Short Overview on Blind Equalization

Philippe Ciblat

Télécom ParisTech
Outline

1. Introduction
   - General problem
   - Problem classification
   - Considered problem

2. Statistical framework

3. High-Order Statistics (HOS) algorithms
   - Constant Modulus Algorithm (CMA)
   - Adaptive versions

4. Second-Order Statistics (SOS) algorithms
   - Covariance matching algorithm
   - (Deterministic) Maximum-Likelihood algorithm
   - Some sub-optimal algorithms

5. Other types of algorithms

Part 1: Introduction
General model

Unknown signal mixture with additive noise

\[ y(n) = \text{fct}(s(n)) + w(n) \]  \hspace{1cm} (1)

with

- \( y(n) \): observations vector at time-index \( n \)
- \( w(n) \): white zero-mean Gaussian noise

Find out the multi-variate input \( s(n) \) given

- only a set of observations \( y(n) \)
- statistical model for the noise

Blind techniques

Unknown \text{fct} without deterministic help of \( s(n) \) to estimate it
Problem classification

- \( s(n) \) belongs to a discrete set: **equalization**
  - Military applications: passive listening
  - Civilian applications: no training sequence
    - Goal 1: remove the header and increase the data rate (be careful: with the same raw data rate)
    - Goal 2: follow very fast variation of wireless channel (be careful: set of observations is small)

- \( s(n) \) belongs to a uncountable set: **source separation**
  - Audio (cocktail party)
    - The cocktail party effect is the phenomenon of being able to focus one’s auditory attention on a particular stimulus while filtering out a range of other stimuli, much the same way that a partygoer can focus on a single conversation in a noisy room.
  - Hyperspectral imaging
  - Cosmology (Cosmic Microwave Background map with Planck data)
In the context of **Blind Source Separation (BSS)**:

- **Instantaneous mixture:**
  \[ y(n) = Hs(n) + w(n) \]
  with a unknown matrix \( H \)

- **Convolutive mixture:**
  \[ y(n) = \sum_{\ell=0}^{L} H(\ell)s(n - \ell) + w(n) \]
  with a unknown set of matrices \( H(\ell) \)

- **Nonlinear mixture:** \textbf{fct} is not linear

**BSS field**

- Vast community mainly working on the instantaneous case
- Goal: find out \( s(n) \) up to scale and permutation operators

Philippe Ciblat (Télécom ParisTech)

Short overview on Blind Equalization
Considered Problem

Go back to **equalization** (done in blindly manner)

Unlike BSS, sources are strongly structured:
- discrete set (often a lattice, i.e., $\mathbb{Z}$-module)
- discrete set with specific properties: constant modulus if PSK
- man-made source (can be even modified to help the blind equalization step)

**Classification problem rather than Regression problem**

**First questions**
- Do we have a Input/Output model given by Eq. (1)?
- If yes, what is the shape of the mixture given by *fct*?
Signal model

- Single-user context
- Single-antenna context
- Multipath propagation channel

Equivalent discrete-time channel model (by sampling EM wave at the symbol rate)

\[ y(n) = \sum_{\ell} h(\ell) s(n - \ell) + w(n), \forall n = 0, \ldots, N - 1 \iff y = Hs + w \]

where \( H \) is a band-Toeplitz matrix, \( N \) is the frame size
Signal model (cont’d)

Sampling at symbol rate leads to

- no information loss on the symbol sequence
- but information loss on the electro-magnetic wave, and probably on the channel impulse response (our goal, here)

Go back to the “true” receive signal...

\[ y(t) = \sum_{k} s(k) h(t - kT_s) + w(t), \quad \forall t \in \mathbb{R} \]

with

- \( s(k) \): symbol sequence
- \( w(t) \): white Gaussian noise
- \( h(t) \): filter coming from the channel and the transmitter

occupied band = \[ \left[ -\frac{1 + \rho}{2T_s}, \frac{1 + \rho}{2T_s} \right] \]

with the roll-off factor \( \rho \in (0, 1] \)
Signal framework

Shannon-Nyquist sampling theorem \( \Rightarrow T = \frac{T_s}{2} \)

- Scalar framework: no filtering anymore
  \[ \tilde{y}(n) = y(nT) = \sum_{k} s(k)h(nT_s/2 - kT_s) + \tilde{w}(n) \]

