SAPIENZA
UNIVERSITÀ DI ROMA

ROBUST AND WIDE-BAND SENSOR ARRAY PROCESSING

Elio D. Di Claudio
DIET
University of Roma «La Sapienza»

Sensor arrays

- Multiple co-located sensors within few wavelengths sample a wave-field in time and space.
- Scalar or vector wave-fields are sampled by proper sensors (electro-magnetic, acoustic pressure and velocity, strain…).
- Antenna is synthetically generated by linearly combining element outputs in reception or feeding properly convolved signal versions to sensors in transmission.
- Massively parallel hardware architectures.

SAPIENZA
UNIVERSITÀ DI ROMA

Sensor array processing

- *Time-synchronized* multi-sensor systems (*arrays*) are widely used today in remote sensing, biomedical, telecommunications and multimedia applications, for their capabilities of exploiting the wavefield properties for the purposes of *localization* of sources and *copy of (re-)radiated signals.*
- Signals and environmental parameters are estimated by processing the received multi-channel data.
- The *array processing* theory encompasses a broad range of specific applications (radar, SAR/ISAR, sonar, MIMO…) into a *compact framework.*

Main advantages of sensor arrays

- Elimination or great reduction of mechanical sensor movements (rotating antennas, near-field microphones, etc…).
- Potentially smaller environmental impact (conformal arrays).
- Electronic *compensation of platform movements* (yaw, pitch and roll).
- *Multiple functions* sharing the same sensors (e.g. search + tracking + radio-communications).
- Simultaneous management of *multiple threads*. This is impossible to do consistently with single sensor systems.
- Higher *transmitting power* and *antenna gain* available.
- *Greater empirical life-span*, fault-tolerance and upgradeability of sensor arrays than conventional systems.
- Hardware *modularity* (lower costs due to *mass production* of key components even for few systems).

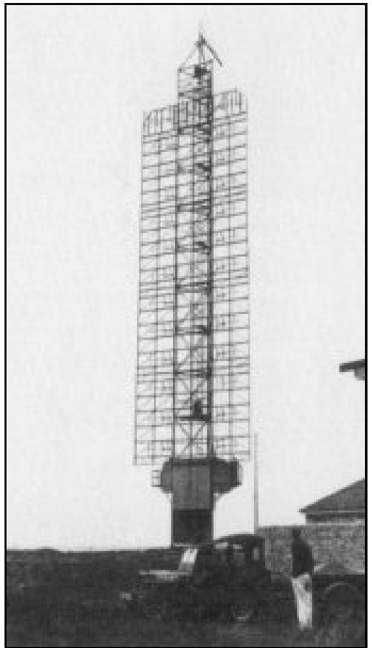
Perceived array disadvantages

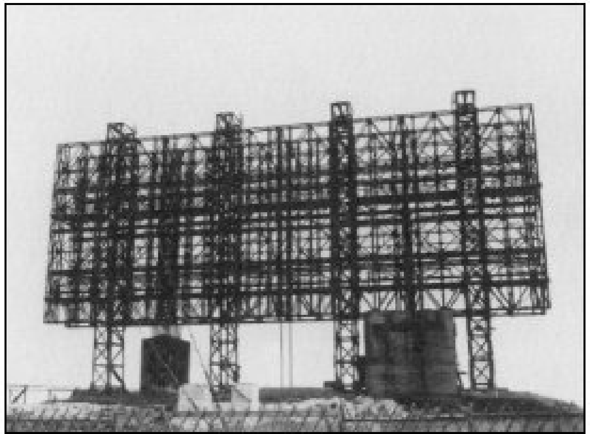
- *Antenna cost* more than proportional to element number.
- *Computational cost* roughly proportional to the cube of the number of sensors and the number of narrowband channels.
- *Higher weight and power* requirements.
- *Reduced gain for the same antenna size* due to inter-sensor gaps and near-fixed orientation.
- *Ambiguity problems* especially at high frequencies (circuitry size obstacles element size reduction w.r.t. the wavelength).
- Superficially attractive *competing solutions* in communications (pico-cells, sensor networks).
- *Difficult performance prediction* in several environments.

A bit of history…

- Most complex living organisms have passive arrays with two or more sensors, like ears, sonars and tactile systems for hunting, food search and risk avoiding.
- At early stages of communications and remote sensing development, sensor arrays were not endorsed, because of technical and scientific knowledge limitations.
- Only pressing of war and the introduction of new services forced switching to arrays.
- Today single-sensor systems are confined to low-performance equipment.

1940-1945: Early ideas


WASSERMANN


MAMMUT

ADVANCED GERMAN RADARS
(US NATIONAL ARCHIVES)

Peak power 100-200 KW, 125 MHz, pulse width 3 μs, PRF 500 Hz, range > 240 Km, accuracy: 0.25° (v) for WASSERMANN, 0.5° (h) for MAMMUT

- Array processing techniques date back to the W.W. II, mainly in Germany (MAMMUT and WASSERMANN ground radars, hydrophone arrays on battleships, cruisers and submarines)
- Evolution of earlier two-sensor interferometers with the introduction of the electronic steering (*beamforming*) of many elementary sensors

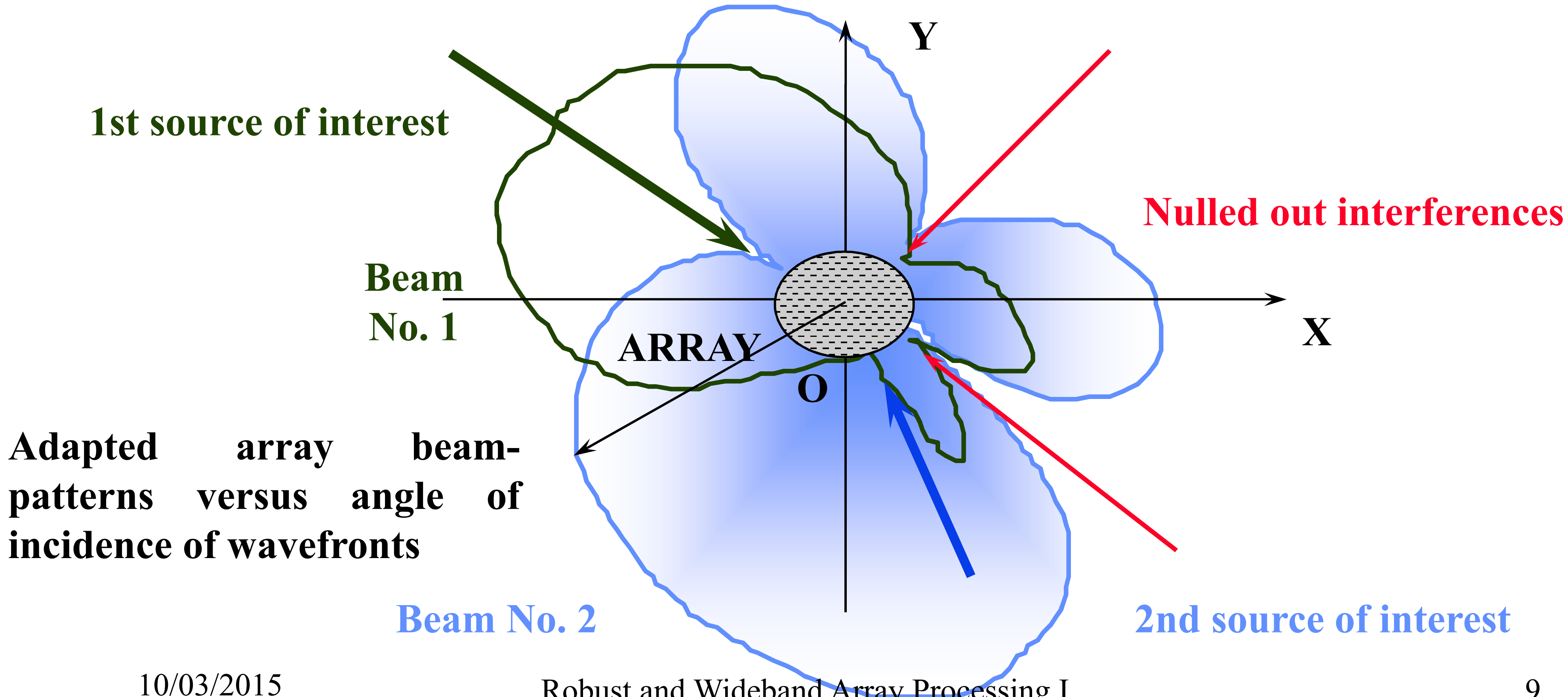
SAPIENZA
UNIVERSITÀ DI ROMA

1945-1968: Fourier-based array processing

- Phase scanned microwave radar
- Monopulse radar for tracking
- Digital beamforming (sonar)
- Early seismic imaging and migration algorithms
- Synthetic Aperture Radar (SAR)
- AR modeling of uniform linear arrays

1969-1972: Adaptive beamforming

- Linear combination of array outputs, optimized to recover signals of interest and cancel out interferences (passive sonar, radar)



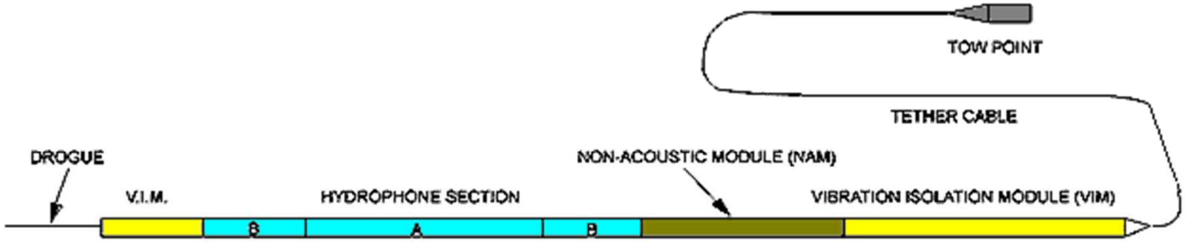
SAPIENZA
UNIVERSITÀ DI ROMA

1972-1977: 3-D applications

- Space-time array processing (STAP) for radar
- Magnetic resonance Fourier-based imaging
- Multi-function phased array radar
- Wave-based seismic processing
- Pisarenko harmonic decomposition: first consistent exploitation of the array signal model!

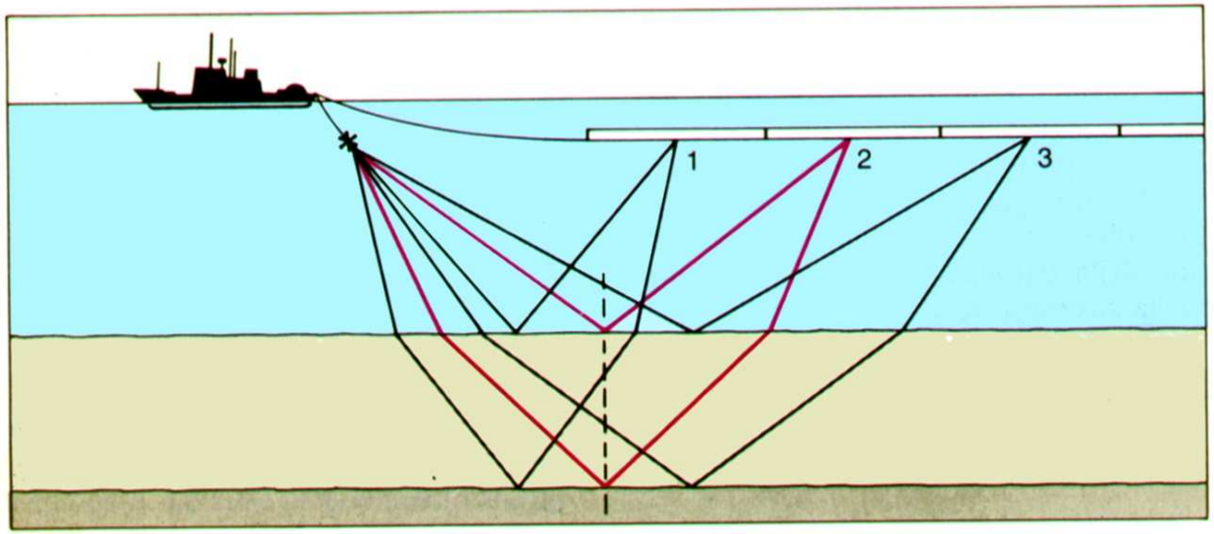
1977-1980: MUSIC and subspace techniques

- Integrated localization and adaptive beamforming of multiple sources by a subspace technique
- Consistent, but slightly suboptimal location estimates
- Initially applied to towed passive sonar and ESM systems



1980-1985: Theoretical advancements

- ML techniques for localization
- Extensive application of MLE to underwater acoustics and seismic prospecting
- Theoretical justification of signal subspace algorithms as approximate MLE.
- Fast and accurate rooting algorithms (ROOT MUSIC, MIN-NORM, ESPRIT) for direction finding



SAPIENZA
UNIVERSITÀ DI ROMA

1985-90: Extensions

- Asymptotically efficient subspace techniques (WSF and MODE)
- Extensive performance analysis of algorithms
- Information Theoretic techniques for model selection (AIC, MDL)
- HOS-based signal subspace identification (blind signal separation)
- Processing of coherent multipath:
  - Spatial smoothing
  - Toeplitz approximation
  - Coherent wideband focusing
  - Adaptive wideband steered beamforming

1990-1996: Rethinking

- Sensitivity analysis of algorithms to modeling errors gave badly surprising results.
- Optimal subspace algorithms replace MLE in critical applications
- Quasi-deterministic signal modeling (instrumental variable fitting)
- Wide-band array interpolation and beamspace processing
- Narrow-band *matched-field* processing for reverberant fields (underwater acoustics, ultrasound, seismics)
- Linearly constrained adaptive beamforming for robust source tracking and signal copy

SAPIENZA
UNIVERSITÀ DI ROMA

1997-2001: Network integration

- Smart-antennas (cellular base stations, satellites)
- Space Division Multiple Access
- Space-time coding
- Adaptive (2^{nd} generation) multi-function, multi-static radar
- Seismic tomography
- MMIC active T/R modules for microwaves
- Array proposals for multimedia applications (audio recording and immersive playback, teleconference, acoustic surveillance)

2001-2005: Robust array processing

- Spread source modeling
- Isotropic and redundant arrays
- Robust Capon-type beamforming
- Robust wide-band parametric localization (WAVES)
- Robust ML wide-band steered beamforming
- Robust wide-band matched-field array processing

2005-Today

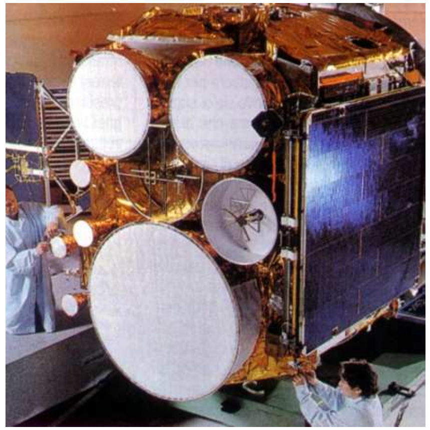
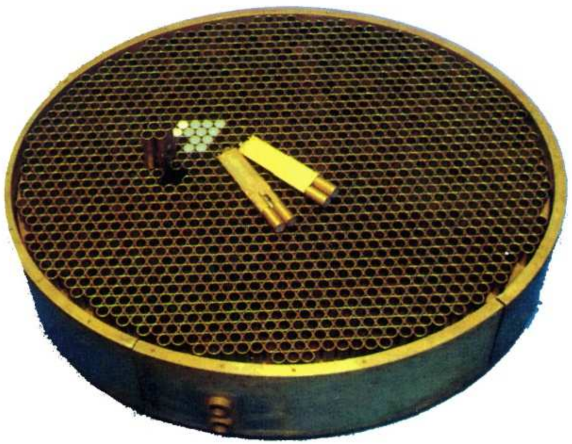
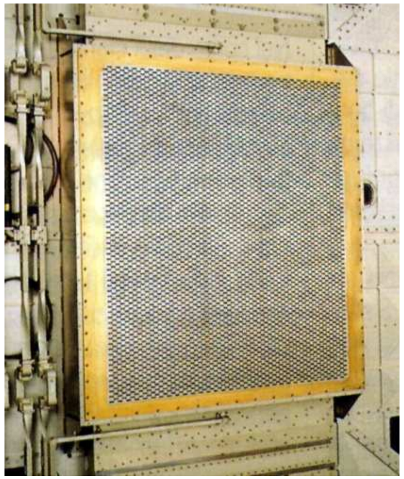
- Attempts to improve finite sample performance with random matrix statistical tools.
- MIMO systems under channel state uncertainties.
- Large scale (suboptimal) covariance estimation for audio.
- Transfer of L1 based compressed sensing techniques to array processing/spectral estimation.
  - But all high resolution parametric array estimators do compressed sensing in a rigorous way.

The future

- More rigorous array response modeling, especially for (ultra-) wide-band systems.
- Overcoming basic inadequacies of classical wide-band approaches:
  – limitations of the binning approach;
  – limitations of time delay approaches for multiple UWB source estimation.
  – Severe inconsistency of estimators w.r.t. the physical model.
- Innovative beamspace processing (e.g., Laguerre-Gauss) for UWB arrays and SAR/ISAR.
- Understanding propagation uncertainties and spatial source distribution effects on MIMO systems.

Some array-based products




Array problem setting

Goals:
•signal source *localization* in space and frequency
•*optimal reconstruction* of (re-)radiated signals
•array response *calibration*
•wave-field *measurements*

Available information:
•wave-field *physical mode*
•*array geometry* (*spatial sampling*)
•*sensor characteristics*
•*probabilistic signal model*

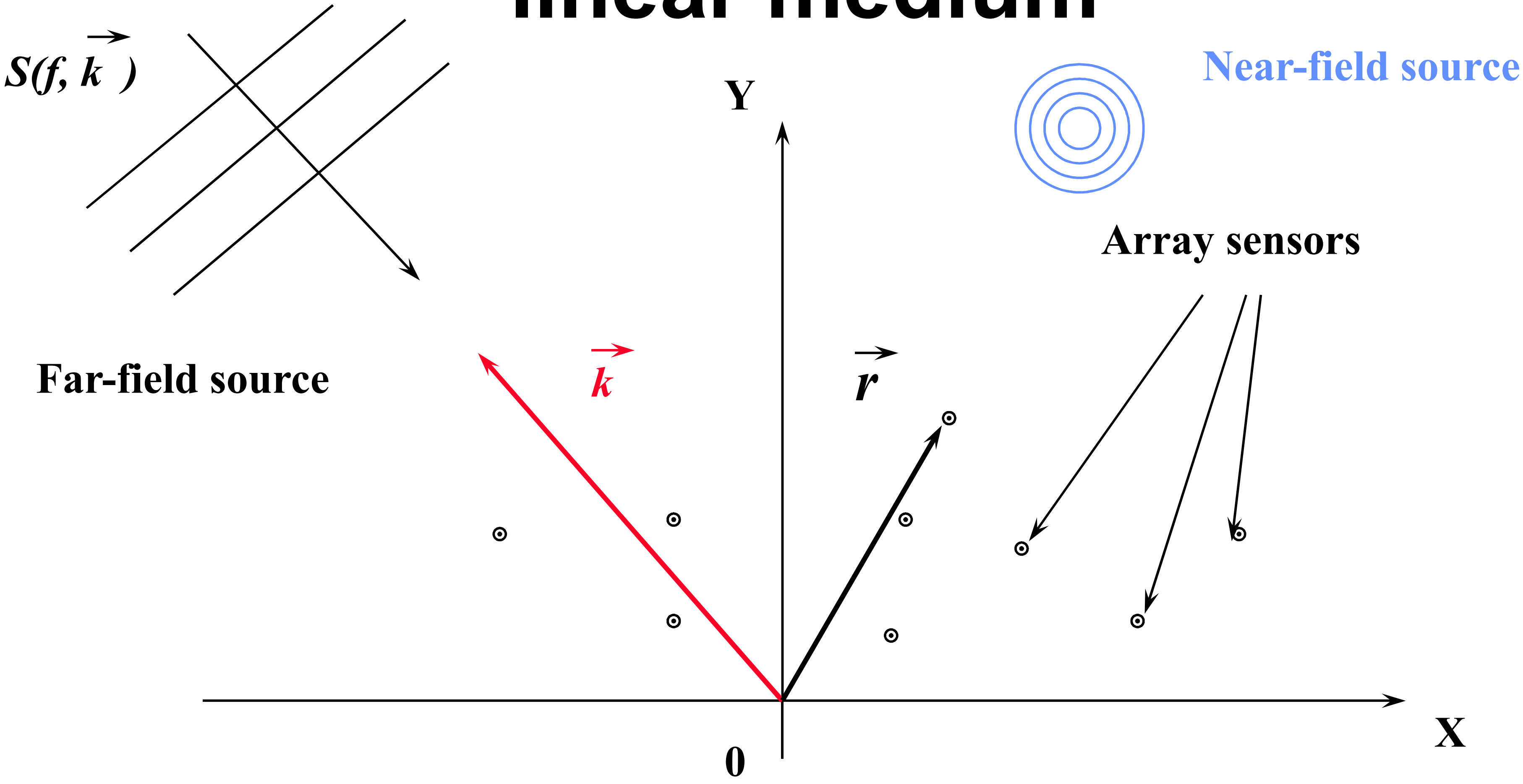
Review of narrowband array classical theory

- Static beamforming
- Adaptive beamforming
- Parametric source localization
- Array geometries
- Mutual coupling effects
- Modal interpolation
- Calibration
- Only state of the art techniques are reviewed
- Blind signal recovery not considered here

Part I

PROPAGATION AND ARRAY MODELING BACKGROUND

Example: EM wave propagation in homogeneous, isotropic and linear medium



Waves impinging on a N sensor array

$$\vec{\mathbf{E}}(f,\vec{\mathbf{P}}) = \iiint S(f,\vec{\mathbf{k}})e^{j\vec{\mathbf{k}}\cdot\vec{\mathbf{P}}}d\vec{\mathbf{k}}$$

$$g(f,\vec{\mathbf{r}}_i) = \iiint_{\vec{\mathbf{P}}\in I(\vec{\mathbf{r}}_i)} K_i(f,\vec{\mathbf{E}}(f,\vec{\mathbf{P}}))d\vec{\mathbf{P}}$$

$$\mathbf{x}(f) = \begin{bmatrix} g(f,\vec{\mathbf{r}}_1) \\ g(f,\vec{\mathbf{r}}_2) \\ \cdots \\ g(f,\vec{\mathbf{r}}_N) \end{bmatrix} + \begin{bmatrix} n_1(f) \\ n_2(f) \\ \cdots \\ n_N(f) \end{bmatrix}$$

SAPIENZA
UNIVERSITÀ DI ROMA

Starting assumptions

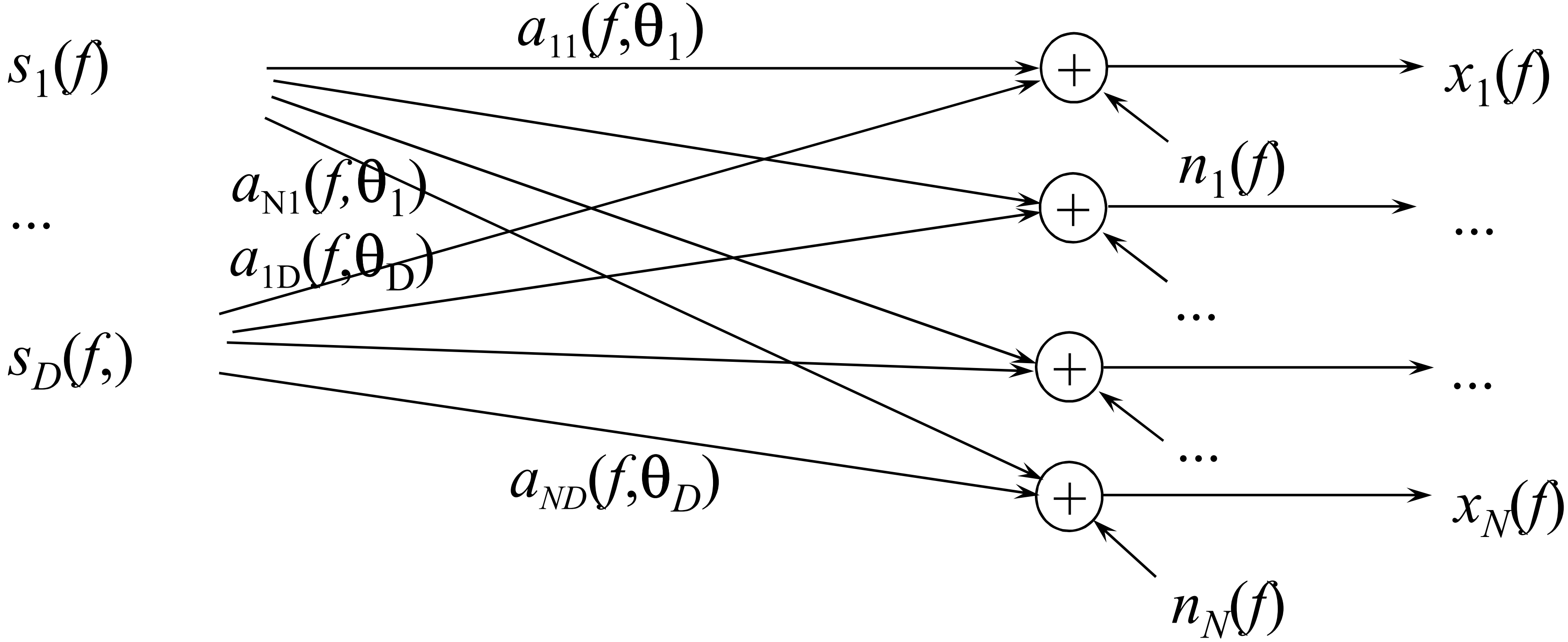
- *Linear sensors* and propagating medium (*linear superposition* of source signals) to ensure superposition of effects.
- *Convolutive,* discrete time *(wide-band)* model
- *Additive* noise, *independent* from source signals.
- *Known array response* for any source location of interest $\mathbf{p}$ (not sufficient for MLE!)
- No *spatial aliasing* (e.g., ambiguity of location)

$$\mathbf{x}(n) = \left[\sum_{k=0}^{K} \sum_{d=1}^{D} \mathbf{h}(\mathbf{p}_d, k) s_d(n-k) \right] + \mathbf{v}(n)$$

Point source approximation

- A *point source* is a source whose dimensions cannot be resolved from array data:
  - Sources located at infinity (plane wave approximation).
  - Small sources in near field (spherical wave).
  - Represented by a Dirac pulse in spatial coordinates.
- Point source assumption allows consistent localization by arrays with sufficient number of sensors.
- Extended sources often modeled by (stochastic) clusters of point sources emitting inter-correlated signals.

Standard frequency domain array model for point sources



$$\mathbf{x}(f) = \mathbf{A}(f,\Theta)\mathbf{s}(f) + \mathbf{n}(f)$$

Narrow-band transfer function approximation

- Approximation of the generic array sensor response around the frequency of interest.

$$a_k(f,\mathbf{p}) = A_k(f,\mathbf{p}) \cdot e^{j\varphi_k(f,\mathbf{p})}; \quad A_k(f,\mathbf{p}) = \left|a_k(f,\mathbf{p})\right|$$

$$A_k(f,\mathbf{p}) = A_k(f_0,\mathbf{p}) + \left.\frac{\partial A_k(f,\mathbf{p})}{\partial f}\right|_{f\to f_0}(f-f_0)+\ldots \cong A_k(f_0,\mathbf{p})$$

Flat filter pass-band or peaked response

$$\varphi_k(f,\mathbf{p}) = \varphi_k(f_0,\mathbf{p}) + \left.\frac{\partial \varphi_k(f,\mathbf{p})}{\partial f}\right|_{f\to f_0}(f-f_0)+\ldots$$

$$\cong \varphi_k(f_0,\mathbf{p}) - 2\pi\tau_k(\mathbf{p})(f-f_0)$$

Group delay

SAPIENZA UNIVERSITÀ DI ROMA

Narrow-band array model approximations

- **Narrow-band phasor model**
  - Signal envelopes transferred as pure sinusoids with frequency f_0.

$$a_k(f,\mathbf{p})s(f) \cong a_k(f_0,\mathbf{p})\hat{s}(f-f_0)$$

$$\tau_k(\mathbf{p}) \ll \frac{1}{2\pi|f-f_0|} \leq \frac{1}{\pi B}$$

- **Time (group) delay model**
  - Delayed signal envelopes transferred as pure sinusoids with frequency f_0

$$a_k(f,\mathbf{p})s(f) \cong a_k(f_0,\mathbf{p})\hat{s}(f-f_0) \cdot e^{-j2\pi f_0\tau_k(\mathbf{p})(f-f_0)}$$

Narrow-band discrete time signal model

$$\mathbf{x}(l) = \left[\mathbf{a}(f,\mathbf{p}_1) \quad \cdots \quad \mathbf{a}(f,\mathbf{p}_D)\right] \begin{bmatrix} s_1(l) \\ \vdots \\ s_D(l) \end{bmatrix} + \mathbf{v}(l)$$

- The bandwidth of the filtered array signal, centred at f, is much smaller than the reciprocal of the *sum of the lengths of the impulse response and of the source correlation*
  - Snapshots can be considered as mutually independent (i.e., the spectra of *sampled* signals and noise are essentially constant within the filter pass-band)
- A wide-band, stationary signal can be decomposed into *approximately independent* narrow-band components by a critically sub-sampled filter bank or a block DFT processing

SAPIENZA
UNIVERSITÀ DI ROMA

Basic array research topics

- Flexible *propagation and array response models* for theoretical analysis and simulation purposes.
- Field *calibration* of arrays based on experimental measurements and/or parametric models (modal interpolation).
- *Beampattern (spatial filter) synthesis* for optimal transmission and reception of signals
- *Model identifiability* condition to prevent aliasing and false source detection;
- *Source location* and *signal estimation* by calibrated or uncalibrated arrays.

Propagation and array response models

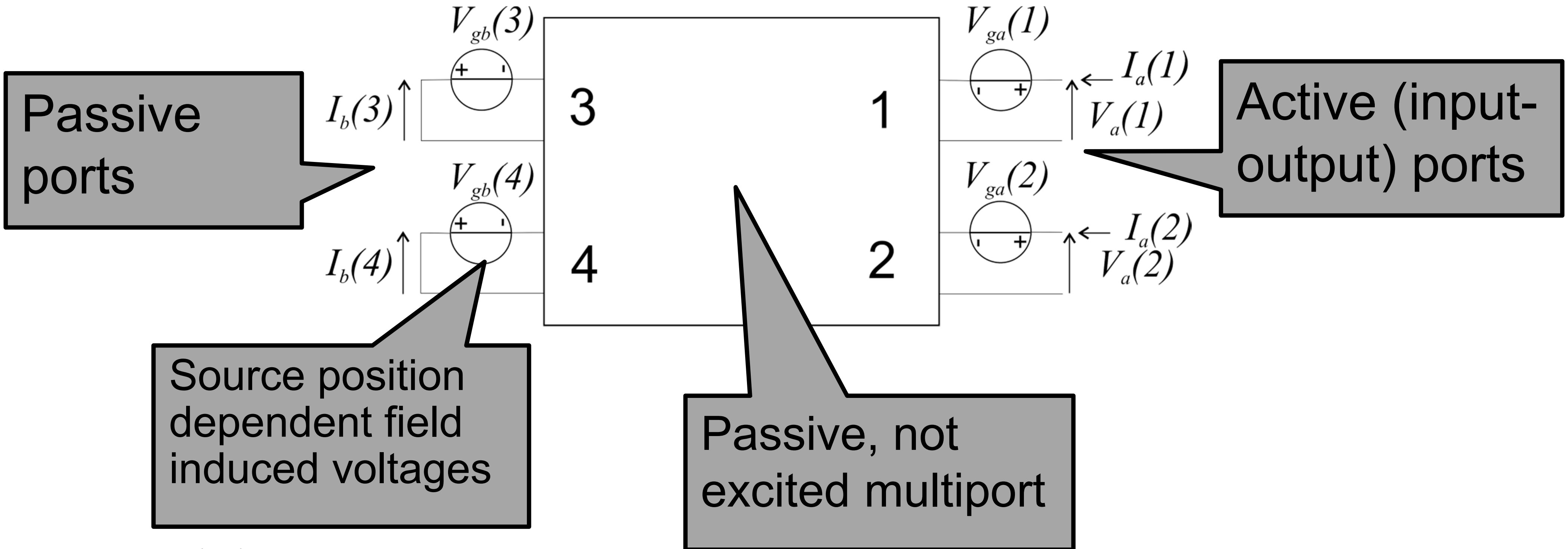
- Establish links between wave propagation equations and steering vector.
- Integral models based on voltage (line integral of a vector field) and current (integral flux through a surface patch) concepts.
- Integral model have very low sensitivity w.r.t. field perturbations, maybe due to imperfect propagation modeling.
- Hyperbolic wave equation solutions luckily not very sensitive to boundary conditions.

Electric array multiport modeling example

- Array element discretization (integral equations based on vector potential), e.g., *thin wire* antennas, subdivided into pieces of about $0.1\ \lambda$.
- Linear medium and electrical loads.
- External electrical field $\mathbf{E}(f,\mathbf{r};\mathbf{p})$ given by a wavefront (linear-shaped or not) radiated at frequency f from the position characterized by the generic coordinate vector $\mathbf{p}$.
- Compact sub-elements w.r.t. wavelength, subdivided into two disjoint subsets:
  - *A.* N antenna sub-elements connected by gaps to input/output waveguides (*active array*) , located around positions $\mathbf{r}_a(n)$ with port currents $I_a(n,\mathbf{p})$ and voltages $V_a(n,\mathbf{p})$, for $n=1,2,..,N$.
  - *B.* M-N sub-elements not connected to waveguides (*passive co-array*) located around positions $\mathbf{r}_b(n,\mathbf{p})$ with currents $I_b(n,\mathbf{p})$, for $n=N+1,2,..,M$. They can also model nearby scatterers.

Multiport array model

- Wavefront excited array seen as a multiport with external voltage sources (generalized Thevenin equivalence).



Impedance type integral model for thin wire antenna array

- Symbolic solution similar to the one obtained by the loop method for lumped circuits.

$$\begin{bmatrix} \mathbf{Z}_{aa} & \mathbf{Z}_{ab} \\ \mathbf{Z}_{ab}^{T} & \mathbf{Z}_{bb} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{a} \\ \mathbf{I}_{b} \end{bmatrix} = - \begin{bmatrix} \mathbf{V}_{ga}(\mathbf{p}) \\ \mathbf{V}_{gb}(\mathbf{p}) \end{bmatrix} + \begin{bmatrix} \mathbf{V}_{a} \\ \mathbf{0} \end{bmatrix}$$

$$\mathbf{V}_a = -\mathbf{Z}_L \mathbf{I}_a + \mathbf{V}_L \qquad V_{gx}(n) \cong \int_{C(\mathbf{r}_n)} K_n(\mathbf{r}) \mathbf{t}_n \bullet \mathbf{E}(\mathbf{r})\, dl_n$$

Passive antenna loads and (optional) Tx excitation

Field excitation voltages about proportional to incident field components at each element (gap)

Network solution

- Off diagonal terms in admittance matrix characterize *mutual coupling* between sensors
- *Load-dependent* mutual coupling.
- No general load optimization procedure unlike the single sensor case.
- Interference by sub-element and scatterer reflections (constructive or destructive).
- For thin wires, impedance matrix blocks are nearly independent from the wavefront arrival angle (DOA).

$$\mathbf{I}_a = \underbrace{\left[\mathbf{Z}_{aa} - \mathbf{Z}_{ab}\mathbf{Z}_{bb}^{-1}\mathbf{Z}_{ab}^{T}\right]^{-1}}_{\text{Multiport open-circuit admittance matrix}} \left[\mathbf{V}_a - \underbrace{\left(\mathbf{V}_{ga} - \mathbf{Z}_{ab}\mathbf{Z}_{bb}^{-1}\mathbf{V}_{gb}\right)}_{\text{Position dependent external field excitation}}\right]$$

$$\mathbf{V}_a = -\mathbf{Z}_L\mathbf{I}_a + \mathbf{V}_L$$

SAPIENZA
UNIVERSITÀ DI ROMA

Steering vector calculation from antenna electrical model

- Loads are connected to array outputs and ADCs sense array port voltages (or currents).
- The array steering vector for position $\mathbf{p}$ is obtained for zero transmission excitation.

$$\mathbf{V}_L = 0; \quad \mathbf{V}_a = -\mathbf{Z}_L \mathbf{I}_a$$

$$\mathbf{V}_a(\mathbf{p}) = \left\{ \mathbf{Z}_L^{-1} + \left[\mathbf{Z}_{aa} - \mathbf{Z}_{ab}\mathbf{Z}_{bb}^{-1}\mathbf{Z}_{ab}^T \right]^{-1} \right\}^{-1} \times$$

$$\left[\mathbf{Z}_{aa} - \mathbf{Z}_{ab}\mathbf{Z}_{bb}^{-1}\mathbf{Z}_{ab}^T \right]^{-1} \left[\mathbf{V}_{ga}(\mathbf{p}) - \mathbf{Z}_{ab}\mathbf{Z}_{bb}^{-1}\mathbf{V}_{gb}(\mathbf{p}) \right]$$

$$\mathbf{a}(\mathbf{p}) \propto \mathbf{V}_a(\mathbf{p})$$

Common approximations

- Sensors compact w.r.t. wavelength.
- Block shaped impedance and admittance matrices (e.g., decoupled wires).
- Response proportional to a projection of the electric field at the sensor gaps (array ports).

$$\mathbf{Z}_{ab} = \mathbf{0}; \quad \mathbf{Z}_{aa} = diag\left(\begin{bmatrix} z_{a1} & \cdots & z_{aN} \end{bmatrix}\right)$$

$$\mathbf{V}_a(\mathbf{p}) = \left(\mathbf{Z}_L + \mathbf{Z}_{aa}\right)^{-1} \mathbf{Z}_L \mathbf{V}_{ga}(\mathbf{p})$$

$$\mathbf{a}(\mathbf{p}) \propto \left[a_1 \mathbf{t}_1 \bullet \mathbf{E}(\mathbf{r}_1, \mathbf{p}) \quad \cdots \quad a_N \mathbf{t}_N \bullet \mathbf{E}(\mathbf{r}_N, \mathbf{p}) \right]^T$$

Baseline far field steering vector

- Each sensor has a directivity pattern given by $g_n(f,\mathbf{p})$ and is located at position $\mathbf{r}_n$.
- Planar impinging wave-front.
- Given field polarization.
- In many cases the steering vector is *normalized*.

$$\mathbf{a}(f,\mathbf{p}) \propto \left[g_1(f,\mathbf{p}) e^{j\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}_1} \quad \cdots \quad g_N(f,\mathbf{p}) e^{j\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}_N} \right]^T$$

$$\vec{\mathbf{k}} = \frac{2\pi}{\lambda}\left[-\sin(\theta)\cos(\varphi)\vec{\mathbf{i}} + \cos(\theta)\cos(\varphi)\vec{\mathbf{j}} + \sin(\varphi)\vec{\mathbf{z}} \right]$$

Mutual coupling reflection-based model

- Mutual coupling can be interpreted as the sum of multiple reflections between sensors.
- Linear transformation of the baseline steering vector by a finite *coupling matrix*.

$$\mathbf{a}(\mathbf{p}) \cong \mathbf{a}_0(\mathbf{p}) + \mathbf{T}\mathbf{a}_0(\mathbf{p}) + \mathbf{T}^2\mathbf{a}_0(\mathbf{p}) + \dots$$

$$\mathbf{T} = \mathbf{W}\begin{bmatrix} \lambda_1 & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & \lambda_N \end{bmatrix}\mathbf{W}^{-1}; \quad |\lambda_k| < 1$$

$$\mathbf{a}(\mathbf{p}) \cong \left[\mathbf{W}\mathbf{D}\mathbf{W}^{-1}\right]\mathbf{a}_0(\mathbf{p})$$

Baseline steering vector

Mutual coupling transformation

Invariances of ideal mutual coupling matrices

- The mutual coupling (MC) matrix of a long, linear, equi-spaced array tends to be Toeplitz (equal elements along diagonals) for a translational invariance argument:
  - the neighborhood of all identical elements is the same, but translated along a line.
- Toeplitz MC matrix can be enforced on a whole short linear array by extending it with passive (dummy), impedance-matched elements.
- Circular array MC matrix ideally is circulant:
  - the neighborhood of all elements is the same, but rotated in angle.

Mutual coupling remarks

- The above model does not consider the effect of nearby scattering and so essentially refers to short/small antenna arrays.
- Mutual coupling matrix mostly depends on the direction (sensors always are somewhat directive).
- Often it is preferable to consider the coupled array as an unknown one, calibrate and interpolate its response with a general model (e.g., circular harmonics, polynomials), rather than enforcing a particular steering vector structure.

Diversity factors in arrays

- The effectiveness of array design depends on the *spatial diversity* of sensor responses w.r.t. arrival angles.
- The directivity gain terms create an angular partition of the space (*beam space diversity*):
  - highly directive, oriented elements reduce the effective number of sensors (i.e., number of resolvable sources) active in some directions, but increase the overall array gain.
  - increasing sensor directivity helps in reducing array size and mutual coupling.
- The phase (delay) components depend on the inter-sensor separation (*element space diversity*):
  - a large separation increases directional and range resolution capabilities of the array at expense of ambiguity risks and sensitivity to sensor position and tolerances.
  - a small separation creates a redundant wave-field sampling (i.e., smaller spatial resolution) and increases mutual coupling.

Polarized plane waves

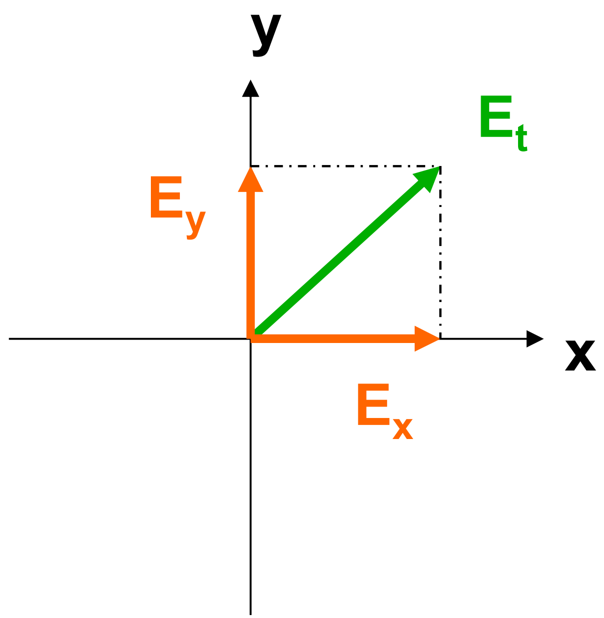
- Polarized incident plane wavefront, orthogonally decomposed on the plane perpendicular to the propagation vector -**k**.
- Linear composition of two plane waves.

$$\mathbf{E}_t(\mathbf{r},t)=E_x(\mathbf{r},t)\mathbf{i}+E_y(\mathbf{r},t)\mathbf{j};\ \mathbf{k}=-\mathbf{i}\times\mathbf{j}$$

$$E_x(\mathbf{r},t)=\mathrm{Re}\left\{E_x e^{j(\omega t+\mathbf{k}\cdot\mathbf{r})}\right\}$$

$$E_y(\mathbf{r},t)=\mathrm{Re}\left\{E_y e^{j(\omega t+\mathbf{k}\cdot\mathbf{r})}\right\}$$

$$\mathbf{E}_t=\begin{bmatrix}E_x\\E_y\end{bmatrix}=\begin{bmatrix}A_x e^{j\varphi_x}\\A_y e^{j\varphi_y}\end{bmatrix}$$



Steering vector of polarized waves

- Array output is the linear combination of the partial outputs for excitations along $\mathbf{i}$ and $\mathbf{j}$.
- Linear combination of non-normalized partial steering vectors.

$$\mathbf{E}_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow \mathbf{a}_x(\mathbf{p}, f); \quad \mathbf{E}_t = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow \mathbf{a}_y(\mathbf{p}, f)$$

$$\mathbf{a}(\mathbf{p}, f) \propto \frac{A_x e^{j\phi_x} \mathbf{a}_x(\mathbf{p}, f) + A_y e^{j\phi_y} \mathbf{a}_y(\mathbf{p}, f)}{\sqrt{A_x^2 + A_y^2}}$$

$$= \begin{bmatrix} \mathbf{a}_x(\mathbf{p}, f) & \mathbf{a}_y(\mathbf{p}, f) \end{bmatrix} \begin{bmatrix} \cos(\phi_J) e^{j\phi_x} \\ \sin(\phi_J) e^{j\phi_y} \end{bmatrix}$$

Vector arrays

- Most wave-fields are characterized by up to three *independent components* (e.g., vector potentials).
- Two or more sensors essentially sensitive to a single polarization or field component and located nearly at the same position form a *vector sensor*.
- More information about the source signal, but limited space diversity addition: most sensors are not excited by common wavefronts.
- In sonar and audio applications up to three orthogonal velocity directional sensors plus one pressure (scalar) sensor are placed at each sampling site.
- Unequal sensor sensitivities and internal noise levels complicate processing.

SAPIENZA
UNIVERSITÀ DI ROMA

Array manifold ambiguity

- The array steering vector describes a *manifold* (i.e., an hypersurface) w.r.t. parameters in $\mathbf{p}$ at constant frequency.
- *Ambiguity* or *spatial aliasing* arises when a steering vector at direction $\mathbf{p}$ is *linearly dependent* from another set of steering vectors at different directions.
- A K*-fold ambiguity* arises when K steering vectors, referred to different angles, become linearly dependent, but all combinations of dimension K-1 are linearly independent.
- In the presence of ambiguity, *infinite scenarios* exist that *cannot be distinguished* by *any* data-based processing technique, leading to false, systematic source detections.

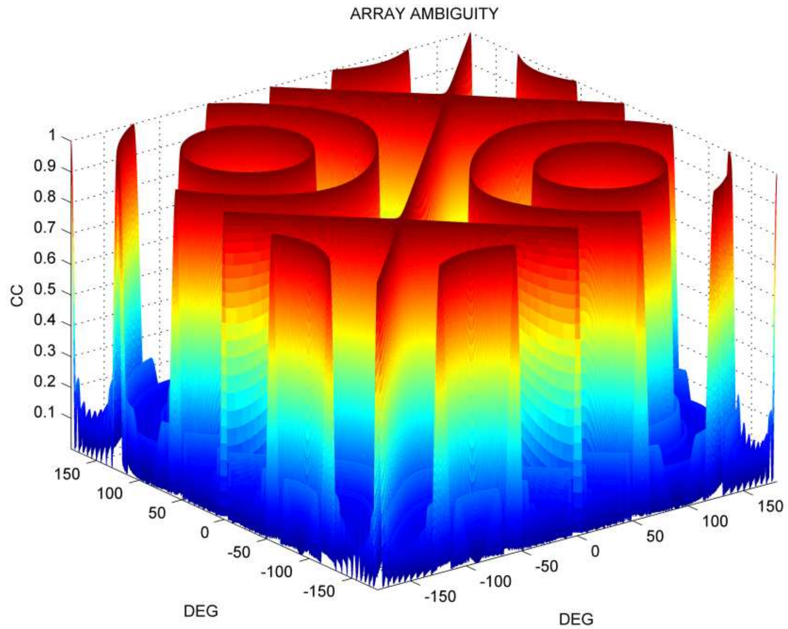
$$\begin{bmatrix}\mathbf{a}(\mathbf{p}_1) & \cdots & \mathbf{a}(\mathbf{p}_K)\end{bmatrix}\mathbf{c}_0 = \begin{bmatrix}\mathbf{a}\left(\mathbf{p}_1^{(1)}\right) & \cdots & \mathbf{a}\left(\mathbf{p}_K^{(1)}\right)\end{bmatrix}\mathbf{c}_1$$

$$\left\{\mathbf{p}_1,\ldots\mathbf{p}_K\right\} \cap \left\{\mathbf{p}_1^{(1)},\ldots,\mathbf{p}_K^{(1)}\right\} = \left\{\varnothing\right\}; \quad \mathbf{c}_1,\mathbf{c}_0 \neq \mathbf{0}$$

Correlation coefficient and simple ambiguity

- Simple ambiguity can be detected by a unit magnitude correlation coefficient (CC) or *grating lobe* among two different steering vectors, whose plot is a kind of *beampattern*.

$$\left|\rho\left(\mathbf{p}_1,\mathbf{p}_2\right)\right| = \frac{\left|\mathbf{a}\left(\mathbf{p}_1\right)^H \mathbf{a}\left(\mathbf{p}_2\right)\right|}{\left|\mathbf{a}\left(\mathbf{p}_1\right)\right|\left|\mathbf{a}\left(\mathbf{p}_2\right)\right|} = 1$$



SAPIENZA
UNIVERSITÀ DI ROMA

Ambiguity sources

- Basic spatial ambiguity is tied to:
  - unavoidable non-uniqueness of sensor relative phases within the FoV (similar to digital signal aliasing);
  - same directionality and mechanical orientation of array sensors;
  - many sensors interspaced by much more than half wavelength.
  - sampling lattices with few, distant sensors.
- To avoid *low order* ambiguities:
  - ensure sensor directivity pattern and/or orientation diversity (non-identical sensors, squinted directional antennas, mutual coupling);
  - insert many sensor interspaced by less than half wavelength;
  - avoid large gaps in sampling lattices (several wavelengths wide);
  - avoid close, similar sensors;
  - prefer irregular lattices.
- High order ambiguity is tied to and measurable by the following modal array decomposition in the directional parameter space.

Modal decomposition of the array response

- The array manifold is approximated by a matrix combination (*expansion*) of linearly independent basis functions of the location coordinates $\mathbf{p}$ at constant frequency.
- The numerical rank of the mixing matrix $\mathbf{A}$ is the *maximum number* of independent steering vectors and of surely detectable sources over the entire field of view (FoV).
- This number is smaller for smaller angular sectors!

$$\mathbf{a}(\mathbf{p},f)=\sum_k \mathbf{a}_k(f)u_k(\mathbf{p})\cong \mathbf{A}(f)\mathbf{u}(\mathbf{p})$$

$$\int_{FOV} u_k(\mathbf{p})^* u_l(\mathbf{p})d\mathbf{p}=\delta_{kl};\ \mathbf{a}_k(f)=\int_{FOV}\mathbf{a}(\mathbf{p},f)u_k(\mathbf{p})^* d\mathbf{p}$$

Some modal decompositions

- *Angular harmonic* decomposition: sinusoidal basis functions over azimuth fit each sensor response.
- *Chebychev decomposition*: stretched Chebychev angular polynomial basis over a sector.
- *Spherical harmonic decomposition* for 3-D arrays (little computational gains…).
- Local (*prolate-like*) decomposition on the orthogonal basis of actual steering vectors over little angular sectors. Requires SVD concepts (see later).

