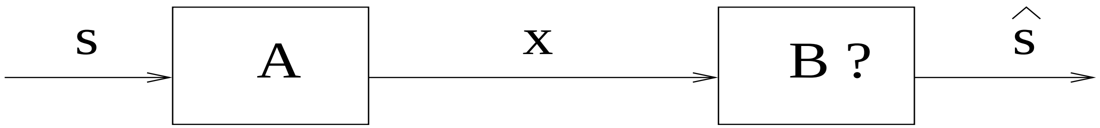
- **Independent component analysis**: seeks directions in feature space that are most *independent* from one another.

BSS approaches



- *HOS-based methods*: Exploit the observations higher order statistics either explicitely by processing their higher order cumulants or implicitely through the optimization of non-linear functions given by information-theoretic criteria.
- *SOS-based methods*: When the sources are ‘temporally colored’, one can achieve BSS using signal decorrelation.
- *FLOM-based methods*: Dedicated to the separation of impulsive signals, e.g. alpha-stable signals (these signals have infinite 2nd and higher order moments).

Information theoretic principles



B is computed such that its outputs are most independent from one another:

- By minimizing the mutual information between the components of $\hat{\mathbf{s}}(t)$.
- By minimizing the Kullbak-Leibler distance in between the pdf of $\hat{\mathbf{s}}(t)$ and the product of its components pdfs, i.e.

$$KL(p(\hat{\mathbf{s}}(t)), \prod_k p_k(\hat{s}_k(t)).$$

- My maximizing the nongaussianity of $\hat{\mathbf{s}}(t)$ (measures of nongaussianity include the Kurtosis -fourth order cumulant- and the Negentropy -differential entropy-).

BSS by decorrelation

Basic assumptions:

- The mixing matrix $\mathbf{A}$ is full column rank.

- The sources are temporally coherent but mutually uncorrelated, i.e.,

$$\begin{aligned}
\mathbf{R}_s(\tau) &\stackrel{\text{def}}{=} E(\mathbf{s}(t+\tau)\mathbf{s}(t)^H) = \begin{bmatrix} \rho_1(\tau) & & 0 \\ & \ddots & \\ 0 & & \rho_n(\tau) \end{bmatrix} \\
\mathbf{R}_x(\tau) &= \mathbf{A}\mathbf{R}_s(\tau)\mathbf{A}^H
\end{aligned}$$

Separation by decorrelation

- *Principle*: $\mathbf{B} = \mathbf{A}^{-1}$ is the linear transform that decorrelate the signal components at all time lags, i.e.

$$\mathbf{B}\mathbf{R}_x(\tau)\mathbf{B}^H = \mathbf{R}_s(\tau)$$

is diagonal for all τ.

- *A two step procedure*:
  - Data *whitening*: The whitening matrix transforms $\mathbf{A}$ into a unitary matrix.
  - *Diagonalization*: Estimate the unitary matrix by diagonalizing the non-zero lag correlation matrices.

Whitening

Whitening Matrix: Let $\mathbf{W}$ denotes a $n \times m$ matrix, such that

$$(\mathbf{WA})(\mathbf{WA})^H = \mathbf{UU}^H = \mathbf{I}$$

$\mathbf{W}$ can be computed as an inverse square root of covariance matrix of the observation vector (assuming unit-power sources).

Whitened correlations: Defined as

$$\underline{\mathbf{R}}_x(\tau) \quad = \quad \mathbf{W}\mathbf{R}_x(\tau)\mathbf{W}^H = \mathbf{U}\mathbf{R}_s(\tau)\mathbf{U}^H$$

Diagonalization

- Diagonalization of **one** single normal matrix $\mathbf{M}$
$\Longleftrightarrow$ Minimizing under unitary transform the sum of squared moduli of the off-diagonal elements. This is equivalent to the maximization under unitary transform $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_n]$ the sum of the squared moduli of the diagonal elements:

$$C(\mathbf{M}, \mathbf{V}) = \sum_i |\mathbf{v}_i^* \mathbf{M} \mathbf{v}_i|^2$$

- For **a set** of d matrices:

$$C(\mathbf{V}) = \sum_{k=1}^{d} C(\mathbf{M}_k, \mathbf{V}) = \sum_{k,i} |\mathbf{v}_i^* \mathbf{M}_k \mathbf{v}_i|^2$$

$\Longrightarrow$ **Joint diagonalization criterion**.

Identifiability

Objectives: Given a set of K correlation matrices $\mathbf{R}_x(\tau_1), \cdots, \mathbf{R}_x(\tau_K)$ answer the following:

- Is it possible to separate the sources given this statistics?
- If no, what it the best we can do (partial identifiability)?
- Is it possible to test the identifiability condition?