- Vector framework: SIMO filtering
  \[
  \begin{align*}
  y_1(n) &= y(nT_s) = h_1 * s(n) + w_1(n) \\
  y_2(n) &= y(nT_s + T_s/2) = h_2 * s(n) + w_2(n)
  \end{align*}
  \]

with \( h_1(n) = h(nT_s) \) and \( h_2(n) = h(nT_s + T_s/2) \)
Problems to be solved

Goals

Estimate

1. **Scalar case**: $h_1$ given $y_1(n)$ only and $h_2$ given $y_2(n)$ only, i.e., working with model of Slide 7

2. **Vector case**: $h = [h_1, h_2]^T$ given $y(n) = [y_1(n), y_2(n)]^T$ jointly

**Glossary:**

- without training sequence
  - $\Rightarrow$ Non-Data-aided (NDA) or blind/unsupervised
- with training sequence
  - $\Rightarrow$ Data-aided (DA) or supervised
- with decision-feedback
  - $\Rightarrow$ Decision-Directed (DD)
Part 2: Statistical framework
Available data statistics

- Only \( \{y(n)\}_{n=0}^{N-1} \) is available to estimate \( H \)
- What is an algorithm here? a function depending only on \( \{y(n)\}_{n=0}^{N-1} \) ...

... a statistic of the random process \( y(n) \)

\[ \Theta \left( \{y(n)\}_{n=0}^{N-1} \right) \]

Choice of \( \Theta \):
- \( P \)-order polynomial: moments of the random process
  Question: which orders are relevant? listen to the talk
- A Deep Neural Network (DNN)
  Question: how calculating the weights? see Slide 37
A not-so toy example

\[ y(n) = Hs(n) + w(n) \]

with

- \( y(n) \) is a vector of length \( L \)
- \( H \) is a \( L \times L \) square full rank matrix
- \( s(n), w(n) \) are i.i.d. circularly-symmetric Gaussian vectors with zero-mean and variances \( \sigma_s^2 \) and \( \sigma_w^2 \) respectively

Results

- \( y(n) \) Gaussian with zero-mean and correlation matrix
  \[ R(H) = \sigma_s^2 HH^H + \sigma_w^2 \text{Id}_L \]
- \( R(H) = R(HU) \) for any unitary matrix \( U \)
- Principal Component Analysis (PCA) is a deadlock

\( s(n) \) has to be non-Gaussian

⇒ Independent Component Analysis (ICA)
Scalar case

Go back to blind equalization

\[ y(n) = h \ast s(n) + w(n) \]

As \( y(n) \) is stationary, second-order information lies in

\[ S(e^{2i\pi f}) = \sum_m r(m)e^{-2i\pi fm} = \sigma_s^2|h(e^{2i\pi f})|^2 + \sigma_w^2 \]

with

- \( r(m) = \mathbb{E}[y(n + m)y(n)] \)
- \( h(\zeta) = \sum_\ell h(\ell) \zeta^{-\ell} \), with \( \zeta = e^{2i\pi f} \)

Results

- Lack of information on the channel impulse response, except if
  - \( h(\zeta) \) is phase minimum (\( h(\zeta) \neq 0 \) if \( |\zeta| > 1 \))
  - non-stationary signal
  - non-Gaussian signal (by resorting to high-order statistics) : OK for PAM, PSK, QAM sources
Scalar case: the pavement of the HOS road

Let \( X = [X_1, \ldots, X_N] \) be a real-valued random vector of length \( N \).

### Characteristic function of the first kind (MGF)

\[
\Psi_X : \omega \mapsto \mathbb{E}[e^{i\omega^T X}] = \int p_X(x)e^{i\omega^T x} dx
\]

**Moments (of order s)** \( \propto \) component of Taylor series expansion of \( \Psi_X \) for \( s \)-th order

**Example**: \( N = 2 \); Second-order means \( \mathbb{E}[X_1^2], \mathbb{E}[X_2^2] \), and \( \mathbb{E}[X_1 X_2] \)

### Characteristic function of the second kind (CGF)

\[
\Phi_X : \omega \mapsto \log(\Psi_X(\omega))
\]

**Cumulants (of order s)** \( \propto \) component of Taylor series expansion of \( \Phi_X \) for \( s \)-th order
Useful properties