Angular harmonic expansion of the delay factor

- Azimuthal harmonic expansion of the delay factor in 2-D is expressed by Bessel J functions of the first kind.
- Expansion coefficients quickly converging to zero for $n > 2\pi R/\lambda$

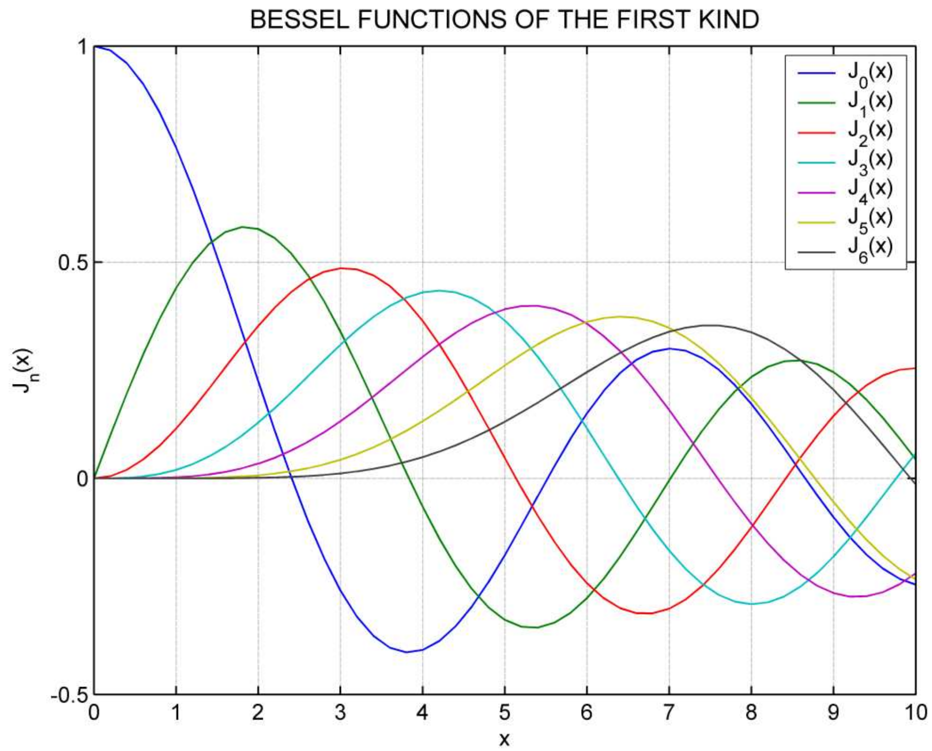
$$\mathbf{r} = R\left[\cos(\varphi)\mathbf{i} + \sin(\varphi)\mathbf{j}\right]; \quad \mathbf{k} = \frac{2\pi}{\lambda}\left[\cos(\theta)\mathbf{i} + \sin(\theta)\mathbf{j}\right]$$

$$e^{j\mathbf{k}\cdot\mathbf{r}} = e^{j\frac{2\pi R}{\lambda}\cos(\theta-\varphi)}; \quad e^{jx\sin(u)} = \sum_{n=-\infty}^{+\infty} J_n(x)e^{jnu}$$

$$e^{j\mathbf{k}\cdot\mathbf{r}} = \sum_{n=-\infty}^{+\infty}\left[j^n J_n\left(\frac{2\pi R}{\lambda}\right)e^{jn\varphi}\right]e^{-jn\theta}$$

Bessel-J function table

- The Bessel functions of the first kind are found also in optics, waveguides and disk antennas.



Spherical harmonic decomposition

- Spherical harmonics
  - Solutions of the Laplace equation in spherical coordinates;
  - Orthogonal basis for 3-D modal decomposition of 3-D arrays.

$$\nabla^2 f = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) + \frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left[\sin(\theta)\frac{\partial f}{\partial\theta}\right] + \frac{1}{r^2\sin(\theta)}\frac{\partial^2 f}{\partial\varphi^2} = 0$$

$$Y_l^m(\theta,\varphi) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m\left[\cos(\theta)\right]e^{jm\varphi}$$

$$P_l^{-m}(x) = (-1)^m\frac{(l-m)!}{(l+m)!}P_l^m(x)$$

$$P_l^m(x) = \frac{(-1)^m}{2^l l!}\left(1-x^2\right)^{\frac{m}{2}}\frac{\partial^{l+m}}{\partial x^{l+m}}\left(x^2-1\right)^l; \quad -l \le m \le l$$

Remarks on orthogonal expansion

- For algorithmic purposes, it is better to expand the *normalized* array steering vector, at the likely expense of the expansion convergence rate due to non-linear distortion of the angular response.
- Only vector direction matters for ambiguity.
- *Phase/delay centering* of the expansion is a critical art to reduce the expansion order and get low errors.
- Basis for a *virtual array* design independent of physical constraints.

$$\frac{\mathbf{a}(\mathbf{p}, f)}{\left|\mathbf{a}(\mathbf{p}, f)\right|_2} = \sum_k \hat{\mathbf{a}}_k(f) u_k(\mathbf{p}) \cong \hat{\mathbf{A}}(f)\mathbf{u}(\mathbf{p})$$

Notes on orthogonal expansions (2)

- Frequency-dependent convergence rate (faster at low frequencies, slower at high frequencies).
- Strong dependence on sensor directivity.
- Non parsimonious for sparse arrays.
- Divergence and or large fitting errors outside the calibration/definition angular sector for a bad basis function selection.
- 3-D array manifold expansions not convenient from any point of view.

Notes on orthogonal expansions (3)

- Harmonic expansion reduces a general array to a transformed harmonic one, fine for algorithmic purposes (e.g., root-finding harmonic retrieval).
- Possible improvements of low SNR thresholds and high SNR robustness for complex reasons
- Expansions reduce random calibration errors by relieving measurement noise (LS or ML fitting)!

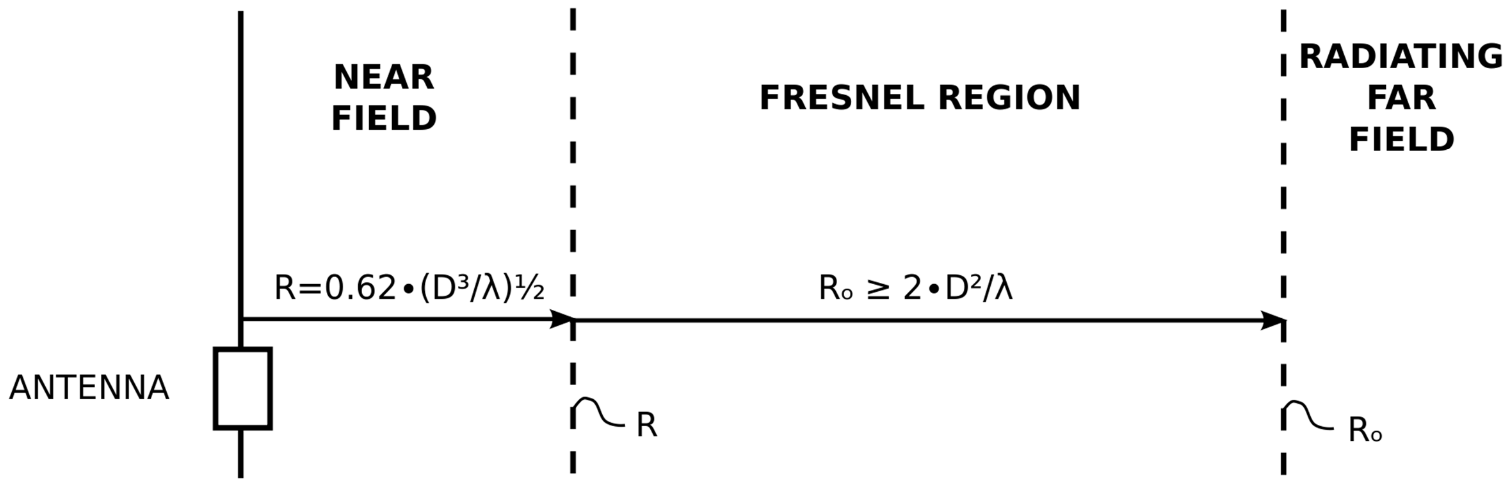
$$\hat{\mathbf{A}}(f)\mathbf{u}(\mathbf{p}) = \frac{\mathbf{a}(\mathbf{p}, f)}{\left|\mathbf{a}(\mathbf{p}, f)\right|_2} - \mathbf{v}(\mathbf{p}, f)$$

Additive calibration noise

Fixed interpolation basis

Near field array behavior

- The definition of near/far field is not very clear for arrays (sensor and array sizes differ by much).
- Conventionally the typical array antenna size D in formulas is made equal to the array (maximum) diameter.
- In thin wire / patch antenna simulators only far field components of the vector potential are retained!



SAPIENZA
UNIVERSITÀ DI ROMA

Near-field source models

- The array model with coupling is valid even for a distance of few wavelengths between the target and the sensors.
- Far field (point source) steering vector model is quickly approximated by increasing the source distance.
- For really near field sources $(<0.1\ \lambda)$, the target itself must be inserted in the array electric model as some additional, coupled and electrically feeded elements.
- Steering vector and signal models remain essentially the same.

SAPIENZA
UNIVERSITÀ DI ROMA

Basic commercial array types

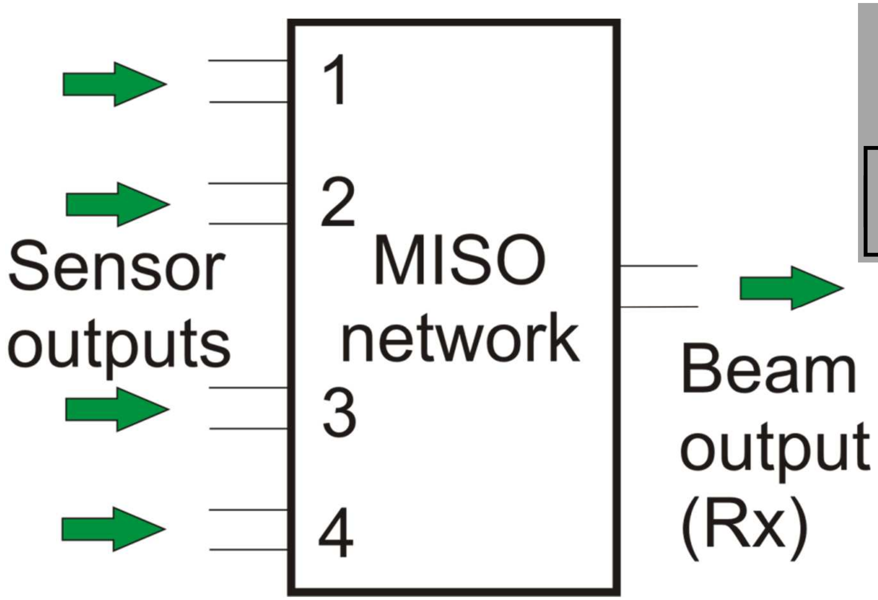
- Passive Electronically Scanned Antenna (PESA):
  - Sensor outputs combined by a passive, distributed or lumped network in Tx and/or Rx;
  - MISO or MIMO models: sequential, programmable generation of synthetic beams.
  - Signal amplification and demodulation only at beam outputs.
- Active Electronically Scanned Antenna (AESA):
  - Sensors combined by an *active network* made containing amplifiers and other active elements (filters, summation nodes);
  - Enables the use of *T/R modules* including amplifiers and matched interfaces behind each antenna patch;
  - Signal demodulation at beam outputs.
- Parallel array receiver/transmitter:
  - one (digital) receiver/amplifier for each sensor;
  - sensor outputs digitized and recorded in parallel batches;
  - digital communication network.
  - real time and batch processing possible even for multiple tasks;
  - high power requirements for multiple receivers.

Performance notes

- PESA, AESA are often trademarks with little influence on intrinsic performance.
- Real technical evaluation elements include:
  - array aperture and ambiguity
  - number of sensors available for flexible/adaptive post-combination
  - noise factor
  - reliability and restorability.
  - antenna losses.
- Real world substitution of PESA with AESA without post-processing advances gave little or no practical advantage (e.g., F18).

Synthetic antenna generation

- Sensor outputs are combined by a Multiple Input Single Output (MISO) multi-port network;
- Antenna gain given by the scalar product in frequency or Laplace domains between the array steering vector (i.e., sinusoidal response to a wavefront) and the network transfer function vector.



$$G(\mathbf{p}, f) = \left[H_1(f) \quad \cdots \quad H_N(f)\right]\mathbf{a}(\mathbf{p}, f)$$

Notes

- Hyperbolic wave propagation is essentially modeled by delays and attenuation.
- Sensor combination that generates a beam represents an *inversion* of the propagation equations to recover the radiated signal.
- Due to multiple, diverse sensors, inversion is much easier than for SISO systems.
- Distributed networks more suited for this task.
- Lumped networks (generally LC) may require a high number of reactive elements for a good wide-band approximation.

Basic analog combining operations

- **Narrow-band**
  - Phase shift and weighted sum of sensor outputs.
  - Valid only at a single frequency.
- **Wide-band**
  - Delay and sum network.
  - Continuous delays allowed in principle.

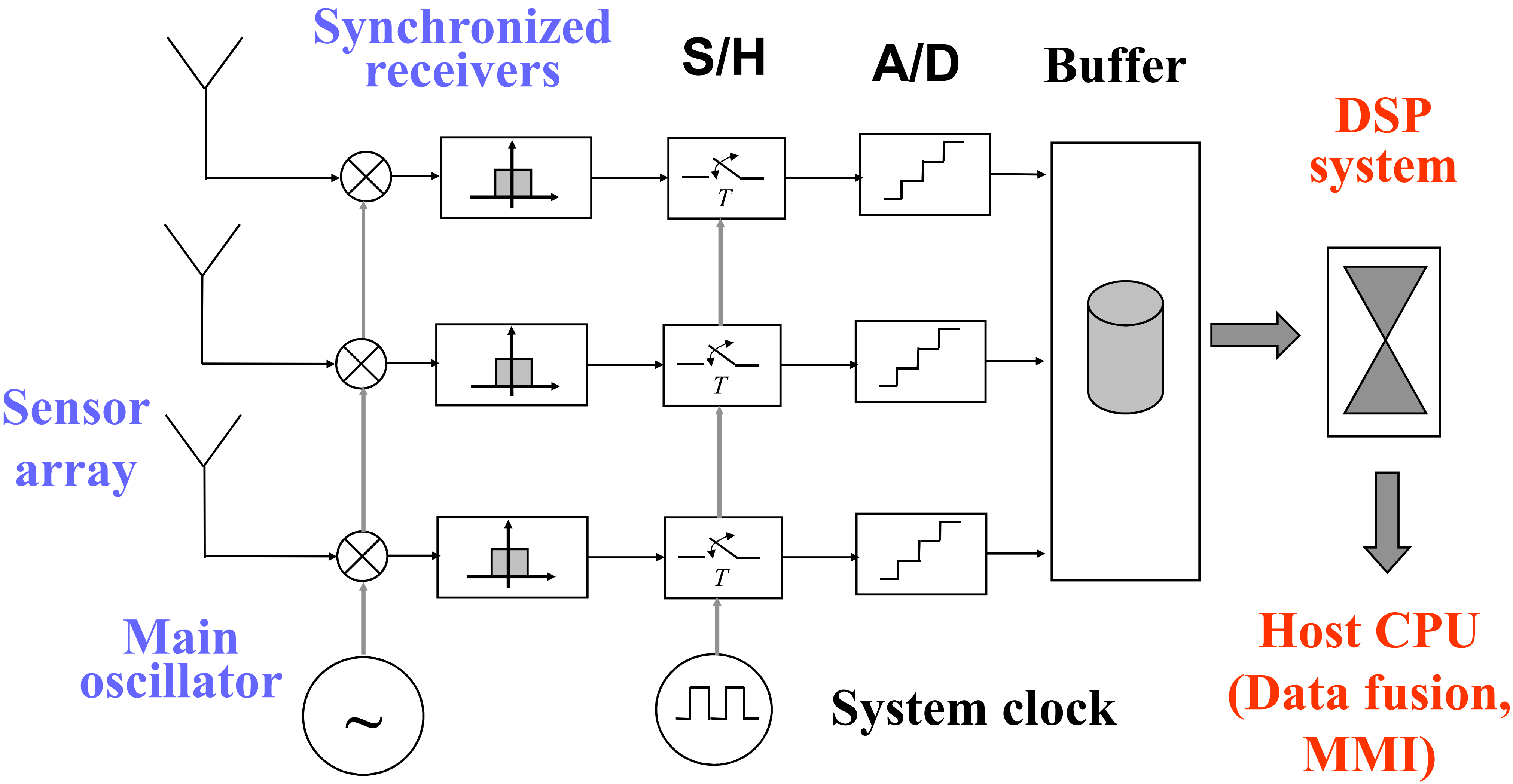
$$y(t) = \sum_{k=1}^{K} A_k e^{j\varphi_k} x_k(t)$$

$$y(t) = \sum_{k=1}^{K} \sum_{p=1}^{P} a_{kp} x_k\left(t - \tau_{kp}\right)$$

Sensor interconnection network

- Early array sensors outputs were simply combined by LC delay lines and amplifiers up to the VHF band (Hubbard, 1921) and by waveguides and *phase shifters* above to create a fixed or programmable synthetic beam (as seen in the beamforming section), giving origin to the *passive phased-array*.
- In the last decades, passive mixing networks were replaced by integrated (MMIC, MEMS), active network for noise behavior and hardware modularity.
- Today each sensor, maybe specialized for transmission or reception, is preferably connected to a *separate receiver* and ADC converter chain (*parallel, active array*) for subsequent DSP.
- *Parallel array receivers* preserve more statistical information, in particular the information about the actual steering vectors.
- Very large arrays (radar, sonar) may still require passive or active combination of some sensor subsets (*sub-arrays*) to reduce hardware and computational costs.

Digital parallel array receiver

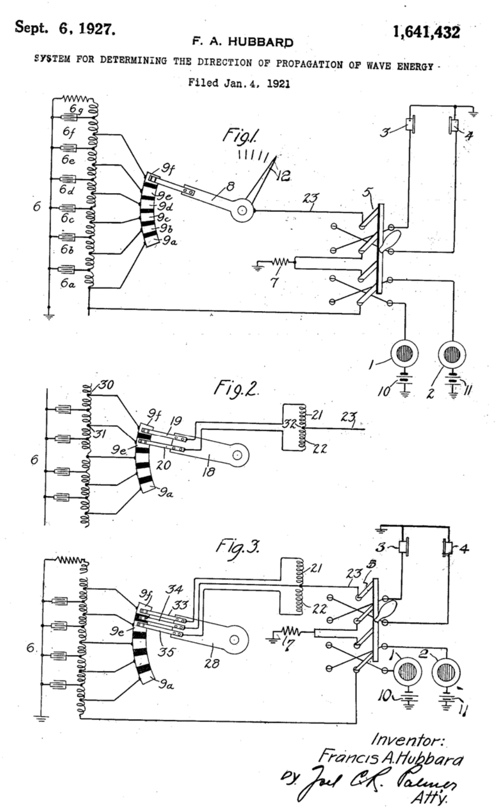


SAPIENZA
UNIVERSITÀ DI ROMA

Receiver building blocks

- *Delay lines* or *phase shifters* modify and/or calibrate the steering vectors and the signal model, and create beams.
- RF *mixers* and *amplifiers* convert sensor outputs into baseband signal envelopes and change instantaneous bandwidth and operating frequency of the array.
- *Sampling and ADC devices* are driven by a *clock* distribution network.
- *Basic DSP* (DFT, filters) is used for signal preconditioning after digital conversion and sample storage.

LC delay lines



- Inductors and capacitors can be inserted or excluded by mechanical switches, relays, PIN diode matrices.
- Still used in electrostatic loudspeakers and up to VHF.
- Similar architecture is used for load matching in HF/VHF radios.

Electro-mechanical delay lines

- A piezo-mechanic sensor transforms the electric signal into acoustic vibrations.
- In a piezo-electric material, the electric field is proportional to local strain.
- Acoustical signal propagation speed is reduced.
- Convenient delay is obtained by acoustic propagation through *surface acoustic wave devices* (SAWs), quartz blocks, aerial baffles.
- At the end, the signal is converted back into electric form by another piezo-mechanic transducer (also used as pressure sensor in acoustic arrays).

Phase shifters

- Waveguides filled with ferroelectric or ferromagnetic material (almost constant group delay, good for UWB) and microstripes slow down wave propagation speed. In many cases a continuously variable delay is possible.
- Fiber optic, single conductor waveguides: no TEM mode, variable group delay (dispersion) with frequency. Suited for narrow-band applications.
- Two-conductors waveguides, commuted by PIN diodes, have very low dispersion for wide-band uses.
- Today in active arrays fixed phase shifters can be easily and preferably used for sub-array combining.

Phase shifter design issues

- Passive phase shifters attenuate signals before amplifiers, bad for sensitivity and SNR!
- Network design similar to that of generalized FIR filters, but with less control on response and sidelobes due to irregular delay lattices.
- Quantized (digital) phase shifters designs, common in radar applications, require complex optimization routines of uncertain convergenge to the global minimum (genetic, simulated annealing).
- Phase shifters reduce cabling costs w.r.t. fully parallel receivers.

Passive vs. active phase shifters

- Active phase shifters contain integrated amplifier stages.
  - Magnitude and phase weighting (more freedom);
  - Non reciprocal, uni-directional (transmission or reception);
  - Signal gain;
  - Power requirements.
- Passive phase shifters;
  - Essentially change phase (like optical lenses).
  - Reciprocal (LC ladders, isotropic, homogeneous filling) for transmission and reception or non-reciprocal;
  - Signal loss;
  - Minimal power requirements.
  - Interferences due to impedance mismatches.

SAPIENZA
UNIVERSITÀ DI ROMA

Mixer and amplifiers

- Main problems for these components are:
  - Noise factor;
  - *Parameter mismatches* and *response fluctuation* with time, especially within analog linear combination networks;
  - Challenging signal ranges must be covered across the array (field peaks and nulls within wavelengths);
  - Non-linear *inter-modulation distortion* introduces spurious and statistically coupled sum-difference frequency components in wide-band receivers.
- Real time checks:
  - Faulty channels provide spurious signals hard to detect in subsequent processing;
  - AGC must be precisely tracked and recorded on every sensor to ensure signal model consistency.

Basic array receiver types

- Homodyne receivers (complex carrier, two ADCs):
  - Low carrier frequency, easy to realize with DDS.
  - Highly sensitive to I/Q calibration errors (overlapped spectral images);
  - Still preferred in the USA for RF communications given tight tolerances.
- Heterodyne receivers (real carrier with IF, one ADC):
  - Higher carrier frequencies;
  - Insensitive to spectral aliases;
  - Require low-cost analog filters, S/H and ADC;
  - Extremely low phase noise oscillators (such as fractional PLLs);
  - Higher signal range;
  - Generally IF digital sampling.
- Baseband receivers in acoustics.

Array receiver requirements

- Carriers for demodulation must be delayed and scaled copies of a single signal.
- Power splitting techniques for small arrays and clock regeneration by PLL for large arrays are generally used.
- For astronomical search operations, that use even sensors on satellites, radio synchronization is adopted.
- The steering vector is modified by the receiver transfer functions and internal reflections, that should be *perfectly known* for accurate operations.

Digitization issues

- Heavy ADC requirements:
  - High conversion ranges and rates: flash, sigma-delta and successive approximation ADC with 6-16 bits, 100K-2Gsample/s in RF applications;
  - Static and dynamic (for any subset of codes spanned within a small time interval) monotonicity of conversion;
  - *Dynamic range much greater* for arrays than for single receivers because of wavefront spatial interference;
  - AGC has to track huge signal changes;
  - High spurious free range (SFR) to avoid false detections;
  - Stability of transfer characteristics with time;
  - Low aperture jitter;
  - High power requirements: main challenge for costs and sizes.

Synchronization

- All ADCs must operate with a *synchronous clock*.
- Single clock source and *power-splitters* work for small array size.
- *PLL regenerated clocks* used for larger arrays.
- Synchronization errors between sensors are equivalent to a movement of the array sensors in the direction of the impinging wavefront.
- Fixed (mean) delay errors can be compensated by calibration.
- Random delay errors (aperture jitter) raise essentially the noise floor with frequency.

SAPIENZA
UNIVERSITÀ DI ROMA

Classical array modeling issues

- Ignoring mutual coupling effects (i.e., assuming *diagonal* admittance matrix) leads to gross errors in processing.
- Actual steering vector much different from the theoretical one, but recoverable by *calibration*, but correction tables are memory and computation time consuming.
- Large changes of mutual coupling with frequency and/or direction of wave-front arrivals.
- Algorithms tailored to specific, simple array geometries and sensor directivity may not work reliably with closely spaced antennas.
- Mutual coupling is small among microphones, piezo-electric sensors and hydrophones.

Part II

ALGEBRAIC AND STATISICAL CONCEPTS REVIEW

Numerical optimization

- Array processing and design is based on the solution of a set of *numerical optimization* problems.
- Optimization means obtaining a set of *system parameters* (direction/time of arrivals, frequency response, interpolation coefficients,…) or *signal parameters* (voltages, variance, mean…) from a set of linear or non-linear equations satisfying a set of conditions.
- A *cost functional* is derived and optimized from theoretical paradigms to find the best solution among the feasible solutions of the equation set.

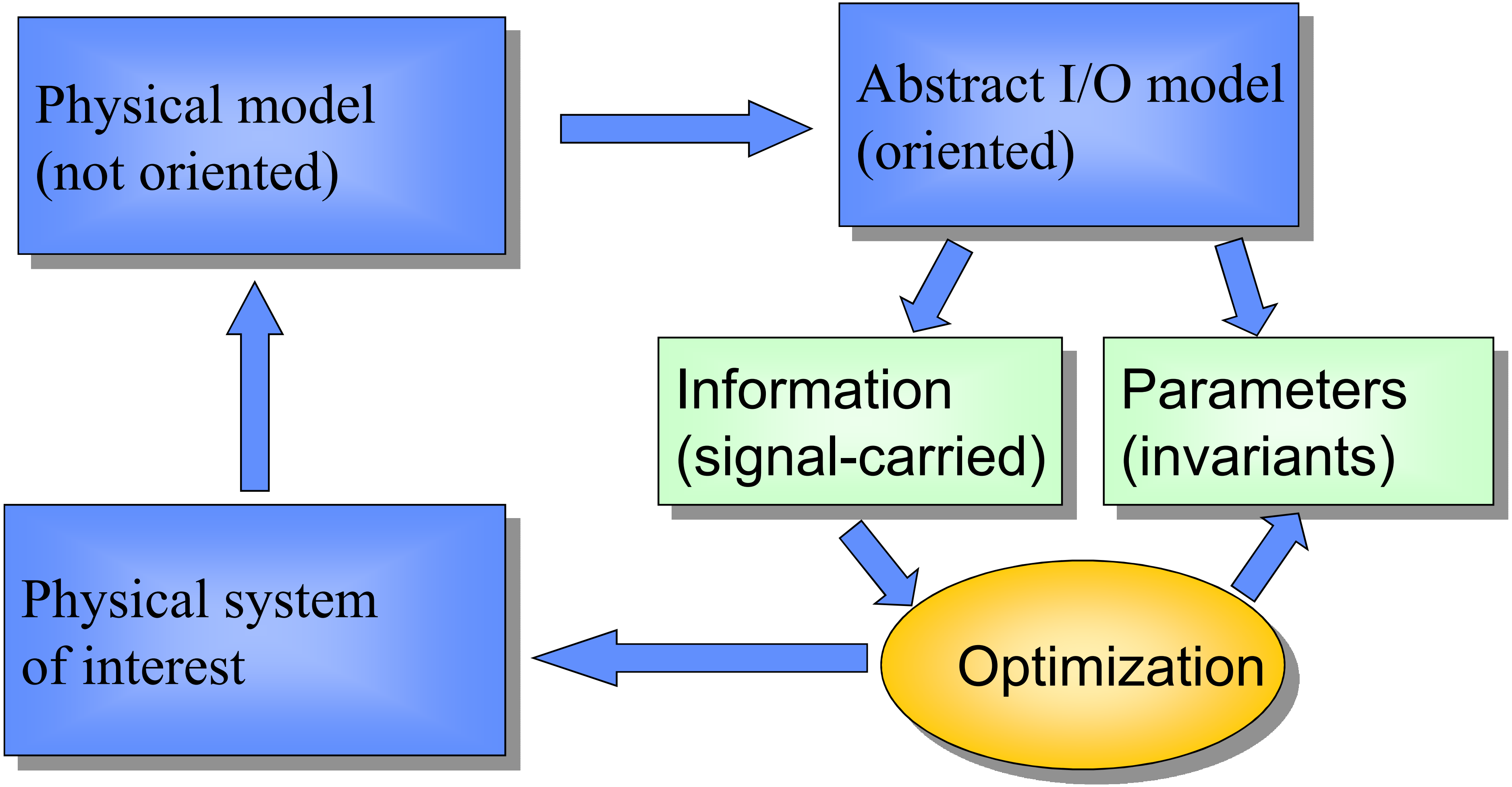
Optimization for signal processing

- Cost functionals are tied to statistical properties of signals and free design parameters of system equations
- If signals are *not stationary* or system equations are *not permanent*, also *optimal parameters (slowly) change with time.*
- The rate of system changes should be much lower than the minimun frequency carried by signals, otherwise the equation set is inconsistent.
- Parameter updating is advantageously based on *information* carried by recently incoming signals.

Optimization problem formulation

- Building of a *non-oriented* system model based on physical equations and individuating information-carrying *input and output signals*.
- Solve the non-oriented equation set so to make input-output relationships explicit, obtaining an *abstract, oriented model*.
- Define a cost functional based on signals, where system parameters are the unknowns;
- Numerically find the global minimum (or maximum) of the cost functional.

Parametric model abstraction



Array parametric model building elements

Acoustic propagation equations

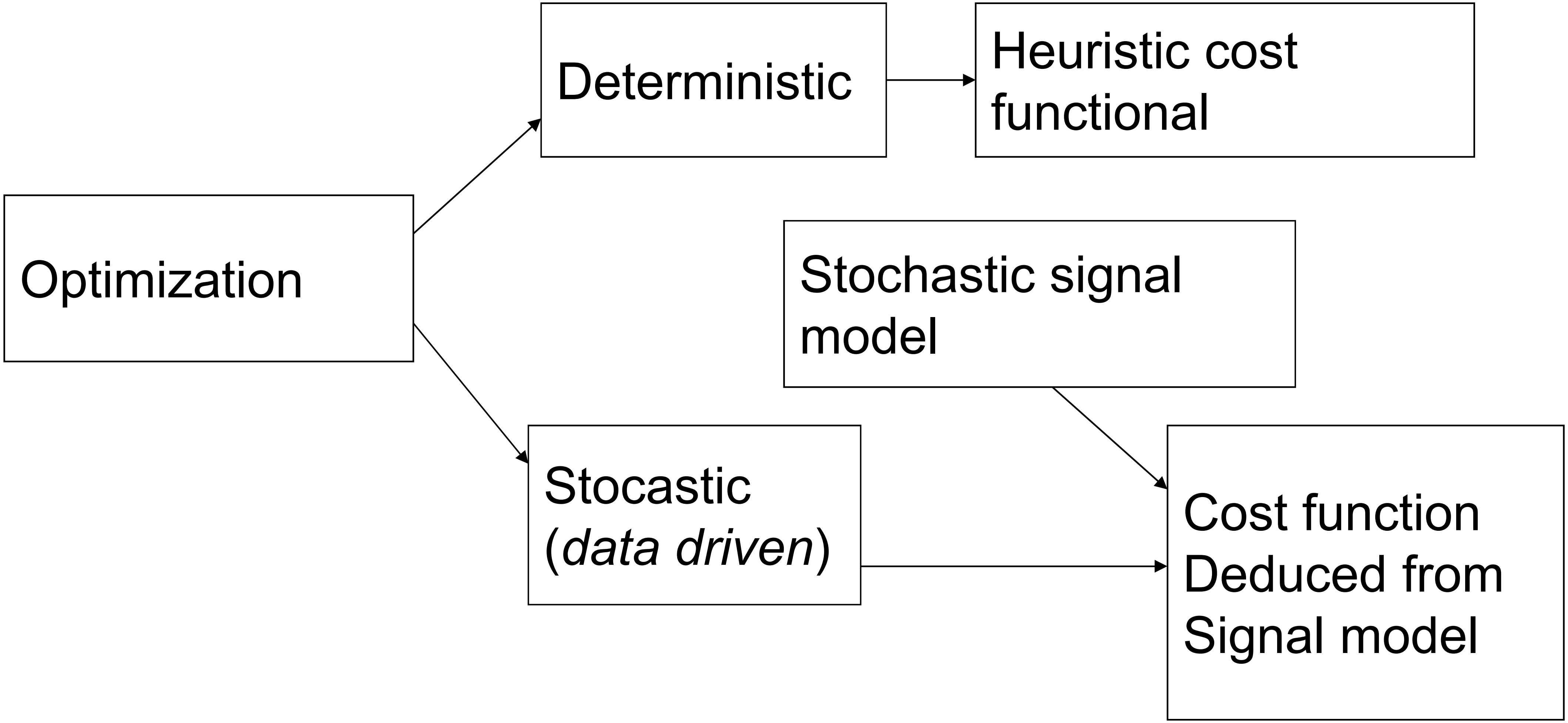
Geometric and physical Constraints (graph)

Electromagnetic propagation equations

Device equations

Signal transfer equations

Problemi di ottimizzazione



Example: deterministic optimization (e.g., FIR filter design)

Cost functional

$$J(\mathbf{w}) = \int_0^{\pi} f\left(\left|H_d(j\omega) - H(j\omega, \mathbf{w})\right|\right) d\omega$$

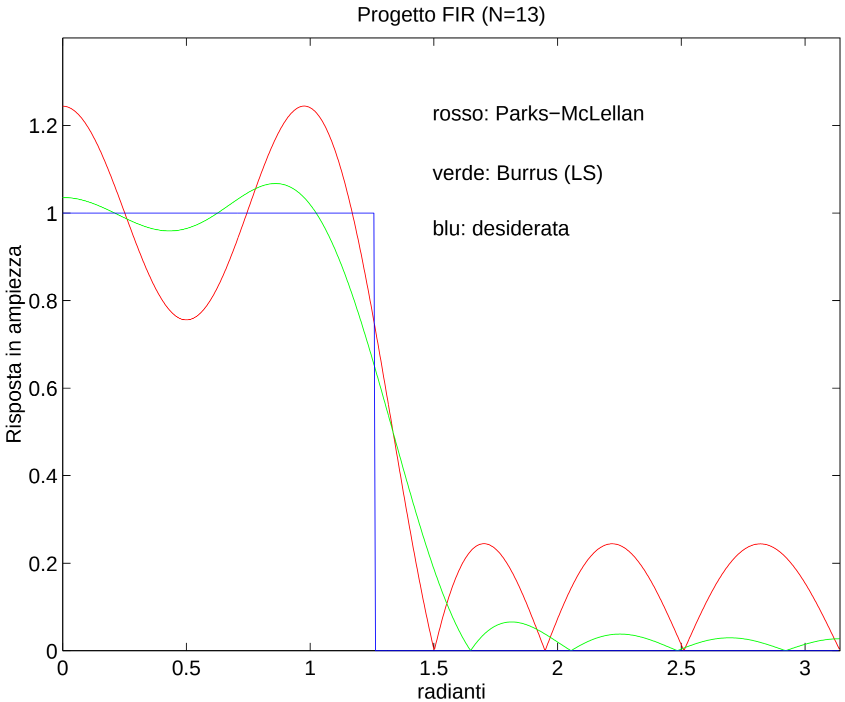
Error

Free parameters

Desired response

Actual response

Error weight function



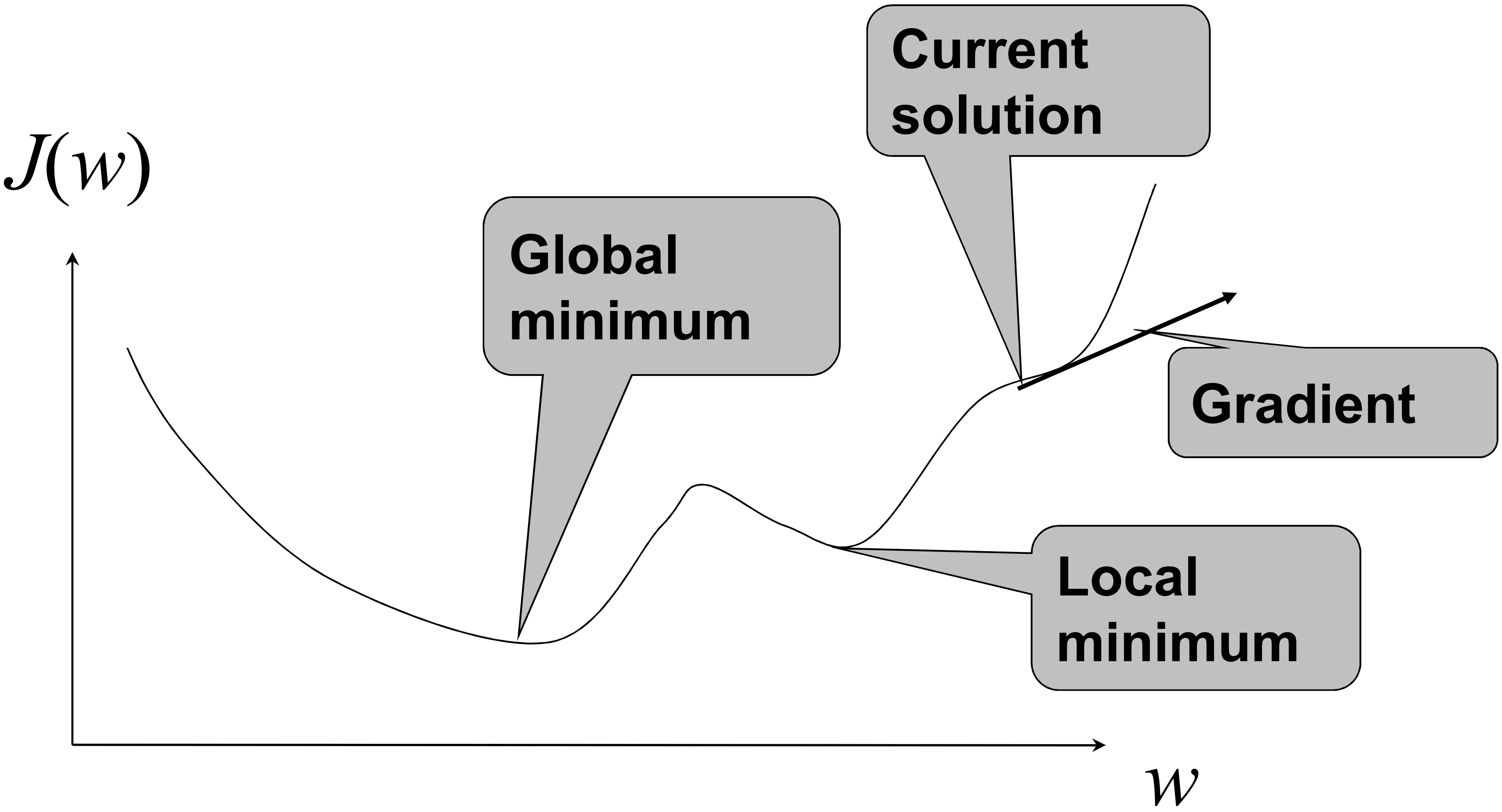
$$\boxed{\mathbf{w}_{opt} = \arg\min_{\mathbf{w}} \left[J(\mathbf{w}) \right]}$$

Stochastic optimization

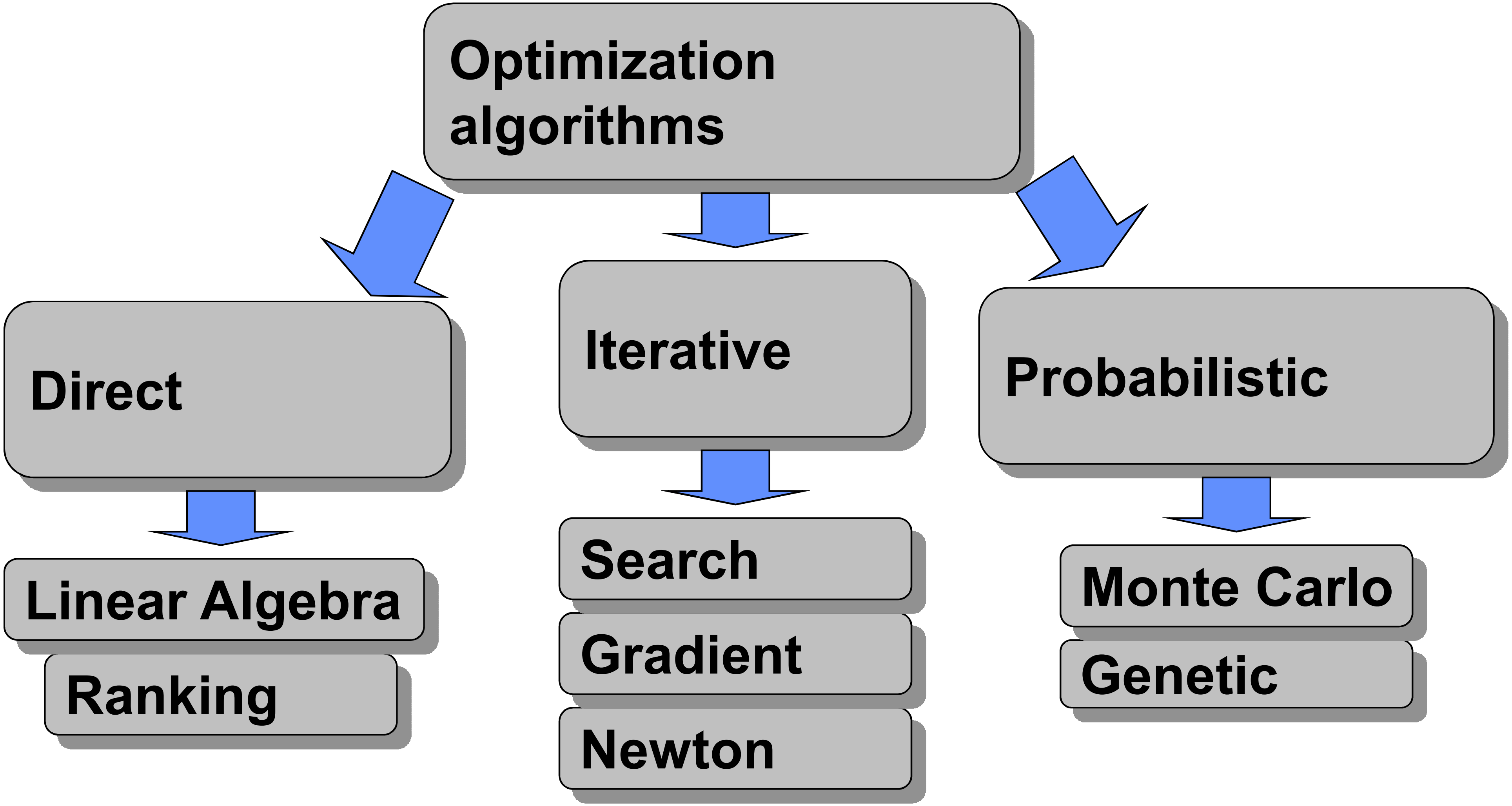
- Stochastic optimization mainly uses input and output signals as a basis.
- Minimize or maximize the cost functional by modifying free model parameters.
- *Generally non-linear injective function* from the signal space to the much lower-dimensional parameter space.
- Some parameters may be *un-observable* from system output identification.
- Ensure *paremeter identifiability* issues before proceeding.

$$\boxed{\Theta_{opt} = T\big(\mathbf{u}(n), \mathbf{y}(n)\big)}$$

Cost functional minimization



Classification



Direct algorithms

- Terminated in a *finite number of steps*.
- Include many *linear algebra* techniques (Gauss, Cholesky, Gram-Schmidt, QR, DCT, FFT, DWT,…) and rank-based ordering (e.g., median filter).
- Solve *few, but important problems*.
- Subject to *round-off* errors.
- Suited for *small-sized problems* (generally less than 100 unknowns).
- Fundamental *building blocks* for more general optimization techniques.

SAPIENZA
UNIVERSITÀ DI ROMA

Probabilistic algorithms

- Ensure a *guaranteed probability* of finding the optimal solution given available time and computing resources.
- Try and err approach based on multiple evaluations of the cost functional for varying parameters.
- Suited for *big data problems* (>100,000 unknowns) or for strongly non-linear functionals.
- *Very slow convergence, not suited for real time* signal processing.

SAPIENZA
UNIVERSITÀ DI ROMA

Iterative algorithms

- Start from a *coarse solution* and refine it on the basis of a *local model* of the functional.
- Include fundamental processing blocks for signal processing, like SVD and EVD.
- Suited for medium to large sized problems (100-10.000 equations).

Search methods

- Start from a tentative coarse solution.
- Functiona evaluated on a local grid of points in the parameter space.
- Find the parameter changes that locally minimize the functional.
- Iterate until solution does not appreciably change (*convergence*).
- *Simple and costly* algorithms.
- Too slow for signal processing.

Descent methods

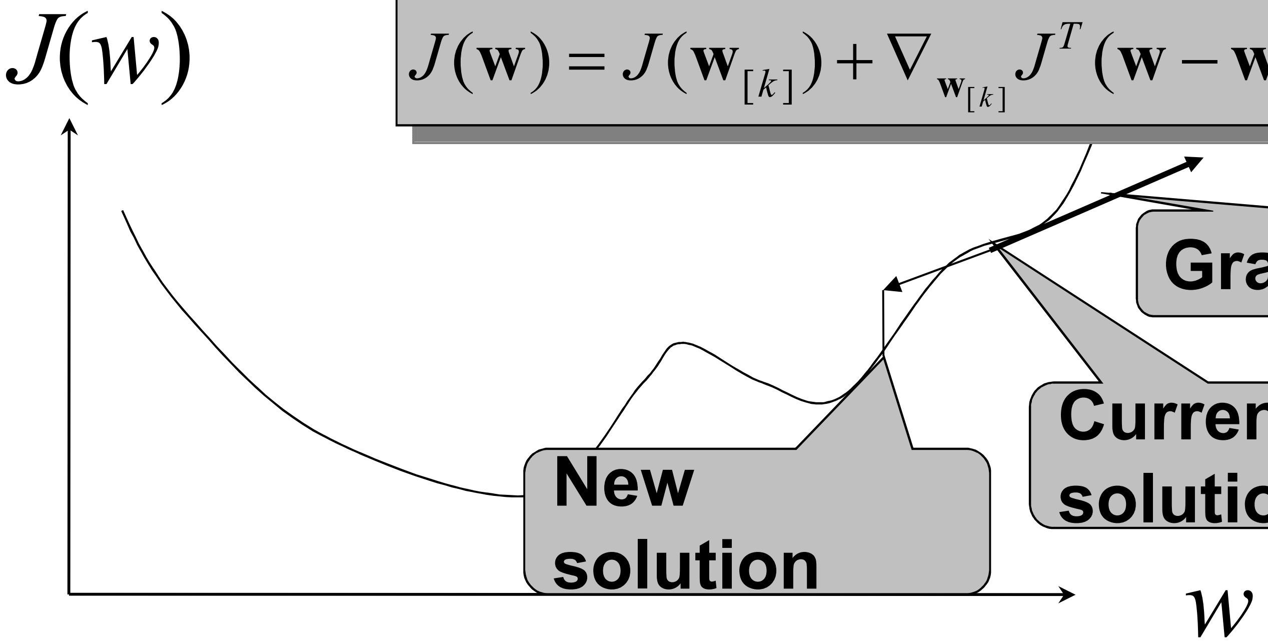
- Heuristic methods based on the *multivariate Taylor* expansion of the cost functional around the (supposed) global minimum.
- Iteratively refine a starting solution on the basis of the local *gradient* of the functional w.r.t. free parameters.
- *Second order faster Newton type* techniques require the additional knowlenge of the *Hessian* (second derivative matrix) of the functional or an its *close approximation (Hessian replacement matrix)*.
- Find the local extremum closest to the starting guess solution.
- The Hessian or its replacement must be *positive definite* within a compact region around the local minimum for convergence.
- Non positive Hessian means (local) parameter identifiability loss!