Theorem 1: Identifiability

- Define for each source i

$$\boldsymbol{\rho}_i \quad = \quad [\rho_i(\tau_1), \rho_i(\tau_2), \cdots, \rho_i(\tau_K)] \quad \text{and} \quad \tilde{\boldsymbol{\rho}}_i = [\Re(\boldsymbol{\rho}_i), \Im(\boldsymbol{\rho}_i)]$$

Then, BSS can be achieved using the output correlation matrices at time lags $\tau_1, \tau_2, \cdots, \tau_K$ iff $\forall \;\; i \neq j$

$$\tilde{\boldsymbol{\rho}}_i \textit{ and } \tilde{\boldsymbol{\rho}}_j \textit{ are } \underline{(\textit{pairwise})} \textit{ linearly independent}$$

- If this condition is satisfied then $\mathbf{B}$ is a separating matrix iff $\forall \;\; i \neq j$

$$r_{ij}(k) = 0 \quad \textit{and} \quad \sum_{k=\tau_1}^{\tau_K} |r_{ii}(k)| > 0 \tag{1}$$

where $r_{ij}(k) \stackrel{\text{def}}{=} E(z_i(t+k)z_j^*(t))$, $\mathbf{z} = \mathbf{B}\mathbf{x}$ and $k = \tau_1, \tau_2, \cdots, \tau_K$.

Discussion

- Theorem 1 gives a **necessary** **and sufficient** condition to achieve BSS.
- It is possible to separate the sources from only **one** correlation matrix.
- $K \longrightarrow \infty \Longrightarrow$ 2 sources are separable iff they have **different spectral shape**.
- It is well known that HOS methods can achieve BSS when no more than one Gaussian source is present. In contrast, SOS methods can achieve BSS when no more than one temporally white source is present.

Theorem 2: Partial Identifiability

Assume there are d distinct groups of sources each of them containing d_i source signals with same (up to a scalar) correlation vector $\tilde{\boldsymbol{\rho}}_i$, $i = 1, \cdots, d$, i.e., $\mathbf{s} = [\mathbf{s}_1^T, \cdots, \mathbf{s}_d^T]^T$.

Let $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$ be an $m \times 1$ random vector satisfying equation (1).

Then, there exists a permutation matrix $\mathbf{P}$ and non-singular matrices $\mathbf{U}_i$ such that

$$\begin{aligned} \mathbf{P}\mathbf{z}(t) &= [\mathbf{z}_1^T(t), \cdots, \mathbf{z}_d^T(t)]^T \\ \mathbf{z}_i(t) &= \mathbf{U}_i \mathbf{s}_i(t) \end{aligned}$$

Moreover, sources belonging to the same group, i.e., having same (up to a scalar) correlation vector $\tilde{\boldsymbol{\rho}}_i$ can not be separated using only the correlation matrices $\mathbf{R}_x(k),\ k = \tau_1, \cdots, \tau_K$.

Theorem 3: Testing of Identifiability Condition

Let $\tau_1 < \tau_2 < \cdots < \tau_K$ be K distinct time lags and $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$. Assume that $\mathbf{B}$ is such a matrix that $\mathbf{z}(t)$ satisfies equation (1). Then there exists a generalized permutation matrix $\mathbf{P}$ such that for $k = \tau_1, \cdots, \tau_K$:

$$\mathbf{R}_z(k) = E(\mathbf{z}(t+k)\mathbf{z}^H(t)) = \mathbf{P}\mathbf{R}_s(k)\mathbf{P}^T$$

In other words, $z_1, \cdots, z_m$ have the same (up to a permutation) correlation factors as $s_1, \cdots, s_m$ at time lags $\tau_1, \cdots, \tau_K$.

Discussion

- Two situations may happen:
  1. For all pairs (i, j), $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ are pairwise linearly independent. Then we are sure that the sources have been separated and that $\mathbf{z}(t) = \mathbf{s}(t)$ up to a scalar and a permutation.
  2. A few pairs (i, j) out of all pairs satisfy $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ linearly dependent. Therefore the sources have been separated in blocks.
- The angle between $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ can be used as a measure of the quality of separation between source i and source j.