- Why cumulants? Let $X$ and $Y$ be independent vectors
  \[ \Psi_{X+Y}(\omega) = \Psi_X(\omega_1) \cdot \Psi_Y(\omega_2) \text{ but } \Phi_{X+Y}(\omega) = \Phi_X(\omega_1) + \Phi_Y(\omega_2) \]

- $X = [X_1, \cdots, X_N]$ and $Y = [Y_1, \cdots, Y_N]$ be independent vectors
  \[ \text{cum}_s(X_{i_1} + Y_{i_1}, \cdots, X_{i_s} + Y_{i_s}) = \text{cum}_s(X_{i_1}, \cdots, X_{i_s}) + \text{cum}_s(Y_{i_1}, \cdots, Y_{i_s}) \]

- $X = [X_1, \cdots, X_N]$ with at least two independent components
  \[ \text{cum}_N(X_1, \cdots, X_N) = 0 \]

- $X = [X_1, \cdots, X_N]$ Gaussian vector
  \[ \text{cum}_s(X_{i_1}, \cdots, X_{i_s}) = 0 \text{ if } s \geq 3 \]

Remarks

- No HOS information for Gaussian vector
- “Distance” to the Gaussian distribution $\Rightarrow$ (normalized) Kurtosis
  \[ \kappa_X = \frac{\text{cum}_4(x, \bar{x}, x, \bar{x})}{(\mathbb{E}[|x|^2])^2} \]
Fourth-order information: the trispectrum

\[ S_4(e^{2i\pi f_1}, e^{2i\pi f_2}, e^{2i\pi f_3}) = \sum_{m_1, m_2, m_3} \text{cum}_4(m_1, m_2, m_3) e^{-2i\pi(f_1m_1 + f_2m_2 + f_3m_3)} \]

\[ = \kappa_S h(e^{2i\pi f_1})h(e^{2i\pi f_2})h(e^{2i\pi f_3})h(e^{2i\pi(-f_1 + f_2 + f_3)}) \]

with \( \text{cum}_4(m_1, m_2, m_3) = \text{cum}(y(n), y(n + m_1), y(n - m_2), y(n - m_3)) \)

Remarks

- Trispectrum provides information enough on channel impulse response
- **Question**: how carrying out algorithms using it (see Part 3)
Vector case

Go back to the signal model

\[ y(n) = h \ast s(n) + w(n) \]

with \( y(n) = [y_1(n), y_2(n)]^T \) and \( h(n) = [h_1(n), h_2(n)]^T \)

Reminder: oversampling or symbol rate sampling with two RX

As \( y(n) \) is stationary, second-order information lies in

\[ S(e^{2i\pi f}) = \sum_m R(m)e^{-2i\pi fm} = \sigma_s^2 h(e^{2i\pi f})h(e^{2i\pi f})^H \]

with \( R(m) = \mathbb{E}[y(n + m)y(n)^H] \) and \( h(e^{2i\pi f}) = \sum_\ell h(\ell)e^{-2i\pi f\ell} \)

Results

- Unique solution if \( h(\tilde{z}) \) is phase minimum (\( h(\tilde{z}) \neq 0 \) if \( |\tilde{z}| > 1 \))
- Unrestrictive assumption since often \( h_1(\tilde{z}) \neq h_2(\tilde{z}), \forall \tilde{z} \), i.e., no common root, i.e., \( h_1(\tilde{z}) \) and \( h_2(\tilde{z}) \) are prime jointly
- Information enough on channel impulse response
Vector case: a cyclostationarity point-of-view

Go back to the continuous-time signal model

\[ y(t) = \sum_{k} s(k) h(t - kT_s) + w(t) \]

Its autocorrelation is periodic with period \( T_s \)

\[ t \mapsto r(t, \tau) = \mathbb{E} \left[ y(t + \tau) y(t) \right] \]

Result

- \( \tilde{y}(n) \) cyclostationary with period \( (T_s / T) = 2 \)
- By denoting \( \tilde{s} = (s(0), 0, s(1), 0, \cdots) \), we have

\[ \tilde{y}(n) = \tilde{h} \star \tilde{s}(n) \]

Remark: Cyclostationary discrete-time signal with period 1 is stationary
Cyclostationary second-order information

Fourier series expansion of the correlation:

\[ n \mapsto r(n, m) = \mathbb{E}[\tilde{y}(n + m)\tilde{y}(n)] = r^{(0)}(m) + r^{(1/2)}(m)e^{2i\pi(1/2)n} \]