Gradient method (1st order)

$$\mathbf{w}_{[k+1]} = \mathbf{w}_{[k]} - \eta_{[k]} \nabla_{\mathbf{w}_{[k]}} J$$

$$\nabla_{\mathbf{w}_{[k]}} J = \left[\left.\frac{\partial J(\mathbf{w})}{\partial w_1}\right|_{\mathbf{w}=\mathbf{w}_{[k]}} \quad \cdots \quad \left.\frac{\partial J(\mathbf{w})}{\partial w_P}\right|_{\mathbf{w}=\mathbf{w}_{[k]}} \right]^T$$

$$J(\mathbf{w}) = J(\mathbf{w}_{[k]}) + \nabla_{\mathbf{w}_{[k]}} J^T (\mathbf{w} - \mathbf{w}_{[k]}) + O(\left|\mathbf{w} - \mathbf{w}_{[k]}\right|^2)$$



Complex derivatives

- Special *complex derivatives* are defined for complex vector variable optimization problems with real valued functionals.
- They must obey the chain, product, quotient rule.
- At an extremum, the *conjugate derivative* must be zero.

$$\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} = \frac{\partial f(\mathbf{w})}{\partial \operatorname{Re}(\mathbf{w})} - j\frac{\partial f(\mathbf{w})}{\partial \operatorname{Im}(\mathbf{w})}; \quad \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}^*} = \frac{\partial f(\mathbf{w})}{\partial \operatorname{Re}(\mathbf{w})} + j\frac{\partial f(\mathbf{w})}{\partial \operatorname{Im}(\mathbf{w})}$$

$$\frac{\partial\left(\mathbf{w}^H\mathbf{A}\right)}{\partial \mathbf{w}^*} = 2\mathbf{A}; \quad \frac{\partial(\mathbf{A}\mathbf{w})}{\partial \mathbf{w}^*} = \mathbf{0}; \quad \frac{\partial\left(\mathbf{w}^H\mathbf{A}\mathbf{w}\right)}{\partial \mathbf{w}^*} = 2\mathbf{A}\mathbf{w}$$

$$f \in \mathbb{R} \rightarrow 2\nabla_{\mathbf{w}} = \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}^*};$$

SAPIENZA UNIVERSITÀ DI ROMA

Newton method (2nd order)

$$\forall \mathbf{v} \neq \mathbf{0} \Rightarrow \mathbf{v}^H \mathbf{H} \mathbf{v} > 0$$

$$\mathbf{w}_{[k+1]} = \mathbf{w}_{[k]} - \eta_{[k]} \mathbf{H}^{-1} \nabla_{\mathbf{w}_{[k]}} J$$

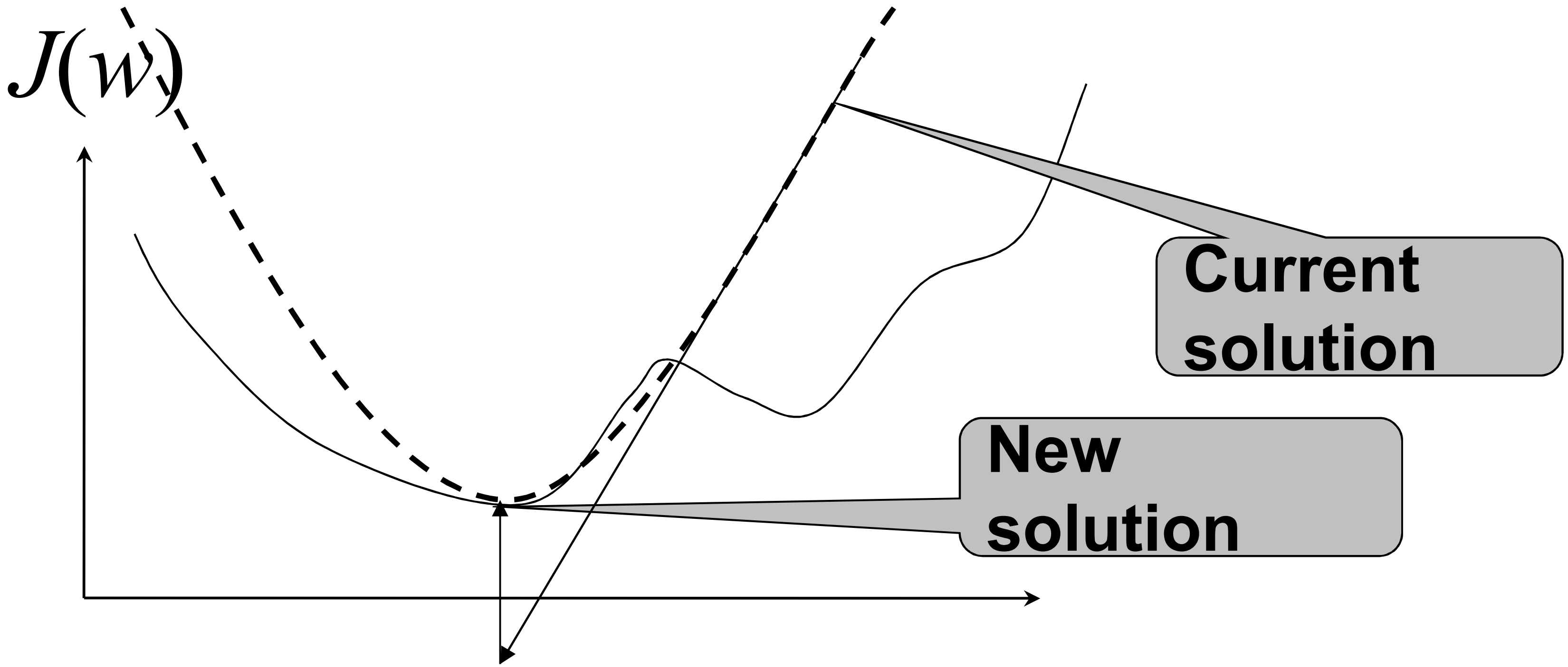
$$\mathbf{H}_{ij} = \left. \frac{\partial J(\mathbf{w})}{\partial w_i \partial w_j} \right|_{\mathbf{w}_{[k]}} \approx const$$

$$J(\mathbf{w}) = J(\mathbf{w}_{[k]}) + \nabla_{\mathbf{w}_{[k]}} J^T (\mathbf{w} - \mathbf{w}_{[k]}) +$$

$$+ \frac{1}{2} (\mathbf{w} - \mathbf{w}_{[k]})^T \mathbf{H} (\mathbf{w} - \mathbf{w}_{[k]}) + O(\left| \mathbf{w} - \mathbf{w}_{[k]} \right|^3)$$

Geometrical intepretation of the Newton method

If the Hessian is *positive definite* (paraboloid convexity toward the graph bottom), the solution *locally converges* very quickly to the global minimum:



Convergence rate

$$\beta = \lim_{k\to\infty} \frac{\left|J(\mathbf{w}_{[k+1]}) - J_{\min}\right|^{\frac{1}{p}}}{\left|J(\mathbf{w}_{[k]}) - J_{\min}\right|} \leq 1$$

- *Convergence rate* in a neighborhood of the global minimum is expressed by the *order* p and the *convergence ratio* β.
- The Newton method has $p = 2$ and $\beta \approx 1$ (*quadratic convergence*).
- Gradient descent has $p = 1$ and $\beta < (\kappa\text{-}1)^2/(\kappa+1)^2$, where κ is the *condition number* of the Hessian (ratio between maximum and minimum eigenvalues):

$$\kappa = \left|\mathbf{H}\right|_2 \left|\mathbf{H}^{-1}\right|_2 \gg 1; \kappa_{typ} \subset \left(10^2, 10^{12}\right)$$

Discussion

- Gradient descent is *very slow* for correlated inputs (i.e., ill-conditioned Hessian) and is used for little, cheap systems.
- Newton method is fast near the global minimum, but the Hessian must be positive definite *everywhere* for a *global convergence* capacity.
- Newton method should be initialized by other techniques in most practical application.
- Functionals employed in array processing are oscillating and non-convex far from the local minimum.
- No real *guarantee of convergence* in application!

Quadratic optimization and linear(ized) systems

- Most array processing problems involve *quadratic* or nearly quadratic cost functionals, so that the underlying problem is *locally linear*.
- Local *linearization* of system equations is essential for *optimization* purposes and *sensitivity and performance analysis* of array processing algorithms.
- The transfer functions between equation perturbations and parameters are sought.
- So principles of Least Squares optimization of linear equation sets is of paramount relevance.

Algebraical and statistical background

- Most math manipulations in array processing involve solving rank deficient, under- and over-determined linear systems and eigenvector-like matrix decompositions (e.g., SVD).
- Processing algorithms development is driven by their resistance to model (systematic or random) and statistical (finite sample) errors:
  - random models;
  - optimal estimation;
  - robust estimation.
- A brief review of linear algebra and statistical concepts and applications will follow.

Least Squares linear systems

- The theory of linear Least Squares is fundamental to understand the links between the *estimation theory* and the *numerical optimization* and to establish sound concepts of *statistical robustness* of sample estimates.
- In particular, Gaussian ML estimation problems and performance analysis can be very often recasted, at least *locally around true parameters*, as the solution and the sensitivity analysis of a linear LS system.

Basic LS parametric estimation problems

- *Interpolation* of a noisy target vector on a fixed vector basis (design matrix).
- *Inversion* of the noisy output of a linear system to recover system inputs.
- *Regression* between two set of noisy signals.
- The three problems have different reasoning behind and application and different statistical properties.

Linear interpolation

- The interpolation problem is the single LS problem which is completely understood.

$$x(n) = \sum_{i=1}^{P} a_{n,i} w_i + v(n); \ldots n = 1, 2, \ldots, N > P$$

System outputs

Additive noise

Sampled interpolating functions

Parameters

Linear inversion

- Tries to recover a known signal by a set of system noisy observations
- Frequently used in engineering, but not theoretically rigorous because of bias and input noise amplification.

$$y(n) = \sum_{i=1}^{P} a_i x_i(n) + v(n); \ldots n = 1, 2, \ldots, N > P$$

Known target

Fitting error

Parameters

Noisy system observations

Linear regression

- Tries to recover a *noisy* signal by a set of system noisy observations
- Used in statistics and stochastic prediction problem (AR, MA, ARMA), but not well understood.

$$y(n)=\sum_{i=1}^{P}a_i x_i(n)+v(n);\ldots n=1,2,\ldots,N>P$$

Noisy target

Parameters

Noisy system observations

Fitting error

Interpolation problem setting

- In compact (matrix) form the following *over-determined* system is derived from the above linear model:

$$\underset{(N\times P)}{\left[\mathbf{A}\right]}\underset{(P\times 1)}{\left[\mathbf{w}\right]}=\underset{(N\times 1)}{\left[\mathbf{x}\right]}-\underset{(N\times 1)}{\left[\mathbf{e}\right]}$$

$$N > P$$

$$\mathbf{A}\mathbf{w}=\mathbf{x}-\mathbf{e}(\mathbf{w})$$

Parameter estimates

Interpolation (*fitting*) error

Gaussian error model

- A stochastic model is assumed for the additive noise (equal to the fitting error, if the estimated parameter vector $\mathbf{w}$ were the true one).
- Often noise is assumed *Gaussian distributed*, *white*, with *zero mean* and *unknown variance* σ^2.
- Under these hypoteses, each interpolating equations constitutes a *statistically independent* observation of the analyzed system.

Gaussian Likelihood

- Problem unknowns are the parameter vector $\mathbf{w}$ and the noise variance.

$$f_v(v) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{v^2}{2\sigma^2}}; \quad e(n,\mathbf{w}) = x(n) - \sum_{i=1}^{P} a_{n,i} w_i; \quad \mathbf{w} \triangleq \begin{bmatrix} w_1 \\ \vdots \\ w_P \end{bmatrix}$$

$$L_N(\sigma,\mathbf{w}) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{e(n,\mathbf{w})^2}{2\sigma^2}}$$

Concentrated Maximum Likelihood (ML) estimation

- The negative Gaussian log likelihood is first minimized w.r.t. $\mathbf{w}$:

$$l_N(\sigma^2, \mathbf{w}) = -\log L_N(\sigma^2, \mathbf{w})$$
$$= const + \frac{N}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{n=1}^{N}\left|e(n, \mathbf{w})\right|^2$$

$$\mathbf{w}_{ML} = \arg\min_{\mathbf{w}}\left(\left|\mathbf{e}(\mathbf{w})\right|_2^2\right) = \arg\min_{\mathbf{w}}\left(\left|\mathbf{x} - \mathbf{A}\mathbf{w}\right|_2^2\right) = \mathbf{w}_{LS}$$

ML and Least Squares

- For this problem the ML estimate of interpolation coefficients coincides with the LS one.
- The estimate of the additive noise variance requires a little more attention:

$$\sigma_{ML}^2 = \frac{1}{N}\left|\mathbf{e}(\mathbf{w}_{LS})\right|_2^2$$

ML estimate, biased for finite N

$$\sigma_{ML2}^2 = \frac{1}{N-P}\left|\mathbf{e}(\mathbf{w}_{LS})\right|_2^2$$

Unbiased estimate

Separable ML estimates

- The LS problem arises from the elimination (concentration) of the parameter vector of the Gaussian log-likelihood.
- The noise variance estimate is *not independent* of the other parameter estimates.
- However the sample error of the variance estimate is a consequence of the parameter estimation errors: this ML problem is therefore said to be *separable*.
- However the LS solution is valid even when $\sigma^2=0$ and the Gaussian log-likelihood becomes singular.

LS algorithms: normal equations

- The LS solution satisfies the following Hermitian system of *normal equations* if the rank of the matrix $\mathbf{A}$ is full.

$$\mathbf{A}^H \mathbf{A} \mathbf{w}_{LS} = \mathbf{A}^H \mathbf{x}$$

$$\mathbf{w}_{LS} = \left(\mathbf{A}^H \mathbf{A}\right)^{-1} \mathbf{A}^H \mathbf{x} = \mathbf{A}^{\dagger} \mathbf{x}$$

Moore-Penrose pseudoinverse of $\mathbf{A}$

LS solution sensitivity

- The LS solution by normal equations has severe risks of numerical instability. In particular, the LS solution insensitivity is expressed through the *condition number* κ (ratio betweem maximum and minimum eigenvalues) of $\mathbf{A}^H\mathbf{A}$, which is of order 10^5-10^{12} for radar and acoustic signals!

$$\left(\mathbf{A}^H\mathbf{A}+\mathbf{E}\right)\left(\mathbf{w}+\Delta\mathbf{w}\right)=\mathbf{A}^H\mathbf{x}$$

$$\varepsilon \triangleq \frac{\left|\mathbf{E}\right|}{\left|\mathbf{A}^H\mathbf{A}\right|} \Rightarrow \frac{\left|\Delta\mathbf{w}\right|}{\left|\mathbf{w}\right|} \leq \varepsilon\cdot\kappa\left(\mathbf{A}^H\mathbf{A}\right)$$

Orthogonalization methods

- To overcome the numerical stability problem, LS techniques based on proper *orthogonalization* of the columns of $\mathbf{A}$ are used.
- These methods act directly on data, without building Hermitian sample correlation matrices and are said of the *square root* type.
- In parallel to the algorithmic aspects, these techniques shed light on the stochastic influence of single observations (equations) on the global estimate.

QR decomposition (QRD)

$\mathbf{Q}_1$ has the same column span of $\mathbf{A}$

$$\underset{(N\times P)}{\left[\mathbf{A}\right]} = \underset{(N\times N)}{\left[\mathbf{Q}\right]}\underset{(N\times P)}{\left[\mathbf{R}\right]}$$

$$\mathbf{Q}^H\mathbf{Q} = \mathbf{I}_N ; \mathbf{Q} = \left[\underset{(N\times P)}{\mathbf{Q}_1} \;\middle|\; \underset{[N\times(N-P)]}{\mathbf{Q}_2}\right]; \quad \mathbf{Q}_2^H\mathbf{A} = \mathbf{0}$$

$$\mathbf{R} = \begin{bmatrix} \underset{(P\times P)}{\mathbf{R}_1} \\ \hline \underset{[(N-P)\times P]}{\mathbf{0}} \end{bmatrix}; \quad \underset{(3\times 3)}{\mathbf{R}_1} = \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}$$

Triangular *Cholesky* factor of $\mathbf{A}^H\mathbf{A}$

SAPIENZA
UNIVERSITÀ DI ROMA

Full rank LS-QR solution

$$\left[\begin{array}{c|c} \underset{(N\times P)}{\mathbf{A}} & \underset{(N\times 1)}{\mathbf{x}} \end{array}\right]\left[\begin{array}{c} \underset{(P\times 1)}{\mathbf{w}} \\ \hline -1 \end{array}\right] = \mathbf{QR}\left[\begin{array}{c} \underset{(P\times 1)}{\mathbf{w}} \\ \hline -1 \end{array}\right] = -\mathbf{e}; \quad \left|\mathbf{Q}^H\mathbf{e}\right|_2 = \left|\mathbf{e}\right|_2$$

$$\mathbf{R}\left[\begin{array}{c} \underset{(P\times 1)}{\mathbf{w}} \\ \hline -1 \end{array}\right] = \left[\begin{array}{c|c} \underset{(P\times P)}{\mathbf{R}_{11}} & \underset{(P\times 1)}{\mathbf{r}_{12}} \\ \hline \underset{(1\times P)}{\mathbf{0}} & \underset{(1\times 1)}{\sigma} \\ \hline \underset{[(N-P-1)\times P]}{\mathbf{0}} & \underset{[(N-P-1)\times 1]}{\mathbf{0}} \end{array}\right]\left[\begin{array}{c} \underset{(P\times 1)}{\mathbf{w}} \\ \hline -1 \end{array}\right] = -\left[\begin{array}{c} \underset{(P\times 1)}{\mathbf{e}_1} \\ \hline \underset{(1\times 1)}{\mathbf{e}_2} \\ \hline \underset{[(N-P-1)\times 1}{\mathbf{e}_3} \end{array}\right]$$

Full rank LS solution by QRD

- Three equation subsets are visible after the QRD.

$$\mathbf{R}_{11}\mathbf{w}_{LS} = \mathbf{r}_{12} \Rightarrow \mathbf{w}_{LS} = \mathbf{R}_{11}^{-1}\mathbf{r}_{12}$$

$$\left[\underset{(1\times P)}{\mathbf{0}}\right]\mathbf{w}_{LS} \neq \sigma$$

$$\left[\underset{[(N-P-1)\times P]}{\mathbf{0}}\right]\mathbf{w}_{LS} = \left[\underset{[(N-P-1)\times 1]}{\mathbf{0}}\right]$$

P equations are satified by the full rank hypothesis.

One equation is satisfied if and only if $\sigma=0$.

N-P-1 equations are *always satisfied.*

SAPIENZA
UNIVERSITÀ DI ROMA

LS solution property

- The LS fitting error is *orthogonal* to the column space of A (*orthogonality principle*).

The LS interpolated signal is the orthogonal projection of $\mathbf{x}$ on the column space of $\mathbf{A}$

$$\underset{(N\times P)}{\mathbf{A}} \overset{R-QRD}{=} \underset{(N\times P)}{\mathbf{Q}_1}\underset{(P\times P)}{\mathbf{R}_{11}}$$

$$\mathbf{x}_{LS} = \mathbf{Q}_1\mathbf{Q}_1^H\mathbf{x} = \mathbf{P}_A\mathbf{x}$$

$$\mathbf{e}(\mathbf{w}_{LS}) = \left(\mathbf{I}_N - \mathbf{Q}_1\mathbf{Q}_1^H\right)\mathbf{x}$$

$$\mathbf{e}(\mathbf{w}_{LS})^H\mathbf{A} = \mathbf{0}$$

Reduced size QRD of $\mathbf{A}$

$\mathbf{P}_A$ is the orthogonal projection matrix of $\mathbf{A}$

Expected value of the LS solution

- If the noise is *zero mean*, the mean of the LS solution is the true $\mathbf{w}$, hence the *LS estimate is unbiased*.

$$\mathbf{x} \triangleq \mathbf{s} + \mathbf{n} = \mathbf{A}\mathbf{w}_0 + \mathbf{v}$$

$$\mathbf{w}_{LS} = \mathbf{A}^{\dagger}\mathbf{x} = \mathbf{w}_0 + \mathbf{A}^{\dagger}\mathbf{v}$$

$$E[\mathbf{w}_{LS} - \mathbf{w}_0] = \mathbf{A}^{\dagger}E[\mathbf{v}]$$

$$E[\mathbf{v}] = \mathbf{0} \Rightarrow E[\mathbf{w}_{LS}] = \mathbf{w}_0$$

$\mathbf{v}$ is the random noise vector superposed on $\mathbf{s}$, assumed as *perfectly interpolable* by $\mathbf{A}$

Estimation covariance of the LS solution

- If the noise is uncorrelated (white) among observations, the LS solution covariance only depends on the noise variance and the structure of the design matrix $\mathbf{A}$.

$$E[\mathbf{v}] = 0 \Rightarrow Cov[\mathbf{v}] = E[\mathbf{v}\mathbf{v}^H] = \mathbf{R}_{vv}$$

$$Cov[\mathbf{w}_{LS}] \triangleq E[(\mathbf{w}_{LS} - \mathbf{w}_0)(\mathbf{w}_{LS} - \mathbf{w}_0)^H]$$

$$Cov[\mathbf{w}_{LS}] = \mathbf{A}^{\dagger}\mathbf{R}_{vv}\left(\mathbf{A}^{\dagger}\right)^H$$

$$\mathbf{R}_{vv} = \sigma^2\mathbf{I}_N \Rightarrow Cov[\mathbf{w}_{LS}] = \sigma^2\left(\mathbf{A}^H\mathbf{A}\right)^{-1} = \sigma^2\left(\mathbf{R}_{11}^H\mathbf{R}_{11}\right)^{-1}$$

Directional properties of the LS solution

- The eigenvectors of $\mathbf{A}^H\mathbf{A}$ define special combinations (*directions*) of the parameters: the estimation error covariance among these combinations is diagonal and inversely proportional to the eigenvalues of $\mathbf{A}^H\mathbf{A}$.

$$\mathbf{A}^H\mathbf{A}\mathbf{c}_k = \lambda_k \mathbf{c}_k; \quad \lambda_k \geq 0$$

$$\lambda_k \neq \lambda_j \Rightarrow \mathbf{c}_j^H \mathbf{c}_k = \delta_{jk}$$

$$Cov[\mathbf{w}_{LS}] = \sum_{k=1}^{P} \left(\frac{\sigma^2}{\lambda_k} \right) \mathbf{c}_k \mathbf{c}_k^H$$

The symmetrized interpolation (or *design*) matrix is *positive semidefinite*

SAPIENZA
UNIVERSITÀ DI ROMA

Design matrix choice issues

- The ideal design matrix $\mathbf{A}^H\mathbf{A}$ must have the lowest possible condition number (i.e., 1), hence $\mathbf{A}$ must be *orthogonal*.
- It is important to discard high variance directions (at the possible *expense of bias*) and/or reparametrize the problem.
- This is the basic idea behind the *Principal Component Analysis (PCA)*.

Best Linear Unbiased Estimator (BLUE)

- The LS estimator is a *linear transformation* (i.e., orthogonal projection) of the observation vector $\mathbf{x}$.
- It is shown that it is also the *best* (minimun variance) *linear unbiased estimator* (BLUE) of $\mathbf{w}$ if the noise has zero mean and finite variance.
- In fact, let us consider an alternate estimator of $\mathbf{w}$ which satisfies:

$$\mathbf{w}_* = \mathbf{B}\mathbf{x} = \mathbf{B}(\mathbf{A}\mathbf{w}_0 + \mathbf{v})$$

Unbiasedness condition

- Any *linear, unbiased estimator* must satisfy the third condition:

$$E[\mathbf{w}_*] = \mathbf{BAw}_0 + \mathbf{B}E[\mathbf{v}]$$

$$E[\mathbf{v}] = \mathbf{0} \Rightarrow E[\mathbf{w}_*] = \mathbf{BAw}_0$$

$$E[\mathbf{w}_*] = \mathbf{w}_0 \Rightarrow \mathbf{BA} = \mathbf{I}_P$$

Linear estimator covariance

- The covariance of a generic linear estimator has the same form as the covariance of the LS estimator.

$$Cov(\mathbf{w}_*) = E[(\mathbf{w}_* - \mathbf{w}_0)(\mathbf{w}_* - \mathbf{w}_0)^H]$$

$$E[\mathbf{v}] = \mathbf{0} \Rightarrow Cov(\mathbf{v}) = E[\mathbf{v}\mathbf{v}^H] = \mathbf{R}_{vv}$$

$$\mathbf{R}_{vv} = \sigma^2 \mathbf{I}_N \Rightarrow Cov(\mathbf{w}_*) = \sigma^2 \mathbf{B}\mathbf{B}^H$$

Best linear unbiased estimator

- Let us consider the following positive semidefinite matrix obtained by the following symmetrization:

$$\mathbf{C} \triangleq \sigma^2 \left[(\mathbf{B} - \mathbf{A}^\dagger)(\mathbf{B} - \mathbf{A}^\dagger)^H \right] \geq \mathbf{0}$$

$$\mathbf{C} = \sigma^2 \left[\mathbf{B}\mathbf{B}^H + \left(\mathbf{A}^H\mathbf{A}\right)^{-1} - \mathbf{B}\mathbf{A}\left(\mathbf{A}^H\mathbf{A}\right)^{-1} - \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\left(\mathbf{B}\mathbf{A}\right)^H \right]$$

$$\mathbf{B}\mathbf{A} = \mathbf{I}_P \Rightarrow \mathbf{C} = \sigma^2 \left[\mathbf{B}\mathbf{B}^H - \left(\mathbf{A}^H\mathbf{A}\right)^{-1} \right] \geq \mathbf{0}$$

$$Cov(\mathbf{w}_*) - Cov(\mathbf{w}_{LS}) \geq \mathbf{0}$$

$$Cov(\mathbf{w}_*) = Cov(\mathbf{w}_{LS}) \Rightarrow \mathbf{B} = \mathbf{A}^\dagger \Rightarrow \mathbf{w}_* = \mathbf{w}_{LS}$$

Comments on the BLUE estimator

- Zero mean and finite variance properties of noise do not imply the assumption of any special probability distribution.
- In any case, the LS estimate is the BLUE.
- For non Gaussian noise (e.g. Laplacian noise), the asymptotically best (more efficient) unbiased estimator is the *Maximum Likelihood* (ML) one and differs from the LS estimator, but it must be *non-linear*!

Statistical impact of each equation

Estimation error normalized w.r.t. all directions (*isotropic*)

$$\mathbf{w}_{LS} - \mathbf{w}_0 = \mathbf{A}^{\dagger} v = \mathbf{R}_{11}^{-1}\mathbf{Q}_1^H \mathbf{v}$$

$$\Delta\mathbf{z} \triangleq \mathbf{R}_{11}(\mathbf{w}_{LS} - \mathbf{w}_0) = \mathbf{Q}_1^H \mathbf{v}$$

$$\left|\Delta\mathbf{z}\right|_2^2 = \sum_{k=1}^{N} \left|\mathbf{Q}_1(k,1:P)\right|_2^2 \left|\mathbf{v}(k)\right|^2$$

Noise sample of the k-th equation

Statistical impact of the k-th equation on the LS estimate after orthogonal re-parametrization

On the statistical impact of equations on the LS estimate

- Noise and other disturbance effects on various LS equations are *widely different*.
- A large noise value affecting a *high impact* equation can cause *gross errors* on the LS estimate.
- The same noise sample affecting a *low impact* equation may have *negligible effects* on the LS estimate.
- It is difficult to optimize the statistical impact in DSP applications. In fact, most present sensor array problems (e.g., ML) are not well posed.
- This is the main source of noise-induced breakdown at low SNR.

Singular value Decomposition (SVD)

$$\underset{(N\times P)}{\begin{bmatrix}\mathbf{A}\end{bmatrix}} \overset{SVD}{=} \mathbf{U\Sigma V}^H = \begin{bmatrix}\underset{(N\times P)}{\mathbf{U}_1} & \underset{[N\times(N-P)]}{\mathbf{U}_2}\end{bmatrix}\begin{bmatrix}\underset{(P\times P)}{\mathbf{\Sigma}_1} \\ \underset{[(N-P)\times P]}{\mathbf{0}}\end{bmatrix}\begin{bmatrix}\underset{(P\times P)}{\mathbf{V}}\end{bmatrix}^H$$

$$\mathbf{U}^H\mathbf{U} = \mathbf{I}_N; \quad \mathbf{V}^H\mathbf{V} = \mathbf{I}_N; \quad \begin{bmatrix}\underset{(P\times P)}{\mathbf{\Sigma}_1}\end{bmatrix} = \begin{bmatrix}\sigma_1 & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sigma_P\end{bmatrix}$$

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_P \geq 0$$

$$\mathbf{A} \overset{R-SVD}{=} \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}^H$$

Singular values

Left singular vectors

Right singular vectors

Reduced size SVD

SAPIENZA
UNIVERSITÀ DI ROMA

Eigenvectors and singular vectors compared

- The *eigenvectors* of the left- and right-symmetrized forms of matrix $\mathbf{A}$ are the *singular vectors* of $\mathbf{A}$.
- The *eigenvalues* of the same symmetric forms are the *squares of the singular values* of $\mathbf{A}$.

$$\mathbf{A}^H\mathbf{A} \overset{SVD}{=} \mathbf{V}\left(\boldsymbol{\Sigma}^H\boldsymbol{\Sigma}\right)\mathbf{V}^H \Rightarrow \mathbf{A}^H\mathbf{A}\mathbf{V}(:,k) = \sigma_k^2\mathbf{V}(:,k)$$

$$\mathbf{A}\mathbf{A}^H \overset{SVD}{=} \mathbf{U}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^H\right)\mathbf{U}^H \Rightarrow \mathbf{A}\mathbf{A}^H\mathbf{U}(:,k) = \begin{cases} \sigma_k^2\mathbf{U}(:,k),\ k \le P \\ \mathbf{0},\ P < k \le N \end{cases}$$

LS solution by SVD

- The linear LS interpolation problem is *completely defined* by the (reduced size) SVD of $\mathbf{A}$.

$$\mathbf{A} \overset{R-SVD}{=} \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}^H \Rightarrow \mathbf{A}^\dagger = \mathbf{V}\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^H$$

$$\mathbf{w}_{LS} = \mathbf{V}\mathbf{\Sigma}_1^{-1}\mathbf{U}_1^H\mathbf{x}$$

$$Cov(\mathbf{w}_{LS}) = \sigma^2\mathbf{V}\mathbf{\Sigma}_1^{-2}\mathbf{V}^H$$

Euclidean matrix norms

$$\left|\underset{(N\times P)}{\mathbf{A}}\right|_2 \triangleq \sup_{\mathbf{u},\mathbf{v}} \frac{\left|\mathbf{u}^H\mathbf{A}\mathbf{v}\right|}{\left|\mathbf{u}\right|_2\left|\mathbf{v}\right|_2} = \sigma_1(\mathbf{A})$$

$$\left|\underset{(N\times N)}{\mathbf{A}^{-1}}\right|_2 \triangleq \sup_{\mathbf{u},\mathbf{v}} \frac{\left|\mathbf{u}^H\mathbf{A}^{-1}\mathbf{v}\right|}{\left|\mathbf{u}\right|_2\left|\mathbf{v}\right|_2} = \sigma_N^{-1}(\mathbf{A})$$

$$\left|\underset{(P\times N)}{\mathbf{A}^{\dagger}}\right|_2 \triangleq \sup_{\mathbf{u},\mathbf{v}} \frac{\left|\mathbf{u}^H\mathbf{A}^{\dagger}\mathbf{v}\right|}{\left|\mathbf{u}\right|_2\left|\mathbf{v}\right|_2} = \sigma_P^{-1}(\mathbf{A})$$

$$\kappa(\mathbf{A}) = \left|\underset{(N\times P)}{\mathbf{A}}\right|_2 \left|\underset{(P\times N)}{\mathbf{A}^{\dagger}}\right|_2 = \frac{\sigma_1(\mathbf{A})}{\sigma_P(\mathbf{A})} = \sqrt{\kappa(\mathbf{A}^H\mathbf{A})}$$

Optimal matrix rank truncation by SVD

$$\underset{(N\times P)}{\mathbf{A}} \overset{R-SVD}{=} \sum_{k=1}^{\min(N,P)} \sigma_k \mathbf{U}_1(:,k)\mathbf{V}(:,k)^H$$

$$\mathbf{A}_r = \sum_{k=1}^{r<\min(N,P)} \sigma_k \mathbf{U}_1(:,k)\mathbf{V}(:,k)^H$$

Reduced r rank approximation of $\mathbf{A}$ by SVD

$$|\mathbf{A}|_F \triangleq \sqrt{\sum_{k=1}^{\min(N,P)} \sigma_k^2}$$

Frobenius norm of $\mathbf{A}$

$$|\mathbf{A}-\mathbf{A}_r|_F = \sqrt{\sum_{k=r+1}^{\min(N,P)} \sigma_k^2}$$

Reduced r rank approximation error

Principal component analysis

- Design matrices with *nearly dependent* columns lead to hyper-sensitive LS solutions.
- It is possible to *prune* some parameters before estimation (still by QR and SVD concepts).
- Best results are obtained by using only the dominant singular triplets (*principal components*) of the design matrix.
- Often truncation is directed by statistical significance tests (MDL, random matrix theory).
- The incomplete rank LS solution must be re-defined. The SVD is optimal even for this task.

Reduced rank LS solution

LS, r rank, minimum norm solution

Excess r rank fitting error

$$M > P > r \Rightarrow \mathbf{A}_r \mathbf{w}_{LS(r)} = \mathbf{x} - \mathbf{e}(\mathbf{w}_{LS(r)})$$

$$\mathbf{w}_{LS(r)} = \sum_{k=1}^{r} \sigma_k^{-1} \mathbf{V}(:,k) \mathbf{U}_1(:,k)^H \mathbf{x}$$

$$\mathbf{e}(\mathbf{w}_{LS(r)}) = \mathbf{e}(\mathbf{w}_{LS}) + \mathbf{U}_1(:,r+1:P)\mathbf{U}_1(:,r+1:P)^H \mathbf{x}$$

$$\left.\begin{array}{c} \mathbf{A}_r \mathbf{w}_{LS(r)} = \mathbf{A}_r \mathbf{w}_0 \\ E[\mathbf{v}] = \mathbf{0} \end{array}\right\} \Rightarrow E[\mathbf{w}_{LS(r)}] = \mathbf{w}_0$$

$$Cov(\mathbf{v}) = \sigma^2 \mathbf{I}_P \Rightarrow$$

$$Cov(\mathbf{w}_{LS(r)}) = Cov(\mathbf{w}_{LS}) - \sum_{k=r+1}^{P} \frac{\sigma^2}{\sigma_k^2} \mathbf{V}(:,r+1:P)\mathbf{V}(:,r+1:P)^H$$

Discussion about reduced rank LS solution

- The rank reduced *fitting error and bias increase* w.r.t. the LS one, but the solution is *statistically more stable*.
- To avoid *estimation bias*, the *target signal must remain in the columns span* of the reduced rank design matrix.
- If the noise is strong, a tradeoff must be exercised between bias and covariance of the parameter estimate.
- Delicate questions about consistency and asymptotical efficiency of the rank-reduced LS estimates arise.

Other SVD uses

- The SVD can be applied to a sample matrix $\mathbf{X}$ *(N observations $\times$ P channels)*, furnishing at the same time:
  - ML Gaussian sample covariance ($\mathbf{R}_{xx}$) for zero-mean signals;
  - Observation location within the eigenvector space of $\mathbf{R}_{xx}$ (Karhunen-Loewe Transform, KLT);
  - Statistical impact analysis of single observations onto the LS solution.

Linear time invariant models

The signal $\mathbf{s}$ and the noise $\mathbf{v}$ are realizations of *independent, multivariate, zero mean Gaussian processes*. $\mathbf{H}$ is a constant transfer matrix, like the array steering matrix.

$$\underset{(P\times 1)}{\begin{bmatrix}\mathbf{x}(k)\end{bmatrix}} = \underset{(P\times D<P)}{\begin{bmatrix}\mathbf{H}\end{bmatrix}}\underset{(D\times 1)}{\begin{bmatrix}\mathbf{s}(k)\end{bmatrix}} + \underset{(P\times 1)}{\begin{bmatrix}\mathbf{v}(k)\end{bmatrix}}$$

$$E[\mathbf{v}(k)] = E[\mathbf{s}(k)] \overset{Hp}{=} \mathbf{0}$$

$$E[\mathbf{s}(k)\mathbf{s}(l)^T] = E[\mathbf{v}(k)\mathbf{v}(l)^T] \overset{Hp}{=} \mathbf{0}; \quad E[\mathbf{s}(k)\mathbf{s}(l)^H] \overset{Hp}{=} \delta_{kl}\mathbf{P}$$

$$E[\mathbf{s}(k)\mathbf{v}(l)^H] = E[\mathbf{s}(k)\mathbf{v}(l)^T] \overset{Hp}{=} \mathbf{0}; \quad E[\mathbf{v}(k)\mathbf{v}(l)^H] = \delta_{kl}\sigma^2\mathbf{I}_P$$

$$E[\mathbf{x}(k)\mathbf{x}(l)^H] = \delta_{kl}\mathbf{R}_{xx}; \quad \mathbf{R}_{xx} = \mathbf{HPH}^H + \sigma^2\mathbf{I}_P$$

Gaussian ML covariance estimate

$$p_x(\mathbf{x}) = \frac{1}{\pi^P \left\|\mathbf{R}_{xx}\right\|} \exp\left(-\mathbf{x}^H \mathbf{R}_{xx}^{-1} \mathbf{x}\right)$$

$$\underset{(N\times P)}{\left[\;\mathbf{X}\;\right]} \triangleq \begin{bmatrix} \mathbf{x}(1)^H \\ \vdots \\ \mathbf{x}(N)^H \end{bmatrix} \Rightarrow \tilde{\mathbf{R}}_{xx} = \frac{1}{N}\mathbf{X}^H\mathbf{X} \overset{EVD}{=} \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^H$$

$$\mathbf{X} \overset{R-SVD}{=} \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{V}^H \Rightarrow \tilde{\mathbf{R}}_{xx} = \mathbf{V}\left(N^{-1}\boldsymbol{\Sigma}_1^2\right)\mathbf{V}^H$$

$$\mathbf{V} = \mathbf{W}; \quad \boldsymbol{\Lambda} = N^{-1}\boldsymbol{\Sigma}_1^2$$

Eigenvector decomposition

Sample covariance completely defined by the SVD

SVD, KLT and statistical impact

$$\tilde{\mathbf{y}}(k) \triangleq \left(\sqrt{N}\Sigma_1^{-1}\right)\mathbf{V}^H\mathbf{x}(k) \xrightarrow{w.p.1} \left(\sqrt{N}\Sigma_1^{-1}\right)\mathbf{x}(k)_{KLT}$$

In probability convergence of SVD to KLT

$$\frac{1}{N}\sum_{k=1}^{N}\tilde{\mathbf{y}}(k)\tilde{\mathbf{y}}(k)^H = \mathbf{I}_P$$

Observations *whitened* by the SVD

$$\sum_{k=1}^{N}\left|\tilde{\mathbf{y}}(k)\right|^2 = NP \triangleq P\sum_{k=1}^{N}N_k$$

$$N_k = \frac{1}{P}\left|\tilde{\mathbf{y}}(k)\right|^2$$

Impact of each observation in terms of *equivalent number of observations*

Wiener–type noise reduction by the SVD

$$\underset{(N\times P)}{\mathbf{X}} \overset{R-SVD}{=} \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{V}^H; \quad \mathbf{R}_{vv} = \sigma^2\mathbf{I}_P$$

$$\mathbf{X}_{ML} = \sum_{k=1}^{P}\sqrt{\left(\sigma_k^2-\sigma^2\right)u_{-1}\left(\sigma_k^2-\sigma^2\right)}\mathbf{U}_1(:,k)\mathbf{V}(:,k)^H$$

$$\mathbf{X}_{LS} = \sum_{k=1}^{P}\frac{\sigma_k^2-\sigma^2}{\sigma_k}u_{-1}\left(\sigma_k^2-\sigma^2\right)\mathbf{U}_1(:,k)\mathbf{V}(:,k)^H$$

$$u_{-1}(x) \triangleq \begin{cases} 1, & x>0 \\ 0, & x\le 0 \end{cases}$$

Suboptimal ML spectral subtraction

Optimal formula based on sample SVD perturbation analysis (De Moore, 1994)

Discussion

- An observation $\mathbf{y}(k)$ with N_k much larger than unity can significantly alter the global estimate.
- This condition is *easily verifiable by the SVD* or derived robust tools (e.g., the robust pseudocovariance).
- These problems raised questions about the robustness of the parametric array processing in real world.
- We need additional math support for assessing the quality and robustness of an estimator.

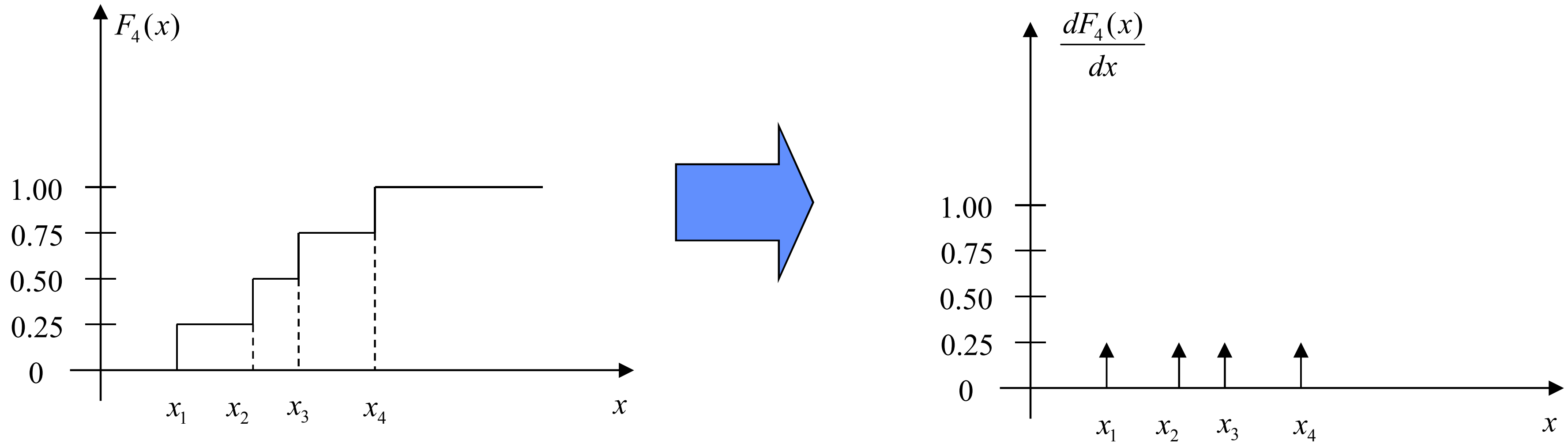
Empirical distribution

- Most common statistics and estimators depend on $\mathbf{x}$ only through the sample empirical distribution $F_M(\mathbf{x})$ obtained from M independent observations.

$$F_M(x) = \frac{1}{M}\sum_{i=1}^{M} u_{-1}(x - x_i); \quad dF_M(x) = \frac{1}{M}\sum_{i=1}^{M} u_0(x - x_i)dx$$

$$u_{-1}(x) = \begin{cases} 0, & x \le 0 \\ 1, & x > 0 \end{cases}; \quad \mathbf{x} = \left\{x_1, x_2, \ldots, x_N\right\}$$

Empirical distribution examples



$$F_N(x) = \int_{-\infty}^{x} dF_N(y) \quad \rightleftarrows \quad F(x) = \int_{-\infty}^{x} dF(y) = \int_{-\infty}^{x} f(y)dy$$

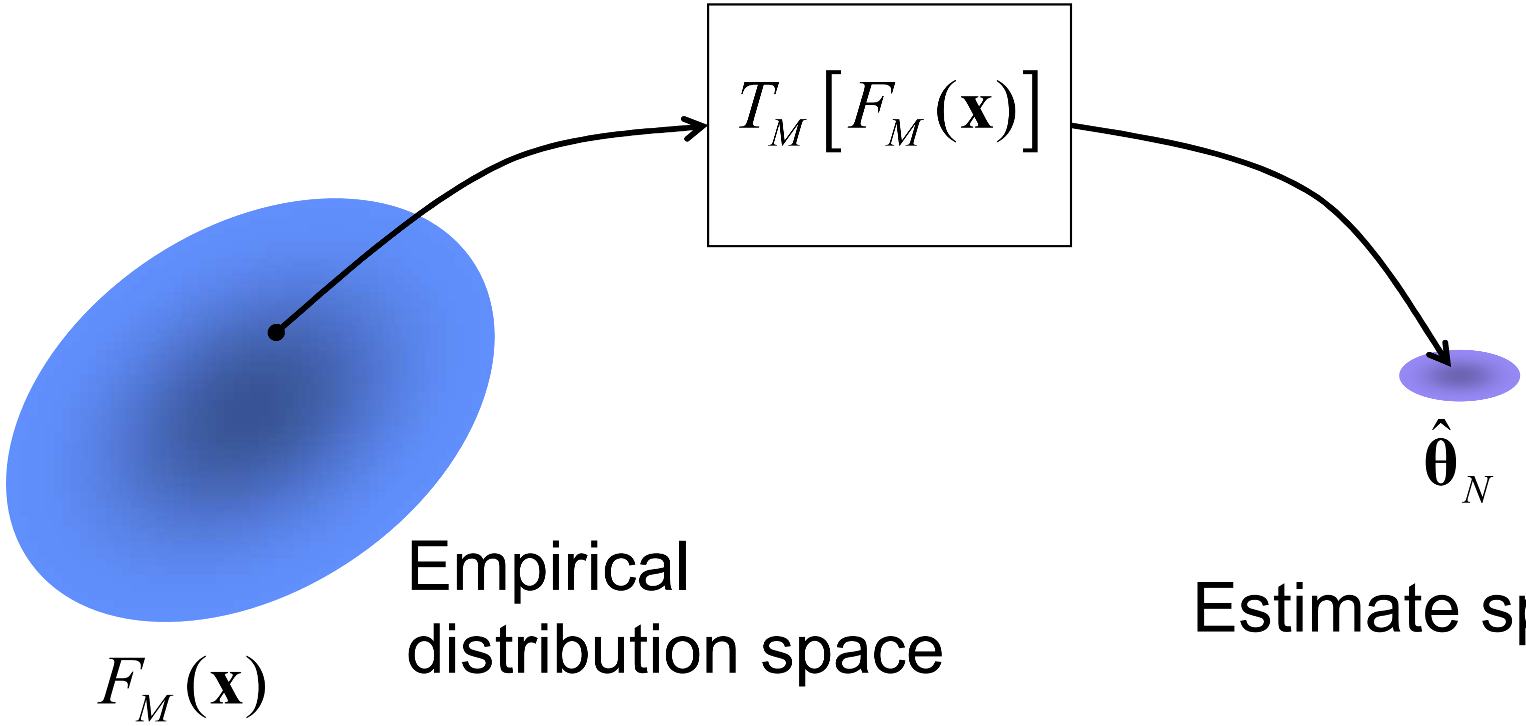
Estimators and empirical distributions

- A generic estimator $T_N(\mathbf{x})$ represents an *injective, deterministic mapping* between the empirical distribution space and the parameter space of θ:

$$\hat{\boldsymbol{\theta}}_M = T_M(\mathbf{x}) = T_M\left[F_M(\mathbf{x})\right]$$

$$\boldsymbol{\theta} = \left\{\theta_1, \theta_2, \ldots, \theta_P\right\}$$

Estimation process



Example: 1-D ML estimator (differential form)

$$\int_{-\infty}^{+\infty} \frac{\partial \log f(x/\theta)}{\partial \theta} \left[\frac{1}{M} \sum_{i=1}^{M} u_0(x - x_i) dx \right]_{\theta=\theta_{ML}} = 0 \Rightarrow$$

$$\int_{-\infty}^{+\infty} \frac{\partial \log f(x/\theta)}{\partial \theta} dF_M(x) \bigg|_{\theta=\theta_{ML}} = 0 \Rightarrow$$

$$\frac{1}{M} \sum_{i=1}^{M} \frac{\partial \log f(x_i/\theta)}{\partial \theta} \bigg|_{\theta=\theta_{ML}} = 0$$

$u_0(x)$ is the *Dirac pulse*, centered on zero.

Consistency in distribution

- An estimator is said *consistent at the distribution* $F(\mathbf{x})$ if, in probability

$$\boldsymbol{\theta} = T\left[F(\mathbf{x})\right] = \lim_{M \to \infty} T_M\left[F_M(\mathbf{x})\right]$$

- *Consistency* means that, for increasing sample size, the estimated parameters converge in probability to the design ones (not always the true ones).
- The *estimator variance vanishes* for infinite sample size.

Asymptotic bias

- Estimators may not converge in probability (for increasing sample size) to the true parameters θ_0 . The limit difference is said *asymptotic bias*.

$$\boldsymbol{\theta}_M = E\left\{T_M\left[\mathbf{x}\left(\boldsymbol{\theta}_0\right)\right]\right\} \neq \boldsymbol{\theta}_0$$

$$\mathrm{Bias}\left\{T_M\left[\mathbf{x}\left(\boldsymbol{\theta}_0\right)\right]\right\} = \boldsymbol{\theta}_M - \boldsymbol{\theta}_0$$

- An estimator is said *asymptotically unbiased* if

$$\lim_{M\to\infty} M\left[\boldsymbol{\theta}_M - \boldsymbol{\theta}_0\right] = 0$$

Log-Likelihood

- By the Central Limit Theorem, the distribution of the (negative) log-likelihood of the sample $\mathbf{x}$ is asymptotically Gaussian and can be expanded in Taylor series around the true parameters θ_0.

$$\bar{l}_M(\mathbf{x}/\boldsymbol{\theta}) = -\frac{1}{M}\sum_{i=1}^{M}\log f(x_i/\boldsymbol{\theta}) \overset{(M\to\infty)}{\to} l(\boldsymbol{\theta}) = \int_{-\infty}^{+\infty} -\log f(\mathbf{x}/\boldsymbol{\theta})\,dF(x)$$

$$\bar{l}_M(\mathbf{x}/\theta) = \bar{l}_M(\mathbf{x}/\boldsymbol{\theta})\Big|_{\theta=\theta_0} + \nabla_\theta \bar{l}_M(\mathbf{x}/\boldsymbol{\theta})\Big|_{\theta=\theta_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) +$$

$$+\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T \mathbf{H}_M(\mathbf{x}/\theta)\Big|_{\theta=\theta_0}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + O\left(\left|\boldsymbol{\theta}-\boldsymbol{\theta}_0\right|^3\right)$$

Regular distributions

- A distribution is said *regular* if:

$$\exists \nabla_\theta f\left(\mathbf{x}/\boldsymbol{\theta}\right)=\left[\frac{\partial f\left(\mathbf{x}/\boldsymbol{\theta}\right)}{\partial \theta_i};i=1,\ldots,P\right] \mid \int_{-\infty}^{+\infty} \nabla_\theta f\left(\mathbf{x}/\boldsymbol{\theta}\right)dx=0$$

- The gradient w.r.t. parameters of the log-likelihood of a regular distribution *is zero at the true parameters*.

$$\int_{-\infty}^{+\infty} \nabla_\theta \log\left[f(\mathbf{x}/\boldsymbol{\theta})\right]dF(\mathbf{x}/\boldsymbol{\theta})=\int_{-\infty}^{+\infty} \nabla_\theta f(\mathbf{x}/\boldsymbol{\theta})\frac{f(\mathbf{x}/\boldsymbol{\theta})}{f(\mathbf{x}/\boldsymbol{\theta})}dx=0$$

Fisher Information Matrix

- The *Fisher Information Matrix* (FIM) $\mathbf{J}[f(\mathbf{x}/\theta)]$ is the expected value of the Hessian (curvature) of the log-likehood functional.

$$\mathbf{J}\left[f\left(\mathbf{x}/\boldsymbol{\theta}\right)\right] = E\left[\mathbf{H}\left(\mathbf{x}/\boldsymbol{\theta}\right)\right] = \int_{-\infty}^{+\infty} \mathbf{H}\left(\mathbf{x}/\boldsymbol{\theta}\right) dF(x) > 0$$

$$\mathbf{H}\left(\mathbf{x}/\boldsymbol{\theta}\right)_{i,j} = -\frac{\partial^2 \log f\left(\mathbf{x}/\boldsymbol{\theta}\right)}{\partial\theta_i \partial\theta_j} = \frac{\partial \log\left(\mathbf{x}/\boldsymbol{\theta}\right)}{\partial\theta_i} \frac{\partial \log\left(\mathbf{x}/\boldsymbol{\theta}\right)}{\partial\theta_j}$$

ML estimate

- Minimizing the sample log-likelihood around the true parameters leads to an approximate formula for the ML estimate, which resembles a single Newton iteration.

$$\nabla_{\boldsymbol{\theta}} \bar{l}_M \left(\mathbf{x} / \boldsymbol{\theta}_{ML} \right) = 0 \Rightarrow$$

$$\nabla_{\boldsymbol{\theta}} \bar{l}_M \left(\mathbf{x} / \boldsymbol{\theta} \right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \mathbf{H}_M \left(\mathbf{x} / \boldsymbol{\theta} \right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left(\boldsymbol{\theta}_{ML} - \boldsymbol{\theta}_0 \right) = O\left(\left| \boldsymbol{\theta}_{ML} - \boldsymbol{\theta}_0 \right|^2 \right)$$

$$\theta_{ML} = \theta_0 - \left[\mathbf{H}_M \left(\mathbf{x} / \boldsymbol{\theta} \right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right]^{-1} \nabla_{\theta} \bar{l}_N \left(\mathbf{x} / \boldsymbol{\theta} \right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + O\left(\left| \boldsymbol{\theta}_{ML} - \boldsymbol{\theta}_0 \right|^2 \right)$$

Log-Likelihood statistics

$$F_M(x) \rightarrow F(x) \Rightarrow E\left[\nabla_\theta \bar{l}_M\left(\mathbf{x}/\boldsymbol{\theta}\right)\right] = E\left[-\nabla_\theta \log f\left(\mathbf{x}/\boldsymbol{\theta}\right)\right] = 0$$

$$\mathrm{Cov}\left[\nabla_\theta \bar{l}_M\left(\mathbf{x}/\boldsymbol{\theta}\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] = \frac{1}{M} E\left\{\nabla_\theta \log f\left(\mathbf{x}/\boldsymbol{\theta}\right)\left[\nabla_\theta \log\left(\mathbf{x}/\boldsymbol{\theta}\right)\right]^T\right\} = \frac{1}{M}\mathbf{J}$$

$$\mathbf{H}_N\left(\mathbf{x}/\boldsymbol{\theta}\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{M}\sum_{l=1}^{N}\mathbf{H}_l\left(\boldsymbol{\theta}_0\right) = \mathbf{J} + O\left(\frac{1}{M}\right)$$

$$\mathbf{H}_l\left(\boldsymbol{\theta}_0\right)_{i,j} = \mathbf{H}_l\left(\boldsymbol{\theta}_0\right)_{j,i} = -\frac{\partial^2 \log f\left(x_l/\boldsymbol{\theta}\right)}{\partial\theta_i\partial\theta_j}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

ML estimator (MLE) statistics

- At this point, it is easy to obtain the main statistics of the ML estimate, following the same passages made with LS fitting estimate.

$$E\left[\boldsymbol{\theta}_{ML}-\boldsymbol{\theta}_0\right]=E\left[-\mathbf{H}_M\left(\mathbf{x}/\boldsymbol{\theta}_0\right)^{-1}\nabla_\theta \bar{l}_M\left(\mathbf{x}/\boldsymbol{\theta}_0\right)\right]=\mathbf{0}+O\left(1/\sqrt{N}\right)$$

$$\lim_{M\to\infty} Cov\left[\boldsymbol{\theta}_{ML}-\boldsymbol{\theta}_0\right]=\mathbf{J}^{-1}\frac{\mathbf{J}}{M}\mathbf{J}^{-1}=\left(M\mathbf{J}\right)^{-1}$$

Local LS system interpretation of the MLE

- The fact that the sample Hessian of the MLE asymptotically converges to the FIM indicates that it arises from the symmetrization of a certain *known underlying system matrix*.
- The *unknown vector is the ML estimation error*.
- The *target is a zero mean, white random* vector.
- A similar system block is appended for each independent observation.
- The resulting LS problem is homoscedastic, so the MLE locally is the *BLUE estimate of a zero vector*.