Simulation Examples

- **Simulation context**:
  - ULA with $M = 5$ sensors, $N = 2$ unit-norm independent sources and $T = 1000$ samples.
- **Criteria**:
  - Rejection level criterion:

$$\mathcal{I}perf_i \stackrel{\text{def}}{=} \sum_{j \neq i} E \left\{ \frac{\rho_j(0) |(\hat{\mathbf{B}}\mathbf{A})_{ij}|^2}{\rho_i(0) |(\hat{\mathbf{B}}\mathbf{A})_{ii}|^2} \right\}$$

  - Identifiability criterion:

$$\vartheta_{\boldsymbol{\rho}} \stackrel{\text{def}}{=} \left| \frac{|\tilde{\boldsymbol{\rho}}_1 \tilde{\boldsymbol{\rho}}_2^T|}{\|\tilde{\boldsymbol{\rho}}_1\| \|\tilde{\boldsymbol{\rho}}_2\|} - 1 \right|$$

Simulation examples

Table 1. Separation performance versus $\vartheta_{\boldsymbol{\rho}}$.

Sources	$\vartheta_{\boldsymbol{\rho}}$	$\mathcal{I}perf$ (dB)
2 AR1 signals	0.213	-26.23
2 CWGP signals	0.007	-5.14

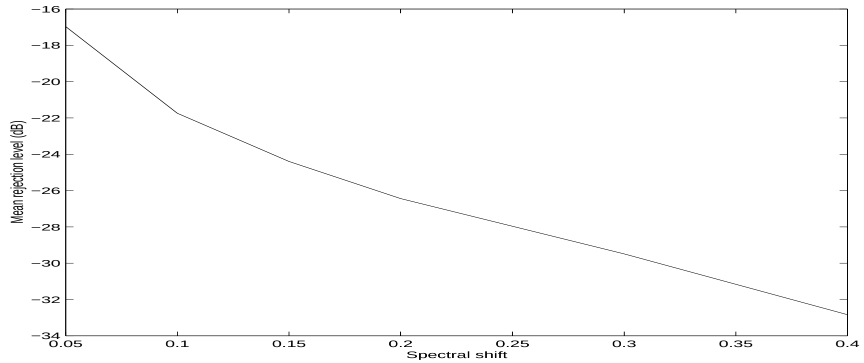


Figure 1. Separation Performance versus $\delta\theta$ (SNR=25dB): 2 AR1 sources with $a_1 = 0.95 \exp(j0.5)$ and $a_2 = 0.5 \exp(j(0.5 + \delta\theta))$.

Simulation examples

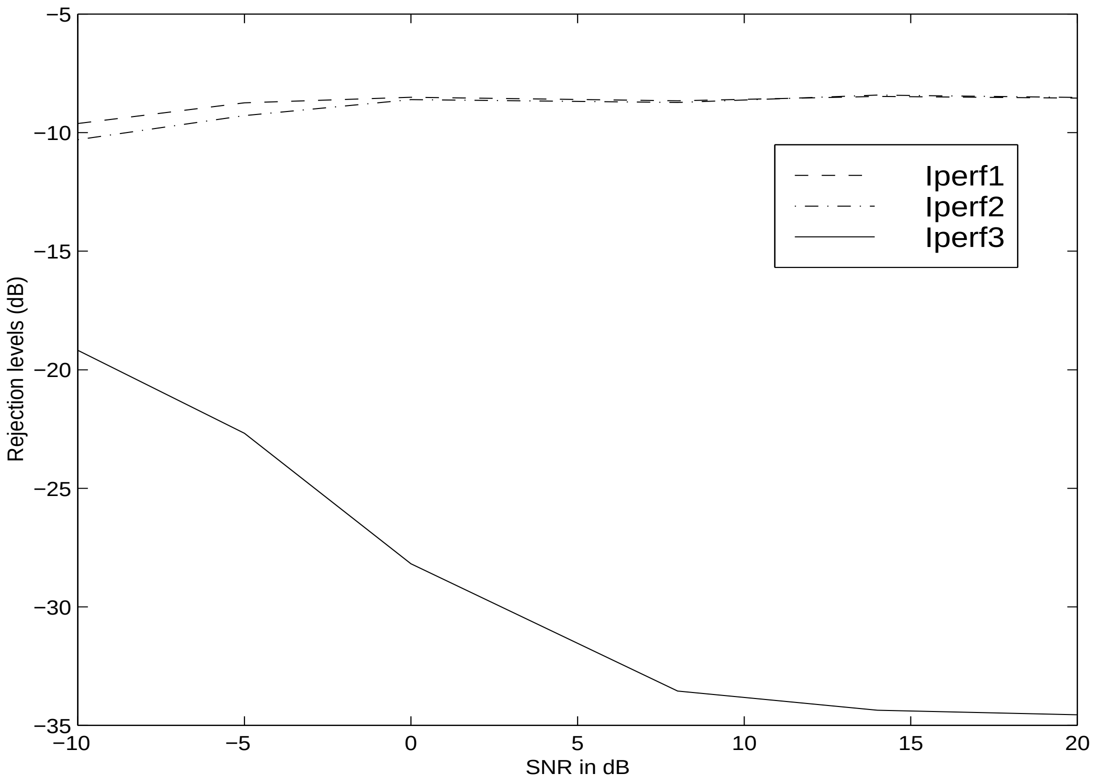


Figure 2. Separation Performance versus the SNRs: $m = 3$ sources 2 of them are CWGP signals and the third is AR1..