- \( \alpha \in \{0, 1/2\} \) : cyclic frequencies
- \( \{r^{(\alpha)}(m)\}_m \) : set of cyclic correlation at cyclic frequency \( \alpha \)
- \( S^{(\alpha)}(e^{2i\pi f}) = \sum m r^{(\alpha)}(m)e^{-2i\pi fm} \) : cyclic spectrum at cyclic frequency \( \alpha \)

Results

\[ S^{(0)}(e^{2i\pi f}) = \sigma_s^2|\tilde{h}(e^{2i\pi f})|^2, \quad S^{(1/2)}(e^{2i\pi f}) = \sigma_s^2\tilde{h}(e^{2i\pi f})\tilde{\tilde{h}}(e^{2i\pi (f + 1/2)}) \]

- Cyclic spectra provide information enough on channel impulse response
- **Question**: how carrying out algorithms using it (see Part 4)
Take-home message

Sampling at $T_s$

Sampling at $T_s/2$

stationary SISO (ante 1991)
\[ y = h \ast s \]

Separate processing

HOS algorithms

SOS algorithms

Stationary SIMO
\[ y = h \ast s \]

Cyclostationary SISO
\[ \tilde{y} = \tilde{h} \ast \tilde{s} \]
Part 3: High-Order Statistics based Algorithms
Usually the algorithms rely on blind deconvolution principle, i.e., retrieving the symbol sequence \( \{s(n)\}_n \) directly from \( \{y(n)\}_n \).

Talk done with the stationary SISO model

\[
\min_P \mathbb{E} [f(z(n))]
\]

with

- \( z(n) = p \star y(n) \)
- \( p \) the equalizer filter
- \( f \) a nonlinear and nonquadratic function
Some algorithms

**Sato Algorithm** [Sato1975]

\[ J = \mathbb{E} \left[ (z(n) - \text{sign}(z(n)))^2 \right] \]

**Constant Modulus Algorithm (CMA)** [Godard1980]

\[ J = \mathbb{E} \left[ (|z(n)|^2 - C)^2 \right] \]

with \( C = \mathbb{E}[|s(n)|^4]/\mathbb{E}[|s(n)|^2] \)

**Kurtosis Minimization (KM)** [ShalviWeinstein1990]

\[ J = |\kappa_z| \]
Implementation issue

How finding the minimum of $J(p) = \mathbb{E}[J_n(p)]$?

**Blockwise processing**

Block of size $N$

$\hat{J}_N(p) = \frac{1}{N} \sum_{n=1}^{N} J_n(p)$

We replace $J(p)$ with $\hat{J}_N(p)$

**Adaptive processing**

We replace $J(p)$ with $J_n(p)$ at time/iteration $n$

- LMS
- Newton

**Gradient algo.**

$p_{i+1} = p_i - \mu \frac{\partial \hat{J}_N(p)}{\partial p} |_{p_i}$

**(Stochastic) Gradient algo.**

$p_{n+1} = p_n - \mu \frac{\partial J_n(p)}{\partial p} |_{p_n}$
Application to CMA

Adaptive implementation

\[
\mathbf{p}_{n+1} = \mathbf{p}_n - \mu \overline{\mathbf{y}_{L_p}(n) z(n)(|z(n)|^2 - C)} \\
= \mathbf{p}_n - \mu \overline{\mathbf{y}_{L_p}(n)(z(n) - F_{cma}(z(n)))}
\]

with

- \( \mathbf{y}_{L_p}(n) = [y(n), \ldots, y(n - L_p)]^T \)
- \( F_{cma}(z(n)) = z(n)(1 + C - |z(n)|^2) \)
- if KM, \( F_{km}(z(n)) = z(n)(1 + \text{sgn}(\kappa_s)|z(n)|^2) \)

Special case: training sequence (known \( s(n) \))

\[
J = \mathbb{E}[|z(n) - s(n)|^2]
\]

Adaptive implementation

\[
\mathbf{p}_{n+1} = \mathbf{p}_n - \mu \overline{\mathbf{y}_{L_p}(n)(z(n) - s(n))}
\]

- \( s(n) \) may be replaced by \( \hat{s}(n) \) after initial convergence (DD)
- \( s(n) \) is replaced by \( F(z(n)) \) which plays the role of “training”
Adaptive trained equalizer scheme