Local LS interpretation of the MLE

$$E\left[\mathbf{v}_k\right]=\mathbf{0};\quad E\left[\mathbf{v}_k\mathbf{v}_l^T\right]=\sigma^2\mathbf{I}\delta_{kl};\quad k=1,2,\ldots,M$$

$$\left(\frac{1}{\sigma\sqrt{M}}\right)\begin{bmatrix}\mathbf{A}\\ \vdots\\ \mathbf{A}\end{bmatrix}\left(\boldsymbol{\theta}_{ML}-\boldsymbol{\theta}_0\right)=\left(\frac{1}{\sigma\sqrt{M}}\right)\begin{bmatrix}\mathbf{v}_1\\ \vdots\\ \mathbf{v}_M\end{bmatrix}-\mathbf{e}$$

$$f=\mathbf{e}^T\mathbf{e};\quad \mathbf{H}\left(f\right)=\frac{1}{\sigma^2}\mathbf{A}^T\mathbf{A}=\mathbf{J};\quad \nabla f=\frac{1}{\sigma^2 M}\mathbf{A}^T\sum_{k=1}^{m}\mathbf{v}_k;$$

$$E\left[\nabla f\right]=0;\quad \mathrm{Cov}\left[\nabla f\right]=\frac{1}{\sigma^2 M}\mathbf{A}^T\mathbf{A}=\frac{\mathbf{J}}{M}$$

$$E\left[\boldsymbol{\theta}_{ML}-\boldsymbol{\theta}_0\right]=\mathbf{0};\quad \mathrm{Cov}\left[\boldsymbol{\theta}_{ML}-\boldsymbol{\theta}_0\right]=\left[M\mathbf{J}\right]^{-1}$$

Cramer-Rao Bound

- The minimum covariance of any unbiased parameter estimator of a regular distribution is given by the Cramer-Rao bound (CRB), which asymptotically coincides with the ML error covariance, if the *number of parameters* of the likelihood is *finite*.
- In this case the MLE is said *asymptotically efficient*.

$$\int_{-\infty}^{\infty} \frac{\partial f\left(\mathbf{x}/\boldsymbol{\theta}\right)}{\partial \theta} dx = 0; \quad \lim_{M\to\infty} E\left[T_M(\mathbf{x})\right] = \boldsymbol{\theta}_0 \Rightarrow$$

$$\mathrm{Cov}\left[T_M\left(\mathbf{x}\right)\right] \geq CRB(\boldsymbol{\theta}) = \lim_{M\to\infty} \mathrm{Cov}\left[\boldsymbol{\theta}_{ML}\right] = M^{-1}\mathbf{J}^{-1}$$

SAPIENZA
UNIVERSITÀ DI ROMA

Efficiency of ML estimates

- The ML estimator is tied to the *specific distribution assumed* for data and *cannot be rigorously generalized* to other distributions.
- The MLE is often difficult or even impossible to realize.
- Different distributions may have *quite different ML estimators*.

$$f(x/m) = \frac{\lambda}{2} e^{-\lambda|x-m|} \Rightarrow \hat{m}_{ML} = \text{median}\left(\begin{bmatrix} x_1 & \cdots & x_M \end{bmatrix}\right);$$

$$f(x/m) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \Rightarrow \hat{m}_{ML} = \frac{1}{M}\sum_{i=1}^{M} x_i$$

SAPIENZA
UNIVERSITÀ DI ROMA

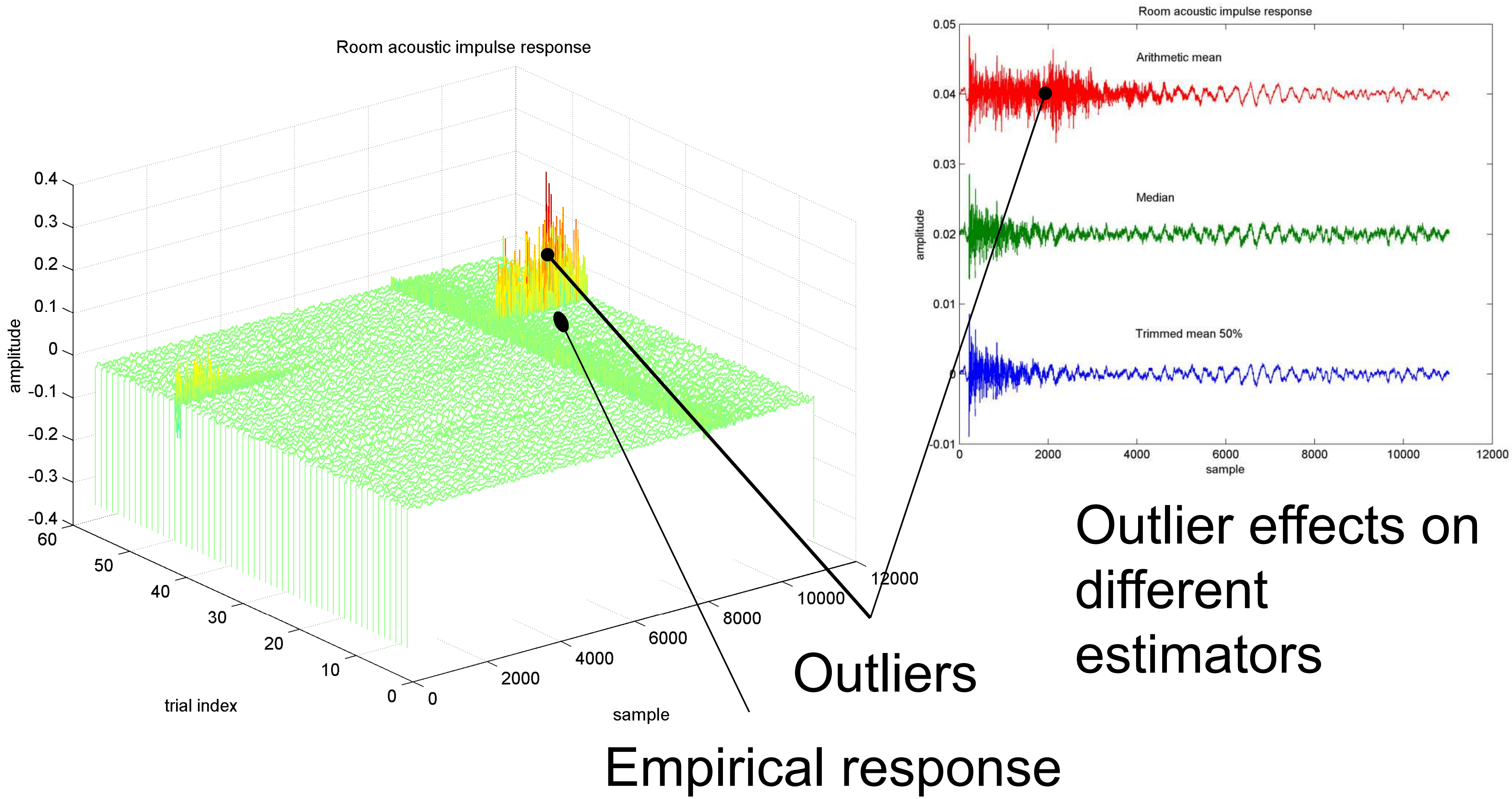
Real world sample problems

- In real world samples many *hidden contamination sources* exist:
  - *Missing data* (*inliers*) by recording defects, arbitrary interpolations, zero padding, etc…;
  - *Non stationarities* (transients, level changes) often introduced by the acquisition system;
  - *Gross errors* (*outliers*), due to e.m. pulses, un-modeled interference, etc…;
  - *Actual distribution different from the assumed one* (quantization, clipping, thresholding, etc…).

Contaminated distributions

- Sometimes, a small, *unknown fraction* of samples comes from a *different distribution*.
- For instance, some return echoes used for radar clutter estimation contain a target: detection threshold is unduely raised.
- *Signal model slightly changes* during acquisition.
- There is the need of a reference contamination model for assessing the robustness of an estimator.

Example: contaminated array impulse response



Outlier effects on different estimators

SAPIENZA UNIVERSITÀ DI ROMA

Contamination model

- *Gross errors*: in the sample worst case *outliers placed at infinity exist* with propability.

$$\frac{dF_{\varepsilon}(x)}{dx} \triangleq f_{\varepsilon}(x) = (1-\varepsilon)f(x) + \varepsilon u_0(\infty); \quad \varepsilon \ll 1$$

- *Sistematically contaminated sample* by a different distribution $H(x)$

$$F_{\varepsilon}(x) = (1-\varepsilon)F_0(x) + \varepsilon H(x); \quad 0 < \varepsilon < 0.5$$

Array robust estimation

- Classical *robust estimators* (median, covariance, MAD, robust regression,…) have several direct applications to array processing, especially in calibration, SCM estimation, fault detection.
- However, array processing mainly uses *functional-type parametric (M-)estimators* and error sources are partly in the equation model and party in the data.
- Most robustness is required to avoid catastrophic intermediate decisions (e.g., source number detection) or mitigate their consequences.
- So array robust approaches follow a different path than classical robust estimators.

SAPIENZA
UNIVERSITÀ DI ROMA

Part III

CALIBRATION AND ARRAY MODELS

Array calibration fundamentals

- The basic form of array *calibration* is essentially an *interpolation* problem.
- A known signal $s(t)$, independent by any possible disturbance, is transmitted by a source located at known position $\mathbf{p}$, for example in an anechoic chamber.
- There is only *additive thermal noise*, independent and equi-powered between sensors.
- In the general wide-band case, the expected array response is a convolution.
- In the narrow-band case, the length of the impulse response is simply one.

LS calibration

- The problem is solved in parallel for all sensors in discrete time.
- Different receiver noise characteristics create estimation problems (hetero-scedastic observations).
- Generally background noise can be assumed Gaussian distributed in controlled environments.

$$\sum_{k=0}^{P} h_{nk} s(t-k) \overset{LS}{=} x_n(t) - v_n(t); \quad n = 1,2,\ldots,N; \quad t = 1,2,\ldots,M$$

$$\mathbf{S}\mathbf{h}_n \overset{LS}{=} \mathbf{y}_n - \mathbf{v}_n; \quad \hat{\mathbf{h}}_n = \mathbf{S}^{\dagger}\mathbf{y}_n$$

Notes on LS calibration

- In general, accurate time alignment of target responses vs. excitation source is required, otherwise:
  - Non-causal or void responses may appear;
  - Steering vectors are affected by unknown phase shifts in narrow-band cases.
- Wide-band responses (assumed FIR) are converted into steering vectors by the vector Discrete Time Fourier Transform, after taking into account the mapping between the analog analysis frequencies and the digital frequencies after demodulation and conversion.
- LS *calibrated vectors are essentially corrupted* by a small additive, independent Gaussian noise vectors, equi-powered between sensors. Its variance is inversely proportional to the number of independent array response observations.
- *Time-uncorrelated test signals* are preferred (e.g., chirp signals, sinusoids modulated by Barker codes), since provide more independent observations.

Empirical modal calibration

- Steering vectors measured at different directions can be interpolated on a finite angular modal basis derived by a priori consideration.
- The resulting interpolated modal matrix provides cleaner steering vectors than the calibrated ones.
- Good calibration is fundamental for the accuracy and robustness of subsequent array processing.

$$\hat{\mathbf{a}}(\mathbf{p}_k, f) \overset{LS}{\cong} \mathbf{G}(f)\left[\varphi_1(\mathbf{p}_k) \quad \cdots \quad \varphi_Q(\mathbf{p}_k)\right] = \mathbf{G}(f)\mathbf{u}(\mathbf{p}_k);$$
$$k = 1, 2, \ldots, K \gg N; \quad Q \geq N$$

Mutual coupling matrix estimation

- Forces estimation of an unique mutual coupling matrix $\mathbf{M}$ for a wide angular sector.
- Extremely careful time alignment is required.
- Always prefer interpolation to inversion whenever possible.
- Solution easily ill-conditioned and diverging outside the sector, requiring further constraints (unitary matrix, regularization).
- Matrix transformation model not valid in general (*closure* problem w.r.t. angular sensor response).

$$\hat{\mathbf{a}}\left(\mathbf{p}_k, f\right) \overset{LS}{\cong} \mathbf{M}\left(f\right)\mathbf{a}\left(\mathbf{p}_k, f\right); \quad k = 1, 2, \ldots, K \gg N; \quad \mathbf{p}_k \in \mathbf{\Omega}$$

LS array calibration in telecommunication devices

- Routinely done in telecommunications devices, such as MIMO Wi-Fi ones, to estimate the actual steering (*mixing* in pattern recognition jargon) matrix from a known periodic training sequence.
- Estimation often done in non-white noise.
- Different signal model than expected, often with dominant model error effects over symbol noise.
- Consequent perturbation of any derived estimate.

$$\mathbf{y}(t) = \left(\mathbf{H} + \dot{\mathbf{H}}\right)\mathbf{s}(t) + \mathbf{v}(t)$$

$$\left\|\dot{\mathbf{H}}\right\|_2 \ll \left\|\mathbf{H}\right\|_2 ; E\left[\dot{\mathbf{H}}\right] \cong \mathbf{0}; \quad Cov\left[\dot{\mathbf{H}}\right] = ??$$

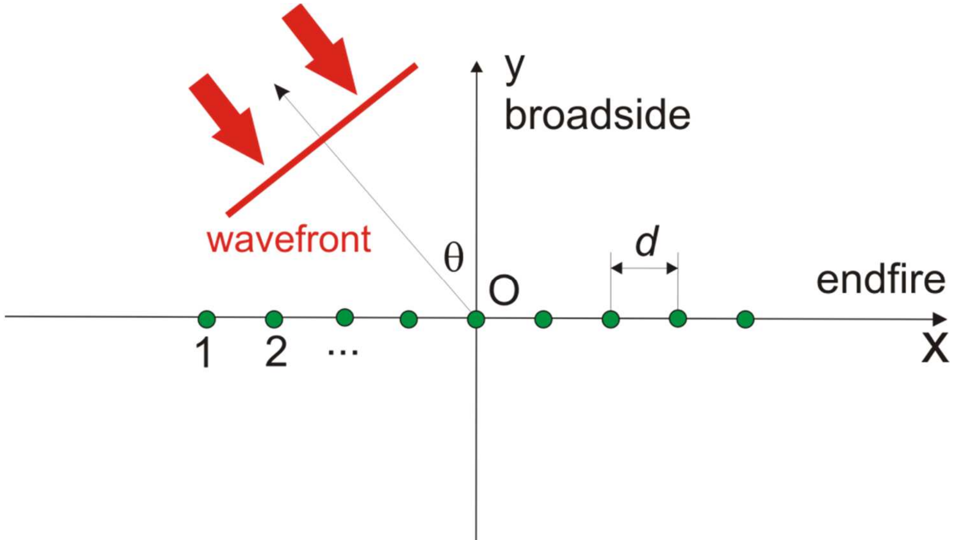
Array geometries

- The importance of the array size (*aperture vs. the wavefront*) and dimensional tolerances are of paramount importance for final performance, exactly as in optics.
- The choice of an array geometry involves tradeoffs between *number of elements*, *ambiguity resistance*, synthetic *gains*, robustness to element mis-calibration and mis-placement.
- Common geometries belong to a few fundamental types, with some modifications (sensor *pruning*, *filling*, *unequal spacing*) and hybridizations.

Classical array geometries: ULA

- The most classical geometry is the Uniform Linear Array (ULA) with N omnidirectional (or identical) sensors, equispaced along a line by d wavelengths.
- Vandermonde type steering vector:

$$\mathbf{a}(f,\theta) \propto g(f,\theta)\begin{bmatrix} 1 & e^{-j2\pi d\sin(\theta)} & \cdots & e^{-j2\pi d(N-1)\sin(\theta)} \end{bmatrix}^T$$



ULA properties

- $N\text{-}1$ steering vectors corresponding to *different wavenumbers* $d\sin(\theta)$ are linearly independent by the polynomial algebra fundamental theorem: therefore $N\text{-}1$ sources can be asymptotically resolved.
- Front-rear ambiguity and no resolution capability in elevation with identical sensors.
- *Ambiguity* (spatial aliasing) within FOV for $d>0.5$.
- *Invisible space* (physically impossible wavenumbers) for $d<0.5$.
- Design techniques and processing derived from the ones of FIR filters (LS, equi-ripple Remez, windowing).

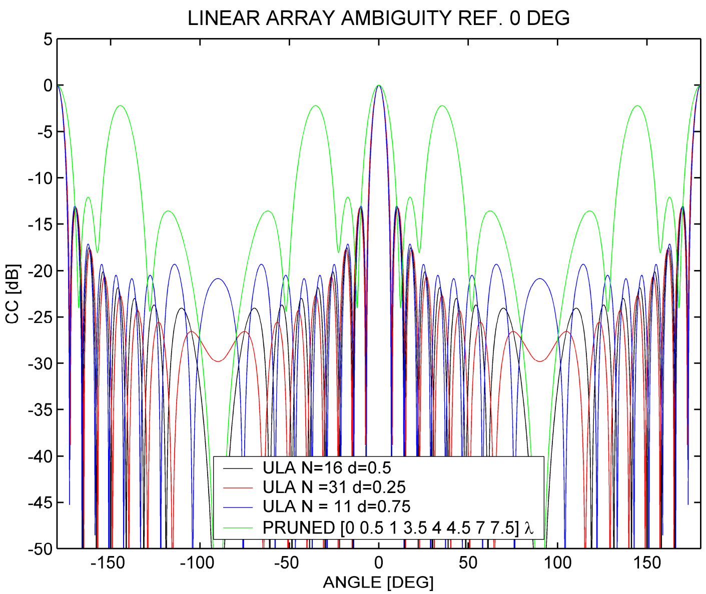
$$g\left(f,\theta\right)=const; \quad z=e^{j2\pi d\sin(\theta)}$$

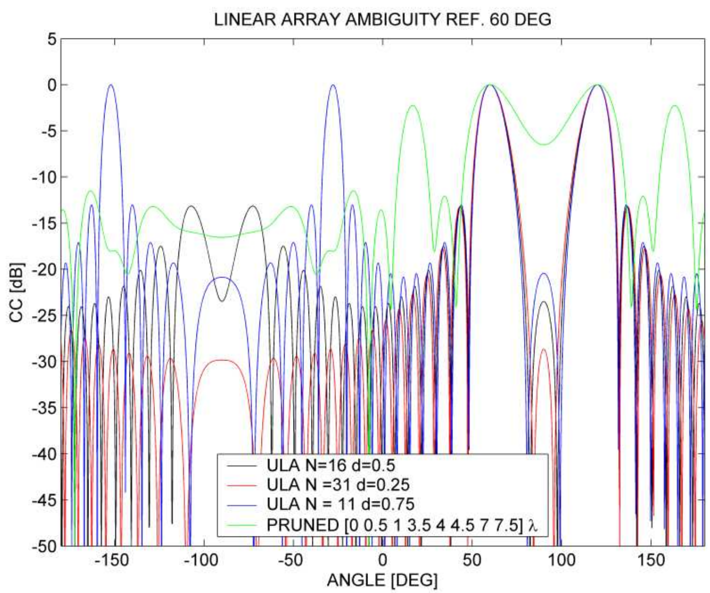
$$\mathbf{a}\left(f,\theta\right)\propto\begin{bmatrix}1 & z^{-1} & \cdots & z^{-(N-1)}\end{bmatrix}^T$$

SAPIENZA
UNIVERSITÀ DI ROMA

Linear array ambiguities

- Enlarging inter-sensor separation or pruning the sensors raise CC sidelobes and create grating lobes of ambiguity for equal effective aperture.





ULA mutual coupling

- Ignoring *end effects*, the mutual coupling ULA matrix is approximatively Toeplitz shaped, with equal diagonal elements, which can be further approximatively diagonalized by the DFT.
- Computational efficiency is retained.
- Dummy, not feeded, but loaded elements can be added at both ULA ends to enforce Toeplitz coupling on active sensors

$$\hat{\mathbf{a}}(z) \cong z^{\frac{N-1}{2}} \begin{bmatrix} m_0 & m_1 & \cdots & m_{N-1} \\ m_1 & m_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & m_1 \\ m_{N-1} & \cdots & m_1 & m_0 \end{bmatrix} \begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-(N-1)} \end{bmatrix}^T$$

SAPIENZA
UNIVERSITÀ DI ROMA

Compound linear arrays

- To overcome the limitations of the basic ULA, multiple linear subarrays pointing to different directions can be employed (L-shaped, X-shaped, grid) .
- Most basic ambiguities can be removed with a reasonable number of additional sensors.
- New sampling lattices have to be re-checked for ambiguity forms.
- Sources impinging from the directions of large sidelobes interfere with the main source and reduce or destroy spatial resolution.
- From these studies, some element can be *moved* (non-uniform LA) or *pruned* to save resources at the expense of the sidelobe structure.

Sensor choices for linear arrays

- Directive sensors aligned in parallel on a linear array do not modify ambiguity properties, except for the attenuation of some sectors.
- *Directional, parallel sensors force large minimum inter-sensor spacing* with great ambiguity risks.
- Horns and apertures are less subject to mutual coupling.
- Introducing diversity in directivity patterns and vector sensors zeroes computational and constructive advantages of linear arrays, but removes most ambiguities.

Isotropic arrays in 2-D

- The family contains circular and multiple ring arrays.
- *Equal statistical performance* over all azimuth angles with omnidirectional sensors.
- *Circular sensor coordinate scatter* matrix (equal singular values along abscissa and ordinate axes).
- *Multi-ring circular arrays* maintain isotropy even with equal, radially oriented directive sensor.
- Compact single-ring circular arrays have *diagonal harmonic decomposition* matrix with scaled Bessel-J coefficients.
- When a Bessel function is zero (*resonance*), the circular array loses one effective sensor.

SAPIENZA
UNIVERSITÀ DI ROMA

Uniform circular arrays

- Uniform circular arrays (UCAs) are the most efficient isotropic arrays for a given physical aperture.

$$\mathbf{r}_n = R\left[\cos\left(\theta_n\right)\mathbf{i} + \sin\left(\theta_n\right)\mathbf{j}\right]; \quad \theta_n = \theta_0 + \frac{2\pi n}{N}; \quad n = 0,1,2,\ldots,N-1$$

$$\mathbf{a}_n\left(\theta, f\right) = e^{j\frac{2\pi R\cos\left(\theta-\theta_n\right)}{\lambda}} = \sum_{q=-\infty}^{+\infty} j^q J_q\left(\frac{2\pi R}{\lambda}\right) e^{jq\theta_n} z^{-q}; \quad z = e^{j\theta}$$

$$\left|J_{N-Q}\left(\frac{2\pi R}{\lambda}\right)\right| \ll \left|J_Q\left(\frac{2\pi R}{\lambda}\right)\right|; \quad Q \le \left(N \operatorname{div} 2\right) + \left(N \bmod 2\right) - 1 \Rightarrow$$

$$\mathbf{a}_n\left(z, f\right) \cong \sum_{q=-Q}^{+Q} j^q J_q\left(\frac{2\pi R}{\lambda}\right) e^{jq\theta_n} z^{-q}$$

SAPIENZA
UNIVERSITÀ DI ROMA

UCA mutual coupling

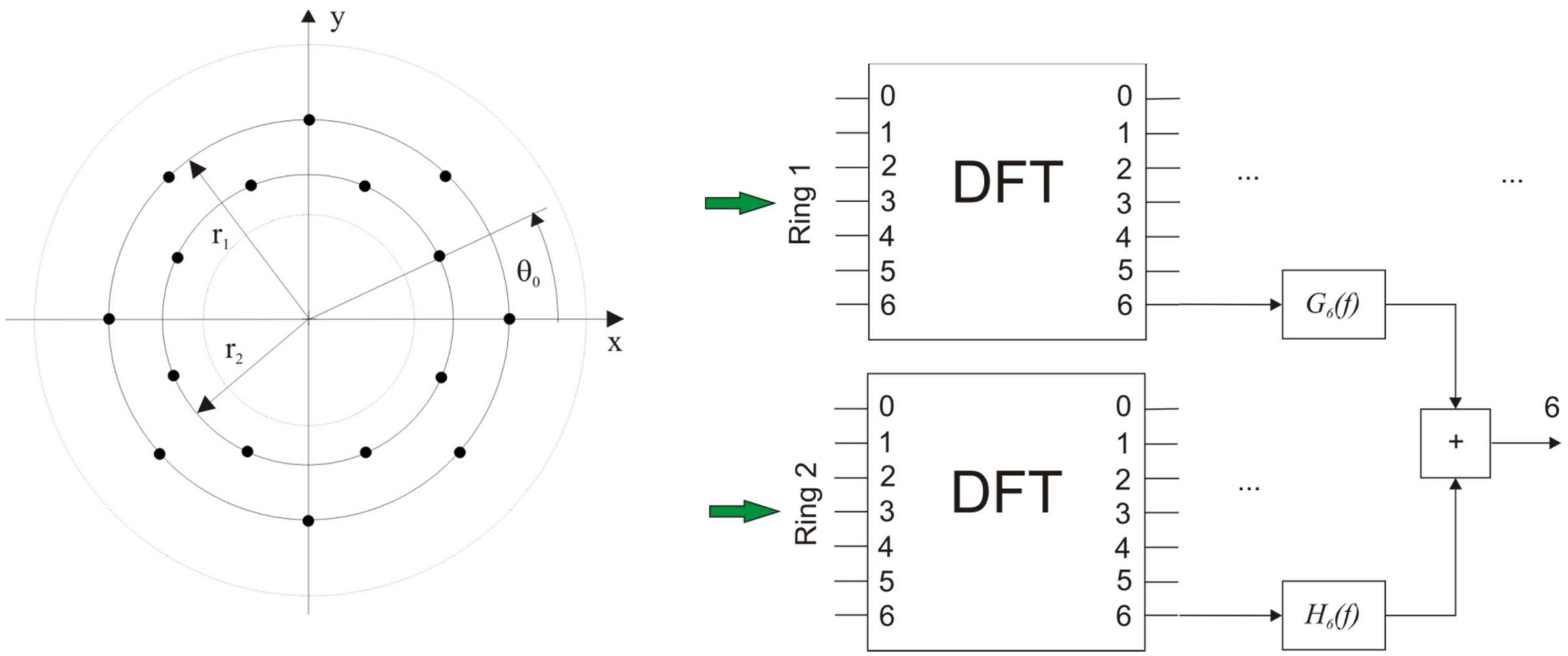
- Because of circular simmetry, the ideal mutual coupling matrix of the UCA is *right circulant* and can be diagonalized by the DFT.
- If the UCA is compact w.r.t. the wavelength, its angular harmonics modal decomposition is also (near) diagonal even under mutual coupling.
- UCA enables processing algorithms in close form in the Z-domain, as the ULA.

Multi-ring array (MRA)

- Two or more rings with the same number of sensors avoid resonances over several octaves.
- Compact multi-ring arrays can be combined by spatial DFT and harmonic filters to give an overall *wide-band diagonal harmonic decomposition equal for all frequencies (Di Claudio, 2005).*
- *Each combination output has a scaled harmonic spatial response.*
- The number of sensors in inner rings *cannot be reduced* to avoid Bessel function aliasing by mutual coupling.
- Maximum number of resolvable sources less than the number of hamonics and the sensors on each ring (*linearly dependent ring spatial responses*).
- Cheap thin wire realizations.

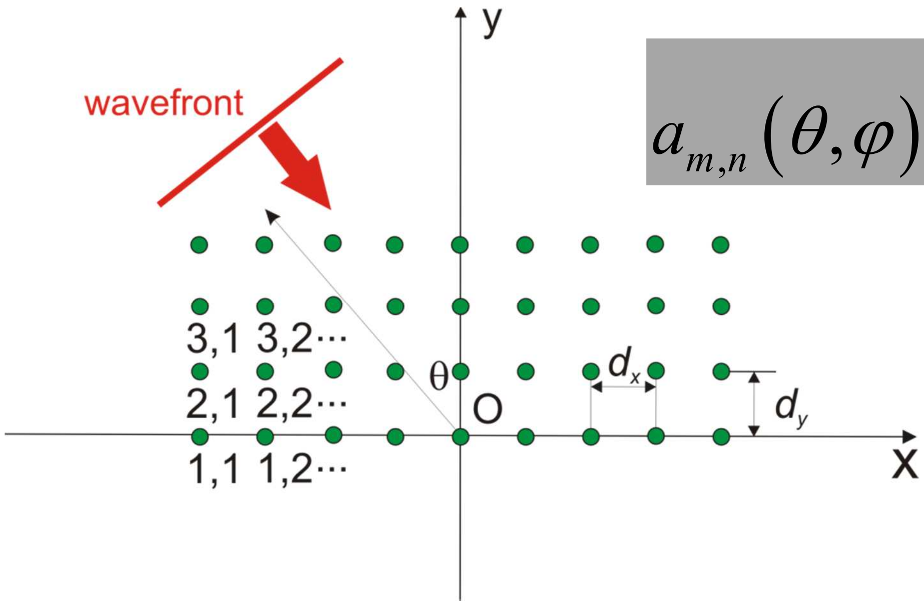
MRA DFT harmonic combining (Di Claudio, 2005)

- Spatial DFT can be realized by four quadrant multipliers and analog Butler matrices made by phase shifters and couplers.



Other planar geometries

- Filled circular disks are often special cases of the MRA and can be viewed in different ways.
- Uniformly filled rectangle array (*meshes*) have two inter-related wavenumbers.

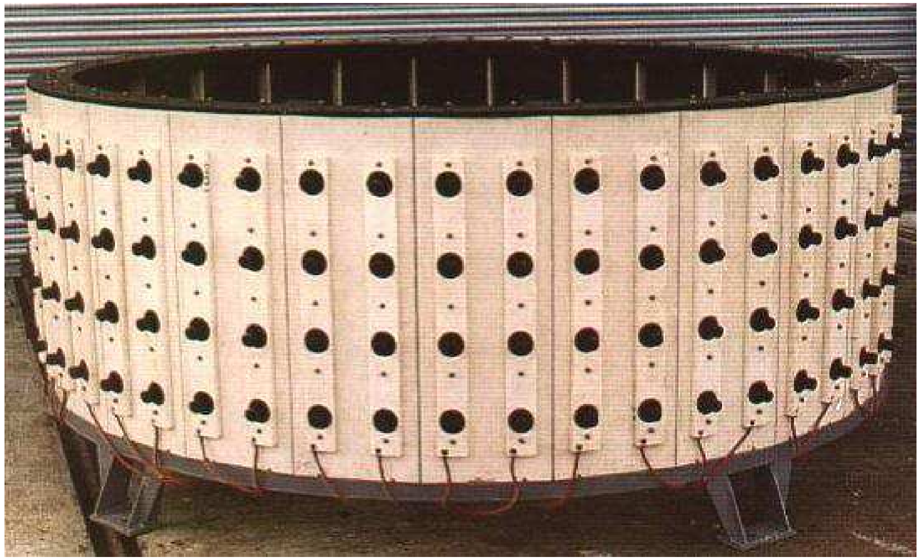


$$a_{m,n}(\theta,\varphi) = g(\theta,\varphi)\, e^{j\frac{2\pi}{\lambda}\cos(\varphi)\left[-d_x m\sin(\theta)+d_y n\cos(\theta)\right]}$$

SAPIENZA
UNIVERSITÀ DI ROMA

3-D geometries

- Primarily used in acoustics:
  - filled cylindric surfaces in sonar.
  - filled spherical surfaces in aerial acoustics.
- Sensor shadowing and interconnection problems.
- Slow, memory hungry 3-D scan techniques.



SAPIENZA
UNIVERSITÀ DI ROMA

Noise sources in arrays

- Internal noise sources:
  - Thermal, shot and Flicker noise from analog channel and electronics;
  - Digital quantization noise from ADC;
  - Intermodulation noise by non-linearity;
  - Tone-like spurious signals.
- External noise sources:
  - Industrial noise;
  - Interference;
  - Atmospheric noise;
  - Spatially un-resolvable source mixture.
  - Background clutter in active systems.

Thermal (Johnson, Nyquist) noise

- Originated in *resistive devices*.
- Zero mean, *Gaussian* distributed.
- White (constant) power spectral density up to some GHz.
- Independent between sensors.
- Only possible correlated exception is the noise generated by a resistive load and coupled (re-radiated) toward other sensors.
- Variance sensitive to AGC status.

$$E\left[v\right]=0; \quad P_{vv}\left(f\right)=4kTR;\ k=1.3806488\ \times\ 10^{-23}\,m^2kg\ s^{-2}\,{}^{\circ}K^{-1}$$

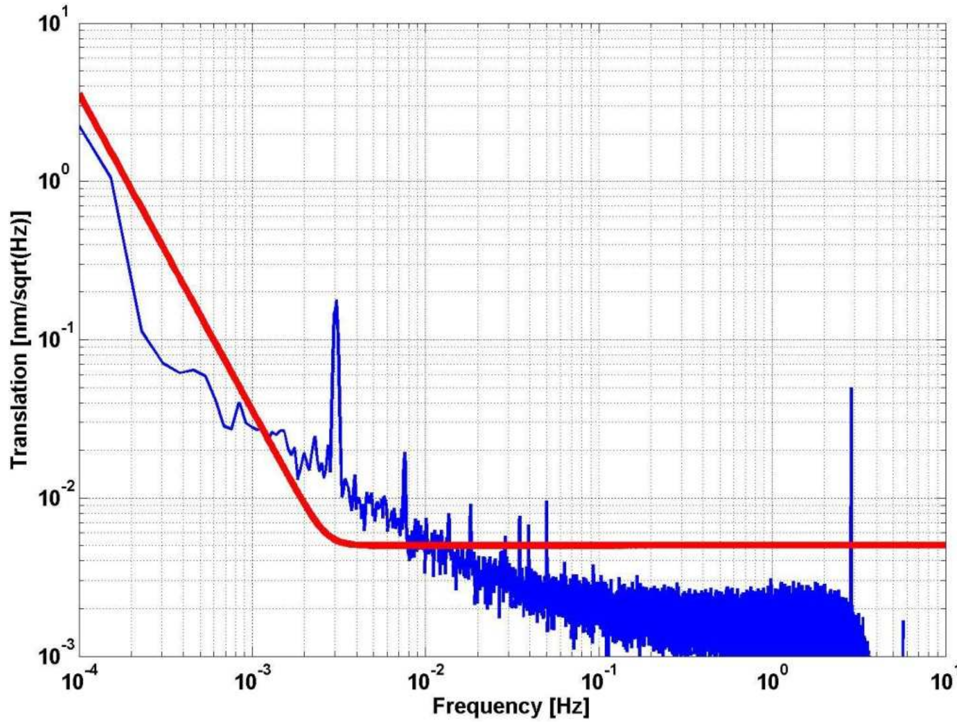
Shot (Schotty) noise

- Due to current inhomogeneities in resistive electric and electronic components (graphite).
- Poisson distributed.
- Heavier distribution tails than Gaussian (leptokurtotic, troublesome for LS fitting).
- Almost white and proportional to DC bias current.
- Almost independent between sensors.

SAPIENZA
UNIVERSITÀ DI ROMA

Flicker noise

- Still not well explained.
- Linked to time correlated releases of particles, energy...
- Lepto-kurtotic (pulse-type) noise.
- (*1/f*) power spectrum.
- Relevant at low frequencies (audio, sonar) below about 100 Hz, masked by other noise sources at higher frequencies.

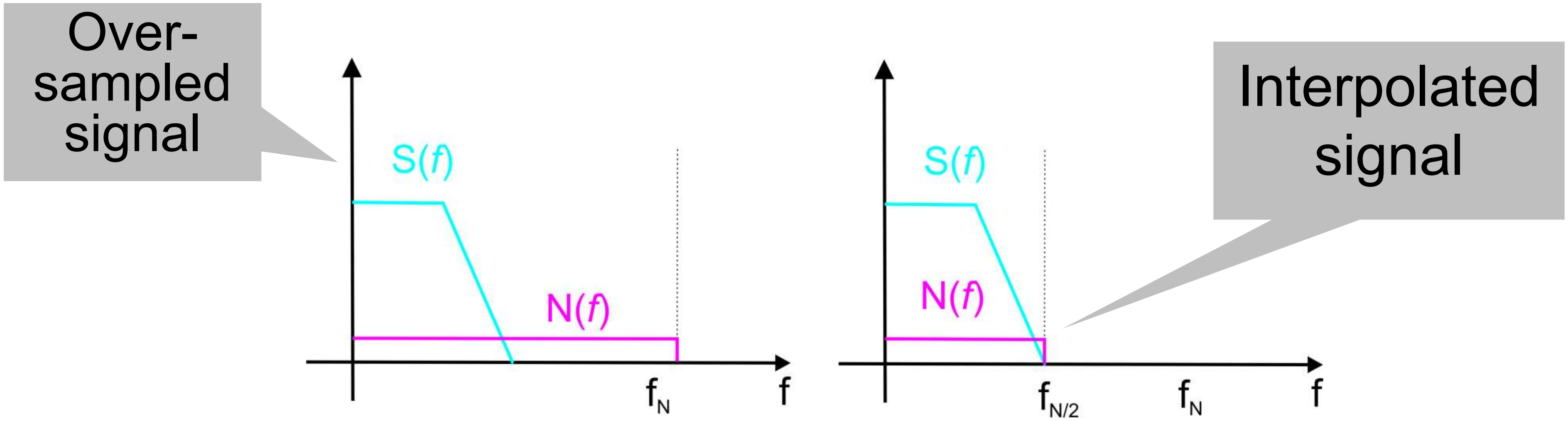


Quantization noise

- Typical of ADC quantization.
- Uniformly distributed.
- *Temporally correlated with simple or narrow-band signals*.
- May take the form of *spurious harmonics*.
- Spurious correlation with useful signal:
  - mitigated by dithering or wide-band, complex signals.
  - reduced by large ADC resolution.
- Nearly uncorrelated (spatially white) between sensors.

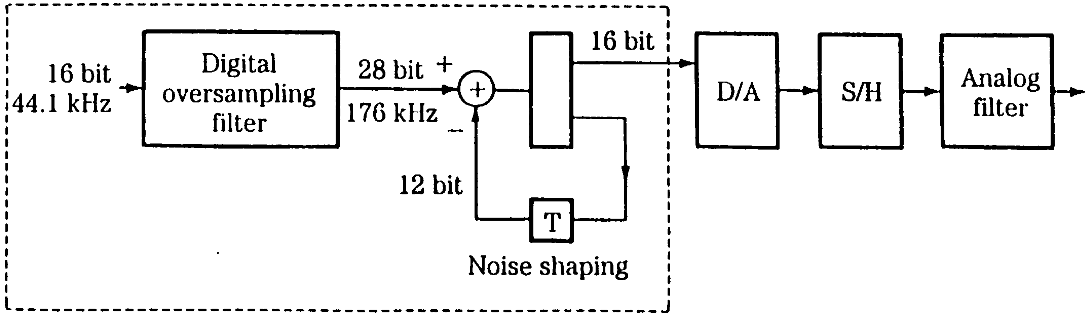
Oversampling

- Oversampling + low pass filtering + decimation
  - more room in frequency for dispersing noise, post filtered by a FIR interpolating filter.
  - increases correlation with signal and create spurious harmonics: moderate advantages without other countermeasures, because of *noise aliasing*.
  - Do not use IIR filters, that concentrate their strong re-quantization noise near system pole frequencies.

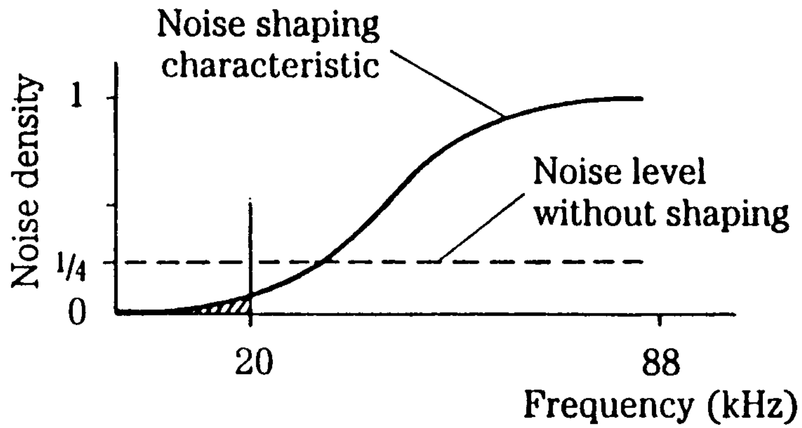


Noise shaping for audio

- In oversampled systems, collect, filter and re-circulate least significant bits to push noise out of the band.
- As an alternative, use Sigma-Delta differential modulation.



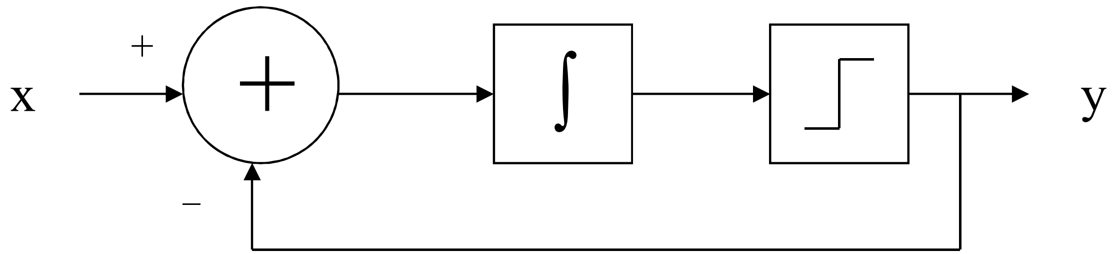
A



B

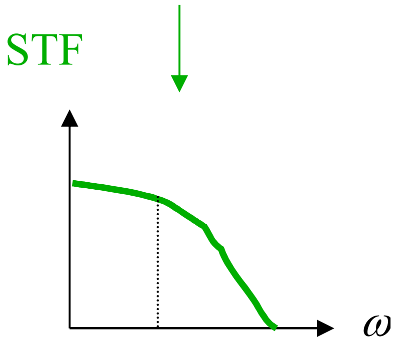
Sigma-Delta loops

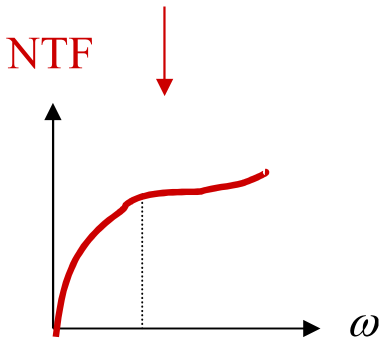
- Assumes that the quantizer generates white noise, uncorrelated with signals



Low-pass signal transfer function (STF)

High-pass noise transfer function (NTF)





$$Y(s) = \frac{X(s) - Y(s)}{s} + N(s)$$

$$Y(s) = \frac{1}{s+1} X(s) + \frac{s}{s+1} N(s)$$

Dithering

- ADC LSBs often excited by injecting *out of band* artificial noise at the input for *dithering* at little expense of in-band noise spectral density.
- Dither essentially shakes the ADC LSB, makes quantization noise uncorrelated with the signal and linearizes integral nonlinearities.
- Mainly done for RF (IF) signals. Digital dithering can be summed to input by a DAC and digitally subtracted or filtered out at the output.
- Enhanced resolution and SFR by more than 10 dB.
- Weak signal detection is also enhanced.

SAPIENZA
UNIVERSITÀ DI ROMA

Aperture jitter model

- Due to sampling time uncertainty in ADCs.
- Delay is multiplied by signal time derivative.
- ADC aperture jitter generates a Gaussian, temporally coloured (high pass or nearly white) noise
- Independent (i.e., spatially white) between sensors.
- The delay $\tau(n)$ is a random process, generally Gaussian and with a low-pass spectrum, mainly due to the random electrical load on clock drivers.

$$s_k\left[nT+\tau_k(n)\right] \cong s_k(nT) + \left.\frac{ds_k(t)}{dt}\right|_{t=nT} \tau_k(n) = s_k(nT) + e_k(nT)$$

$$|\tau_k| \ll 1; \quad E\left[e_k(\omega)e_l^*(\omega)\right] \cong \delta_{kl}\omega^2 P_s(\omega) * P_{\tau_k}(\omega)$$

SAPIENZA
UNIVERSITÀ DI ROMA

Intermodulation noise

- Addition of small non-linear distortion components originated by wide-band signals whose spectrum extends over more than one octave.
- Increases with signal power.
- Non stationary.
- Complex spectrum, flattens while increasing signal bandwidth.
- Maybe correlated between sensors (since they receive about the same signal mixture…).

External noise sources

- Sum of *many low-power signals* (machinery, shrimps) or *few strong impulsive sources* (lightning, jamming, NEMP, dolphins).
- Strong sources often better regarded as Signals of Interest (SOI), because of non-Gaussianity and nulling chance, sometimes excised from the sample.
- Spatially correlated between sensors.

$$\mathbf{v}(f) = \sum_{k=1}^{K} \mathbf{a}_k(\mathbf{p}_k, f) v_k\left(f\right); K \gg N$$

$$\mathbf{R}_{vv}(f) = E\left[\mathbf{v}(f)\mathbf{v}^H(f)\right]$$

$$= \sum_{k=1}^{K}\sum_{l=1}^{K} \mathbf{a}_k(\mathbf{p}_k, f) E\left[v_k\left(f\right) v_l^*\left(f\right)\right] \mathbf{a}_l^H(\mathbf{p}_l, f)$$

Clutter returns

- Active systems radiate e.m. energy versus targets.
- A large number of background objects reflect incoming signals toward the array receiver (clutter).
- Same return structure as signals of interest.
- Non Gaussian, heavy tailed clutter distribution.
- Clutter is correlated in space and with target returns.
- Clutter can be consistently estimated only in the absence of targets of interests.

Part IV

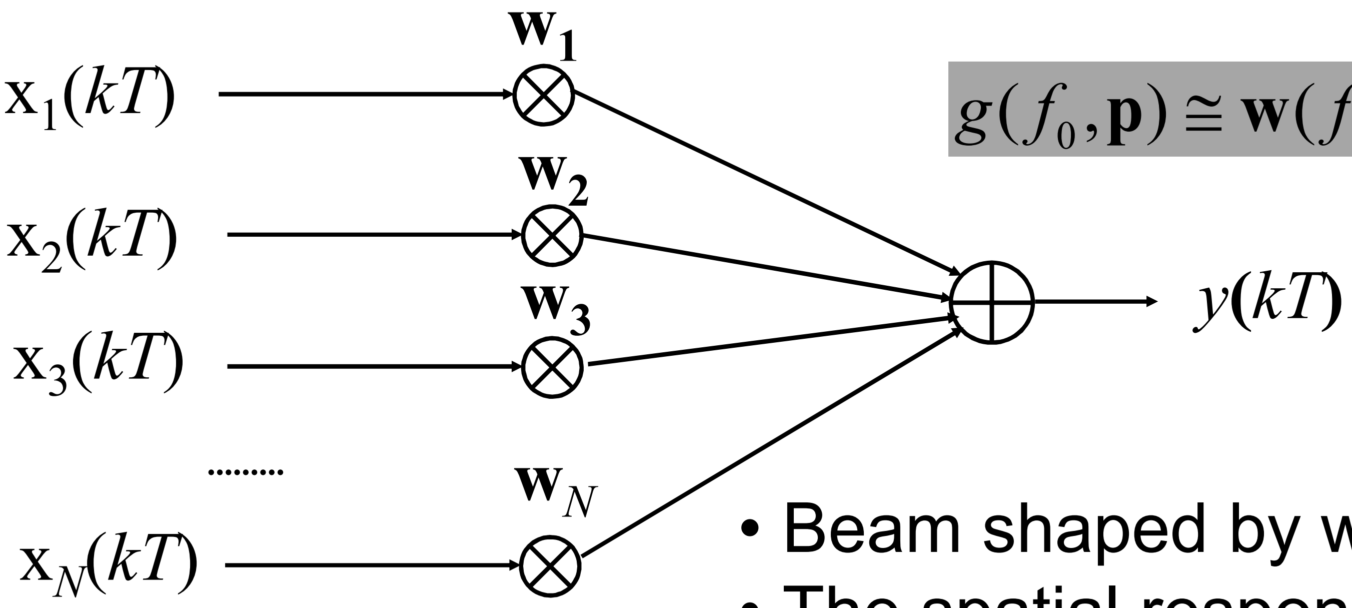
NARROWBAND ARRAY PROCESSING REVIEW

SAPIENZA
UNIVERSITÀ DI ROMA

Digital beamforming

- A convolutive combination of sensor outputs realizes a synthetic beam.
- Static and adaptive narrow-band beamforming is realized with a *single, generally complex, weight* for each sensor outputs and one summing node.
- On-line analog beamforming
  - *One beam* at a time;
  - *Transients* when switching weights;
  - High operating frequency;
  - Low latency.
- Off-line digital beamforming.
  - Multiple beams computed in parallel on stored data;
  - No transients on switching;
  - Low operating frequencies (IF, baseband);
  - High latency.