Simulation examples

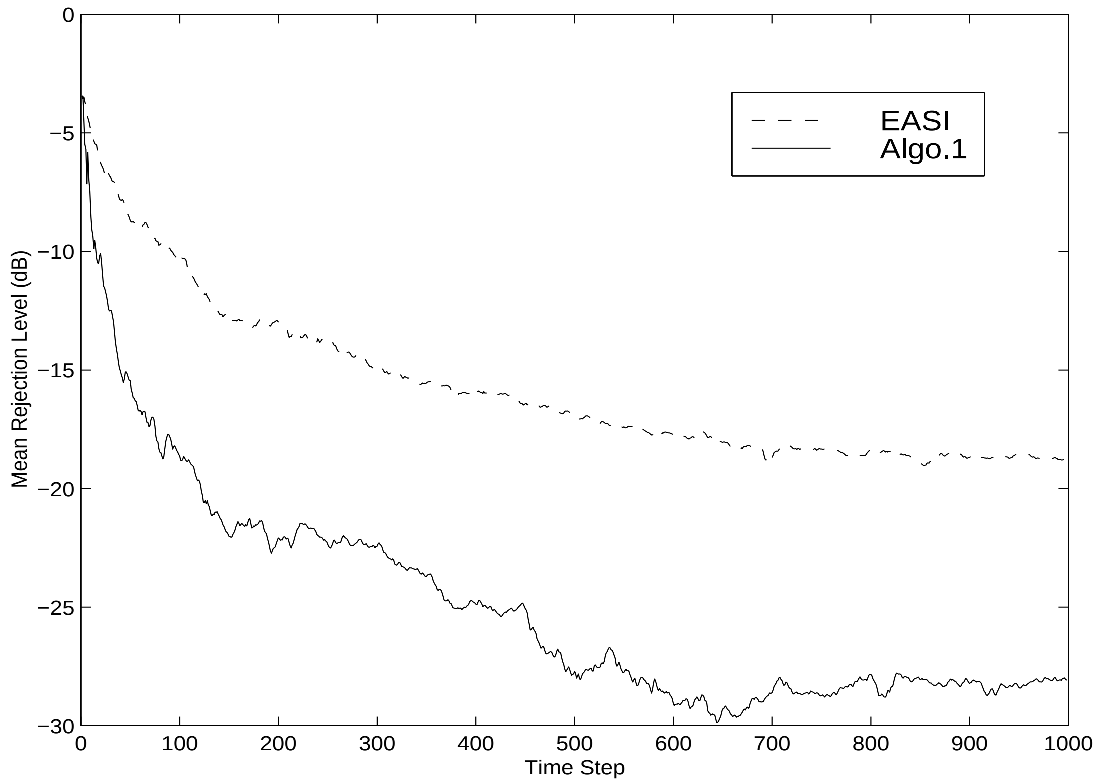


Figure 5. Comparison with EASI (Laheld & Cardoso 1996): 2 AR1 sources with QAM4 innovation processes & $SNR = 30dB.$.

Conclusions ...

Concluding remarks

- A common feature of all presented methods is the use of time and/or space **diversity**.
- **Extension** to IIR case or multiple input case is possible.
- **Partial knowledge** of the channels can be incorporated in the blind criteria, e.g., DOA of multi-paths, pulse-shape filters, spreading sequence in CDMA systems, etc.
- **Robustness** to channel order estimation errors: The last 3 methods (SS, MRE, LP) are more robust than the first 3 methods (ML, CR, CS).

Some hot topics & perspectives

- **Semi-blind** methods: i.e., combining blind and non-blind criteria (i.e., training sequence) to improve the estimation accuracy.

- **Induced cyclostationarity or pre-filtering**: i.e., modify the signal modulation at the transmission side in such a way to suit and simplify the blind system identification (BSI).

- **Space time coding**: BSI is a tool to exploit the diversity at the reception. The space-time coding is to create the diversity at the transmission.

Some hot topics & perspectives

- **Application-oriented BSI methods**: Derive or adapt blind system identification (BSI) methods for specific applications (this allows to exploit a maximum of side-information).

- **Robustness**: Improve the robustness of BSI methods against noise and modellization errors.

- **Under-determined case**: BSI for systems with more sources than sensors.