Adaptive blind equalizer scheme
Part 4: Second-Order Statistics based Algorithms
Usually the algorithms rely on blind identification principle, i.e., retrieving the filter $h = [h(0)^T, \cdots, h(L)^T]^T$

Talk done with the stationary SIMO model

$$
\begin{bmatrix}
    y(n) \\
    \vdots \\
    y(n-N)
\end{bmatrix}
= 
\begin{bmatrix}
    h(0) & \cdots & h(L) & \cdots & 0 \\
    \vdots & \ddots & \vdots & \ddots & \vdots \\
    0 & \cdots & h(0) & \cdots & h(L)
\end{bmatrix}
\begin{bmatrix}
    s(n) \\
    \vdots \\
    s(n-N-L)
\end{bmatrix}
$$

with $\mathcal{T}(h)$ a $2(N+1) \times (N+L+1)$ Sylvester matrix

**Result**

If $h(\zeta) \neq 0, \ \forall \zeta$ and $N > L$, then $\mathcal{T}(h)$ is full column rank and left-invertible
Covariance matrix algorithm

**Question:** what is the best second-order algorithm?

Let

1. \( \mathbf{R}(h) = \mathbb{E}[\mathbf{Y}_N(n)\mathbf{Y}_N(n)^H] \) and \( \hat{\mathbf{R}}_{N_{\text{obs}}} = \frac{1}{N_{\text{obs}}} \sum_{n=0}^{N_{\text{obs}}-1} \mathbf{Y}_N(n)\mathbf{Y}_N(n)^H \)
2. \( \mathbf{r}(h) = [\Re\{\text{vec}(\mathbf{R}(h))\}, \Im\{\text{vec}(\mathbf{R}(h))\}]^T \)
3. \( \hat{\mathbf{r}}_{N_{\text{obs}}} = [\Re\{\text{vec}(\hat{\mathbf{R}}_{N_{\text{obs}}})\}, \Im\{\text{vec}(\hat{\mathbf{R}}_{N_{\text{obs}}})\}]^T \)

**Result**

\[
\sqrt{N_{\text{obs}}} (\hat{\mathbf{r}}_{N_{\text{obs}}} - \mathbf{r}(h)) \xrightarrow{D} \mathcal{N}(0, \Gamma_h),
\]

i.e.,

\[
\hat{\mathbf{r}}_{N_{\text{obs}}} \approx \mathbf{r}(h) + \mathbf{w}_{N_{\text{obs}}}
\]

with \( \mathbf{w}_{N_{\text{obs}}} \) zero-mean Gaussian noise with covariance matrix \( \Gamma_h / N_{\text{obs}} \)
Covariance matching algorithm (cont’d)

Maximum-Likelihood based on $\hat{r}_{N_{\text{obs}}}$ instead of data $Y = Y_{N_{\text{obs}}}(N_{\text{obs}})$

$$\frac{1}{N_{\text{obs}}} \log(p(\hat{r}_{N_{\text{obs}}}|h)) \approx -(\hat{r}_{N_{\text{obs}}} - r(h))^T \Gamma_h^{-1} (\hat{r}_{N_{\text{obs}}} - r(h))$$

$$- \frac{\log(\det(\Gamma_h))}{2N_{\text{obs}}} + \text{constant}$$

Result

$$\hat{h}_{\text{cm}} = \arg\min_h \left\| \Gamma_h^{-\frac{1}{2}} (\hat{r}_{N_{\text{obs}}} - r(h)) \right\|^2$$

with $\left\| W^{\frac{1}{2}} x \right\|^2 = x^H W x$

Ping-pong procedure for update $\Gamma_h$
Question: Maximum Likelihood based on \( Y \)

\[
Y = \mathcal{T}(h)S + W
\]

with \( W \) white zero-mean Gaussian noise and unknown \( S \)

<table>
<thead>
<tr>
<th>Method</th>
<th>Equation</th>
<th>Description</th>
</tr>
</thead>
</table>
| TRUE ML | \[
\max_h p(Y|h) = \int p(Y|h, S)p(S)dS
\] | almost always untractable |
| GAUSSIAN ML | \[
\max_h p(Y|h) = \int p(Y|h, S)e^{-S^H\Gamma^{-1}S}dS
\] | tractable but not optimal |
| DETERMINISTIC ML | \[
\max_{h,S} p(Y|h, S)
\] | tractable but not optimal |