Narrowband non-adaptive beamforming (Haykin 1984)



$$g(f_0,\mathbf{p}) \cong \mathbf{w}(f_0)^H \mathbf{a}(f_0,\mathbf{p}) = b(f_0,\mathbf{p})$$

- Beam shaped by weight vector $\mathbf{w}$.
- The spatial response $b(f,\theta)$ is called *beampattern*.
- Sensor spatial responses constitute the linear basis for design, similar to digital FIR filters.

Typical beamforming and FIR filter optimization

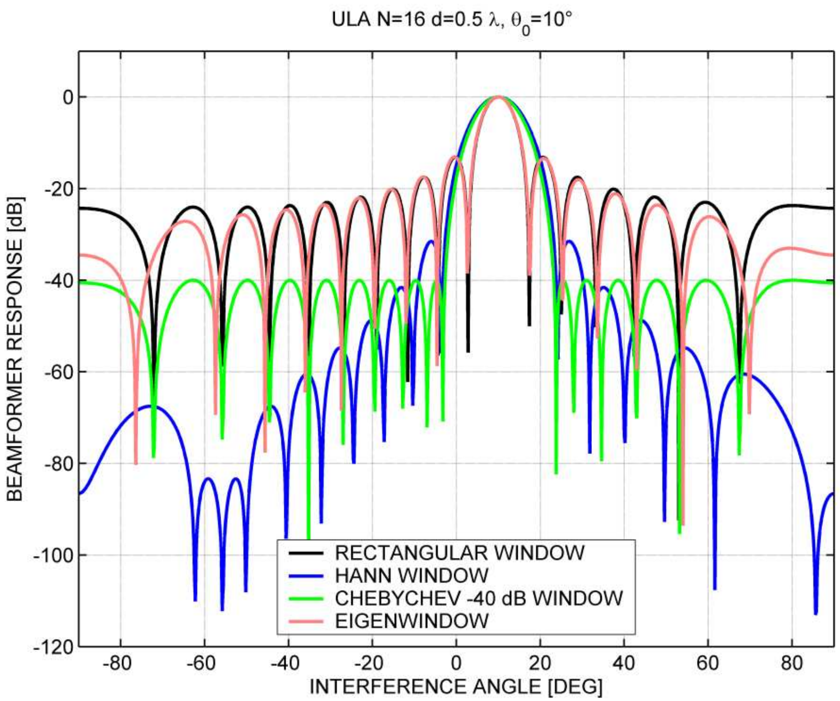
- High order error norms are used (we require smooth solutions).
- LS optimization, good for noise and aliasing filtering, can be performed also on-line.
- Equiripple or L-∞ norm or minimax designs for coefficien economy: complex and unstable algorithms only for batch works.

$$\hat{\mathbf{w}}(f) = \arg\min_{\mathbf{w}} \int_{\mathbf{p}\in \mathrm{FoV}} \left| g(f,\mathbf{p}) - \mathbf{w}^H \mathbf{a}(f,\mathbf{p}) \right|_p d\mathbf{p}$$

SAPIENZA
UNIVERSITÀ DI ROMA

ULA beamforming examples

- *Windows directly applied to the steering vector* to suppress sidelobes (rectangular or intrinsic, Hann, Chebichev,…).
- *Dominant pencil eigenvector* design $\mathbf{w}_{ev}$ optimized for uniform interference loading in angle space.



$$\mathbf{w}_{ev} \propto \arg\max_{\mathbf{w}} \left[\frac{\left|\mathbf{w}^H \mathbf{a}(\theta_0)\right|^2}{\mathbf{w}^H \mathbf{R}_{aa} \mathbf{w}} \right]$$

$$= \frac{\mathbf{R}_{aa}^{-1}\mathbf{a}(\theta_0)}{\mathbf{a}^H(\theta_0)\mathbf{R}_{aa}^{-1}\mathbf{a}(\theta_0)}$$

$$\mathbf{R}_{aa} = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}(\theta)\mathbf{a}^H(\theta)\,d\theta$$

Narrowband adaptive LCMV beamforming

•The weight vector $\mathbf{w}$ linearly combines array ouputs optimizing a specific criterion and satisfies $K<M$ linear constraints

$$y(l) = \mathbf{w}^H \mathbf{x}(l) \quad l = 1,\ldots L; \quad \text{s.t.} \quad \mathbf{C}^H \mathbf{w} = \mathbf{d}$$

•In this example, the signal arriving from $\mathbf{p_0}$ passes without distortion, the signal from $\mathbf{p_1}$ is cancelled, while the beamformer output energy is minimized:

$$\mathbf{C} = \left[\mathbf{a}(f,\mathbf{p}_0) \quad \mathbf{a}(f,\mathbf{p}_1)\right]$$
$$\mathbf{d} = \left[1 \quad 0\right]^T$$

$$\mathbf{w}_{opt} = \arg\min_{\mathbf{w}} \left\{ E\left[\left| y(l) \right|^2 \right] \right\}$$

SAPIENZA
UNIVERSITÀ DI ROMA

LCMV beamformer optimization

•The (M-K)-dimensional, orthogonal basis $\mathbf{U}$ (blocking matrix) is defined by

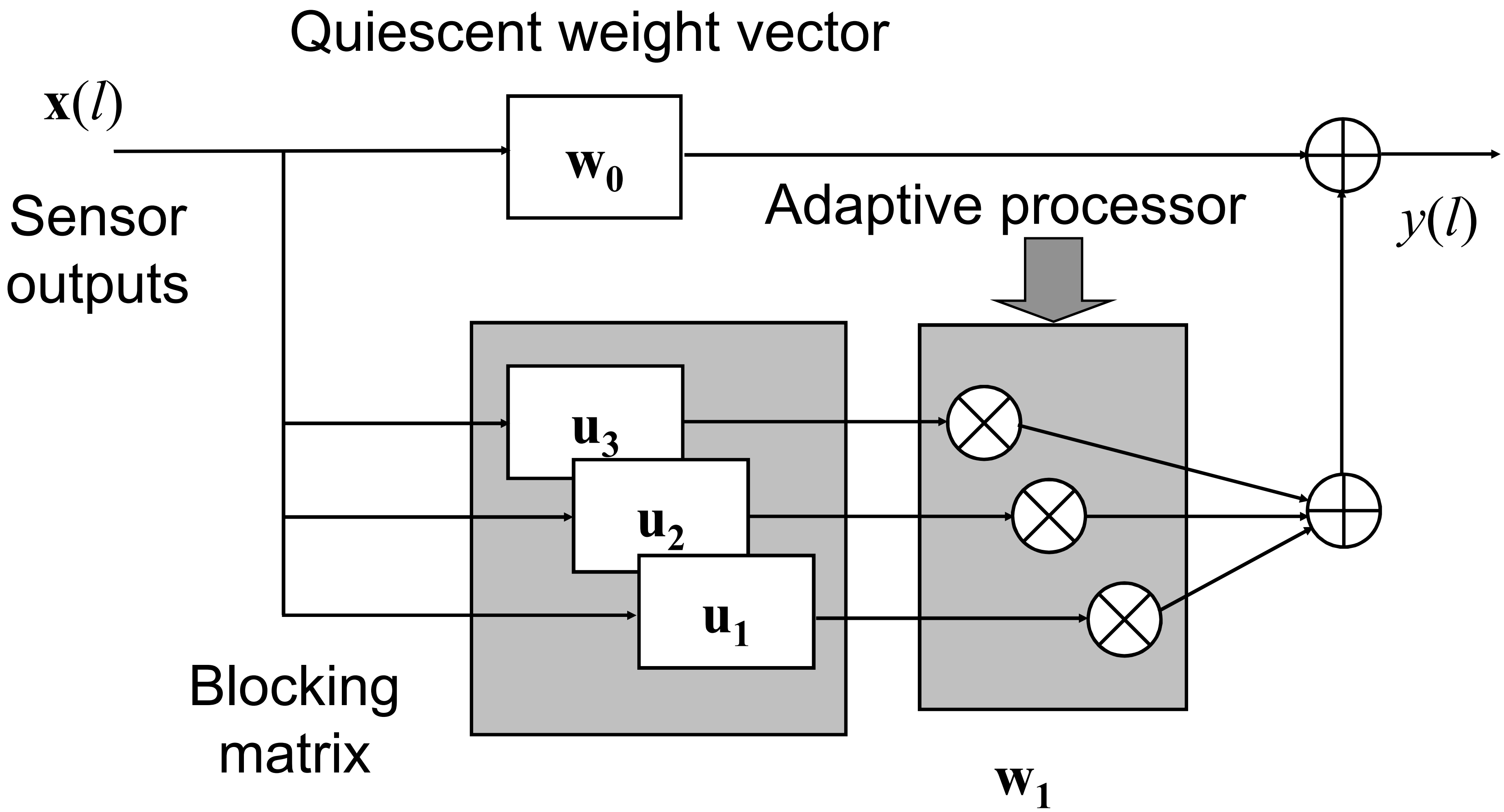
$$\mathbf{U}^H\mathbf{C} = \mathbf{0}$$

•The vector $\mathbf{w}$ is decomposed into a *fixed quiescent vector* $\mathbf{w}_0$, which satisfies the linear constraints, and an *adaptive vector* $\mathbf{w}_1$, which solves an *unconstrained* optimization problem.

$$\mathbf{w} = \mathbf{w}_0 + \mathbf{U}\mathbf{w}_1 \qquad\qquad \mathbf{C}^H\mathbf{w}_0 = \mathbf{d}$$

$$\mathbf{w}_{1opt} = \arg\min_{\mathbf{w}_1}\left\{E\left[\left|\mathbf{w}_0^H\mathbf{x}(l) + \mathbf{w}_1^H\mathbf{U}^H\mathbf{x}(l)\right|^2\right]\right\}$$

Narrow-band adaptive beamformer



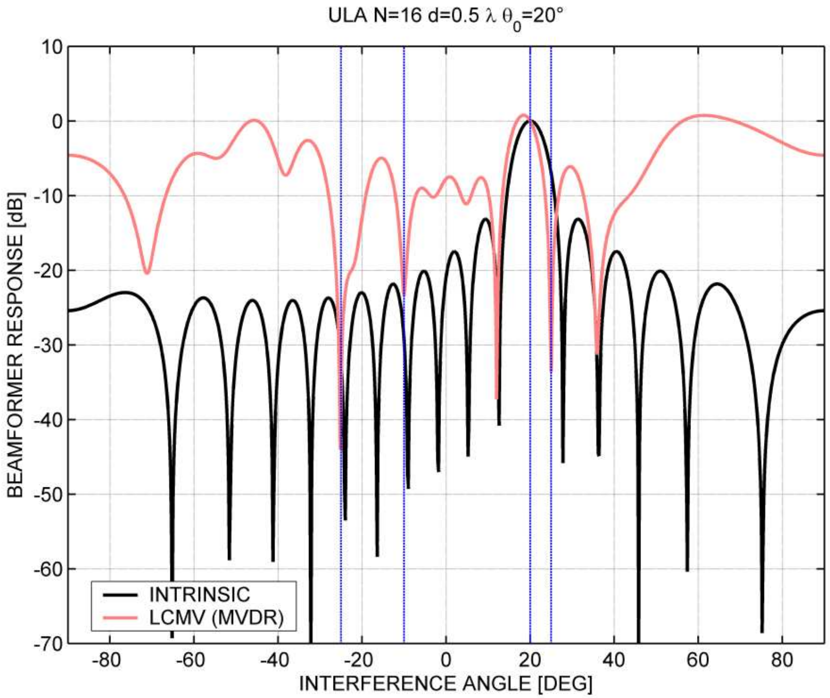
LCMV properties

- Assumes *interferences uncorrelated* with the signal of interest.
- Correlated interference (e.g., specular multipath, repeater jammer) entering the blocking matrix (auxiliary beams) produce *cancellation* of the useful signal.
- LCMV beamforming generalizes earlier and basic Applebaum and *Capon Minimum Variance Distortion-Less* (MVDR) beamformers.
- Assumes Gaussian white signals: the Least Squares error is the recovered signal. Cost function should be adapted to the actual signal distribution (Laplacian, pulsed,…), but the LCMV beamformer architecture is unchanged.

SAPIENZA
UNIVERSITÀ DI ROMA

LCMV (MVDR) example

- Four Gaussian uncorrelated sources at -25°, -10°, 20° and 25°, with SNR = 26, 20, 14 and 26 respectively.
- *Single unit gain constraint* at 10° (i.e., *Minimum Variance Distortion-Less Beamformer*, MVDR).



$$\mathbf{w}_{MVDR}(\mathbf{p}) = \frac{\hat{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\mathbf{p})}{\mathbf{a}(\mathbf{p})^H \hat{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\mathbf{p})}$$

$$\mathbf{a}(\mathbf{p})^H \mathbf{w}_{MVDR}(\mathbf{p}) = 1$$

Notes on LCMV

- Rather *high noise gain*.
- *Deep nulls at strong interfering* sources.
- Unit gain versus the source of interest.
- Numerically delicate and non-regular sidelobe level, partly due to finite sample SCM errors.
- If environment changes a little during adaptation (e.g., source direction, pointing constraint), *high level of noise and interferences can be re-injected* from the mainlobe.

Virtual array (*beamspace*)

- A bank of W di $K<N$ fixed beamformers applied to an array simulates a *virtual array* with K directive sensors.

$$\mathbf{b}(f,\theta) = \mathbf{W}^H \mathbf{a}(f,\theta) = \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_K \end{bmatrix}^H \mathbf{a}(f,\theta)$$

- A virtual array generally works well only within a *limited angular sector* Ω, but offers a bunch of statistical and computational advantages.

- K is essentially limited by the numerical rank of the sector matrix:

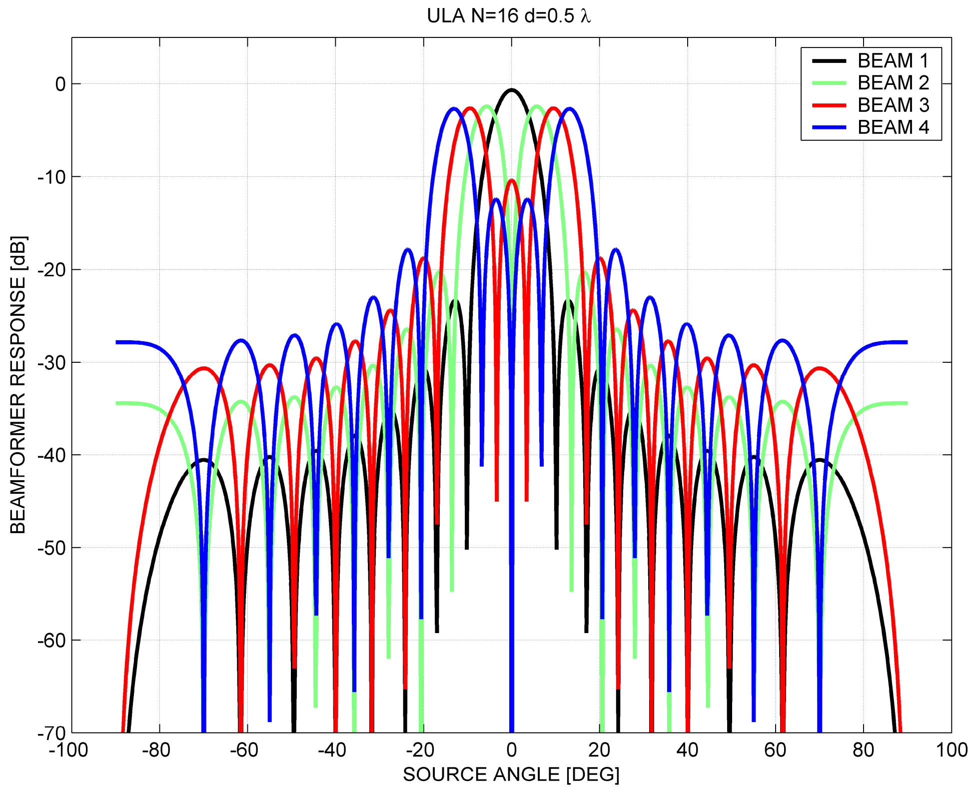
$$\mathbf{A}_\Omega = \int_{\theta\subset\Omega} p(\theta)\mathbf{a}(f,\theta)\mathbf{a}(f,\theta)^H d\theta$$

Virtual array uses

- *Compresses the size* of very large arrays (e.g., phased array radars, pruned antennas) while retaining maximum gain within a limited angular sector.
- *Interpolation and recalibration* of a real array onto a virtual one with special, simple geometry for reducing computations.
- *Focusing of a near-field source* onto a well-known virtual far-field response (distortion and curvature of field correction).
- *Focusing from multiple frequencies* onto a fixed virtual array (chromatic aberration correction).
- *Fault tolerance* (elimination of faulty sensors and subsequent response interpolation).

Prolate-like virtual array beams

- Four real, orthogonal beams concentrated within an angular sector centered at ULA broadside.



$$\left[\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathbf{a}(\theta)\mathbf{a}(\theta)^H \, d\theta\right]\mathbf{w}_k = \left[\int_{-\Delta\theta}^{\Delta\theta} \mathbf{a}(\theta)\mathbf{a}(\theta)^H \, d\theta\right]\mathbf{w}_k \lambda_k$$

Beamforming in trasmission

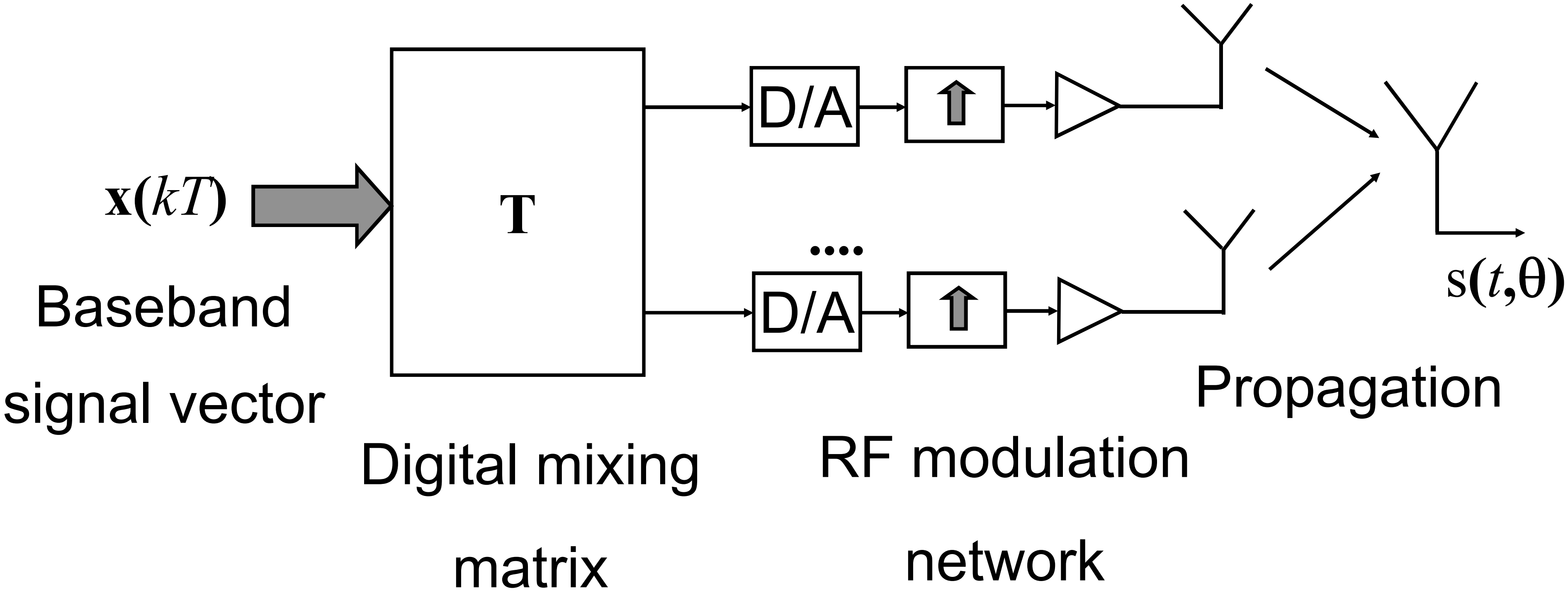
- A filtered version of the desired signal is applied at each transmitting transducer.
- Beam is reconstructed by the Huygens principle.
- Transmission beamforming *concentrates power* in the desired directions and minimizes the power radiated in undesired directions.
- Reduces EM pollution and evidence.
- *Multiple beams* can be simultaneously generated (*space-time coding*), for high capacity MIMO communications, sonar and multi-function radar.

Notes on transmit beamforming

- Basic rules for ambiguity are *identical* between transmit and receive beamforming.
- Ambiguity in capture means that *it is not possible to avoid radiation toward grating lobe* directions.
- So called *super-directive transmit beamforming* with widely spaced sensors allows grating lobes, at least in element space.
- Grating lobes are suppressed by a certain sensor directivity in beamspace.
- In transmission there is not the issue of steering vector normalization: in such applications only radiation intensity at the target location counts!

Narrow-band array transmitter

$$s(f,\theta) = e^{-j2\pi f \tau_p}\,\mathbf{a}(f,\mathbf{p})^T\,\mathbf{T}\mathbf{x}(f)$$



SAPIENZA UNIVERSITÀ DI ROMA

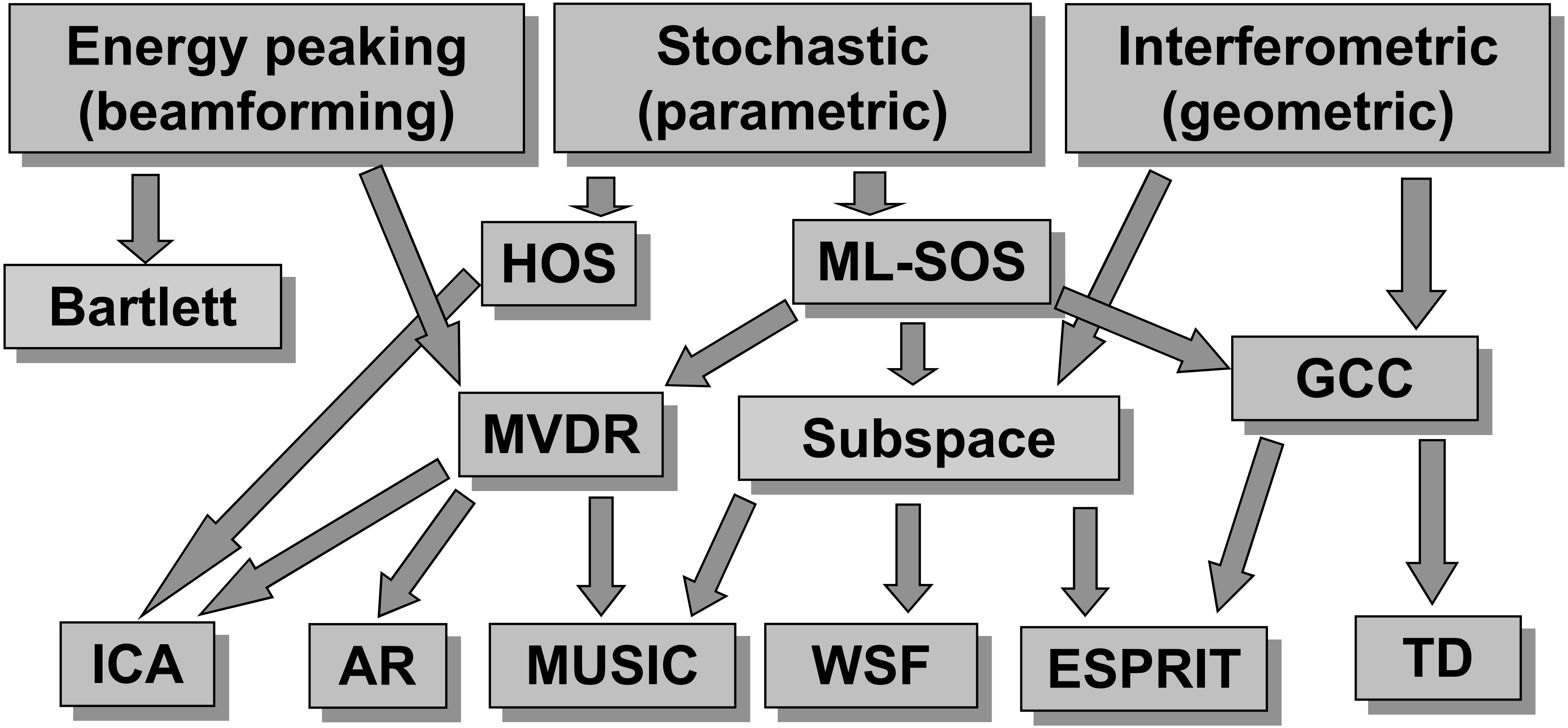
Multi-source localization

- Sonar and radar should detect and locate multiple, closely spaced targets in typical environments.
- Data batch available: *track before detect. No information loss accepted before localization* (i.e., no lossy compressed sensing).
- Signal copy to validate detection.
- Only arrays with DSP offer sufficient information and flexibility for all these tasks.
- High computational costs, but affordable today.

Early narrow-band source localization techniques

- Earlier parametric direction finding narrow-band techniques were based on interferometry from two to four sensors.
  - Phase /amplitude real time analysis.
  - Special antenna design and accurate calibration.
- More measurements are available in modern arrays with DSP.
  - Spatial AR regression over ULAs to detect wavenumbers;
  - Beamforming on a grid of spatial coordinates and output power or feature peaking (*spatial imaging*).
  - Anyway, *not a correct exploitation* of the array signal model.

Source localization approaches



The spatial covariance approach

- All existing localization algorithms are based on the *structure* of the Spatial Covariance Matrix (SCM).
- The SCM collects all available stationary second order information from narrow-band arrays and circular signals (uncorrelated I/Q components).
- High-order statistics (HOS) are sometimes used, but they are built so to share the basic SCM structure.
- Other second-order (e.g., ciclostationary), approaches are more rigorously viewed within a wide-band framework, since they observe *temporal correlations* between snapshots.

$$\mathbf{R}_{xx}(f) = E\left[\mathbf{x}(l)\mathbf{x}^H(l)\right]$$

SAPIENZA
UNIVERSITÀ DI ROMA

Spatial covariance (SCM)

$$\mathbf{x}(l;f)=\mathbf{A}(\mathbf{p},f)\mathbf{s}(l)+\mathbf{v}(l),\quad l=1,2,\ldots,L$$
$$\mathbf{p}=\left\{p_1\quad\cdots\quad p_D\right\}$$

$$E[\mathbf{s}(l)]=E[\mathbf{v}(l)]=\mathbf{0};\ E[\mathbf{s}(k)\mathbf{v}(l)^H]=E[\mathbf{s}(k)\mathbf{v}(l)^T]=\mathbf{0};$$
$$E[\mathbf{s}(k)\mathbf{s}(l)^T]=E[\mathbf{v}(k)\mathbf{v}(l)^T]=\mathbf{0};\quad E[\mathbf{s}(k)\mathbf{s}(l)^H]=\delta_{kl}\mathbf{P}(f)$$
$$E[\mathbf{v}(k)\mathbf{v}(l)^H]=\delta_{kl}\lambda_v(f)\mathbf{R}_{vv}(f);\quad E[\mathbf{x}(k)\mathbf{x}(l)^H]=\delta_{kl}\mathbf{R}_{vv}(f)$$

$$\begin{aligned}\mathbf{R}_{xx}(f)&=E\left[\mathbf{x}(l;f)\mathbf{x}^H(l;f)\right]\\&=\mathbf{A}(\mathbf{p},f)\mathbf{P}(f)\mathbf{A}^H(\mathbf{p},f)+\lambda_v(f)\mathbf{R}_{vv}(f)\end{aligned}$$

SAPIENZA
UNIVERSITÀ DI ROMA

Gaussian SCM estimate

- Zero mean variables (previously remove DC offsets and demodulated sinusoidal carriers).
- Sample mean of *independent snapshots* (i.e., critically Nyquist sampled vector array outputs).
- Non-uniform time weighting is possible.

$$\mathbf{x}(l;f) = \left[x_1(lT;f) \quad \cdots \quad x_N(lT;f) \right]^T$$

$$\mathbf{X}(f) = \left[\mathbf{x}(1;f) \quad \cdots \quad \mathbf{x}(L;f) \right]; \quad L \gg N$$

$$\hat{\mathbf{R}}_{xx}(f) = \frac{1}{L} \sum_{l=1}^{L} \mathbf{x}(l;f)\mathbf{x}^H(l;f) = \frac{1}{L}\mathbf{X}(f)\mathbf{X}^H(f)$$

Remarks on SCM

- Sufficient statistics only for *temporally white, circular* Gaussian noise and signals (unconditional model) and near optimal for most practical environments.
  - Gaussian ML estimators are functions of the SCM only.
  - Energy-type (bolometric) detection and estimation.
  - Needs further information on $\mathbf{R}_{vv}$.
- Standard array signal model assumes isotropic noise with unknown variance ($\mathbf{R}_{vv} = \lambda_v \boldsymbol{I}$).
- *Whitening* required for general noise covariance.
- Modified (real-valued) SCM required for non-circular (even Gaussian) signals and noise.

Noise whitening

- Algorithms are initially formulated for white noise, i.e., with SCM proportional to the identity matrix.
- The *known* or the *independently estimated* noise SCM is factorized into Cholesky or Hermitian square root factors.
- The inverse of these factors is applied to signal snapshots before computing SCM or to the SCM estimate.
- The transformed noise covariance is now white.

$$\mathbf{R}_{vv}\left(f\right)=\mathbf{R}_{v}\left(f\right)\mathbf{R}_{v}^{H}\left(f\right);\quad \mathbf{y}\left(l;f\right)=\mathbf{R}_{v}^{-1}\left(f\right)\mathbf{x}\left(l;f\right)$$

$$\begin{aligned}\mathbf{R}_{yy}&=E\left[\mathbf{y}\left(l;f\right)\mathbf{y}^{H}\left(l;f\right)\right]=\mathbf{R}_{v}^{-1}\left(f\right)\mathbf{R}_{vv}\left(f\right)\mathbf{R}_{v}^{-H}\left(f\right)\\&=\left[\mathbf{R}_{v}^{-1}\left(f\right)\mathbf{A}\left(\mathbf{p},f\right)\right]P\left(f\right)\left[\mathbf{R}_{v}^{-1}\left(f\right)\mathbf{A}\left(\mathbf{p},f\right)\right]^{H}+\lambda_{v}\mathbf{I}\end{aligned}$$

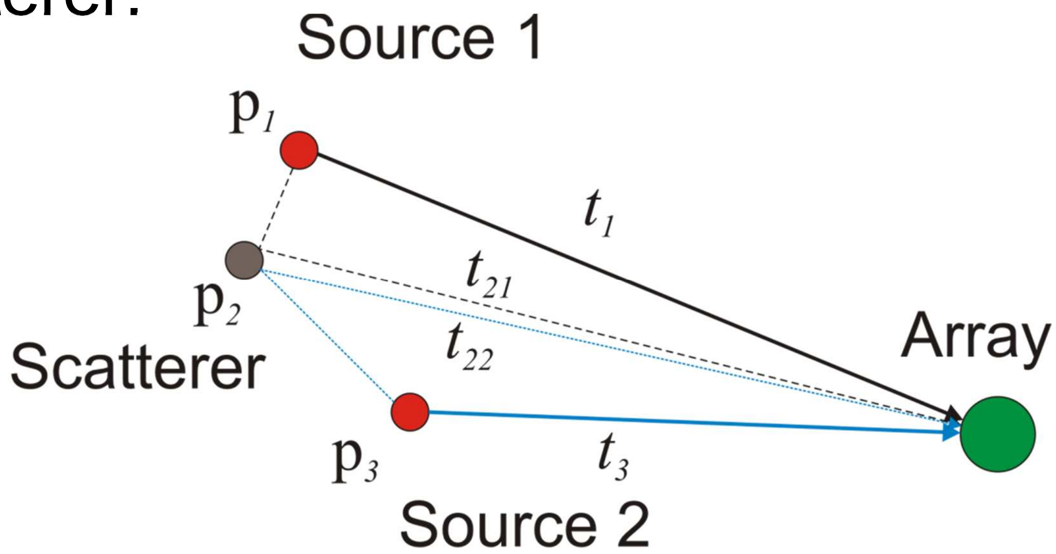
SAPIENZA
UNIVERSITÀ DI ROMA

Signal coherence

- If some signal can be at *least partially reconstructed by other ones*, it is said to be *coherent* with them.
- Coherence strongly degrades source localization and signal analysis by arrays.
- Coherence is originated by *specular multipath* over smooth surfaces and *jamming*.
- Sometimes modeled by a curved wavefront.
- The steering vector of coherent sources is a linear combination of the steering vectors of all rays.
- In the fully coherent case, $\mathbf{P}(f)$ is *rank deficient*.
- Partial coherence is observed in scattering and clutter.

Coherence example

- Two radiating sources and a common (point) scatterer.



$$y(t)=\left[\mathbf{a}\left(\mathbf{p}_1\right)\quad \mathbf{a}\left(\mathbf{p}_2\right)\quad \mathbf{a}\left(\mathbf{p}_3\right)\right]\begin{bmatrix} s_1(t-t_1) \\ \gamma_1 s_1(t-t_{21})+\gamma_2 s_2(t-t_{22}) \\ s_2(t-t_2) \end{bmatrix}+\mathbf{v}(t)$$

Coherent signal covariance

- The signals of sources 1 and 2 are uncorrelated.
- Scattered signal is a mix of the two source signals.
- Covariance is Hermitian and positive semi-definite.

$$\mathbf{P} = \begin{bmatrix} p_1 & \gamma_1^* p_1 \rho_1 \left(t_{21} - t_1\right) & 0 \\ \gamma_1^* p_1 \rho_1 \left(t_1 - t_{21}\right) & \left|\gamma_1\right|^2 p_1 + \left|\gamma_2\right|^2 p_2 & \gamma_2 p_2 \rho_2 \left(t_{22} - t_2\right) \\ 0 & \gamma_2^* p_2 \rho_2 \left(t_2 - t_{22}\right) & p_2 \end{bmatrix}$$

Beamforming approach (Bartlett)

- A directive beam $\mathbf{w}(\mathbf{p})$ is pointed toward each location $\mathbf{p}$ of interest;
- Locations are estimated by the local maxima of the beam output power.

$$\hat{\mathbf{p}}_{BF} = \arg\max_{\mathbf{p}} \left\{ \frac{\mathbf{w}(\mathbf{p})^H \hat{\mathbf{R}}_{xx} \mathbf{w}(\mathbf{p})}{\mathbf{w}(\mathbf{p})^H \mathbf{w}(\mathbf{p})} \right\}$$

- This estimate is ML iff a *single source* is present, $\mathbf{w}(\mathbf{p})=\mathbf{a}(\mathbf{p})$ and the *noise is white*. ($\mathbf{R}_{nn}=\lambda\mathbf{I}_M$).

Beamforming based localization

- Generalization of periodogram techniques.
- Non-parametric estimates (*imaging*).
- Tolerant to signal coherence and model mismatches.
- Easy to compute in real time with precomputed beamformers on a grid of location parameters.
- Source direction of arrival (DOA) estimated by picking the angle of *maximum beamformer output power*.
- *Non consistent* estimation in the multi-source case.
  - *bias* due to beam sidelobes.
  - some sources may be *masked* or *cancelled*.
  - Rayleigh resolution limit dictated by CC ambiguity function.

Minimum Variamce Distortionless beamformers (Capon or MVDR)

- $\mathbf{w}(\mathbf{p})$ minimizes the array output power without linearly distorting *the signal of interest* coming from $\mathbf{p}$ (i.e., $\mathbf{w}(\mathbf{p})^H\mathbf{a}(\mathbf{p},f)=1$) ;
- $\mathbf{w}(\mathbf{p})$ *suffers of cancellation of the useful signal* if the array is mis-calibrated and/or sources are coherent, but it has *low sensitivity to SCM estimation* errors *and to the presence of multiple closely spaced sources*.

$$\mathbf{w}(\mathbf{p}) = \frac{\hat{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\mathbf{p},f)}{\mathbf{a}(\mathbf{p},f)^H\tilde{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\mathbf{p},f)}$$

$$\hat{\mathbf{p}}_{MVDR} = \arg\max_{\boldsymbol{\theta}}\left\{\frac{\mathbf{a}(\mathbf{p},f)^H\mathbf{a}(\mathbf{p},f)}{\mathbf{a}(\mathbf{p},f)^H\hat{\mathbf{R}}_{xx}^{-1}\mathbf{a}(\mathbf{p},f)}\right\}$$

Other beamformers for localization

- More sophisticated beamformers can be devised:
  - APES uses a normalized MVDR vector for constant noise power output.
  - More complex LCMV beamformers for particular cases (fixed sources/jammers).
  - Robust beamformers to cope with model uncertainties and tailored to *signal copy*.
- Beamformers are superficially attractive for localization, because of simplicity and parallel processing, but are very costly, require large arrays and furnish quantized, non consistent, location estimates.
- Beamformers may be useful for quick data inspection, but the performance of MVDR with coherent arrivals is marginal.

Narrow-band parametric models

$$\hat{\mathbf{p}} = T\left\{\mathbf{x}(l), l = 1, \ldots, M; D\right\}$$

- Fully exploit the multi-source array signal model.
- Mostly based on a probabilistic data description.
- The number of sources (say D) must be *known a priori or estimated* from a set of snapshots.
- Gaussian parametric models demonstrated good adherence to real world scenarios (at least for noise) and have nice geometrical interpretations.

Maximum Likelihood parameter estimation

- If a *probabilistic characterization of signals* is available and the parametric model equations are *exact*, a Maximum Likelihood (ML) estimator under mild conditions furnishes a set of noteworthy advantages:
  - The ML approach casts parameter estimation as the search of the local or global extrema of a real-valued functional;
  - *ML estimates are asymptotically unbiased;* bias vanishes as $O(1/M)$ for a number of independent observations M;
  - *ML estimates are consistent*, i.e., estimation variance asymptotically vanishes as $O(1/M)$;
  - If the number of parameters remains finite w.r.t. M, the ML estimator is *asymptotically efficient*, i.e., approaches the variance Cramer Rao Bound (CRB) for unbiased estimators.

Mainstream Gaussian models

- *Conditional* Gaussian model:
  - *Independent signal and noise* processes.
  - *Gaussian circular noise field* with known SCM, up to a scale factor;
  - *Deterministic, but unknown signals*, that are estimated together with location parameters.
  - *Consistent, but not asymptotically efficient* (though near optimal) vs. location for finite SNR and array size.
  - *Non-consistent (noisy) though useful signal estimates*.
  - *Insensitive to signal* distribution.
- *Unconditional* Gaussian model:
  - *Independent, circular, Gaussian signal and noise* processes;
  - Estimates locations, signal covariance, noise variance.
  - *Asymptotically efficient vs. stochastic CRB*, which is higher than deterministic CRB.
  - *Sensitive to actual signal and noise* distributions.

Some comments on ML estimators

- The *partition* between noise and signals (of interest) is rather arbitrary and flexible.
- The essential limit to the partition is the statistical independence between the two processes.
- The steering vectors of all sources of interest must be known.
- The noise SCM estimate must be consistent.
- Some signals may be considered as *deterministic* or *subject to known parametric modifications* (dispersion, Doppler shifts).
- In many practical cases, performance tradeoffs of different models are not clear.

Identifiability conditions (Schmidt, 1980)

- The general localization problem with sensor arrays is *well posed* (not ambiguous) only when:
  - The sources of interest are *point sources* (Dirac pulses in the location parameter space), *narrow-band* (the source has a *rank-one representation* in the SCM) .
  - The source number D is smaller than the sensor number (i.e., N).
  - The number of *free model parameters to estimate* (location+signal+noise related ones) is smaller than the number of free real parameters of the sample SCM (equal or less than N^2). This is not obvious at all in 3-D or in some physical problems (e.g., light speed estimate).
  - D steering vectors, related to different directions, are always linearly independent (no *spatial aliasing*);

Conditional ML estimates

- The *noise is spatially white* (or pre-whitened).
- Often a *single snapshot* is avaliable (*monopulse* estimation).
- *Signals* are unknown and estimated *together with location parameters* (no a priori distribution is assumed).
- Raw array ouputs can be replaced by *matched filter sampled outputs* without information loss.
- *Single source* of interest in many cases.
- *Simplified steering vector* by *special array geometries* (Watson-Watts, proportional 2-D e 3-D monopulse).
- *Non linear* multi-source estimation. Sub-optimal estimators often preferred for initialization and less stringent requisites.

Gaussian conditional ML monopulse model

- Single source, monopulse case.
- Signal estimated by LS fitting, conditioned to the candidate steering vector.

$$\mathbf{e}(k) = \mathbf{x}(k) - \mathbf{a}\left(\mathbf{p}, f\right) s(k)$$

$$E\left[\mathbf{e}(k)\mathbf{e}(k)^H\right] = \lambda_v \mathbf{R}_{vv}$$

$$\mathrm{L}\left[\mathbf{x}(k) \big/ \left(\mathbf{p}, \left\{\mathbf{s}(k)\right\}\right)\right] = \frac{1}{\pi^N \det\left(\lambda_v \mathbf{R}_{vv}\right)} e^{-\lambda_v^{-1} \mathbf{e}(k)^H \mathbf{R}_{vv}^{-1} \mathbf{e}(k)}$$

$$\mathrm{L}\left[\mathbf{x}(k)/\mathbf{p}\right] \propto N \log\left(\lambda_v\right) + \lambda_v^{-1} \operatorname{trace}\left[\mathbf{e}(k)^H \mathbf{R}_{vv}^{-1} \mathbf{e}(k)\right]$$

ML monopulse estimator in white noise

- Signal estimate can be eliminated and the resulting *concentrated* ML is a special beamforming search in the location parameter space.
- Neyman-Pearson detection test is set up after ML estimate.

$$\hat{\mathbf{p}}_{CML} = \arg\min_{\mathbf{p},s} \left| \mathbf{a}(\mathbf{p}, f)s - \mathbf{x} \right|_F$$

$$= \arg\min_{\mathbf{p}} \left\| \left[\mathbf{I} - \frac{\mathbf{a}(\mathbf{p}, f)\mathbf{a}(\mathbf{p}, f)^H}{\left| \mathbf{a}(\mathbf{p}, f) \right|^2} \right] \mathbf{x} \right|_F$$

$$\hat{s}_{CML} = \frac{\mathbf{a}^H(\hat{\mathbf{p}}_{CML}, f)\mathbf{x}}{\left| \mathbf{a}(\hat{\mathbf{p}}_{CML}, f) \right|^2}; \quad \hat{\lambda}_v = \left| \mathbf{a}(\hat{\mathbf{p}}_{CML}, f) s_{CML} - \mathbf{x} \right|_F$$

Beamspace 3-D proportional monopulse

- Generally formed by four identical, directive sensors colocated on the vertices of a square on the vertical yz plane, recombined into three orthogonal beams
- Elevation ϕ and azimuth θ are zero along the x axis.
- "Sum" beam (*sigma*) no. 1 with a single directional mainlobe
- Gain of "difference" (*delta*) beams no. 2 e 3 about proportional to θ and ϕ through *monopulse gains*.

$$\mathbf{a}(\theta,\phi,f) \cong g(\theta,\phi,f)\begin{bmatrix} 1 \\ K_\theta(f)\theta \\ K_\phi(f)\phi \end{bmatrix}; \quad \theta,\phi \ll 1$$

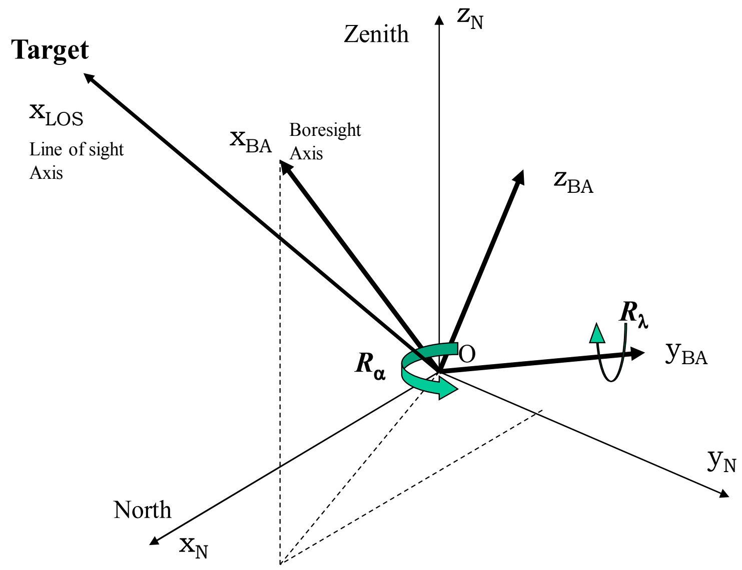
Watson-Watts array

- Four identical and slightly directive sensors at the vertices of ha square with diagonal length d and sides aligned along x e y axes.

$$\mathbf{a}(\theta,\phi,f) \cong g(\theta,\phi.f)\begin{bmatrix} e^{j\frac{\pi d}{\lambda(f)}\cos(\theta)\cos(\phi)} \\ e^{-j\frac{\pi d}{\lambda(f)}\cos(\theta)\cos(\phi)} \\ e^{j\frac{\pi d}{\lambda(f)}\sin(\theta)\cos(\phi)} \\ e^{-j\frac{\pi d}{\lambda(f)}\sin(\theta)\cos(\phi)} \end{bmatrix}$$

Polar source 2D angle parametrization and monopulse

- The source position is parametrized by the *off-boresight* angle θ and by the *revolution* angle φ around the boresight axis.



Polar separable Gauss-Laguerre functions (GLF)

Generalized radial Laguerre polynomials and tangential circular harmonics.

$$L_{k,n}(r,\varphi) = (-1)^k 2^{(|n|+1)/2} \pi^{|n|/2} \sqrt{\frac{k!}{(|n|+k)!}}$$

$$\times r^{|n|} P_{k,n}\left(2\pi r^2\right) e^{-\pi r^2} e^{jn\varphi}$$

$$P_{k,n}(x) = \sum_{h=0}^{k} (-1)^h \binom{n+k}{k-h} \frac{x^h}{h!}$$

Polar Amplitude Comparison Monopulse (P-ACM) receiver

- For any pair of GLFs having radial order 0 and consecutive radial orders, it is possible to build a P-ACM as follows:

$$L_{0,n}(r,\varphi) \propto r^{|n|} e^{-\pi r^2} e^{jn\varphi}$$

Circular harmonic

GL radial profile

$$R_n(r,\varphi) \propto \frac{L_{0,n+1}(r,\varphi)}{L_{0,n}(r,\varphi)} \propto r e^{j\varphi}$$

Monopulse ratio

First-order P-ACM

- The simplest polar ACM is characterized by the following three beampatterns:

$$\gamma = \vartheta / \sqrt{2}\beta$$

Angular width of the sum beam

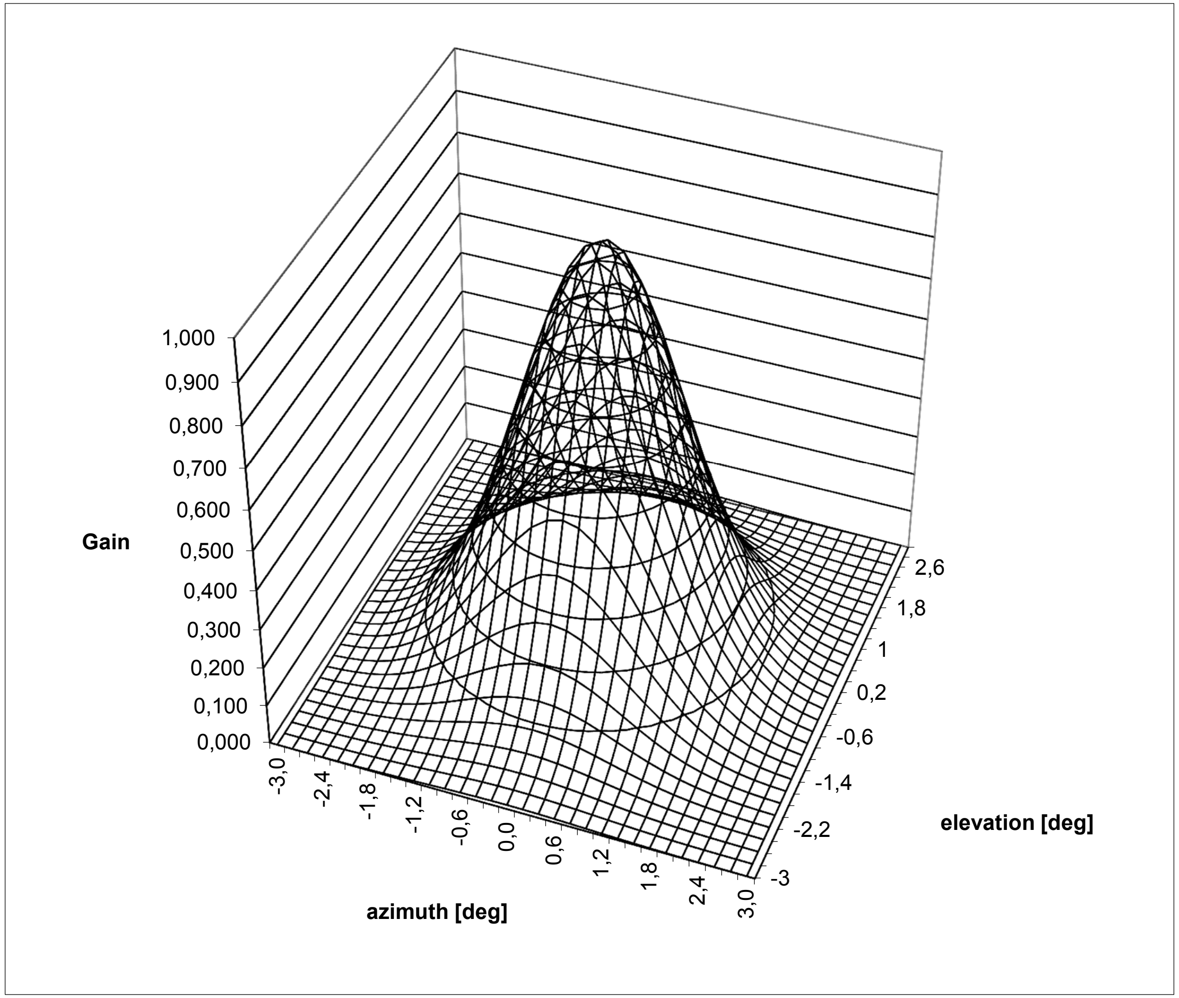
$$\begin{cases} G_1(\gamma) = \mathrm{e}^{-\gamma^2} \\ G_2(\gamma,\varphi) = \mathrm{Re}\left[\gamma\, \mathrm{e}^{-\gamma^2}\, \mathrm{e}^{j\varphi}\right] \\ G_3(\gamma,\varphi) = \mathrm{Im}\left[\gamma\, \mathrm{e}^{-\gamma^2}\, \mathrm{e}^{j\varphi}\right] \end{cases}$$

Sum beam

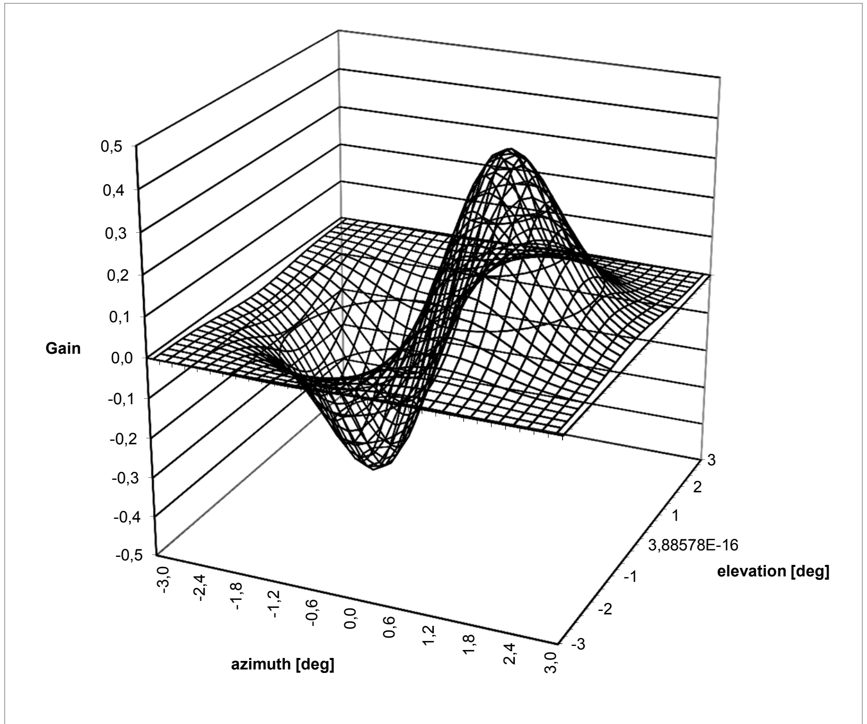
In-phase beam

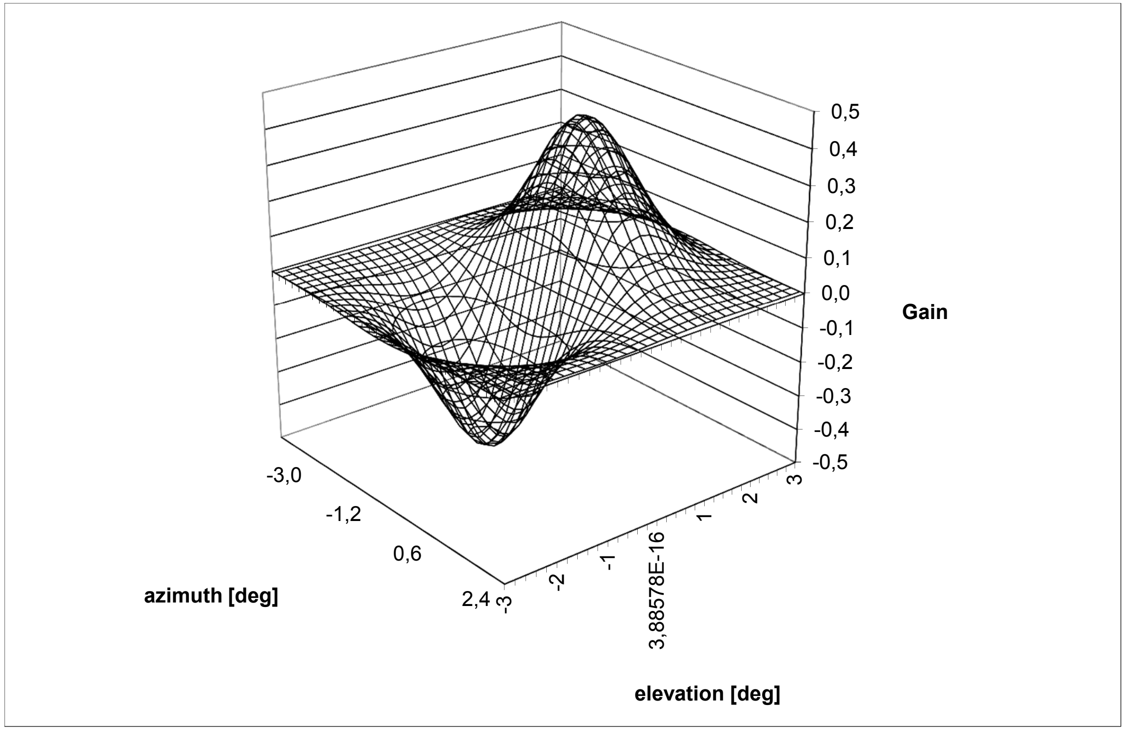
Quadrature beam

P-ACM sum beampattern



P-ACM difference beampatterns





Unconditional Gaussian model assumptions

- In the *stochastic model* for *passive arrays*, signals and noise are generally assumed as realizations of independent, zero mean, *ergodic*, multivariate circular Gaussian processes, since:
  - the PDF of narrowband filtered signals approaches a Gaussian PDF (Whittle, 1956), under mild assumptions.
  - typical deviations from Gaussianity (impulse noise, sub-Gaussian signals) can be treated as *outlier contaminations* (Johnson, 1993) and/or handled by simple modifications of techniques derived in a Gaussian framework.
  - ML parameter estimators for general non-Gaussian array signals are generally *direction-dependent* and require information unavailable in practice.
- $M>N$ independent snapshot available.
- Location, signal covariance and noise variance are the unknowns.

Unconditional Gaussian ML estimation

- The SCM is a *sufficient statistic* for ML parameter estimation, if:
  - the array response is a *continuous and derivable* function of the directional parameters
  - the array response is *perfectly known* for any possible combination of parameters (a real issue).
  - the noise-only SCM is known up to a real, positive scalar.
  - the number of unknown parameters to be estimated is less than the number of the degrees of freedom of the SCM
- Array or signal redundancies (e.g., *ciclo-stationarity*) can be exploited to achieve identifiability in special cases.
- Under these hypotheses, ML estimation is asymptotically ($M \rightarrow \infty$) unbiased and efficient
- Estimation of directional parameters is *statistically decoupled* from that of signal cross-spectra in most cases.

Unconditional ML localization

$$\hat{\mathbf{R}}_{xx} = \frac{1}{M}\sum_{l=1}^{L}\mathbf{x}(l)\mathbf{x}(l)^{H} \xrightarrow{w.p.1} \mathbf{R}_{xx} = \mathbf{A}\left(\mathbf{p}\right)\mathbf{P}\mathbf{A}^{H}\left(\mathbf{p}\right) + \lambda_{v}\mathbf{I}$$

$$\hat{\mathbf{p}}_{UML} = \arg\max_{\mathbf{p},\mathbf{P},\lambda_v}\left\{\mathrm{trace}\left[\hat{\mathbf{R}}_{xx}\mathbf{R}_{xx}^{-1}\right]\right\} \text{ s.t. } \mathbf{R}_{xx} = \mathbf{A}\left(\hat{\mathbf{p}}\right)\hat{\mathbf{P}}\mathbf{A}^{H}\left(\mathbf{p}\right) + \hat{\lambda}_{v}\mathbf{I}$$

- Results in a non-linear *covariance fitting* which requires a joint search in the directional parameter space.
- After 1992, ML algorithms have been replaced by more flexible and still asymptotically efficient *signal subspace fitting* algorithms (WSF, MODE) based on the *equivalent* signal subspace concept.

Subspace fitting estimators

- An alternate, but *statistically and numerical equivalent* characterization of the SCM is in terms of its *eigenvectors*, partitioned into signal and noise subspaces, and *eigenvalues*.
- ML algorithms can be manipulated to work on this alternate SCM characterization.
- A priori estimation of the source number (or, better, the number of uncorrelated, impinging signals) can be made from the eigen-spectrum.
- If the choice is correct, asymptotical efficiency can be retained.

Signal subspace

- For *spatially white* noise and D sources, only $\eta \leq D$ eigenvalues of $\mathbf{R}_{xx}(f)$ are *larger* than the noise power λ_v.
- Remaining ones are all equal to λ_v.
- The η dominant eigenvectors define the *signal subspace basis* $\mathbf{E}_s$.
- The other eigenvectors define the complementary *noise subspace basis* $\mathbf{E}_n$.

$$\mathbf{R}_{xx} = \mathbf{APA}^H + \lambda_v \mathbf{I}_N \overset{EVD}{=} \begin{bmatrix} \mathbf{E}_s & \mathbf{E}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_s & \mathbf{0} \\ \mathbf{0} & \lambda_v \mathbf{I}_{N-\eta} \end{bmatrix} \begin{bmatrix} \mathbf{E}_s & \mathbf{E}_n \end{bmatrix}^H$$

$$\hat{\mathbf{R}}_{xx} \overset{EVD}{=} \begin{bmatrix} \hat{\mathbf{E}}_s & \hat{\mathbf{E}}_n \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\Lambda}}_s & 0 \\ 0 & \hat{\boldsymbol{\Lambda}}_n \end{bmatrix} \begin{bmatrix} \hat{\mathbf{E}}_s & \hat{\mathbf{E}}_n \end{bmatrix}^H$$

GLS Paige equations for subspace fitting algorithms

- *Direct* (ML-type) fitting (WSF, MODE):

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \left(\left|\mathbf{E}(\mathbf{p})\right|_F, \left|\mathbf{C}(\mathbf{p})\right|_F \right) : \mathbf{A}(\mathbf{p}, f)\mathbf{C}(\mathbf{p}) + \mathbf{R}_v \mathbf{E}(\mathbf{p}) = \hat{\mathbf{E}}_s \mathbf{W}_s$$

$$M \to \infty \Rightarrow \left|\mathbf{E}(\mathbf{p})\right|_F \overset{w.p.1}{\to} 0$$

- *Inverse* (MUSIC type) fitting, only valid for non-coherent sources ($\eta = D$):

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \left(\left|\mathbf{e}(\mathbf{p})\right|_2, \left|\mathbf{c}(\mathbf{p})\right|_2 \right) : \tilde{\mathbf{E}}_s \mathbf{c}(\mathbf{p}) + \mathbf{R}_v \mathbf{e}(\mathbf{p}) = \mathbf{a}(\mathbf{p})$$

$$L \to \infty \Rightarrow \left|\mathbf{e}(\mathbf{p})\right|_2 \overset{w.p.1}{\to} 0$$

SAPIENZA
UNIVERSITÀ DI ROMA

Equation comparison

- MUSIC type estimators estimate one source at time after suppressing the others.
  - The desired steering vector is the *target* of a regression from sample signal subspace.
  - Breakdown when other source residuals greater then signal coming from the analyzed direction.
  - Fitting error is generally not detailed: robustness to array errors but performance limitations.
- ML/WSF estimators model all sources at the same time.
  - The weighted signal subspace or the SCM itself is the target of interpolation by a set of trial steering vectors.
  - Risk of curse of parameters.

Subspace estimators

- Direct and inverse subspace modeling are two faces of the same medal.
- *Direct signal subspace fitting* is an *heteroscedastic regression* of a set of *supposed true* steering vectors onto the weighted sample signal subspace.
- *Inverse subspace fitting,* like *MUSIC* (Multiple Signal Interception and Classification) tries to recover a *single* steering vector from the sample signal subspace:
  - *Steering vector errors are accounted* for in some manner;
  - MUSIC-type beamforming nulls all sources except one at a time.

SAPIENZA
UNIVERSITÀ DI ROMA

Weighted Subspace Fitting

- Assumes *no error in steering vectors*.
- Subspace weighting *optimized for finite sample* errors only.
- Asymptotically efficient under the central model even for coherent scenarios.
- Optimal Wiener type subspace weighting *amplifies calibration errors* and whitens spurious projections of signal eigenvectors onto the noise subspace.
- *All locations must be calibrated*.
- Attempts of inserting effects of mis-calibration resulted in loss of consistency, bias and efficiency.

SAPIENZA
UNIVERSITÀ DI ROMA

Weighted signal subspace SNR threshold

- Wiener WSF weights indicate the existence of a a tight *estimation threshold* at low SNR.
- The threshold depends on the separation of the minimum SCM eigenvalue from noise.
- In academic simulations, ML may exploit signal eigenvector leakage in the noise subspace, but at the cost of high SNR instabilities.

$$w_{WSF}^2 \propto \frac{\left(\hat{\lambda}_k - \hat{\lambda}_v\right)^2}{\hat{\lambda}_k \hat{\lambda}_v}; \quad Cov\left(\hat{\lambda}_k\right) \simeq \frac{\lambda_k^2}{N} \Rightarrow \left|\hat{\lambda}_k - \hat{\lambda}_v\right| \gg \frac{\lambda_v + \lambda_k}{\sqrt{N}}$$

SAPIENZA
UNIVERSITÀ DI ROMA

MODE

- For ULAs a clever rooting technique is possible for WSF, based on the property of Toeplitz Sylvester matrices generated by a polynomial.
- Unit modulus root angles of the polynomial directly furnish azimuth estimates.

$$\mathbf{A}_{ULA} = \begin{bmatrix} 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_D \\ \alpha_1^2 & \cdots & \alpha_D^2 \\ \vdots & \cdots & \vdots \\ \alpha_1^{N-1} & \cdots & \alpha_D^{N-1} \end{bmatrix}; \quad \underset{(N-D)\times N}{\mathbf{B}} = \begin{bmatrix} b_0 & \cdots & b_D & 0 & \cdots & 0 \\ 0 & b_0 & \cdots & b_D & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & b_0 & \cdots & b_D \end{bmatrix}$$

$$\mathbf{BA}_{ULA} = 0 \rightarrow b_0 + b_1\alpha_d + b_2\alpha_d^2 + \ldots + b_D\alpha_d^D = 0; \quad d = 1,2,\ldots,D$$

SAPIENZA
UNIVERSITÀ DI ROMA

MODE GLS equation

- Solved iteratively until (uncertain) convergence.
- The Sylvester matrix asymptotically spans the noise subspace orthogonal to steering vectors.
- Polynomial roots *radii modifications w.r.t. unity* account for little steering vector mis-calibration.
- However MODE remains a rather weak estimator.

$$\mathbf{B}^{(k)}\mathbf{A}(\mathbf{p})\mathbf{C}(\mathbf{p})+\mathbf{B}^{(k-1)}\mathbf{R}_v\mathbf{E}^{(k)}=\mathbf{B}^{(k)}\hat{\mathbf{E}}_s\mathbf{W}_s$$

$$\mathbf{B}^{(k-1)}\mathbf{R}_v\mathbf{E}^{(k)}=\mathbf{B}^{(k)}\mathbf{E}_s\mathbf{W}_s$$

$$\mathbf{B}^{(k)}\to\hat{\mathbf{B}}$$

SAPIENZA
UNIVERSITÀ DI ROMA

Comments on MODE

- MODE exploits the fact that ULA subarrays share the same steering vectors with a power-type phase shift.
- Implicit averaging of subarray SCMs (*spatial smoothing*) to a certain degree restores the maximum signal subspace rank D even in the case of full coherence.
- This method was used to enhance MUSIC, but at the expense of the overall array effective aperture.
- MODE fully and correctly exploits spatial smoothing for asymptotical efficiency and improved WSF iteration complexity and safety.
- However forward-backward conjugate relationships of ULA steering vectors can still provide statistical advantages by destroying source coherence.

MUSIC as beamforming

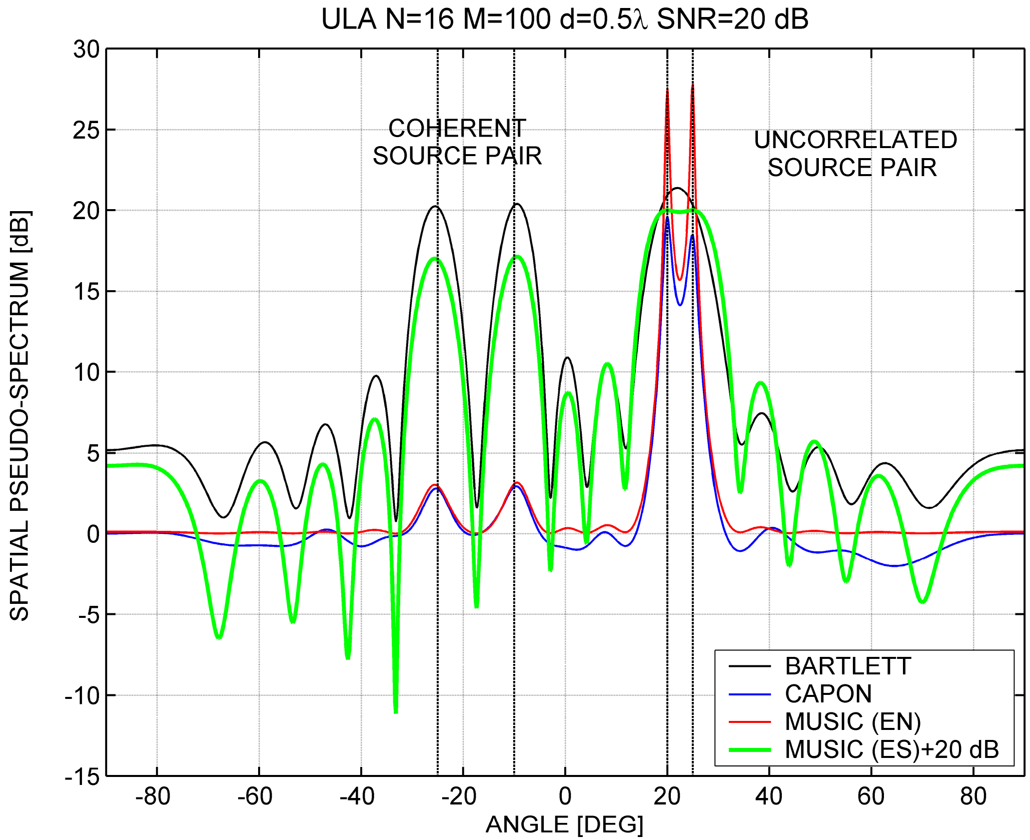
- Equivalent MUSIC functionals (*pseudo-spectra*) for white noise establish links with Capon and Bartlett estimators:
  - Bartlett estimator for the uniformly weighted signal subspace;
  - MVDR estimator in the limit of infinite source SNR.

$$\hat{\mathbf{p}} = \arg\max_{\mathbf{p}} \left[\frac{\mathbf{a}(\mathbf{p}, f)^H \mathbf{E}_s \mathbf{E}_s^H \mathbf{a}(\mathbf{p}, f)}{\left|\mathbf{a}(\mathbf{p}, f)\right|^2} \right]$$

$$= \arg\max_{\mathbf{p}} \left[\frac{\left|\mathbf{a}(\mathbf{p}, f)\right|^2}{\mathbf{a}(\mathbf{p}, f)^H \mathbf{E}_n \mathbf{E}_n^H \mathbf{a}(\mathbf{p}, f)} \right]$$

Spatial spectra comparison

- Cost functions of Bartlett, Capon and MUSIC (signal and noise subspace versions) are compared.



SAPIENZA
UNIVERSITÀ DI ROMA

MUSIC properties

- Near asymptotically efficient for uncorrelated sources.
- Optimal w.r.t. i.i.d. Gaussian errors in normalized steering vectors.
- Calibration needed only within the angular sector of interest.
- Small angle *finite sample bias*, but not negligible.
- Spurious sources and rather high estimation threshold at low SNR.
- *Classical source selection criteria* cannot deal with strongly coherent sources, but proper MUSIC spectra *do show peaks near coherent sources*, but with bias and Rayleigh-limited resolution as a Bartlett beamformer!
- Suggested workarounds for coherent sources (Toeplitz constraints, spatial smoothing) impair performance in the uncorrelated cases: SCM eigenstructure distortion, bias, reduced aperture, non-white background noise).

MUSIC SNR threshold

- MUSIC must invert the steering vector linearly combining a *noisy* signal subspace.
- Signal subspace perturbations within the signal subspace itself ($\mathbf{G}_s$) do modify the problem.
- Signal subspace perturbations within the noise subspace ($\mathbf{G}_v$) regularize the system at low SNR, but at the expense of DOA bias and additional error variance.
- When error is excessive, *MUSIC spectrum fails* to resolve spatially close sources (nulls merge):

$$\mathbf{E}_s\mathbf{w} = \mathbf{a}(\theta); \quad \hat{\mathbf{E}}_s\hat{\mathbf{w}} = \mathbf{a}(\theta) - \mathbf{v}(\theta)$$

$$(\mathbf{E}_s\mathbf{G}_s + \mathbf{E}_v\mathbf{G}_v)(\mathbf{w} + \delta\mathbf{w}) = \mathbf{a}(\theta) - \hat{\mathbf{v}}(\theta)$$

ROOT MUSIC

- For arrays with an harmonic response (ULA or circular harmonic azimutal fittings) the steering vector is Vandermonde.
- A polynomial in Z domain can be formed, simulating an ideal Toeplitz noise subspace projector.
- Angles of roots close to the unit circle furnish DOA estimates (beware of quadrantal polinomial symmetry).
- Highly reduced SNR threshold problems.
- Fast computation.
- Effective display of the low SNR/sample size MUSIC inconsistency.
- Mestre and al. try to approach ROOT MUSIC performance with modified spectral forms.

ROOT MUSIC (2)

$$\frac{\mathbf{a}(\theta)}{\left|\mathbf{a}(\theta)\right|} = \mathbf{A}\begin{bmatrix} 1 \\ e^{jf(\theta)} \\ e^{j2f(\theta)} \\ \vdots \end{bmatrix} = \mathbf{A}\begin{bmatrix} 1 \\ z \\ z^2 \\ \vdots \end{bmatrix} = \mathbf{A}\mathbf{e}(z); \quad z = e^{jf(\theta)}; \quad z^* = e^{-jf(\theta)} = z^{-1}$$

$$P_{MUS}(z) = \mathbf{e}^T\left(z^{-1}\right)\mathbf{A}\mathbf{E}_v\mathbf{E}_v^H\mathbf{A}^H\mathbf{e}(z) = \mathbf{e}^T\left(z^{-1}\right)\mathbf{P}_v\mathbf{e}(z) = \frac{N(z)}{z^M}$$

Toeplitz
Matrix equivalent

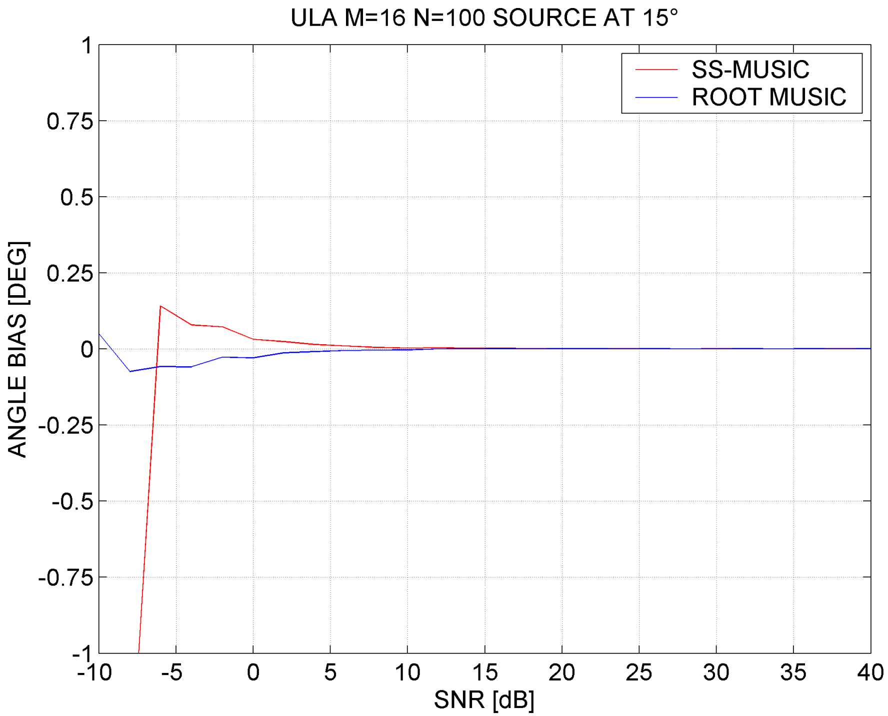
SAPIENZA
UNIVERSITÀ DI ROMA

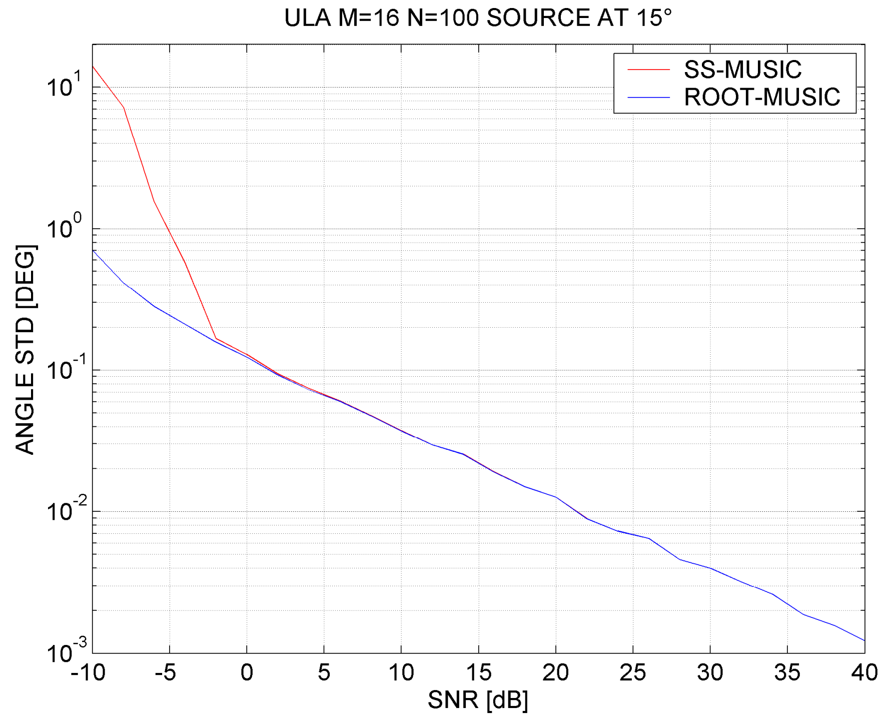
ROOT MUSIC (3)

- ROOT MUSIC creates a Toeplitz modification of the noise projector.
- The projector exactly intersects a slightly modified (i.e., exponentially weighted) steering vector:
  - Superresolution possible by root and residue analysis.
- Spectral MUSIC does not exactly intersect the array manifold for finite sample and calibration errors:
  - no viable technique for distinguishing and separately converge to relatively close sources.

ROOT-MUSIC PERFORMANCE

- 16 sensors ULA, 100 samples, two uncorrelated equi-powered sources at 10° and 15° w.r.t. broadside.
- SS-MUSIC is an advanced variant.





MUSIC and coherent sources

- The *asymptotical consistency of MUSIC is tied to the perfect asymptotical nulling of all sources* operated by noise subspace eigenvectors.
- Nulls can be placed well within the Rayleigh resolution limit *(super-resolution)*.
- *Coherent sources cannot be individually nulled*, but *still appear* in the beampattern form of the MUSIC pseudo-spectrum: a literature tale has been (in part) disproved.
- However peaks produced by coherent arrivals appear separate only because of the *intrinsic array beampattern* (good for ULA!) and they are anyway *slightly biased,* due to the partial superposition of other sources contributes.
- So, while coherent sources can be detected from a MUSIC pseudo-spectrum, *only partially uncorrelated sources can be consistently detected in super-resolution.*
- Coherent sources can still be detected by MUSIC spectra better than using Bartlett beamforming, because of the implicit *source power equalization* in the orthogonal signal subspace which reduces sidelobe interference.

Subspace DOA algorithm dataflow summary

- Collect and a set of independent snapshots
- Perform noise pre-whitening.
- Estimate SCM and its eigendecomposition from snapshots or compute SVD of the snapshot matrix.
- Estimate the numerical rank of the SCM (i.e., number of uncorrelated signals) by Information Theoretic criteria.
- Only for ML/WSF, get initial, rough DOA estimates by beamforming or other suboptimal techniques.
- Compute or locally refine DOA estimates.
- Optionally, estimate incoming signals by LS/SVD fitting or constrained beamforming on original snapshots.
- Validate signals.

Comparison of subspace techniques

WSF type

- Asymptotically efficient for finite sample errors.
- Handles coherent sources.
- Critical calibration.
- Multi-dimensional search.
- Amplify modeling errors.
- Gross errors expected at *high* SNR.
- Tolerant to *under-estimation* of D.

MUSIC type

- Near asymptotically efficient only for uncorrelated sources.
- Optimal for i.i.d. steering vector perturbations.
- Difficult handling of coherent sources.
- 1-D search.
- Higher *low SNR* threshold.
- Tolerant to *over-estimation* of D.

SAPIENZA
UNIVERSITÀ DI ROMA

(Almost-)deterministic model

- Two simultaneous and temporally aligned snapshots $\mathbf{x}(k)$ e $\mathbf{y}(k)$ are available from an array and another *auxiliary system* (AS), whose outputs are considered as *reference signals* (e.g., *stored replicas*);
- Noise and interference components of the array and AS must be independent;
- It is *not necessary* to explicitly know noise SCM and the details of the AS.
- The AS may even be a *non-linear* system!
- Synchronization time shift must be *lower than the signal correlation time*.

SAPIENZA
UNIVERSITÀ DI ROMA

(Almost-)deterministic model

- Includes as a limit case the *purely deterministic model* with known training signals (active radar).

$$\mathbf{z}(k) = \begin{bmatrix} \mathbf{x}(k) \\ \mathbf{y}(k) \end{bmatrix} = \begin{bmatrix} \mathbf{A}(f_0, \Theta)\mathbf{s}(k) + \mathbf{n}_x(k) \\ \mathbf{B}\mathbf{u}(k) + \mathbf{n}_y(k) \end{bmatrix}$$

$$E[\mathbf{n}_x(k)\mathbf{n}_y(k)^H] = E[\mathbf{n}_x(k)\mathbf{n}_y(k)^T] = \mathbf{0}$$

$$E[\mathbf{s}(k)\mathbf{u}(l)^T] = \mathbf{0}; E[\mathbf{s}(k)\mathbf{u}(k)^H] = \mathbf{P}_{su}$$

$$\mathbf{R}_{zz} = E\left[\mathbf{z}(k)\mathbf{z}(k)^H\right] = \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xy} \\ \mathbf{R}_{xy}^H & \mathbf{R}_{yy} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_x\mathbf{C}_x^H & \mathbf{R}_{xy} \\ \mathbf{R}_{xy} & \mathbf{C}_y\mathbf{C}_y^H \end{bmatrix}$$

Generalized cross-correlation (GCC)

- The GCC is the sufficient statistic for the quasi-deterministic ML estimate. Array and AS signals are *individually spatially whitened*.

$$GCC(\mathbf{x},\mathbf{y}) = \mathbf{C}_{xy} \triangleq \mathbf{C}_x^{-1}\mathbf{R}_{xy}\mathbf{C}_y^{-H} = \left[\mathbf{C}_x^{-1}\mathbf{A}(\mathbf{p},f)\right]\mathbf{P}_{su}\left(\mathbf{B}^H\mathbf{C}_y^{-H}\right)$$

$$\hat{\mathbf{R}}_{xx} = \hat{\mathbf{C}}_x\hat{\mathbf{C}}_x^H; \quad \hat{\mathbf{R}}_{yy} = \hat{\mathbf{C}}_y\hat{\mathbf{C}}_y^H; \quad \hat{\mathbf{R}}_{xy} = \frac{1}{M}\sum_{k=1}^{M}\mathbf{x}(k)\mathbf{y}(k)^H$$

$$\hat{\mathbf{C}}_{xy} = \hat{\mathbf{C}}_x^{-1}\hat{\mathbf{R}}_{xy}\hat{\mathbf{C}}_y^{-H}$$

SAPIENZA
UNIVERSITÀ DI ROMA

Parameter concentration

- Estimated location parameter are asymptotically ($M \to \infty$) neraly decoupled by those of the signal covariances;
- The Gaussian ML estimates of signal-related parameters are extracted as a function of location parameters and back-substituted (imperfect *concentration*);
- The orthogonal signal subspace is obtained by the left singular vectors of the GCC, having singular values close to one.
- Optimal subspace weighting is given by singular values itself.

GCC location estimates

- WSF, MUSIC and MODE are applicable, taking into account steering vector prewhitening.

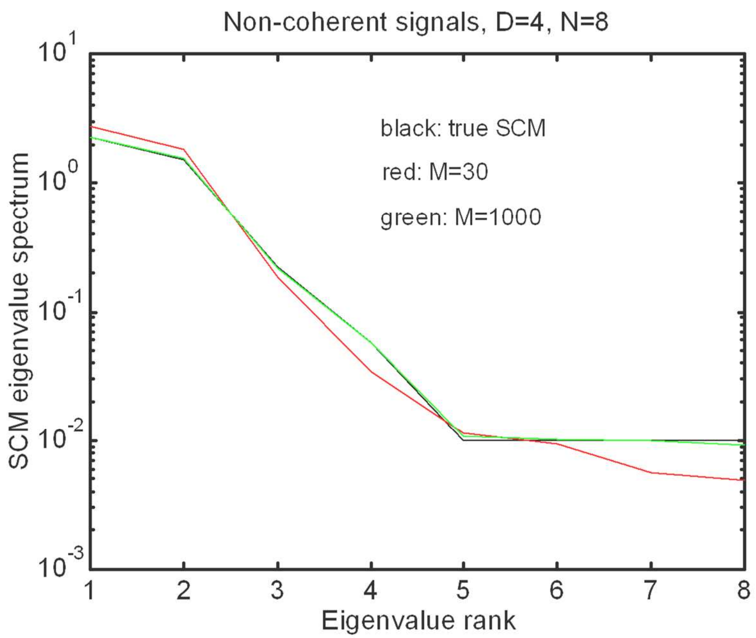
$$\hat{\mathbf{C}}_{xy} = \begin{bmatrix} \mathbf{U}_s & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_n \cong \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_s & \mathbf{V}_n \end{bmatrix}^H$$

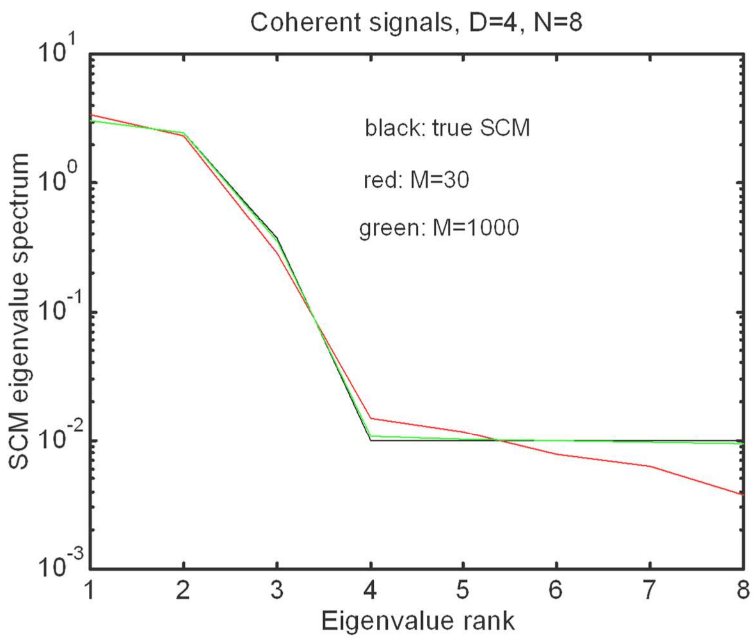
$$\mathbf{A}(\mathbf{p}, f)\mathbf{C}(\mathbf{p}) + \hat{\mathbf{C}}_x \mathbf{E}(\mathbf{p}) = \hat{\mathbf{C}}_x \mathbf{U}_s \boldsymbol{\Sigma}_s$$

$$\hat{\mathbf{C}}_x \mathbf{U}_s \mathbf{w}(\mathbf{p}) + \hat{\mathbf{C}}_x \mathbf{e}(\mathbf{p}) = \mathbf{a}(\mathbf{p}, f)$$

Sample SCM eigenspectra

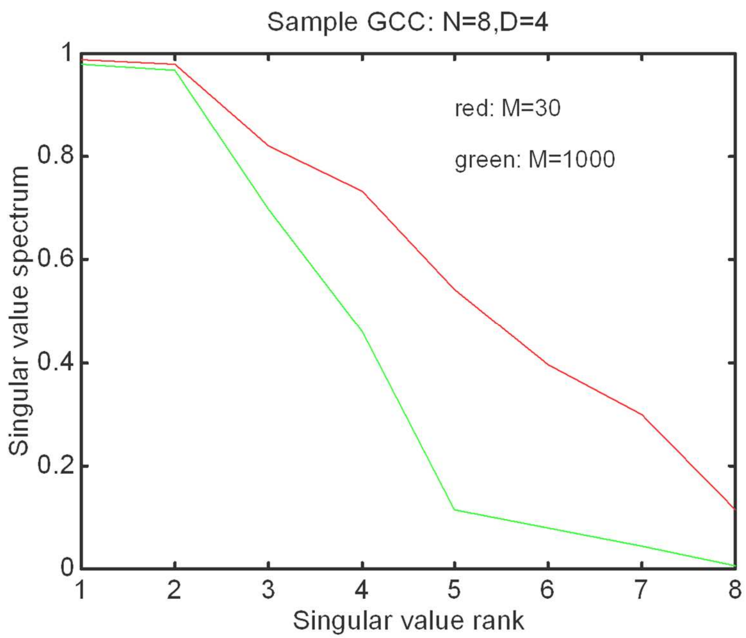
- If $\eta<D$, signals are coherent. The number of identifiable coherent sources diminishes.
- Sample noise eigenvalues are spreaded around the true value (asymptotical Chi-squared distribution).





Sample GCC singular values

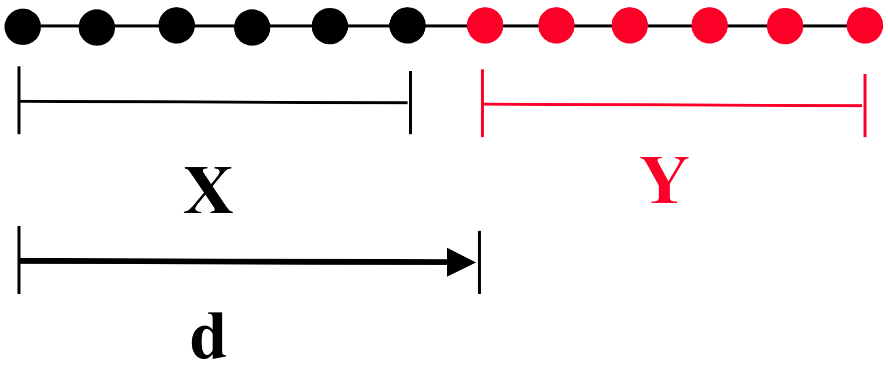
$$\tilde{\mathbf{C}}_{xy} \overset{SVD}{=} \begin{bmatrix} \tilde{\mathbf{E}}_s & \tilde{\mathbf{E}}_N \end{bmatrix} \begin{bmatrix} \tilde{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & O(1/\sqrt{M}) \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_s & \tilde{\mathbf{V}}_N \end{bmatrix}^H$$



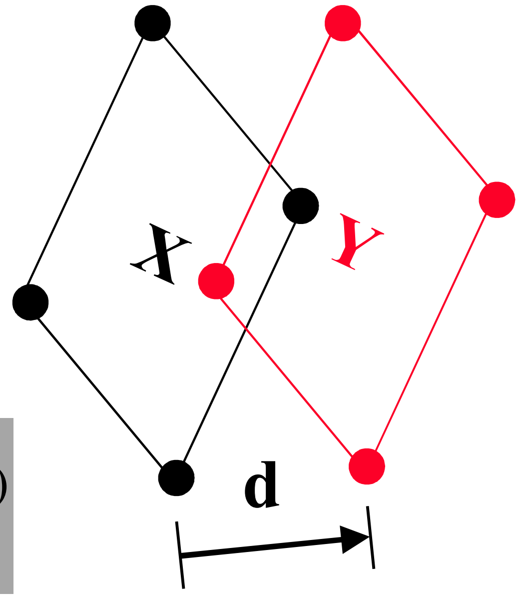
- GCC sample *singular values* are close to one for sources, while the others decay to zero while increasing M.

Rotational invariance

- Some algorithms (ESPRIT, MODE) exploit the *rotational invariance* (phase factor) existing among steering vectors of two or more subarrays characterized by a translational invariance.



$$\mathbf{a}_Y(\theta,f)=\mathbf{a}_X(\theta,f)\,e^{-j2\pi\frac{|\mathbf{d}|}{\lambda}\sin(\theta)}$$



SAPIENZA
UNIVERSITÀ DI ROMA

ESPRIT (Roy e Kailath, 1985)

- Exploits GCC and rotational invariance concepts.
- Makes it easier the selection of the sources of interest from subarray outputs.
- Irremediably sensitive to mutual coupling between close subarrays.

$$\hat{\mathbf{E}}_s \triangleq \begin{bmatrix} \hat{\mathbf{E}}_X \\ \hat{\mathbf{E}}_Y \end{bmatrix} \Rightarrow [\phi, \mathbf{c}] = \arg\min_{|\mathbf{e}|, \text{s.t.}\{|\mathbf{c}|=1\}} \left\{ \hat{\mathbf{E}}_Y^H \mathbf{c} = \hat{\mathbf{E}}_X^H \mathbf{c}\phi + \mathbf{Re} \right\}$$

$$g(\theta) = e^{-j\varphi(\theta)} \Rightarrow \hat{\theta}_{ESPRIT} = -\varphi^{-1}\left[\arg(\phi)\right]$$

Source number estimatiom

- From SCM and GCC only the *number of uncorrelated impinging signals* can be estimated from the eigenvalue spectra.
- Coherent source resolution can be only asserted by the value of fitting residuals compared to the noise floor estimate (search for one more source…).
- The data projection on the noise subspace asymptotically approaches a (M by N-D) *random matrix* with i.i.d. Gaussian entries for which several eigenvalue bounds exist.
- This is a relevant consequence of the *asymptotical independence of eigenvector and eigenvalues* estimates.

Likelihood ratio for signal detection in the SCM

- For η signals, N sensors and M samples:
  - $(N\text{-}\eta)$th power of the *ratio between the geometrical and arithmetical means* of supposed noise eigenvalues (equality test);
  - upper bounded by one.

$$\hat{\lambda}_v(\eta) = \frac{1}{N-\eta}\sum_{n=\eta+1}^{N}\hat{\lambda}_n$$

$$LR(\eta) = \frac{\prod_{n=\eta+1}^{N}\hat{\lambda}_n}{\left[\hat{\lambda}_v(\eta)\right]^{N-\eta}}$$

SAPIENZA
UNIVERSITÀ DI ROMA

Theoretical Information Criteria

- Minimize a cost function given by the scaled log-Likelihood Ratio LR plus a penalty term dependent on the sample size and the number of free SCM parameters *for all hypotheses* between zero and N-1 signals present.
- MDL is *asymptotically* consistent, but tends to *underestimate* the number of signals in finite sample, while AIC always tends to *overestimate* it.

$$AIC(\eta) = -M \ln\left[LR(\eta)\right] + 2\eta(N-\eta)$$

$$MDL(\eta) = -M \ln\left[LR(\eta)\right] + \frac{\ln(M)}{2} 2\eta(N-\eta)$$

Random matrix approach

- Tries to estimate a bound for the *largest sample noise eigenvalue* for each hypothesis.
- Sets a *lower threshold for signal eigenvalues*.
- Results available under several assumptions.
- For Gaussian noise, sample eigenvalues can be approximated in large sample by ranked Chi-Square or even Gaussian independent random variables.

$$E\left[\hat{\lambda}_n\right]=\lambda_v;\quad \mathrm{Cov}\left[\hat{\lambda}_n,\hat{\lambda}_m\right]\simeq\delta_{mn}\frac{\lambda_v^2}{M};\quad \eta+1\le n,m\le N$$

Source copy and validation

- A heteroscedastic GLS fitting is finally performed to extract source signals of interest.
- Specially designed LCMV or even robust beamformers may advantageously replace the LS fitting in many cases.
- If the last condition on LS error is not fulfilled in a statistical test, there are other (maybe coherent) sources to find!

$$\mathbf{R}_{vv} = \mathbf{C}_v \mathbf{C}_v^H; \quad \mathbf{A}(\hat{\mathbf{p}}, f)\hat{\mathbf{s}}(k) + \mathbf{C}_v \mathbf{e}(k) = \mathbf{x}(k)$$

$$k = 1, 2, \ldots, M; \quad \frac{1}{M-D} \sum_{k=1}^{M} \left| \mathbf{e}(k) \right|^2 \cong \hat{\lambda}_v$$

Statistical test breakdown

- *Temporal aliasing* and *spectral leakage* from receiver filters create *non-white*, *non-stationary* noise fields.
- Finite receiver bandwidth creates a multi-rank source, i.e., *ghost sources*.
- Any existing academic statistical model will break down! Simple, regularized tests may perform acceptably and even better than sophisticated ones.
- Wrong number of sources implies some *loss of location accuracy of all sources*, starting from the weakest ones, or even a catastrophe at high SNR.
- Consistency may be lost due to the use of an automatic source detection criteria: in some trials bias or over-fitting may spoil the overall estimates.

Remarks

- All these detection techniques are based on the assumption that the signal covariance rank is *finite* and equals the number of arrivals (rays) rather that the number of uncorrelated source signal, as it is.
- ML techniques are robust (softly degrade their consistency and may lose weak sources) in the presence of covariance rank *underestimation*, but too weak detected source may catastrophically impact estimation of stronger sources at high SNR (the Fisher Information Matrix becomes ill-conditioned).
- MUSIC is robust to rank *overestimation* instead, but finds non-existing weak arrivals to be pruned later.

Self calibration

- Theoretically feasible only for lightly loaded arrays (i.e., $D<<N$) with sufficient degrees of freedom in the SCM, in addition to location and covariance parameters.
- Requires a *parametric, compact model* for steering vector errors.
- ML or subspace functional optimized for all parameters.
- In most cases, some parametrizations lead to singular Fisher Information Matrix (hence some parameters are not uniquely identifiable).
- Several scenarios are anyway required to fully identify a valid re-calibration matrix.

Compressed sensing for arrays

- Any near-optimal technique examined here realizes a near-optimal *compressed sensing scheme* (minimal number of sources and their parametrization).
- Especially for coherent scenarios and ML/WSF initialization one may attempt to estimate the number of sources by a semi-parametric, *general purpose* compressed sensing algorithm.
- A codebook of tentative steering vectors is fitted to the signal subspace.
- A solution involving the minimal number of parameters is sought, generally obtained by regularized L1 norm fitting.

Basic compressed sensing solutions

- The codebook subspace fitting has infinite solutions.
- The minimum L2 norm solution is very smooth with spatially distributed amplitudes.
- Solutions however exhibit *beamforming type* peaks in the vicinity of sources using appropriate spectral measures.

Codebook

$$\mathbf{A} = \left[\mathbf{a}(\theta_1) \quad \mathbf{a}(\theta_2) \quad \cdots \quad \mathbf{a}(\theta_n) \right]$$

Solution set

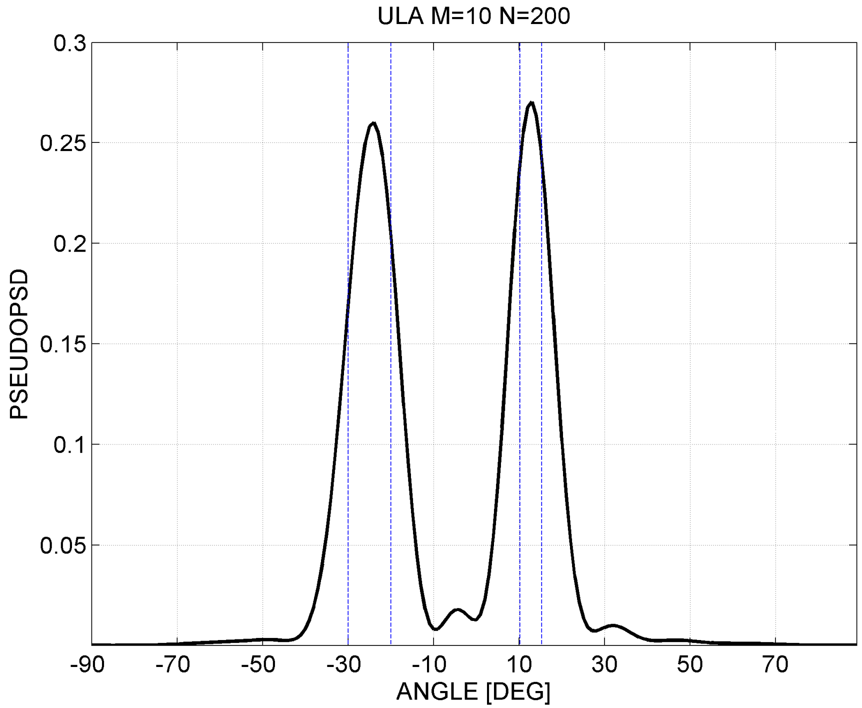
$$\mathbf{AC} = \mathbf{E}_s \mathbf{W}; \quad \mathbf{C}_{MN} = \mathbf{A}^H \left(\mathbf{AA}^H \right)^{-1} \mathbf{E}_s \mathbf{W}$$

Spectral measure

$$\mathbf{C} = \mathbf{C}_{MN} + \mathbf{A}_{\perp} \mathbf{C}_{\perp}; \quad P_{L2} = \mathrm{diag}\left(\mathbf{CC}^H \right)$$

Minimun norm L2 solution

- 10 sensors ULA, 4 sources (10.2°,15.3°,-20°,-30°), the first two are coherent, SNR about 20 dB.
- Zero fitting error.
- No compression, no super-resolution!