**Deterministic Maximum Likelihood**

\[
(\hat{h}, \hat{S})_{ML} = \arg \min_{h,S} \| Y - \mathcal{T}(h)S \|^2
\]
Maximum Likelihood algorithm (cont’d)

- Minimization on $S$ (without constraint):
  \[
  \hat{S}_{ML} = (T(h)^H T(h))^{-1} T(h)^H Y
  \]

- Then minimization on $h$:
  \[
  \hat{h}_{ML} = \arg \min_h \| (\text{Id} - T(h)(T(h)^H T(h))^{-1} T(h)^H) Y \|_2^2
  \]
  with $P_h^\perp$ the projection on $\text{sp}(T(h))^\perp$

  \[
  \hat{h}_{ml} = \arg \max_h h^H Y^H (T(h)^H T(h))^{-1} Y h
  \]

- Quadratic cost function / $Y \Rightarrow$ Second order is fine
- Non-quadratic cost function / $h \Rightarrow$ Ping-pong procedure
Subspace algorithm: principle

Signal model:
\[ y(n) = A(\theta)s(n) \]

Main required property:
\[ \text{sp}(A(\theta)) = \text{sp}(A(\theta')) \iff \theta = \theta' \]

Algorithm main step:
\[ \hat{\theta} = \arg \min_{\theta} \text{distance}(\text{sp}(y(n)), \text{sp}(A(\theta))) \]

Example 1: source localization (MUSIC)

\[ A(\theta) = [a(\theta_1), \ldots, a(\theta_p)] \]

with
- \( a(\theta) = [1, e^{2i\pi\theta}, \ldots, e^{2i\pi(M-1)\theta}]^T \) (steering vector)
- \( M > p \)
Subspace algorithm: application to blind equalization

\[ Y_N(n) = T(h)S_{N+L}(n), \]

i.e.,

\[ A \leftrightarrow T(h) \quad \text{and} \quad \theta \leftrightarrow h \]

**Result**

Let \( T(h') \) be a Sylvester matrix associated with \( h' \)

If \( N \geq L \) and \( h(z) \neq 0 \ \forall z \in \mathbb{C} \), then

\[ \text{sp}(T(h')) = \text{sp}(T(h)) \iff h' = \alpha h \]

up to a constant \( \alpha \)

**Proof:** using rational space or \( \mathbb{C}[X] \)-module
**Subspace algorithm: practical implementation**

White source \( \Rightarrow \mathbf{R} = \mathbb{E}[\mathbf{Y}\mathbf{Y}^H] = \mathcal{T}(\mathbf{h})\mathcal{T}(\mathbf{h})^H \Rightarrow \text{sp}(\mathbf{R}) = \text{sp}(\mathcal{T}(\mathbf{h})) \)

- Let \( \Pi \) be the projector on \( \text{Ker}(\mathbf{R}) \Rightarrow \Pi \mathbf{x} = 0 \quad \text{iff} \quad \mathbf{x} \in \text{sp}(\mathbf{R}) \)

- Then \( \mathbf{h} \) is the unique vector such that \( \Pi \mathcal{T}(\mathbf{h}) = 0 \)

- In practice, \( \mathbf{R} \) (resp. \( \Pi \)) is estimated by \( \hat{\mathbf{R}} \) (resp. \( \hat{\Pi} \)).

\[
\hat{\mathbf{h}}_{ss} = \arg \min_{\|\mathbf{h}\|=1} \|\hat{\Pi} \mathcal{T}(\mathbf{h})\|^2 = \arg \min_{\|\mathbf{h}\|=1} \mathbf{h}^H \mathbf{Q} \mathbf{h}
\]
If \( h_1(\hat{z}) \) and \( h_2(\hat{z}) \) have no common root, Bezout’s theorem holds: \( \exists [g_1(\hat{z}), g_2(\hat{z})] \) polynomials such that \( g_1(\hat{z})h_1(\hat{z}) + g_2(\hat{z})h_2(\hat{z}) = 1 \)

### Result

- Finite-degree MA = Finite-degree AR
- \( y(n) \) AR process of order \( L \) with innovation \( i(n) = h(0)s(n) \), i.e.,

\[
y(n) + \sum_{\ell=1}^{L} A(\ell)y(n-\ell) = i(n)
\]