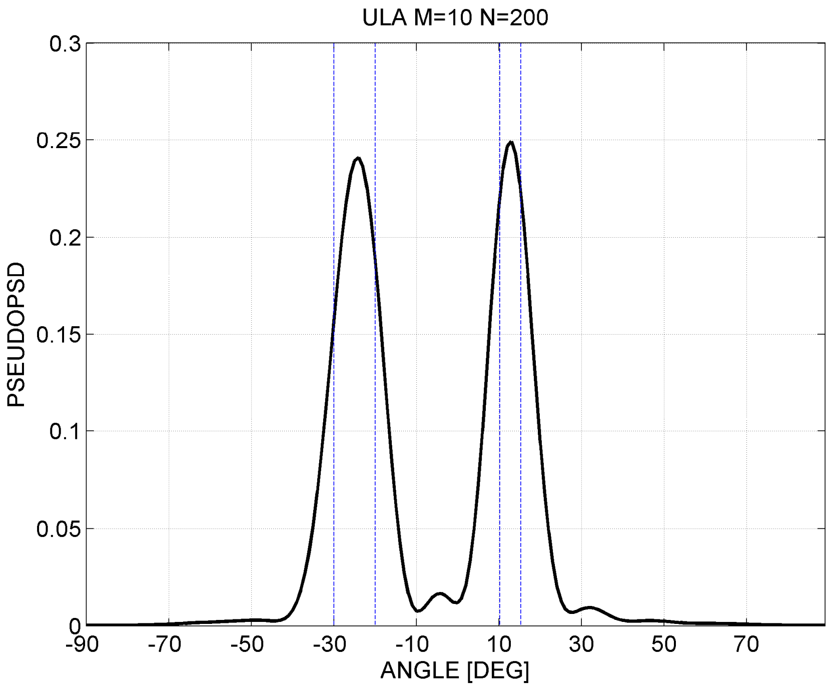


SAPIENZA
UNIVERSITÀ DI ROMA

Tikhonov regularization

$$P=\sum_{k=0}^{\eta}\left[\left|\mathbf{AC}(:,k)-\mathbf{E}_s\mathbf{W}(:,k)\right|_2^2+\lambda\left|\mathbf{C}(:,k)\right|_2^2\right]$$

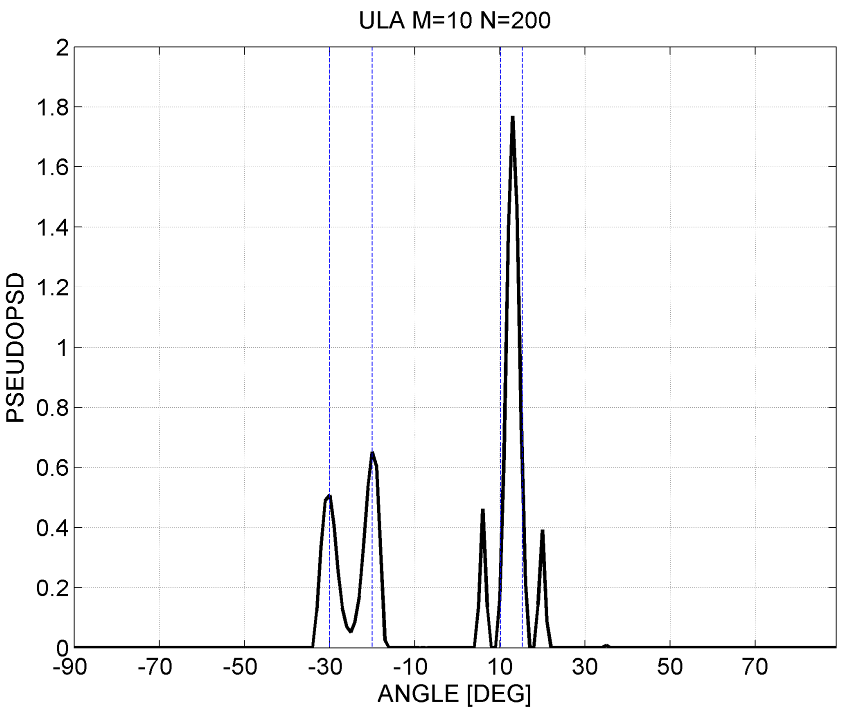
- A penalty is added to the cost functional to suppress small solution coefficients.
- Tikhonov regolarization makes the error non-zero, but the solution is still smooth and full.



L1 penalized solution

$$P=\sum_{k=0}^{\eta}\left[\left|\mathbf{AC}(:,k)-\mathbf{E}_s\mathbf{W}(:,k)\right|_2^2+\lambda\left|\mathbf{C}(:,k)\right|_1\right]$$

- Very simple version
- Five sources detected
- *Sparse* solution with many zeros
- Heavy computations



Other CS type solutions

- More sophisticated L1 penalized functionals.
- Orthogonal Matching Pursuit (Cadzow 1990):
  - sequentially adds tentative sources minimizing the WSF LS fitting error from a steering vector codebooks;
  - first used in WSF initialization;
  - re-proposed by many computer science authors in a simplified, less performing version!
- Direct pseudo-spectrum optimization.

General purpose CS drawbacks

- *No consistency claims is possible* if all the source DOAs are not in the codebook!
  - source splitting.
  - source excision.
  - ghost sources.
- *Unclear detection and threshold* strategies.
- *Uncertain convergence*: the important Restricted Isometry Property (RIP) is *never* satisfied for increasingly close DOA angles!

$$\forall \mathbf{A}_s = \mathbf{A}(:,\mathbf{s}) \quad \exists \delta_s > 0$$

$$(1-\delta_s)|\mathbf{y}|_2^2 \le |\mathbf{A}_s\mathbf{y}|_2^2 \le (1+\delta_s)|\mathbf{y}|_2^2$$

SAPIENZA
UNIVERSITÀ DI ROMA

Conclusions about CS

- General purpose compressed sensing is only interesting for data exploration and optimal algorithm initialization.
- CS is a Mathematical, non-physical viewpoint.
- Uncertain interpretation of results.
- High system complexity.
- No theoretical guarantees!

SAPIENZA
UNIVERSITÀ DI ROMA

PART V

PERFORMANCE AND ROBUSTNESS OF ARRAY PROCESSING

Error sources in array processing

- **Signal-related errors**
  - Finite sample errors
  - Heavy-tailed distributions
  - Outliers
  - Non-stationarity
  - Spectral leakage
  - Temporal aliasing
  - Non-linearity
- **Model-based errors**
  - Coherent sources
  - Calibration errors
  - Multipath & multimode propagation
  - Reverberation
  - Ambiguity
  - Environmental changes

First-order perturbation analysis

- Statistical analysis of array processing algorithms is not easy or productive with traditional techniques, because they are strongly non linear.
- The so called (first-order) *perturbative method* has a widespread relevance for this task.
- It works calculating the *sensitivity w.r.t. parameters of the linearized model*, a kind of Newton derivative.
- Valid for moderate to high SNR and sample size and *sufficiently small perturbations*.

Perturbative techniques

- These techniques are strongly tied to the CRB concept and can analyze the statistical impact of various parameters and give a *clear geometrical significance* to estimation errors.
- Asymptotically, the linearized estimator around *true* parameters assumes the form of the *interpolation of a random vector on a fixed basis*, where the unknown are the parameter estimation errors.
- If the random vector is asymptotically Gaussian distributed, as often it is, the local ML estimator is the local linear LS fitting.
- Anyway, the local BLUE estimator is still the LS one!

SCM eigenvector perturbation (Golub 1989)

- The EVD of a Hermitian matrix $\mathbf{R}$ *perturbed* by another Hermitian matrix $\varepsilon\mathbf{R}^{(1)}$ is expanded in *Taylor series* w.r.t. the dummy variable ε (scale factor of the perturbation itself).
- A *deterministic relationship (transfer function)* is obtained between terms of the same order in ε :

$$\left(\mathbf{R}+\varepsilon\mathbf{R}^{(1)}\right)\left(\mathbf{V}+\varepsilon\mathbf{V}^{(1)}+\ldots\right)=\left(\mathbf{V}+\varepsilon\mathbf{V}^{(1)}+\ldots\right)\left(\mathbf{\Lambda}+\varepsilon\mathbf{\Lambda}^{(1)}+\ldots\right)$$

$$\varepsilon\propto\frac{\left|\tilde{\mathbf{R}}-\mathbf{R}\right|_2}{\left|\mathbf{R}\right|_2};\mathbf{V}^{(1)}\triangleq\mathbf{V}\mathbf{G}\Rightarrow\mathbf{G}=-\mathbf{G}^H;\mathrm{diag}\left(\mathbf{G}\right)=\mathbf{0};g_{ij}=\frac{\mathbf{v}_i^H\mathbf{R}^{(1)}\mathbf{v}_j}{\lambda_j-\lambda_i}$$

Eigenvalue perturbation rewind

- Eigenvalue perturbation is *asymptotically independent* from that of eigenvectors.
- Sample eigenvalue distribution is asymptotically Chi-Squared with $2M$ degrees of freedom in the Gaussian case, hence converging in large sample to a Gaussian distribution.
- Eigenvalues estimates are asymptotically *mutually indipendent*.
- IMHO, a *more detailed stochastic model is not desirable* for most applications, because of the *dominance of systematic and random array model perturbations on the SCM*.

$$E\left[\hat{\lambda}_n\right]=\lambda_v; \quad \mathrm{Cov}\left[\hat{\lambda}_n,\hat{\lambda}_m\right]\simeq\delta_{mn}\frac{\lambda_v^2}{M}; \quad \eta+1\le n,m\le N$$

Stochastic finite sample perturbation

- The random cosine matrix $\mathbf{G}$ can be statistically characterized w.r.t. a dummy variable vanishing with M.
- Perturbation strength is reduced by increasing the SNR;
- Since *a good SCM estimator is asymptotically Gaussian* and unbiased for large M, EVD perturbation results are still valid for a wide class of signal distributions (*asymptotical robustness*) and ML or robust SCM estimators.
- Perturbations of signal eigenvectors into noise subspace are mutually uncorrelated!

$$\varepsilon \triangleq M^{-1/2} \Rightarrow E[g_{ij}] = 0; E[g_{ij} g_{kl}] = 0$$

$$E[g_{ij} g_{kl}^{*}] = \delta_{ik} \delta_{jl} \frac{\lambda_i \lambda_j}{\left(\lambda_i - \lambda_j\right)^2}$$

Asymptotical analysis of WSF (Viberg 1991)

A deterministic relationship between parameter estimation errors and finite sample signal subspace perturbation is found by equating first order tems in ε:

$$\varepsilon \triangleq M^{-1/2}; \mathbf{A}(\mathbf{p})\mathbf{C}(\mathbf{p}) = \mathbf{E}_s\mathbf{W} \Rightarrow \mathbf{C}(\mathbf{p}) = \mathbf{A}(\mathbf{p})^{\dagger}\mathbf{E}_s\mathbf{W}$$

$$\mathbf{p}(\varepsilon) = \mathbf{p} + \varepsilon\mathbf{p}^{(1)} + \ldots; \mathbf{W}(\varepsilon) = \mathbf{W} + \varepsilon\mathbf{W}^{(1)}$$

$$\mathbf{C}(\varepsilon) = \mathbf{C} + \varepsilon\mathbf{C}^{(1)} + \ldots; \mathbf{A}(\mathbf{p}) + \varepsilon(\mathbf{p})\dot{\mathbf{A}}(\mathbf{p})\boldsymbol{\Delta}\left(\mathbf{p}^{(1)}\right) + \ldots$$

$$\mathbf{A}(\mathbf{p})\mathbf{C}^{(1)} + \dot{\mathbf{A}}(\mathbf{p})\boldsymbol{\Delta}\left(\mathbf{p}^{(1)}\right)\begin{bmatrix}\mathbf{C} \\ \cdots \\ \mathbf{C}\end{bmatrix} \cong \begin{bmatrix}\mathbf{E}_s & \mathbf{E}_n\end{bmatrix}\begin{bmatrix}\mathbf{G}_s \\ \mathbf{G}_n\end{bmatrix}\mathbf{W} + \mathbf{E}_s\mathbf{W}^{(1)}$$

The local linear system

- *Location errors are separable* from spectral errors in $\mathbf{C}^{(1)}$ by projecting both term of first order expansion onto the orthogonal complement of $\mathbf{A}(\mathbf{p})$, defined by the orthonormal basis $\mathbf{U}$:

$$\mathbf{U}^H\mathbf{U} = \mathbf{I}_{M-D};\mathbf{U}^H\mathbf{A}(\mathbf{p}) = \mathbf{0} \Rightarrow$$

$$\Rightarrow \mathbf{U}^H\dot{\mathbf{A}}(\mathbf{p})\boldsymbol{\Delta}\left(\mathbf{p}^{(1)}\right)\begin{bmatrix}\mathbf{C}\\ \cdots\\ \mathbf{C}\end{bmatrix} \cong \mathbf{U}^H\mathbf{E}_n\mathbf{G}_n\mathbf{W} \Rightarrow$$

$$\Rightarrow \mathbf{F}\mathbf{p}^{(1)} \cong \mathbf{r};\left\{\mathbf{F},\mathbf{r} \in \mathbb{R}\right\}$$

Asymptotical considerations

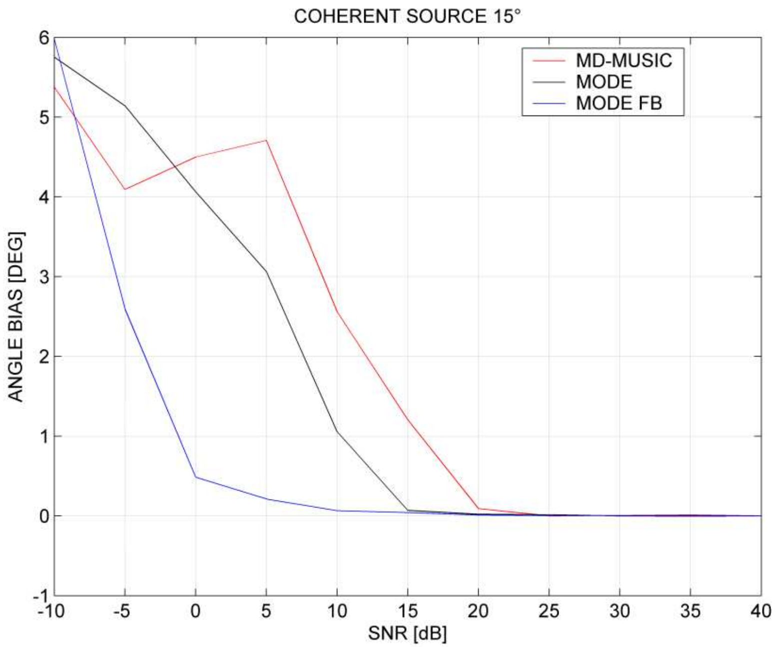
- The design matrix $\mathbf{F}$ is *fixed* and characteristic of the array and environment.
- The target vector $\mathbf{r}$ is instead *random* and depend on the elements of $\mathbf{G}$;
- If $\mathbf{r}$ is Gaussian and the array is perfectly calibrated, the MLE (and BLUE) is locally given by a local LS interpolation (i.e., by the classical WSF).

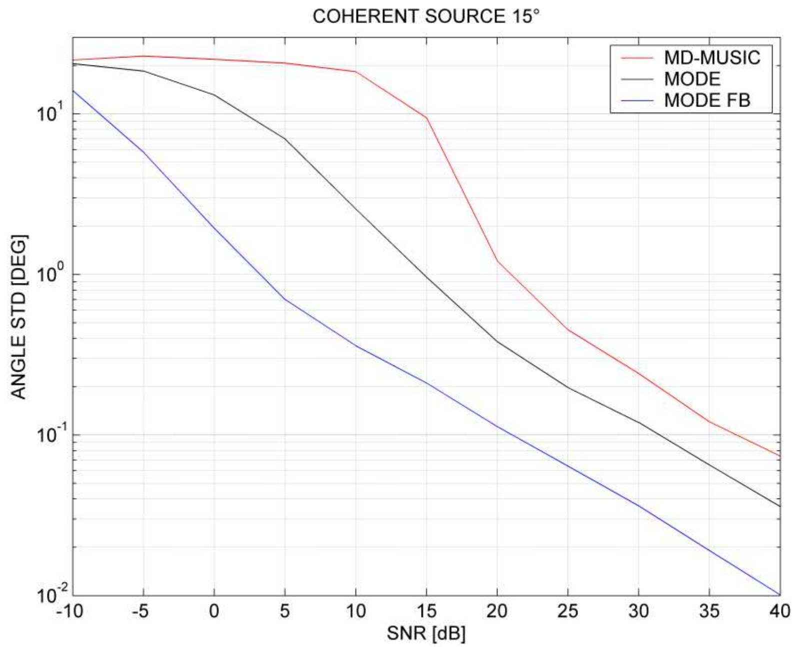
$$\mathbf{p}^{(1)} = \mathbf{F}^{\dagger}\mathbf{r} \Rightarrow E[\mathbf{p}^{(1)}] = \mathbf{0}; \quad Cov[\mathbf{p}^{(1)}] = \mathbf{F}^{\dagger} Cov[\mathbf{r}] \left(\mathbf{F}^{\dagger}\right)^{T}$$

$$\mathbf{W}_{WSF} \propto \mathbf{\Lambda}_s^{-1/2}\left(\mathbf{\Lambda}_s - \lambda_v \mathbf{I}_\eta\right) \Rightarrow Cov[\mathbf{p}^{(1)}] = \frac{\lambda_v}{2}\left(\mathbf{F}^{T}\mathbf{F}\right)^{-1}$$

MODE vs. MD-MUSIC

- Two coherent sources at 0° and 15° in AWGN.
- Eight sensor ULA, *d=0.5*.



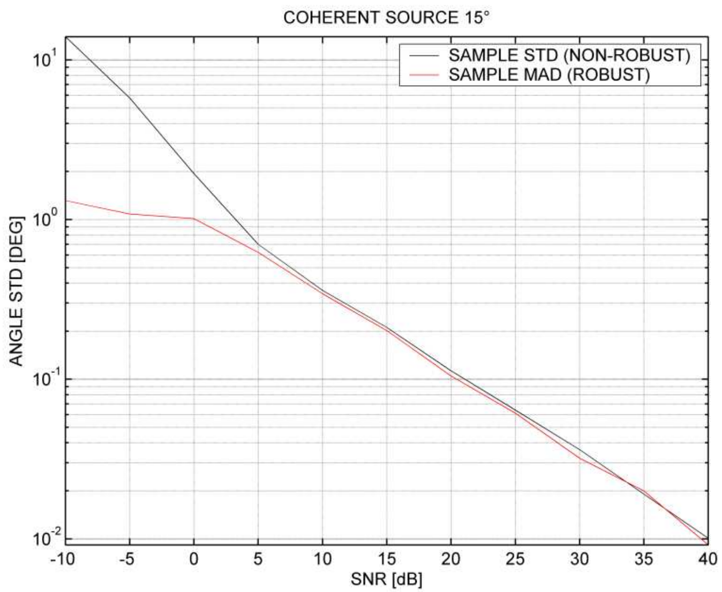


SNR thresholds

- Non-linear parametric estimators are typically affected by *gross estimation errors* whenever the square root CRB approaches about one half of the angular separation of closely spaced sources (at low SNR) or model uncertainty becomes of the same order of magnitude as finite sample errors (at high SNR).
- For medium SNR asymptotically efficient estimators do approach the *corresponding* CRB.
- Introducing *further information* about SCM structure, such as forward-backward relationships for ULAs, *may improve bounds and estimator performance*, at least in the absence of model errors.
- However, if model errors do not satisfy assumed symmetries, performance may be worsened.

Gross errors presence and experiment validation

- Gross errors put into evidence by comparing the non-robust *sample standard deviation* plot vs. the robust *Median Absolute Deviation* (MAD) plot, trimmed to a Gaussian distribution.
- However, even MAD has a non-regular behavior.



Discussion

- This perturbation analysis gives also origin to a practical, state of the art, Newton type *local optimization* of the difficult WSF functional.
- The first order perturbation is valid for a *large SNR range* and can be adapted in principle to any kind of error.
- It demonstrates that no other algorithm based on the SCM can have better large sample estimation variance than WSF, which must equate the ML one, i.e., approach the CRB, under mild conditions.
- The asymptotical perturbative setting makes it easy to analyze the *statistical impact* of any local equation.

Model errors effects on the SCM

A. Eigenvalue bounds based on a single deterministic perturbation.

B. Statistical analysis based on a deterministic or random steering vector perturbation model.

$$\mathbf{R}_{xx}(\mathbf{E}) = (\mathbf{A}+\mathbf{E})\mathbf{P}(\mathbf{A}+\mathbf{E})^H + \lambda_v \mathbf{I}$$
$$= \mathbf{R}_{xx}(\mathbf{0}) + \mathbf{EPA}^H + \mathbf{APE}^H + \mathbf{EPE}^H$$

$$\mathbf{R}_{xx}(\mathbf{0}) = \sum_{k=1}^{N} \lambda_k \mathbf{v}_k \mathbf{v}_k^H; \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$$

$$\lambda_{\eta+1} = \lambda_{\eta+2} = \ldots = \lambda_N = \lambda_v$$

Eigenvalue deterministic bound

- Steering vector errors modify eigenvector directions and eigenvalues of the SCM.
- Static mis-calibration *does not affect noise eigenvalues* (the signal subspace rank is not changed).
- Time varying errors during acquisition (e.g., originated by scattering, motion) modify noise subspace eigenvalues and eigenvectors.

$$\mathbf{E} = \mathbf{E}\mathbf{A}^{\dagger}\mathbf{A}; \quad \dot{\lambda}_k = \lambda_k\left(\mathbf{E}\right) - \lambda_k \cong \mathbf{v}_k^H\left[\mathbf{R}\left(\mathbf{E}\right) - \mathbf{R}\left(\mathbf{0}\right)\right]\mathbf{v}_k$$

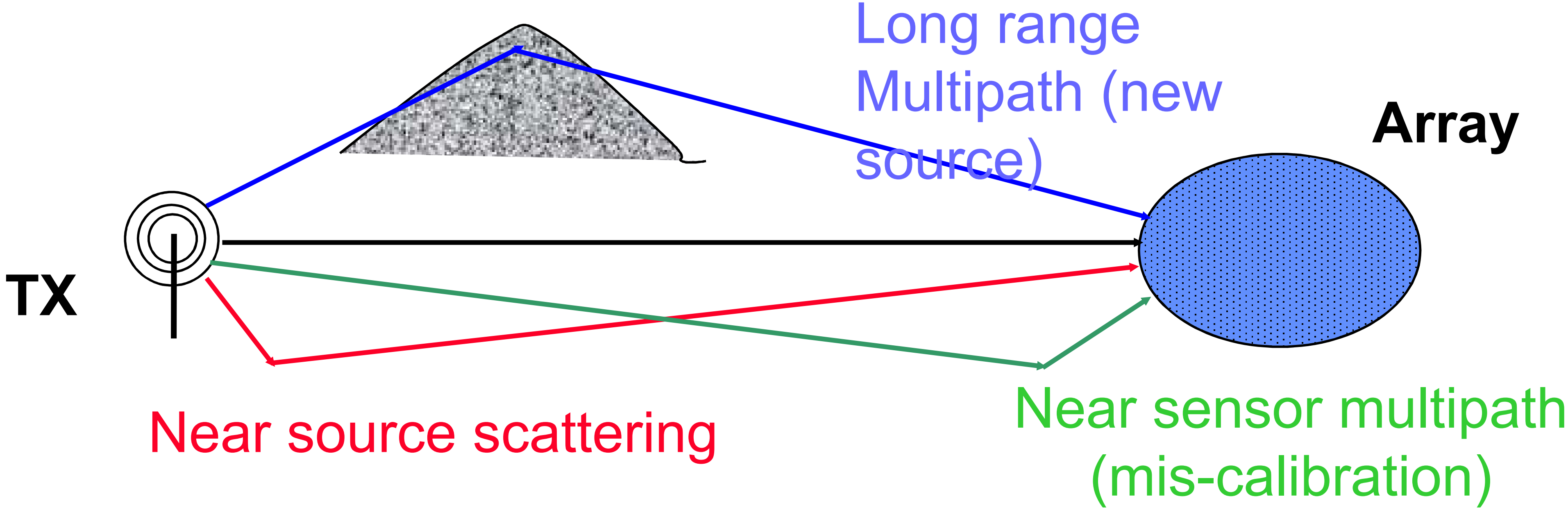
$$\left|\mathbf{E}\right|_2 = \varepsilon\left|\mathbf{A}\right|_2; \quad \varepsilon \ll 1; \quad \kappa\left(\mathbf{A}\right) = \left|\mathbf{A}\right|_2\left|\mathbf{A}^{\dagger}\right|_2$$

$$\left|\dot{\lambda}_k\right| \le 2\varepsilon\kappa\left(\mathbf{A}\right)\left(\lambda_k - \lambda_v\right) + \left[\varepsilon\kappa\left(\mathbf{A}\right)\right]^2\left(\lambda_1 - \lambda_v\right)$$

Notes on the deterministic bound

- Typical values of ε between 0.01 and 0.1.
- It is a pessimistic bound in general, but shows that *closely spaced sources may inflate eigenvalue changes through an high condition number* of the steering matrix.
- Noise subspace is not spherical anymore with time-varying errors.
  - Large sources may mask weaker ones;
  - Source number detection must take this fact into account.
- Noise pre-whitening (not analyzed here) may further inflate mis-calibration effects.

Eigenvector perturbations



$$\mathbf{a}_l(\mathbf{p}_l,k) \cong \mathbf{a}_l(\mathbf{p}_l) + \begin{bmatrix} \mathbf{a}_l(\mathbf{p}_l) & \dot{\mathbf{a}}_l(\mathbf{p}_l) & \ddot{\mathbf{a}}_l(\mathbf{p}_l) \end{bmatrix} \mathbf{h}_l(kT)$$

$$\tilde{\mathbf{A}}(\mathbf{p},k) = \mathbf{A}(\mathbf{p}) + \mathbf{A}^{(1)}(\mathbf{p},k); \quad \left|\mathbf{A}^{(1)}(\mathbf{p},k)\right|_2 \ll \left|\mathbf{A}(\mathbf{p})\right|_2$$

$$g_{ij}(k) = \mathbf{v}_i^H \mathbf{A}^{(1)}(\mathbf{p},k)\mathbf{A}^{\dagger}(\mathbf{p})\mathbf{v}_j, i > \eta, j \le \eta,$$

The scale of the perturbation scale does not depend upon the SNR!

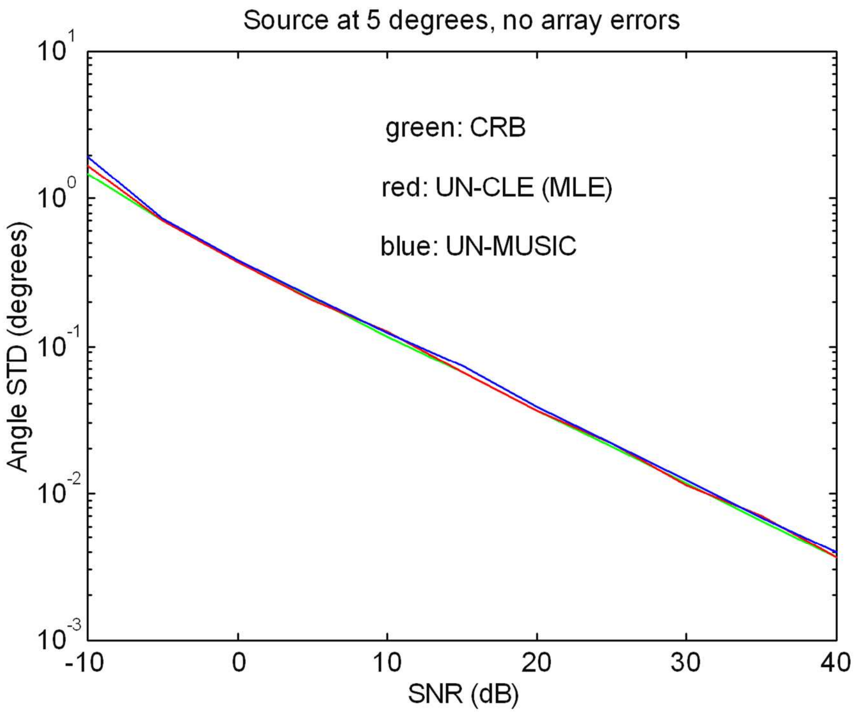
Sensitivity to model errors

- SCM signal eigenvectors act as *independent, cleaned snapshots*, asymptotically summarizing all available information of the SCM.
- The finite sample WSF optimal weighting *amplifies model errors for high SNR sources and small eigenvectors originated by coherent ones*.
- An uniformly weighted signal subspace (used by the so-called MD-MUSIC) improves results in many of these cases at a significant expense of statistical efficiency:

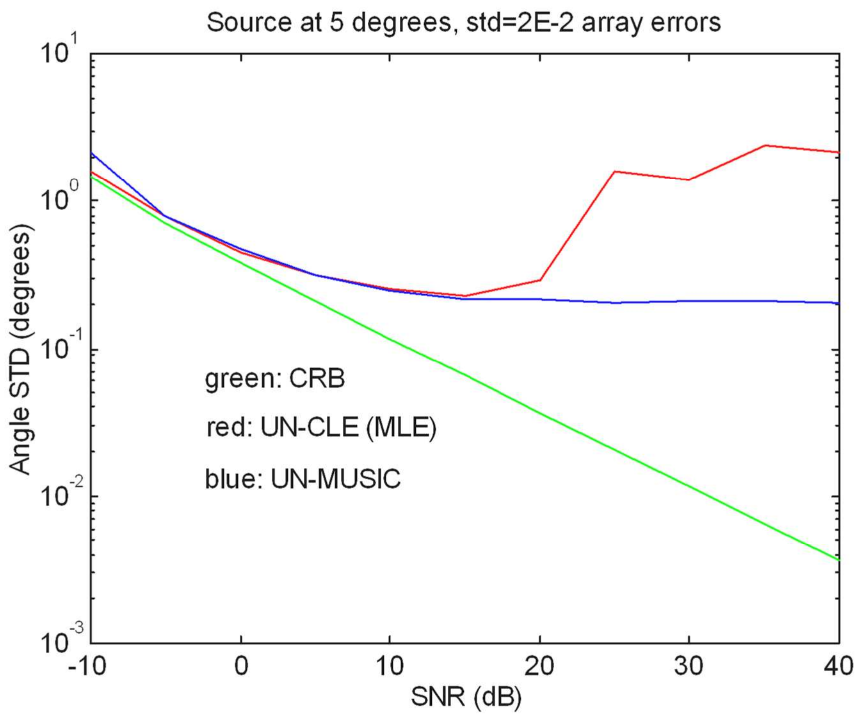
$$\mathbf{C}_{MUSIC} = \mathbf{I}_D \Rightarrow \mathbf{W}_{MUSIC} \propto \mathbf{E}_s^H \mathbf{A}(\mathbf{p})$$

Typical model error effects

Calibrated array



Random 2% RMS array errors



Statistical robustness

- An estimator is said *qualitatively robust* if:
  - it is *near optimal* (i.e., near asymptotically efficient) to the *central* (nominal) model, often a Gaussian one;
  - a moderate perturbation (on the data or model side) produces a statistically bounded estimation error;
  - a large contamination acting on few samples (e.g., lightning) cannot significantly impair the entire estimate (bounded influence function).
- A truly robust estimator has bounded impact of any SCM estimator and is a *minimax* estimator of some kind (minimizes the maximum risk of bias and variance).
- Array parameter estimators are essentially M-estimators, i.e., they are based on non-linear functional optimizations, difficult to make *robust* and reasonably *efficient* from a statistical point of view.

Robustness hints

- Gaussian ML estimation is intrinsically non-robust, because tied to a single, well behaved distribution, leading to an unbounded statistical impact.
- Subspace algorithms instead work through an *intermediate statistic* (SCM, GCC), which can be made robust in principle against data contamination.
- *Gaussian-based estimators applied to a robust statistic* are essentially robust to data distribution.
- However on the *equation side* it is possible an asymptotical mismatch: while increasing the sample size, local equations remain related to a wrong model, while the (robust) statistic converges to the true SCM!

Signal-related errors

- Finite sample effects have been analyzed in depth for localization and beamforming. In many cases they are not the primary source of impairment in practical applications
- Outliers, non-stationarities and heavy-tailed distributions are rather dangerous and can be afforded within the framework of *robust SCM estimation*.
- In any case, the reference (*central*) statistical signal model should be Gaussian, or, at least, *elliptical*. Without this assumption, existing array processing algorithms must be *completely re-designed.*

Model-related errors

- They constitute the main thread to array processing algorithms even for small perturbations.
- Consequences on direction finding algorithms are:
  - Non-consistent detection of the number of sources (imperfect noise whitening, weak signal masking);
  - Bias;
  - Gross errors in parameter estimation;
  - Estimation variance plateaux at high SNR (the classical CRB is not valid anymore).

Qualitative robustness

- WSF and ML weight model errors about proportionally to the sum of signal powers:
  - Masking of weak signals;
  - Errors of all sources contribute to the DOA estimate error.
  - Subspace fitting inconsistent if only some steering vectors are poorly calibrated w.r.t. the true environment.
- MUSIC is only sensitive to the *relative* steering vector mismatch of the *analyzed* source (other sources are cancelled anyway by inversion).
- MUSIC pseudo-spectrum is insensitive (i.e. *qualitatively robust*) to zero-mean reasonable random steering vector perturbations.

Model errors as hidden coherence effects

- Most model errors can be viewed as a result of *mis-calibration* and some unexpected form of *signal coherence*.
- *Model errors at low SNR and/or moderate sample size are masked* by finite sample errors. Thus the ML estimator still essentially minimizes the estimation error at low SNR.
- After a certain threshold SNR, model errors show up as *irreducible bias* and *excess estimation variance*.
- Weak algorithms exhibit a breakdown at high SNR with a *non consistent* behavior (variance may even increase with SNR!).
- Excluding very low eigenvectors in coherent cases raises the low SNR threshold, but may avoid catastrophe.

Random i.i.d. steering vector errors

- Steering vector error often assumed i.i.d. between sensors and directions.
- SCM *diagonal loading* depending on error strength and overall signal power.
- Formula useful for modified MMSE signal subspace estimation and LCMV beamforming.

$$E\left[\mathbf{E}\right]=0;\quad E\left[\mathbf{E}_{ij}\mathbf{E}_{kl}^{H}\right]=\sigma_e^2\delta_{ik}\delta_{jl};\quad \left|\mathbf{A}\right|_2\cong 1\Rightarrow \sigma_e^2\cong\frac{\varepsilon^2}{N}$$

$$E\left[\mathbf{R}_{xx}\left(\mathbf{E}\right)\right]=\mathbf{R}_{xx}\left(\varepsilon^2\right)=\mathbf{R}_{xx}\left(0\right)+\varepsilon^2\frac{\mathrm{trace}\left(\mathbf{P}\right)}{N}\mathbf{I}_N$$

Coherent multipath model

- By hypothesis, all the multipath arrivals happen *well within the duration of the temporal impulse response* of the analysis filter
- Otherwise, *SCM aliasing* results and a new *uncorrelated* source appears.
- The actual array response (*steering vector*) is a linear combination of *elementary* steering vectors, one for each multipath ray.

$$\mathbf{x}(l) = \sum_{d=1}^{\eta} \begin{bmatrix} \mathbf{a}_{1,d} & \cdots & \mathbf{a}_{Q(d),d} \end{bmatrix} \begin{bmatrix} \gamma_{1,d} \\ \vdots \\ \gamma_{Q(d),d} \end{bmatrix} s_d(l) + \mathbf{v}(l)$$

Classical ML approach

- Elementary steering vectors individually belong to a *calibrated manifold* (e.g., plane waves parametrized by their propagation directions).
- The SCM rank is *less* than the overall number of multipath arrivals.
- The combination coefficients are *further unknowns* to be estimated from the sample SCM.
- High number of unknowns and few independent observations are the receipt for bad parameter estimation.
- In addition, multipath coefficients often are *time varying (fast fading)*, impairing the rank limited SCM source signature.

Fast fading model

- Simplicistic scattering models are often used in telecommunications (circle, independent fading coefficient at each sensor).
- The mean SCM source signature is multi-rank, but is spread essentially within a rather *small angular sector*, *not filling the entire space, and badly conditioned*.

$$\mathbf{A}_d = \begin{bmatrix} \mathbf{a}_{1,d} & \cdots & \mathbf{a}_{Q(d),d} \end{bmatrix}$$

$$\boldsymbol{\gamma}_d = \begin{bmatrix} \gamma_{1,d} & \cdots & \gamma_{Q(d),d} \end{bmatrix}^T$$

$$E\left[\mathbf{x}(l)\mathbf{x}^H(l)\right] \propto \mathbf{A}_d E\left[\boldsymbol{\gamma}_d \boldsymbol{\gamma}_d^H\right]\mathbf{A}_d^H = \mathbf{A}_d \mathbf{P}_d \mathbf{A}_d^H$$

Drawbacks

- It is practically impossible to calibrate for all possible multipath sources (far field, near field, spatially extended sources, etc…).
- The overall number of resolvable sources diminishes, because the number of unknowns and the CRB increases.
- Rays (or small “difference type” eigenvectors) below a threshold SNR are badly estimated and *damage the overall performance*.
- Only critical ML searches and some suboptimal subspace techniques (*spatial smoothing*) can be used.
- The overall number of sources can be estimated by information theoretic criteria only by sequentially fitting a set of *competing models* on the data.

Finite bandwidth source SCM signature

- The SCM source model is *multi-rank*.
- The source subspace is confined within the subspace of the steering vector, plus its low-order derivatives vs. frequency.
- SCM structure is rather involved for *ciclostationary* sources.
- Aliasing and spectral leakage from filters further complicate this formula.
- Multi-rank signature places a severe *upper limit to the detectability* of weak sources spatially close to stronger ones.
- The dominant eigenvector of a properly selected single source SCM is a better replacement for the steering vector at the central bin frequency (i.e., a best rank-one SCM approximation).

SAPIENZA
UNIVERSITÀ DI ROMA

Finite bandwidth source SCM

$$\mathbf{x}_d(\mathbf{p},f)=\mathbf{a}_d(\mathbf{p},f)s_d(f)H(f)$$

$$\mathbf{R}_d(\mathbf{p})=\int_{f_1}^{f_2}\int_{f_1}^{f_2}H(f)H^*(u)E\left[s_d(f)s_d^*(u)\right]\mathbf{a}_d(\mathbf{p},f)\mathbf{a}_d^H(\mathbf{p},f)\,df\,du+\dots$$

$$\mathbf{R}_d^{(stat)}(\mathbf{p})=\int_{f_1}^{f_2}\left|H(f)\right|^2P_s(f)\mathbf{a}_d(\mathbf{p},f)\mathbf{a}_d^H(\mathbf{p},f)\,df+\dots$$

$$\cong\sum_{k=0}^{\eta\ll N}\lambda_{k,d}\mathbf{u}_{k,d}(\mathbf{p})\mathbf{u}_{k,d}^H(\mathbf{p})\cong\quad\lambda_{k,1}\mathbf{u}_{k,1}(\mathbf{p})\mathbf{u}_{k,1}^H(\mathbf{p})$$

$$\mathbf{u}_{k,d}(\mathbf{p})\in\mathrm{span}\left(\left[\mathbf{a}_d(\mathbf{p},f)\quad\frac{\partial\mathbf{a}_d(\mathbf{p},f)}{\partial f}\quad\cdots\quad\frac{\partial^\eta\mathbf{a}_d(\mathbf{p},f)}{\partial f^n}\right]\right)$$

Matched-field processing

- In many cases, the multipath structure is not completely *arbitrary* (seismics, low-flying aircrafts on the sea, ducts, shallow waters, microphone arrays within rooms) or it can be accurately estimated (quasi-static radio links).
- Solutions of the hyperbolic wave equation are quite insensitive to reasonable changes of boundary conditions and losses in the medium.
- The array response is made by a linear combination of *modes*, whose coefficients depend upon source location.

Matched-field signal model

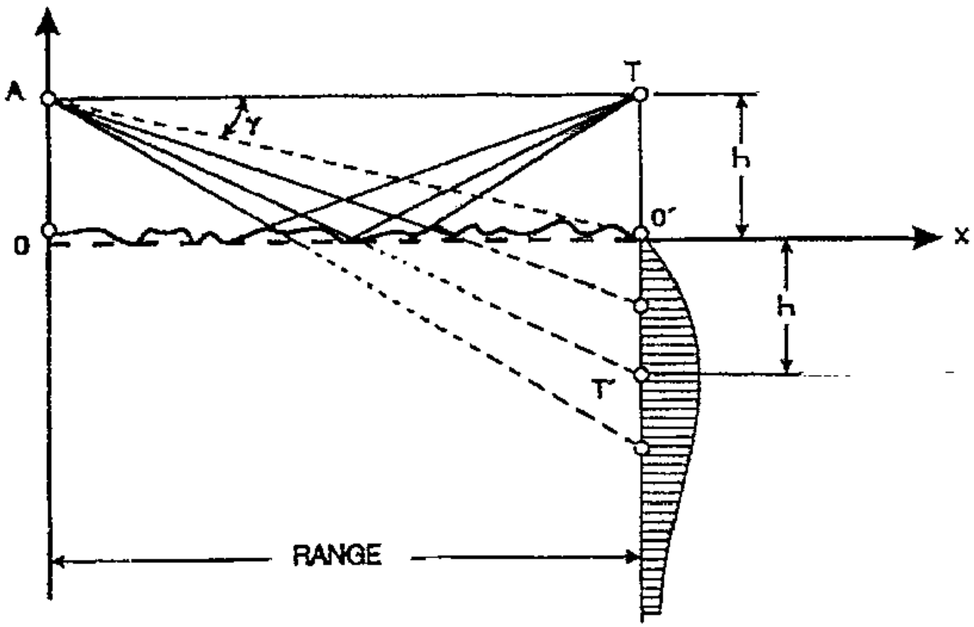
- A *baseline solution* is computed by a wavefield simulator for a given *environmental parameter* vector $\mathbf{c}$ (propagation speed, reflection coefficients at the boundaries, surface roughness indexes,…).
- A *random error vector* takes into account wavefield approximations and uncertainties.

$$\mathbf{a}(f,\mathbf{p},\mathbf{c}) = \begin{bmatrix} \mathbf{b}_1(f,\mathbf{p},\mathbf{c}) & \cdots & \mathbf{b}_Q(f,\mathbf{p},\mathbf{c}) \end{bmatrix} \begin{bmatrix} \gamma_1(\mathbf{p},\mathbf{c}) \\ \vdots \\ \gamma_Q(\mathbf{p},\mathbf{c}) \end{bmatrix} + \mathbf{e}(f,\mathbf{p})$$

Properties of matched field model

- The matched-field model is often capable of 3-D localization and beamforming, together with the identification of some environmental parameters (as in seismic migration).
- The number of array sensors must be higher than the number Q of significant modes for an *unique* matched-field source representation.
- Ambiguity in the presence of multiple sources increases in reverberant environments.
- There is some freedom in choosing Q for the best bias/variance trade-off of the overall matched-field modelling.

Example: Low-flying beacon



- The *main contribute* to the steering vector is due to the *direct* path and the reflected path from the image source
- Phase relationships are rather stable
- *Diffuse multipath* is also present, depending on the sea state
- Height and distance of the beacon are both identifiable with a *sufficient vertical array aperture*

Matched field MUSIC for specular multipath

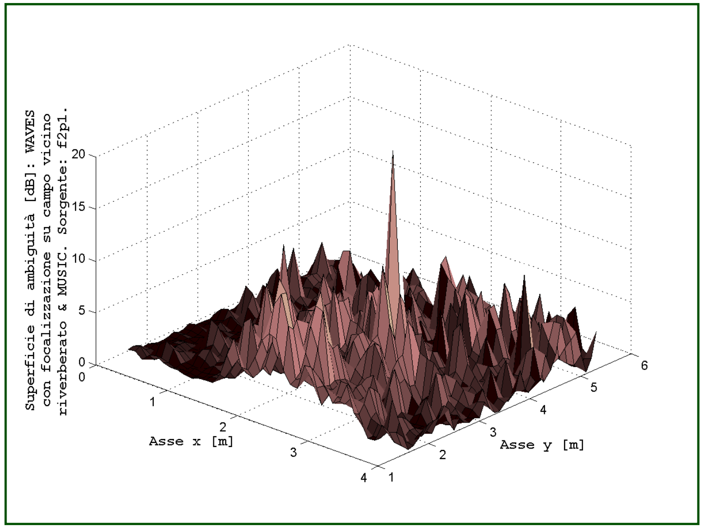
- Steering vector is an *unknown combination* of the steering vectors of the target and of its reflection(s) from spread virtual sources, plus a random term.
- MUSIC can robustly cope with random array errors. Do not ever try this with ML or WSF!
- *Constrained minimum eigenvalue problem* at each candidate target location.

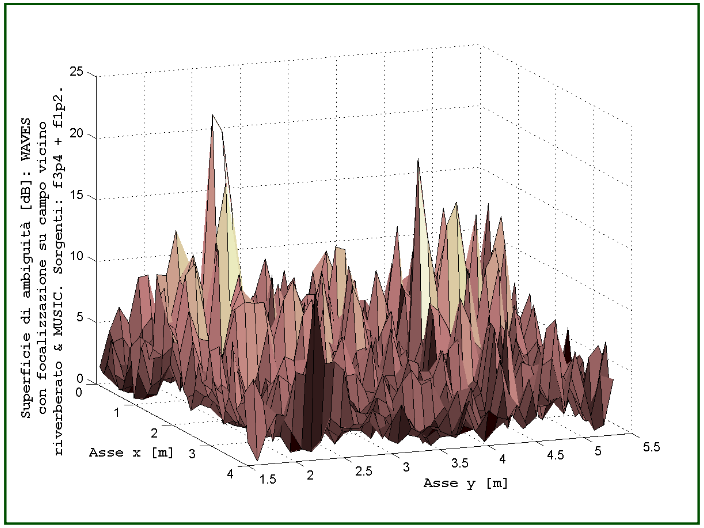
$$\tilde{\mathbf{a}}(\mathbf{p}) = \begin{bmatrix} \mathbf{a}(\mathbf{p}) & \mathbf{a}(\mathbf{p}_v) \end{bmatrix} \mathbf{c}(\mathbf{p}) + \mathbf{e}(\mathbf{p}) = \mathbf{A}(\mathbf{p})\mathbf{c}(\mathbf{p}) + \mathbf{e}(\mathbf{p})$$

$$\mathbf{A}(\mathbf{p})\mathbf{c}(\mathbf{p}) \overset{QRD}{=} \mathbf{Q}(\mathbf{p})\left[\mathbf{R}(\mathbf{p})\mathbf{c}(\mathbf{p})\right] = \mathbf{Q}(\mathbf{p})\mathbf{c}_R(\mathbf{p})$$

$$\left[\hat{\mathbf{p}}_{MUSIC}, \hat{\mathbf{c}}_R(\hat{\mathbf{p}})\right] = \underset{\mathbf{p},\mathbf{c}_R}{\operatorname{argmin}} \left\{ \frac{\mathbf{c}_R^H \mathbf{Q}(\mathbf{p})^H \hat{\mathbf{E}}_N \hat{\mathbf{E}}_N^H \mathbf{Q}(\mathbf{p}) \mathbf{c}_R}{\left|\mathbf{c}_R\right|^2} \right\}$$

Real-world data examples





- Narrow-band matched-field localization experiment at the INFOCOM Dpt.
- Wave-field *simulated* by the virtual source Matlab program
- Recorded data well matched to simulated responses
- One sharp peak exists in the MUSIC pseudo-spectrum for a single source (above).
- Many spurious peaks arise in the two-source case (below)

Robust matched-field array processing

- Matched-field processing *inherently calls for robust estimators* due to practical wave-field approximations.
- Capon-type adaptive beamforming is the preferred basic matched-field technique for both signal copy and source localization tasks in uncertain environments.
- Both the SCM and the signal model can be modified.
- Robust adaptive LCMV beamformers are required because of the risk of common errors in *pointing*, *calibration*, *bad constraint setup*, etc…

SAPIENZA
UNIVERSITÀ DI ROMA

Robust matched-field adaptive beamforming approaches

- Statistically robust SCM estimation against finite sample errors and outliers.
- Mutual coupling correction.
- Optimal quiescent vector selection in a random environment.
- Direction-dependent linear constraints.
- Quadratic robustness constraints.
- Wideband ML steered beamforming.
- Matched-field focusing.

Robustness to signal distribution

- Noise is often Gaussian, but not always.
- Non Gaussian signals of interest may appear when using the unconditional model.
- A set of distributions *close in some norm* to the Gaussian one or a *contaminated Gaussian* distribution is assumed as reference.
- Robust SCM estimation is all what is needed for subspace estimators.
- Rotation properties of the scatter matrix (implicit by MUSIC, ML, WSF…) are not always maintained by robust SCM estimators.
- Bias and nonnegative SCM estimates may result. Use *regularization* (e.g., add a scaled identity matrix)!

SAPIENZA
UNIVERSITÀ DI ROMA

Main basic approaches to SCM robustification

- Independent *robust estimate* of each cross-sensor correlation (or cross-correlation coefficient).
- Robust *scale estimation along many different slices* of the SCM (as in Bartlett beamforming), followed by a non-negative covariance fitting.
- Rotationally invariant robust SVD (PCA) by iterative re-weighting of the norm of each observation.
- Clustering of snapshots based on some heuristic criteria (norm, alignment), followed by outlier excision and robust SCM estimation of remaining ones.

Robust cross-correlation coefficient

- A robust correlation coefficient is separately computed among all real and imaginary part of sensor output signals.
- Quarter-square identity is used.
- SCM for circular signals is reconstructed.

$$\rho_k(\mathbf{x}_i,\mathbf{x}_j)\frac{2c_k s_k}{c_k^2+s_k^2}=\frac{S(c_k\mathbf{y}_i+s_k\mathbf{y}_j)^2-S(c_k\mathbf{y}_i-s_k\mathbf{y}_j)^2}{S(c_k\mathbf{y}_i+s_k\mathbf{y}_j)^2+S(c_k\mathbf{y}_i-s_k\mathbf{y}_j)^2}$$

$$\hat{\rho}(\mathbf{x}_i,\mathbf{x}_j)=\mathrm{median}\left\{\rho_k(\mathbf{x}_i,\mathbf{x}_j);k=1,2,\ldots,K\right\}$$

$$\mathbf{y}_i=\frac{\mathbf{x}_i-\mathrm{median}(\mathbf{x}_i)}{S(\mathbf{x}_i)};\quad S(\mathbf{x})=\mathrm{median}\left[\left|\mathbf{x}-\mathrm{median}(\mathbf{x})\right|\right]$$

$$\mathbf{R}(i,j)=S(\mathbf{x}_i)\hat{\rho}(\mathbf{x}_i,\mathbf{x}_j)S(\mathbf{x}_j);\quad i,j=1,\ldots,P$$

$$i=1,\ldots,P-1;\quad j=i+1,\ldots,P$$

Robust element-wise SCM estimate

- Quality of the estimate is tied to the robust scale estimator $S(x)$ used.
- The SCM estimate may not be positive definite (regularize it…).
- SCM estimate is not invariant over orthogonal/unitary transformations (conceptual issue for PCA).
- The presented robust SCM has *high resistance to outliers* (~25%), but it is far from efficiency at the multivariate Gaussian distribution.