### Algorithm implementation:

- Solve Yule-Walker equations (to obtain \( A(\ell) \) then \( h(\ell) \))

\[
\mathbb{E}[i(n)[y(n-1)^H, \cdots, y(n-L)^H]] = 0
\]

- Estimate \( h(0) \) with the covariance matrix of the innovation
Part 5: Other types of algorithms
Semi-blind approach

Combining both criteria

- DA (with training sequence)
- blind/NDA (without training sequence)

as follows

\[ J(h) = \alpha J_{NDA}(h) + (1 - \alpha) J_{DA}(h) \]

Criteria selection (as an example):

- \( J_{DA}(h) \): ML
- \( J_{NDA}(h) \): Subspace algorithm

Result

Improve the estimation performance, or decrease the training duration
Decision directed approach

DA approach followed by

- NDA well initialized
- DD
  - with hard decisions
  - with soft decisions (turbo-estimation)
An other way: clustering based approach (or a step towards Machine Learning)

\[ y(n) = \mathbf{h}^T \mathbf{s}(n) + \mathbf{w}(n) \]

with \( \mathbf{s}(n) = [s(n), \ldots, s(n - L)]^T \) and \( \mathbf{h} = [h(0), \ldots, s(L)]^T \)

- \( y(n) \) is a point in \( \mathbb{C} \), and belongs to the cluster labelled by one \( c \)
- \( K \) clusters to characterize (where \( K = \text{card}(c) \) is known)
- Apply \textit{unsupervised clustering} algorithm: \( K \)-means
- Now, given \( c \), how retrieving \( \mathbf{s}(n) \) (with unknown \( \mathbf{h} \))

**Hidden Markov Model (HMM) approach**

- \( \mathbf{s}(n) \) is a Markov Chain:
  \[ \Pr(\mathbf{s}(n)|\mathbf{s}(n - 1), \ldots) = \Pr(\mathbf{s}(n)|\mathbf{s}(n - 1)) \]
- \( \mathbf{c}(n) \) observation coming from an unknown Markov Chain state
- Forward-Backward algorithm to retrieve \( \mathbf{h} \)
An other way: clustering based approach
(or a step towards Machine Learning) (cont’d)

\[ y(n) = \text{fct}(s(n)) + w(n) \Rightarrow \hat{s}(n) = \text{threshold} (\Theta(y(n))) \]

with

- threshold: activation function
- \( \Theta(\bullet): \text{DNN weights}(\bullet) \)

Questions:

- One DNN per channel?
- If yes, training step (so it is not a blind approach)
- Gain in performance or less complex?
- Some papers on Optical-Fiber communications (trained for one fiber configuration)
- One DNN available for a large set of \text{fct}?
Part 6: Numerical illustrations
Second-order vs high-order algorithms

- Random multipath channel
- SIMO with oversampling of factor 2
- Observation window $1000T_s$

Figure: MSE vs SNR for 4QAM (left) and 16QAM (right) (courtesy of L. Mazet)
High-order algorithm (CMA)

\[ y(n) = \begin{bmatrix} 1 & \beta_1 \\ \beta_2 & 1 \end{bmatrix} \cdot s(n) + w(n) \]

Figure: BER vs SNR with 4QAM (warmup step of 1000 samples)
Time-varying channels

- Stationary SISO model
- 4QAM
- 6-tap equalizer filter $p$

**Figure:** BER vs iteration: $h = [0.3, 0.86, 0.39]^T$ (left), $h \leftarrow h + \text{std} \times \mathcal{N}(0, 1)$ at time index 500, 1000 and 1500 (right)
Use-case: optical-fiber (simulations)

- PolMux 16QAM, 112Gbits/s, range 1000km
- CD=1000ps/nm
- DGD=50ps
- OSNR=20dB

Blockwise algorithm converges with \( N = 1000 \) and few iterations
Adaptive algorithms need more samples to converge
BER target (\( \leq 10^{-3} \)) satisfied
Use-case: optical-fiber (experimentation)

- PolMux 8PSK, 60Gbits/s, range 800km
- SSMF fiber
- OSNR=23.7dB

It works!
Conclusion

- Blind equalization works in practice

- **HOS:**
  - No in-depth theoretical analysis
  - Drawback: large observation window (not civilian application yet, except optical fiber)

- **SOS:**
  - In-depth theoretical analysis (when $N$ large enough)
  - Easy to use, especially when SIMO coming from spatial diversity

- **DNN?**
References