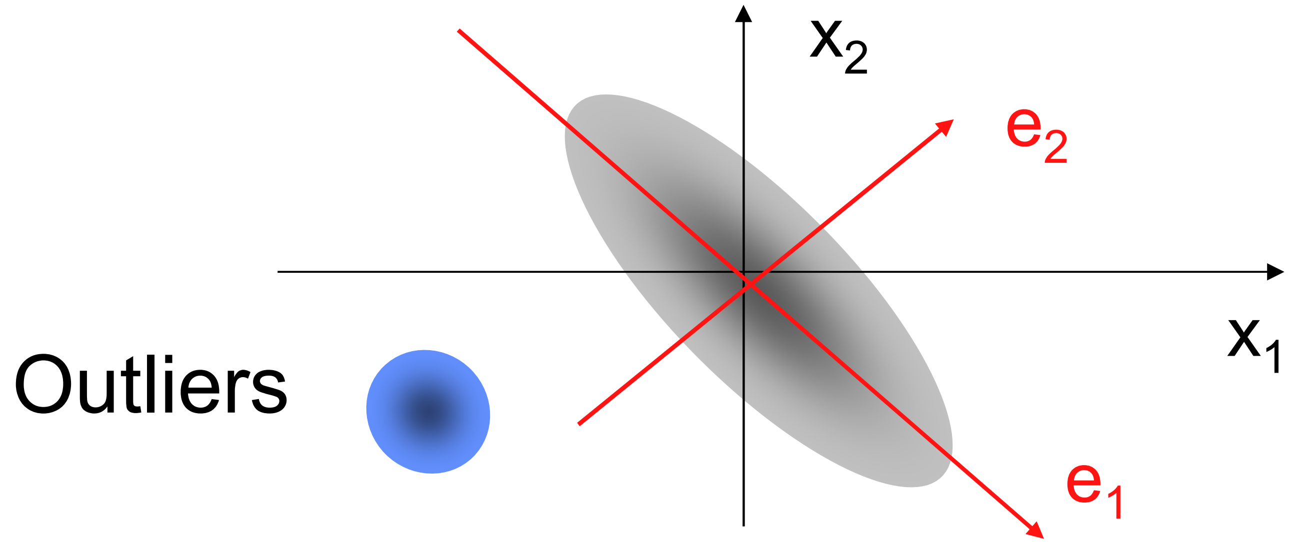
A clever orthogonally invariant use of robust SCM

- Compute robust SCM as indicated.
- Compute EVD and save eigenvectors.
- Rotate the data matrix with these eigenvectors.
  a) Compute a *robust scale estimate* of rotated columns (i.e., robust singular values) and reconstruct a positive definite SCM.
  b) Identify, remove or clip outliers, then compute a Gaussian or pseudo-vcovariance SCM estimate.
- Constrain SCM estimate for circular signals.
- Bias might result, not a cheap approach, but useful for accurate data inspection.

SAPIENZA
UNIVERSITÀ DI ROMA

Robust SCM estimation

- Processing strictly tied to the sample SCM eigen-structure.
- Excision of anomalous snapshots (outliers).
- Adaptive re-weighting of remaining snapshots for SCM estimation (pseudo-covariance).
- Pseudo-SCM eigenvalue regularization.



SAPIENZA
UNIVERSITÀ DI ROMA

Pseudo-covariance

- Robust ML estimation of a *least informative*, *elliptical* and circular multivariate PDF.
- PDF choice limited by numerical considerations.
- An *improper* PDF often results.
- Used for clutter SCM estimation in radar.

$$\underset{(P\times N)}{\mathbf{X}} = \left\{ \mathbf{x}_1, \ldots, \mathbf{x}_N \right\}; \quad f_x(\mathbf{x}) \propto \frac{1}{\det(\mathbf{C}\mathbf{C}^H)} g\left(\left| \mathbf{C}^{-1}\mathbf{x} \right| \right); \quad s(r) = \frac{1}{r + a\sqrt{P/N}}$$

$$\begin{cases} \tilde{\mathbf{C}}\mathbf{y}_i = \mathbf{x}_i \\ \dfrac{1}{N}\sum_{i=1}^{N} s\left(\left| \mathbf{y}_i \right| \right)^2 \left(\mathbf{y}_i \mathbf{y}_i^H - \mathbf{I}_P \right) = 0 \end{cases} \Rightarrow \sum_{i=1}^{N} \frac{s\left(\left| \mathbf{y}_i \right| \right)^2}{\left[\sum_{j=1}^{N} s\left(\left| \mathbf{y}_j \right| \right)^2 \right]} \mathbf{y}_i \mathbf{y}_i^H = \mathbf{I}_P$$

$$\mathrm{Pseudocov}(\mathbf{x}) = \tilde{\mathbf{C}}\tilde{\mathbf{C}}^H = \begin{bmatrix} \tilde{\mathbf{E}}_s & \tilde{\mathbf{E}}_n \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{\Lambda}}_s & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{\Lambda}}_n \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{E}}_s & \tilde{\mathbf{E}}_n \end{bmatrix}^H$$

Notes on pseudo-covariance

- Pseudo-covariance preserves the orthogonal transformation property of the Gaussian ML SCM estimate, so *common subspace fitting estimators can be applied* without modifications.
- Preserves the *independence* of observations.
- It is able to *equalize by construction the statistical impact of all snapshots on the SVD*, so it is useful in clutter scatter estimation, where samples are taken ad differente ranges and some bins can contain targets.
- It is not able to locate even one faulty sensor.
- Fine for moderately long-tailed noise distributions (e.g., *shrimp noise* in underwater sonar).

Robust beamforming

- In extreme cases multi-source high resolution estimators substantially fail in locating signals of interest.
- A Capon based robust beamformer can be a better choice than WSF or MUSIC, even for signal copy.
- Antenna movements, near-field scattering, time varying multipath, pointing errors, uncertain propagation medium (non-homogeneous, stratified) are always present.
- *Standard Capon is surely weak* under coherent environments.

Signal cancellation in adaptive beamforming

- The *signal of interest leaks in the blocking subspace*.
- The quiescent path and the adaptive sidelobe canceller both contain the signal of interest.
- If the SNR in the blocking subspace is greater than about 0.5, the signal at the beamformer output is almost completely *suppressed.*

$$\mathbf{C}_{\perp}^{H}\mathbf{x}(l)=\mathbf{b}s(l)+\mathbf{v}_{\perp}(l); \quad \mathbf{w}_{0}^{H}\mathbf{x}(l)=as(l)+\mathbf{v}_{0}(l)$$

$$\left(\mathbf{w}_{0}^{H}+\mathbf{w}_{1}^{H}\mathbf{C}_{\perp}^{H}\right)\mathbf{x}(l)=\left(a+\mathbf{w}_{1}^{H}\mathbf{b}\right)s(l)+v(l)$$

$$\left|a+\mathbf{w}_{1}^{H}\mathbf{b}\right|\ll\left|a\right|\Rightarrow\text{cancellation}$$

Diagonal loading

- The *finite sample variance* of adaptive beamformers mainly depends on the amplitude and *spread* of the smallest (noise) eigenvalues of the empirical SCM.
- Small eigenvalues can be *fully suppressed* in a Principal Component Analysis framework or *increased* by a constant quantity, or raised above a threshold (*eigenvalue thresholding*).
- The simple *diagonal loading* applies the same amount of *regularization noise* to all eigenvalues.

$$\hat{\mathbf{R}}_{xx}(f,\lambda) = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}(l)\mathbf{x}^H(l) + \lambda\mathbf{I}$$

Linear constraints on LCMV

- Static linear constraints were often inserted on the LCMV weight vector design.
- In particular, gradient constraints are useful to mitigate effects of pointing errors.
- The *beamformer mainlobe is however enlarged* and any constraint strategy can be defeated by certain array perturbations.

$$\begin{bmatrix} \mathbf{a}(\mathbf{p}) & \dfrac{\partial \mathbf{a}(\mathbf{p})}{\partial p_1} & \cdots & \dfrac{\partial \mathbf{a}(\mathbf{p})}{\partial p_q} \end{bmatrix}^H \mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

SAPIENZA
UNIVERSITÀ DI ROMA

Minimax robust beamforming

- A true robust beamforming procedure *should not be tied to an exact error modeling*.
- Robust beamforming should withstand array perturbations *of any kind up to a pre-specified leverage point*.
- So it should be a *minimax* procedure of some type.
- All developed robust beamformers make use of *adaptive diagonal loading* in various forms.
- A finer error model specification is un-desired!

Optimal matched-field quiescent vector

- It represents the *best fit* of a vector to a set of steering vectors simulated within a *ball* in the space of environmental parameters, centred around a set of nominal parameters
- The optimal quiescent vector is *proportional* to the dominant eigenvector of the ball scatter matrix.

$$\mathbf{A} = \int_{\mathbf{c} \in B(\mathbf{c}_0)} \mathbf{a}(f, \mathbf{p}, \mathbf{c}) \mathbf{a}(f, \mathbf{p}, \mathbf{c})^H \, p_c(\mathbf{c}, \mathbf{c}_0) \, d\mathbf{c}$$

$$\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^H + \sum_{k=2}^{r} \lambda_k \mathbf{e}_k \mathbf{e}_k^H + \sum_{k=r+1}^{\dim(\mathbf{A})} \lambda_k \mathbf{e}_k \mathbf{e}_k^H$$

$$\lambda_1 > \lambda_2 > \ldots \lambda_{\dim(\mathbf{A})}; \quad \lambda_{r+q} < \varepsilon, \quad q = 1, \ldots, \dim(\mathbf{A}) - r$$

Linear constraints for robust matched field beamforming

- The eigenvectors corresponding to the *negligible eigenvalues* of $\mathbf{A}$ define the blocking subspace
- The weight vector should not have significant projections along remaining eigenvectors to avoid *cancellation* of the useful signal in adaptive beamforming.
- These constraints are globally expressed by the undetermined set of linear equations:

$$\begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_r\end{bmatrix}^H \mathbf{w} = \begin{bmatrix} a^* & 0 & \cdots & 0\end{bmatrix}; \quad \mathbf{C}_\perp = \begin{bmatrix}\mathbf{e}_{r+1} & \cdots & \mathbf{e}_{\dim(\mathbf{A})}\end{bmatrix}$$

SAPIENZA
UNIVERSITÀ DI ROMA

Quadratic constraints

- Cancellation can be mitigated in a minimax approach by keeping the norm of the adaptive weight vector below a certain threshold.
- This is a *ridge regression* problem, equivalent to an *optimal diagonal loading.*
- Requires the solution of a *secular equation* involving sample SCM eigenvalues.

$$\left.\begin{array}{c} \left|a+\mathbf{w}_1^H\mathbf{b}\right| > (1-\delta)\left|a\right|; \quad 0<\delta\ll 1 \\ \forall\left|\mathbf{b}\right| : \left|\mathbf{b}\right| < \varepsilon\left|a\right|; \quad 0<\varepsilon\ll 1 \end{array}\right\} \Rightarrow \left|\mathbf{w}_1\right| < \left|a\right|\frac{\delta}{\varepsilon}$$

SAPIENZA
UNIVERSITÀ DI ROMA

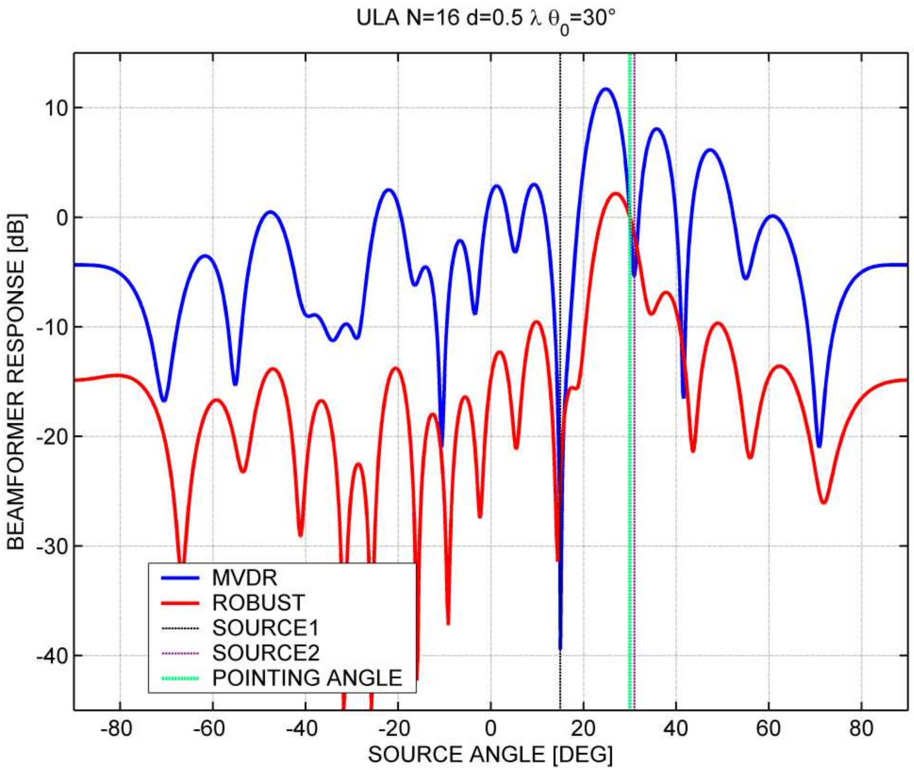
Equivalent constraint formulations

- The SNR threshold approach leads to a similar norm bound.
- The MV beamformer output power can be *maximized* for a steering vector in a *ball around the nominal one* (*Robust Capon beamforming* by Luo, Li and Stoica)
- The resulting *ridge regression* is equivalent to a *convex optimization* problem (Vorobyov), but it is far more computationally efficient.
- It is equivalent to a *non-linear compression of dominant SCM eigenvalues*.

$$\min_{\mathbf{a}}\left[\mathbf{a}^H\tilde{\mathbf{R}}_{xx}^{-1}(f)\mathbf{a}\right]\quad s.t.\quad \left|\mathbf{Q}^{-1}\left(\mathbf{a}-\mathbf{a}_0\right)\right|<\varepsilon$$

Robust beamforming example

- Two uncorrelated, equi-powered Gaussian sources at 15° (no. 1) and 31° (no. 2), SNR = 20 dB.
- Source no. 2 with five *fully coherent scatterers* with SNR around 0 dB, located at 28°, 29°, 30°, 31° and 33°.
- Beamformer aimed at 30°.



Comments on robust Capon beamforming

- The MVDR beamformer places several lobes on coherent scatterers to cancel the main source no. 2.
- The MVDR weight vector has very high norm (i.e., noise amplification).
- The light robustification herein performed by *norm limitation* reduces the weight vector norm, stabilizes the mainlobe and *avoids significant cancellation* by nearby coherent scatters.
- The null placed on the interference source no. 1 is however widened in the robust beamformer.

SAPIENZA
UNIVERSITÀ DI ROMA

Space-time coding robustness

- It can be viewed as a special matched-field beamforming problem.
- Lack of robustness constraints with respect to the array response $\mathbf{H}$ can lead to:
  - part of signals of interest creates a *nearly white background noise field*.
  - SNR and upper rate bound reduction of each extracted stream due to uncorrelated *friendly noise* superposition.
  - high risk of complete link loss.

$$\mathbf{W}^H\left[\hat{\mathbf{H}}\mathbf{U}\mathbf{s}(l)+\mathbf{v}(l)\right]\approx\mathbf{s}(l)\quad s.t.\quad E\left[\left|\mathbf{s}(l)\right|_2^2\right]=C;$$

$$E\left[\mathbf{s}(l)\mathbf{s}^H(m)\right]=\delta_{lm}\,\mathrm{diag}\left\{\left[p_1\quad\cdots\quad p_D\right]\right\};\quad p_i\geq 0$$

$$E\left[\mathbf{s}(l)\mathbf{s}^T(m)\right]=\mathbf{0}$$

Space-time coding issues

- Limited accuracy of array steering vector estimation, due to *non-stationarity* and short *training sequences*.
- Non-independent, but *low-rank* sector correlated fading.
- Some approaches taken by matched field models, such as picking only the largest eigenvectors of the array transfer matrix.
- Not much space for improvements, except adopting diagonally loaded signal copy for robust MMSE signal estimation.

Narrow-band model limitations

- *Low number of independent snapshots* available for a given observation time in reverberant environments.
- Very *fine frequency resolution* required in most applications:
  - Analysis must be repeated for several close frequencies, leading to a *wide-band*, but *incoherent* processing.
  - Excessive number of free parameters (*curse of dimensionality*) and high estimation variance.
  - Analysis unable to really separate contributions of various propagation modes.
- High wavefield ambiguity in reverberant environments.
- Much superior efficiency of robust algorithms in wide-band scenarios.

Part VI

WIDEBAND ARRAY PROCESSING

Wide-band array processing

- Narrow-band model conditions for identifiability are not fullfilled in relevant applications (sonar, acoustics, seismics, UWB communications).
- Wide-band signal model radically changes with frequency.
- The common research goal is of *extending the appealing geometrical narrow-band approaches to wide-band environments*.

The boundary between narrow- and wide-band settings

- The general array convolutional model must surely replace the instantaneous narrow-band mixing model when the SCM is not anymore the sufficient statistic for Gaussian ML identification.
- This happens when the effective length of the array output multi-channel correlation (the largest impulse response length plus the signal correlation length) exceeds one sampling period.
- In this case the sufficient statistic for Gaussian ML identification in the stationary case becomes the *space-time covariance matrix* (STCM).

SAPIENZA
UNIVERSITÀ DI ROMA

Wide-band and UWB paradigms

- Array design and performance are tied to the *ratio between physical distance/aperture and the operating wave-length*.
- Below $\lambda/4$ inter–sensor spacing, element responses become too similar (redundant) and mutual coupling effects quickly increase.
- Over $\lambda/2$ spacing, ambiguity appears in various forms.
- Over one octave signal bandwidth, amplifier and receiver linearity becomes critical (IMD products).
- So it makes sense to speak about an *Ultra-Wide Band (UWB) paradigm* when bandwidth is greater than about *one octave*.

Space-time covariance matrix

- The STCM is the *multi-channel covariance matrix* between all (zero mean) sensor outputs and their delayed versions up to a certain order P.
- The *space time snapshot* is also introduced.

$$\underset{(NP\times NP)}{\mathbf{R}_{ST}} = E\left[\mathbf{x}_{ST}\mathbf{x}_{ST}^{H}\right]$$

$$\mathbf{x}_{ST}(t)\Big|_{t=mT} = \left.\begin{bmatrix} \mathbf{x}\left[t-(P-1)T\right] \\ \mathbf{x}\left[t-(P-2)T\right] \\ \vdots \\ \mathbf{x}(T) \end{bmatrix}\right|_{t=mT} = \begin{bmatrix} \mathbf{x}(m-P+1) \\ \mathbf{x}(m-P+2) \\ \vdots \\ \mathbf{x}(m) \end{bmatrix}$$

STCM block structure

- The STCM has a *block Toeplitz* structure in the stationary case, useful for building fast estimation algorithms.

$$
\mathbf{R}_{ST}=\begin{bmatrix}
\mathbf{R}_{xx}(0) & \mathbf{R}_{xx}(-1) & \mathbf{R}_{xx}(-2) & \cdots & \cdots & \mathbf{R}_{xx}(-P+1)\\
\mathbf{R}_{xx}(1) & \mathbf{R}_{xx}(0) & \ddots & \ddots & \cdots & \vdots\\
\mathbf{R}_{xx}(2) & \mathbf{R}_{xx}(1) & \ddots & \ddots & \ddots & \vdots\\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\
\vdots & \cdots & \cdots & \ddots & \ddots & \mathbf{R}_{xx}(-1)\\
\mathbf{R}_{xx}(P-1) & \cdots & \cdots & \cdots & \mathbf{R}_{xx}(1) & \mathbf{R}_{xx}(0)
\end{bmatrix}
$$

$$
\mathbf{R}_{xx}(p)=E\left[\mathbf{x}(m)\mathbf{x}^{H}(m+p)\right]
$$

STCM hints

- In some cases, the narrow-band *steering vector might be considered as constant* across the signal bandwidth, but *source signals and/or the background noise are temporally correlated*. It is easy to verify that the STCM must still be used instead of the narrow-band SCM, which implies $P=1$.
- STCM can be built with either (complex) pass-band or (real) low-pass signals.
- *Integer delays have uncertain and slow convergence to the array response* across the full digital bandwidth.
  - Signals can be oversampled and then low-pass filtered.
  - By Papoulis theorem, the integer delay expansion can be replaced by more quickly converging expansions, based on Laguerre generalized FIR (dispersive transmission lines) or 1-D Gauss-Hermite filters banks.

Array model representation accuracy by STCM

- Many wide-band signal properties can be *better assessed in continuous time*.
- Even under the finite bandwidth hypothesis, discrete time FIR (MA) type array models have inherent approximations, to be evaluated for each case.
- Sample Fourier transforms and integrals have often to be performed by high order *quadrature and interpolation formulas*.
- Estimation algorithms must take into account these basic approximation errors.

Delay of arrivals

- The simplest Ultra-Wide Band systems cannot afford resources for operating on the huge STCM and operate in non-critical environment (single dominating source, white noise and uncorrelated sensors).
- So it makes sense to operate on selected STCM slices on pairs of sensors.
- *Non dispersive* propagation implies a *constant group delay* between sensor pairs, which can be the target of a localization procedure instead of angles.
- Time Difference of Arrivals (TDOA) localizes sources in a non-dispersive medium *on one branch of a hyperboloid of revolution with two sensors placed in the foci*, approaching a *conical surface* at long ranges.
- Al least two sensors are needed for far-field localization on a plane, three in 3-D space, using *cone or hyperbolic least mean intersection*.

SAPIENZA
UNIVERSITÀ DI ROMA

Time-Delay of Arrival (TDOA) estimation

- Assume that medium and sensor responses are matched so that output array signals are *delayed copies of the same source signal* immersed into independent noise realizations.
- This model is often used in radio-astronomy, radio-navigation, seismics, aerial acoustics and UWB short-range communications.
- Fractional delay estimation implies oversampling.
- TDOA is measured between sensor pairs.
- In most cases, delays are clustered afterwards.

SAPIENZA
UNIVERSITÀ DI ROMA

Two-sensor 2-D TDOA model

- Angle θ measured from the broadside of the *measurement basis line* passing through a sensor pair.
- Single baseband (real valued) source signal.
- Weakly temporally correlated source signals.
- *Independent,* temporally white, additive noise between sensors.
- Wave-front direction is estimated from inter-sensor delays.

$$x_1(t) = A_1 h(t,\mathbf{p}) * s(t) + v_1(t)$$

$$x_2(t) = A_2 h(t,\mathbf{p}) * s\left[t - \tau(\mathbf{p})\right] + v_2(t)$$

$$\tau(\mathbf{p}) = \frac{\vec{\mathbf{k}} \cdot (\vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_2)}{\left|\vec{\mathbf{k}}\right| c} = \frac{\left|\vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_2\right| \sin(\theta)}{c}; \; A_1, A_2 > 0$$

Cross-correlation based TDOA techniques

- Regression between two sensor signals.
- Optional estimation of sensor gain ratio.
- Maximization of the cross-correlation function.
- AMDF (i.e., MPEG-like motion compensation) does the same thing in the L1 norm, more robust to impulsive signals.

$$\left[\hat{\tau}_{CC}, \hat{a}_{CC}\right] = \arg\min_{\tau,a} E\left[\left|x_1(t-\tau) - a x_2(t)\right|^2\right]$$

$$\hat{\tau}_{CC} = \arg\max_{\tau} \left\{E\left[x_1(t-\tau)\, x_2(t)\right]\right\}$$

$$\left[\hat{\tau}_{AMDF}, \hat{a}_{AMDF}\right] = \arg\min_{\tau,a} E\left[\left|x_1(t-\tau) - a x_2(t)\right|\right]$$

SAPIENZA
UNIVERSITÀ DI ROMA

Notes on cross-correlation techniques

- Popular in UWB communications and acoustics.
- *Spurious peaks* (delay ambiguity) for temporally correlated signals.
- *Biased cross-correlation* estimates lead to biased delay (angle) estimates.
- Discrete-time cross-correlation estimates interpolated around peaks by a parabola (Jacovitti-Scarano).
- AMDF better performing in many experiments.
- Really suboptimal and weak estimators to sensor mis-matching, reverberation and coloured source spectra.

SAPIENZA
UNIVERSITÀ DI ROMA

GCC TDOA models

- The two sensor model is essentially a *quasi-deterministic-model*.
- The GCC, concentrated w.r.t. sensor output spectra, is the sufficient statistic for this problem in the frequency domain.

$$X_1(f) = A_1 H(f) S(f) + V_1(f)$$

$$X_2(f) = A_2 H(f) S(f) e^{-j2\pi f \tau(\mathbf{p})} + V_2(f)$$

$$\rho_{GCC}(f) = \frac{E\left[X_1(f) X_2^*(f)\right]}{\sqrt{E\left[\left|X_1(f)\right|^2\right]}\sqrt{E\left[\left|X_2(f)\right|^2\right]}}$$

$$= \frac{A_1 A_2^* \left|H(f)\right|^2 P_s(f) e^{j2\pi f \tau(\mathbf{p})}}{\sqrt{\left|A_1\right|^2 \left|H(f)\right|^2 P_s(f) + P_{v1}(f)} \cdot \sqrt{\left|A_2\right|^2 \left|H(f)\right|^2 P_s(f) + P_{v2}(f)}}$$

GCC properties

- The GCC of two delayed signals contains a *disturbed harmonic* vs. the analysis frequency.
- Whitening of sensor signals equalizes estimation variance (a necessary condition for MLE…).
- *Very insensitive* statistic to signal and noise spectra.
- Difficult consistent and unbiased estimation of the GCC because periodogram estimates are affected by finite bin width and spectral aliasing.

Pre-whitened GCC

- Two-stage pre-whitening is highly preferable:
  - A *compactly parametrized pre-whitener* (e.g., an AR one) is preferably estimated from both sensor signals, averaging the sum (LS) or the logarithmic sum (Approximate ML) of signal prediction errors.
  - Both sensor signals are *pre-whitened in parallel*, may be in a block-wise fashion (as done in LPC).
  - GCC is estimated by a very long DFT applied to whitened signals.
  - GCC is regularizes at (near-)zeroes of the spectra: however these points will be *missing data* for harmonic estimation!
- Harmonic modulation is highly disturbing for parametric delay estimation:
  - PHAT: use only the phase component of the non-zero sample GCC sequence.

AR Pre-whitener

$$x_1(m) = -\sum_{p=1}^{P} a_p x_1(m-p) + y_1(m)$$

$$x_2(m) = -\sum_{p=1}^{P} a_p x_2(m-p) + y_2(m)$$

$$\left\{\hat{a}_p\right\}_{LS} = \underset{\{a_p\}}{\arg\min}\left\{E\left[\left|y_1(m)\right|^2\right] + E\left[\left|y_2(m)\right|^2\right]\right\}$$

$$\left\{\hat{a}_p\right\}_{AML} = \underset{\{a_p\}}{\arg\min}\left\{\sqrt{E\left[\left|y_1(m)\right|^2\right] \cdot E\left[\left|y_2(m)\right|^2\right]}\right\}$$

Prediction errors = pre-whitened signals

TDOA estimation

- DFT applied to PHAT: very common and highly sensitive to low SNR regions.
- Weighted LS fitting in the AML spirit, applied to GCC: robust (bounded frequency sample impact), but apparently unknown in the community.
- AR regression applied to GCC, robust to disturbed (by delay spread) harmonics.
- MUSIC can often detect *multiple delays* even in reverberant fields, but TDOA pairing is difficult and require lots of sensor pairs (Di Claudio, Parisi, 2000).
- *Cepstral pre-processing* may improve PHAT estimation (Di Claudio, Parisi, 2003).

Final TDOA estimation

$$\hat{\rho}(k\Delta f) = \left|\hat{\rho}(k\Delta f)\right| e^{j\arg\left[\hat{\rho}(k\Delta f)\right]} \cong \left|\hat{\rho}(k\Delta f)\right| e^{j2\pi\left[\Delta f \tau(\mathbf{p})\right]k}$$

$$\Delta f \cdot \max\left[\left|\tau(\mathbf{p})\right|\right] < 0.5$$

$$\hat{\tau}(\mathbf{p})_{PHAT} = \arg\max_{\tau} \left| \sum_{k=k_1}^{k_2} e^{j\arg\left[\hat{\rho}(k\Delta f)\right]} e^{-j2\pi\Delta f \tau k} \right|^2$$

$$\hat{\tau}(\mathbf{p})_{AML} = \arg\max_{\tau} \left| \sum_{k=k_1}^{k_2} \left|\hat{\rho}(k\Delta f)\right| \hat{\rho}(k\Delta f) e^{-j2\pi\Delta f \tau k} \right|^2$$

TDOA pooling for arrays

- In 3-D localization, seismics (distorted wave-fronts!) and multi-sensor estimation, *several TDOAs are estimated by different sensor pairs*.
- If the TDOA pattern is *unique* for a certain location, it can be compared with the estimated TDOA pattern (Bienati, Spagnolini, 2001).
- TDOA sample errors by GCC or PHAT can be assumed asymptotically Gaussian with zero mean and variance inversely proportional to the SNR.
- Some gross error in TDOA estimation should be expected: use robust LS interpolators (e.g., Huber IRLS).

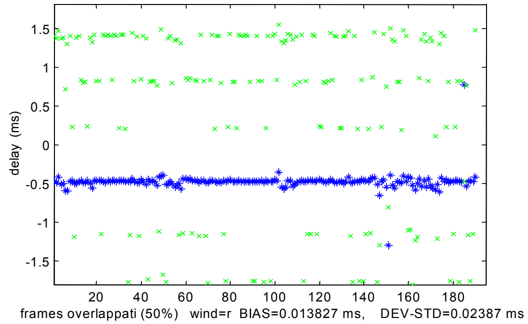
LS (CC) TDOA pattern matching

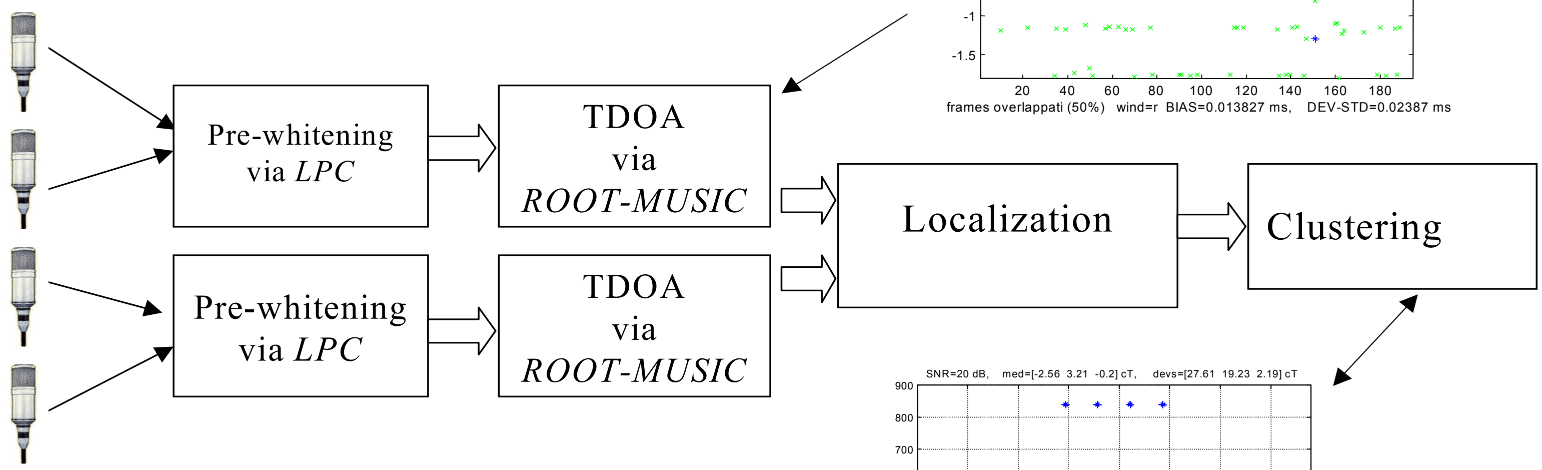
- Very expensive in seismics, but quite acceptable for ULAs, where multiples of a single delay are searched for.
- Balanced (LCC) type functionals for TDOA pattern matching may be preferable, since *errors are expected in both reference and measurement* sides.

$$\hat{\mathbf{p}} = \arg\max_{\mathbf{p}} \frac{\left| \sum_{(k,l)} w_{(k,l)}^{2} \hat{\tau}_{(k,l)} \tau_{(k,l)}(\mathbf{p}) \right|}{\left| \sum_{(k,l)} w_{(k,l)}^{2} \tau_{(k,l)}^{2}(\mathbf{p}) \right|^{\frac{1}{2}} \left| \sum_{(k,l)} w_{(k,l)}^{2} \hat{\tau}_{(k,l)}^{2} \right|^{\frac{1}{2}}}$$

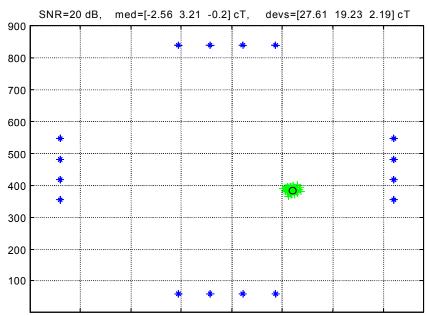
Multi-source localization by TDOA spatial clustering

Reverberated sources are not linearly separable by the GCC





Centroids of most dense cluster give the estimated speaker positions



TDOA actual performance

- The TDOA CRB in white noise is proportional to the energy of the temporal derivative of the sequence, but many error sources are present.
- The Bienati method approaches the single source CRB in a vast SNR range.
- While GCC based approaches demonstrated good results in *ad hoc* simulations and/or non-critical environments, they cannot effectively afford:
  - correlated noise (longer baseline needed, more ambiguity);
  - multiple sources and specular reflections;
  - sensor pair matching errors;
  - signal copy (low gain and minimal interference suppression).

New trends in TDOA estimation

- For moving targets, change rates of TDOA increase overall accuracy.
- Complex non-linear ML fitting have been tried (Estimate-Maximize procedures).
- Noise sensor inter-correlation and sensor matching remain unsolved issues.
- STCM-based models were historically preferred because of the similarity with their narrow-band counterparts, higher flexibility and rigour of the theoretical development.
- However *any wide-band array model is intrinsically approximated* to a certain degree.

STCM sinusoidal response

- Only a complex sinusoid has an *exact rank one signature* on the STCM and should be considered as the basis for further developments.
- The discrete time sinusoid frequency is tied to the continuous time one by a mapping determined by demodulation and sampling operations.
- The *space-time steering vector* (STSV) is a function of the continuous and discrete time frequencies and of location parameters.

SAPIENZA
UNIVERSITÀ DI ROMA

Space-time steering vector

$$s(t) = e^{j2\pi ft}; \quad s(m) = e^{j2\pi \upsilon m}; \quad \upsilon = \upsilon(f);$$

$$\underset{(NP\times 1)}{\mathbf{x}_{ST}(\mathbf{p}, f)} = \begin{bmatrix} \mathbf{a}(\mathbf{p}, f) \\ \mathbf{a}(\mathbf{p}, f) e^{j2\pi\upsilon} \\ \vdots \\ \mathbf{a}(\mathbf{p}, f) e^{j2\pi(P-1)\upsilon} \end{bmatrix} e^{j2\pi\upsilon m} = \mathbf{a}_{ST}(\mathbf{p}, f) e^{j2\pi\upsilon m}$$

$$\underset{(NP\times 1)}{\mathbf{a}_{ST}(\mathbf{p}, f)} = \begin{bmatrix} \mathbf{I}_N \\ \mathbf{I}_N e^{j2\pi\upsilon} \\ \vdots \\ \mathbf{I}_N e^{j2\pi(P-1)\upsilon} \end{bmatrix} \mathbf{a}(\mathbf{p}, f) \overset{def}{=} \mathbf{E}(\upsilon)\mathbf{a}(\mathbf{p}, f)$$

Narrow-band steering vector

Frequency subspace

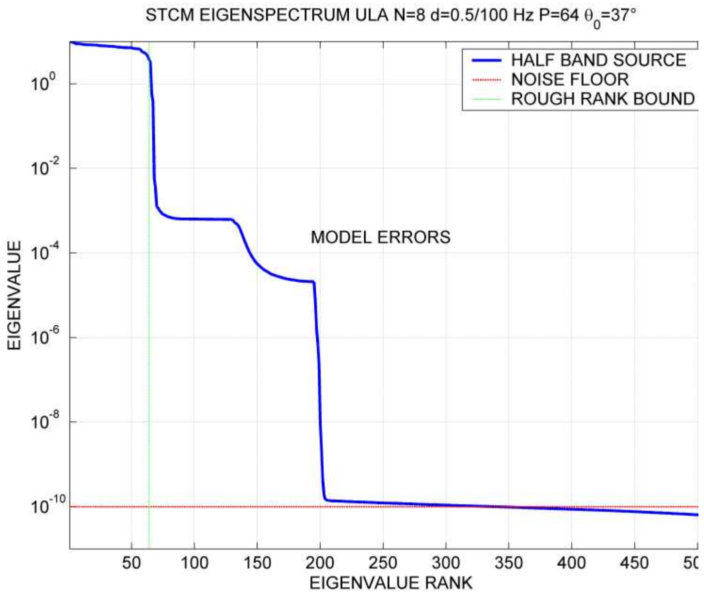
The frequency subspace

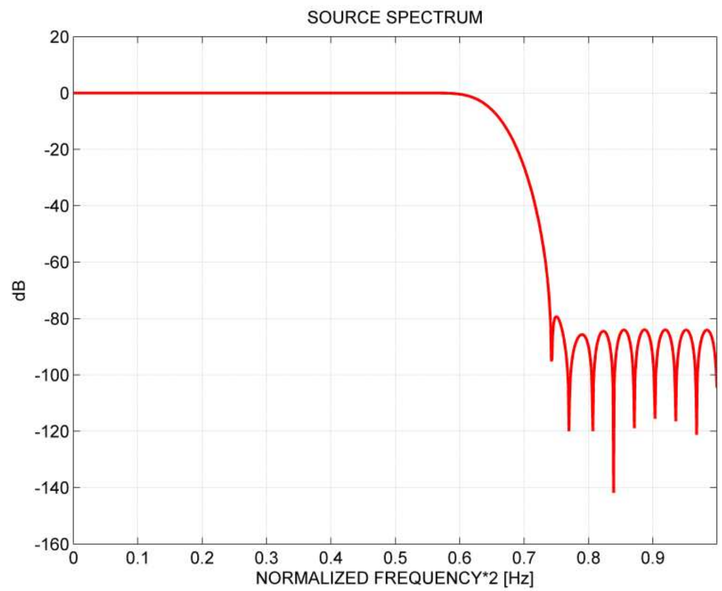
- $\mathbf{E}(\upsilon)$ is a basis for the STCM subspace *potentially spanned* by all the steering vectors at a given frequency, in absence of invisible space.
- Frequency subspaces and space time steering vectors obey the following relevant DFT-like orthogonality property.
- The narrow-band steering vector can be evidently recovered by applying a multi-channel (block) P-point DFT to the STSV.

$$\mathbf{E}^H\left(\frac{2\pi}{P}k\right)\mathbf{E}\left(\frac{2\pi}{P}q\right) = P\mathbf{I}_N\delta_{kq}$$

Signal subspace of a random source

- A stationary, random source has a multi rank signature on the STCM with rank bound equal or greater than P.
- Some signal eigenvalues can be really low.





STCM source subspace rank

- *No sharp rank bound*: the source cannot be completely characterized by P subband steering vectors of a DFT (closure problem)!
- *A well defined noise subspace still exists.*

$$\mathbf{R}_{ST} = \int_{-\pi}^{\pi} P_S(\upsilon)\mathbf{a}_{ST}(\mathbf{p},\upsilon)\mathbf{a}_{ST}^H(\mathbf{p},\upsilon)d\upsilon$$

$$\left[\begin{array}{cccc} \mathbf{a}_{ST}(\mathbf{p},0) & \mathbf{a}_{ST}\left(\mathbf{p},\dfrac{2\pi}{P}\right) & \cdots & \mathbf{a}_{ST}\left[\mathbf{p},\dfrac{2\pi(P-1)}{P}\right] \end{array}\right] =$$

$$\mathbf{a}_{ST}(\mathbf{p},\upsilon) - \mathbf{e}_{ST}(\mathbf{p},\upsilon)$$

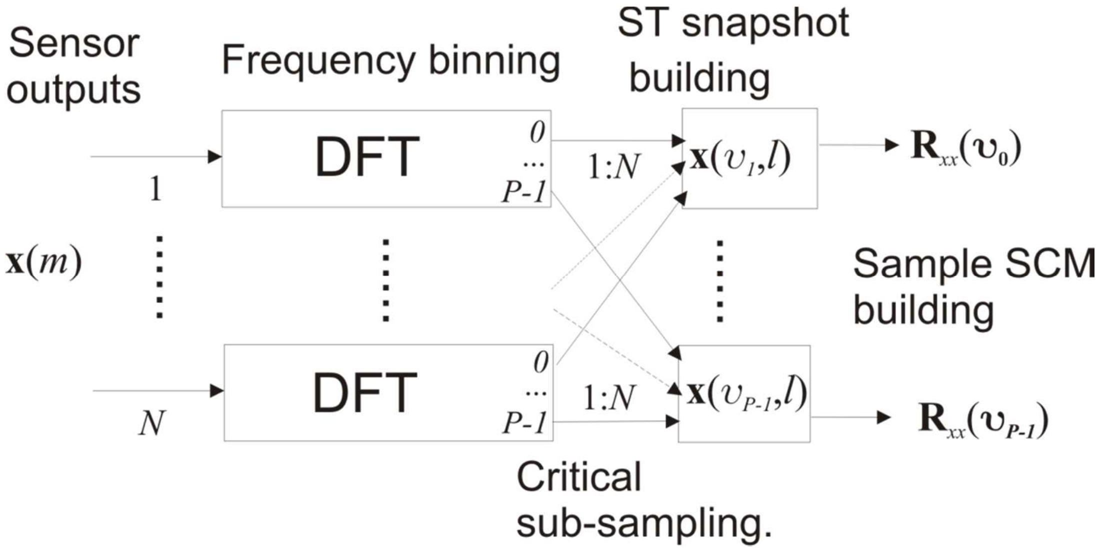
STCM based techniques

- Frequency domain *binning* approach:
  - A bank of *critically sampled* FIR filters (or a DFT) creates up to P subband signals from the space-time snapshots.
  - Each subband signal approximates the narrow-band model.
  - A SCM is estimated for each subband.
  - Almost Gaussian subband signals.
  - Difficult signal reconstruction.
  - Signal non- (ciclo-) stationarity is smoothed out.
  - Aliasing and spectral leakage.
  - High latency design for stationary signals.
- Time domain *delay and sum* approach:
  - Tapped delay filters or beamformers are applied to each sensor.
  - Intrinsically UWB.
  - Easy signal reconstruction, even not stationary.
  - Low latency.
  - High costs.
  - Difficult multi-source location estimate.

Welch DFT periodogram STCM estimate

- Estimates a *constrained frequency transformed, block diagonal STCM* from independent space-time snapshots.
- Assumes *stationary signals,* i.e.,enforcing *zero correlation among different frequencies*.
  - Divides each sensor output sequence into *non-overlapping* segments of length P , whose stacking gives origin to the ST snapshots.
  - Applies a (windowed) DFT to each segment, forming P narrowband frequency domain snapshots.
  - Estimate subband SCMs from narrowband snaphots.
- Rather *slow temporal convergence*.
- Plagued by spectral leakage (*bias*) and temporal aliasing.
- Asymptotically efficient w.r.t. observation time.

Welch binning data flow



$$\mathbf{x}(k,l)=\mathbf{x}\left(\upsilon_k=\frac{2\pi}{P}k,l\right)=\begin{bmatrix} X_1(k,l) \\ \vdots \\ X_N(k,l)\end{bmatrix};k=0,1,\ldots,P-1;l=1,2,\ldots,L$$

$$\hat{\mathbf{R}}_{xx}(k)=\hat{\mathbf{R}}_{xx}(\upsilon_k)=\frac{1}{L}\sum_{l=1}^{L}\mathbf{x}(k,l)\mathbf{x}^H(k,l);\quad \hat{\mathbf{R}}_{ST}^{(Welch)}=\begin{bmatrix} \hat{\mathbf{R}}_{xx}(0) & \mathbf{0} & \cdots \\ \mathbf{0} & \ddots & \vdots \\ \vdots & \cdots & \hat{\mathbf{R}}_{xx}(P-1)\end{bmatrix}$$

Temporal correlation based STCM estimate

- Separately computes cross- and auto-correlation sequences for each sensor pair and put them in the sample STCM.
  - Biased temporal correlation estimator gives positive semidefinite STCM estimates (in the absence of numerical errors).
  - Can employ FFT for fast aperiodic correlation.
  - Costly memory shuffling.
  - Minimal degrees of freedom (*Toeplitz blocks*) for maximum stability.
  - Long triangular lag window;
  - *Lower bias* than Welch periodogram in STCM for typical STCMs.
  - Fast statistical convergence for $M>NP$.
  - Optional temporal window for non-stationary signal tapering.

$$\hat{R}_{xy}(p) = \frac{1}{M}\sum_{m=0}^{M-p} x(m)x^*(m+p); \quad \hat{R}_{xy}(p) = \hat{R}_{xy}^*(-p)$$

Covariance type STCM estimate

- Averages ST snapshots or builds intermediate Toepliz/Hankel data matrices.
- Positive definite, *unbiased* STCM estimates.
- Fast SVD computation by conjugate gradient *Lanczos type algorithms* with low condition number,
- Employs fast aperiodic convolution algorithms.
- Optional temporal *snapshot tapering* window.
- Huge memory requirements.

$$\hat{\mathbf{R}}_{ST} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{x}_{ST}(m)\mathbf{x}_{ST}^{H}(m) = \frac{1}{M}\mathbf{X}_{ST}\mathbf{X}_{ST}^{H}$$

$$\mathbf{X}_{ST} = \left[\mathbf{x}_{ST}(1) \quad \cdots \quad \mathbf{x}_{ST}(M)\right]$$

Links between subband SCM estimates

- Welch periodogram subband SCM estimate is a Bartlett type estimator applied to the STCM.
- Capon MVDR SCM estimate is possible (Krolik):
  - Spectral *leakage suppression*.
  - Fast convergence.
  - *Signal cancellation* issues for non- (or cyclo-) stationary signals.
  - Background noise SCM *distortion*.

$$\hat{\mathbf{R}}^{(Bartlett)}_{(N\times N)}(k) = \mathbf{E}^H(\upsilon_k)\hat{\mathbf{R}}_{ST}\mathbf{E}^H(\upsilon_k)$$

$$\hat{\mathbf{R}}^{(MVDR)}_{(N\times N)}(k) = \left[\mathbf{E}^H(\upsilon_k)\hat{\mathbf{R}}_{ST}^{-1}\mathbf{E}^H(\upsilon_k)\right]^{-1}$$

Work in progress

- Use 2-D *Hermite-Laguerre expansion* of the array response, which turns out to be a sophisticated beamspace transformation of the STCM for ULAs.
  - More accurate response modeling.
  - Fast convergence (low degrees of freedom).
  - Finite rank signal subspaces.
- Optimal rank detection of the sample STCM.
- Understanding observed loss of consistence of Bartlett and Velch SCM estimators from STCM.

Array design issues

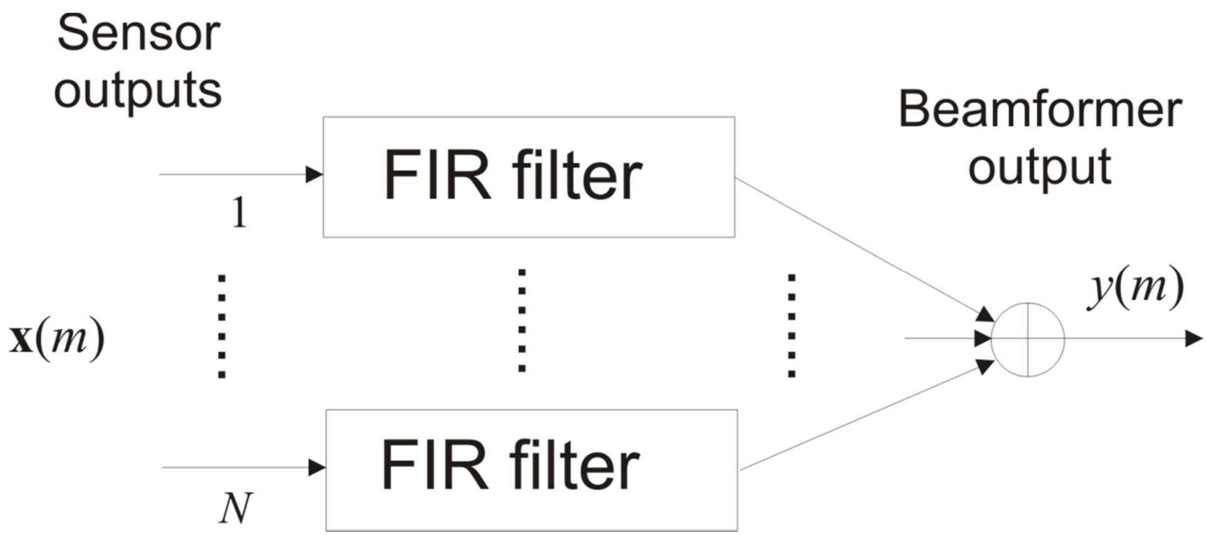
- A single array cannot efficiently cover a bandwidth larger than about *one octave*:
  - Small aperture and high mutual coupling at low frequencies.
  - Excessive spacing and ambiguity at high frequencies.
- *Telescopic arrays with scaled, nested subarrays* are required for demanding UWB applications (sonar, audio, ESM).
  - Multi-ring circular arrays;
  - Irregularly spaced arrays.
  - Pruned arrays.
  - Interpolated arrays.
  - Difficult signal copy from sparse subarrays.

Wideband beamforming on the STCM

- Mainly used in sonar, radar, seismic inversion and ultrasound.
- *Multi-channel convolution* applied to the space-time snapshots in time or frequency domain.
- Array model heavily changes with frequency.
- Huge data: heavy compromises between computational power and global efficacy.
- *Huge number of free parameters*: constrained architectures are searched.
- Delay and sum (adaptive) beamforming useful for wide-band signal copy.

Delay and sum beamformer

- Tapped delay lines (FIR filters) connected to each sensor.
- Adaptive LCMV possible.
- Distortion-less response constraint inverts array response to a desired response hd.
- Typically operates with two to four time oversampling.



Delay-and-sum MVDR beamformer

- Noise and independent interference collected in an additive vector $\mathbf{v}$.
- DR constraint plus optional ones.

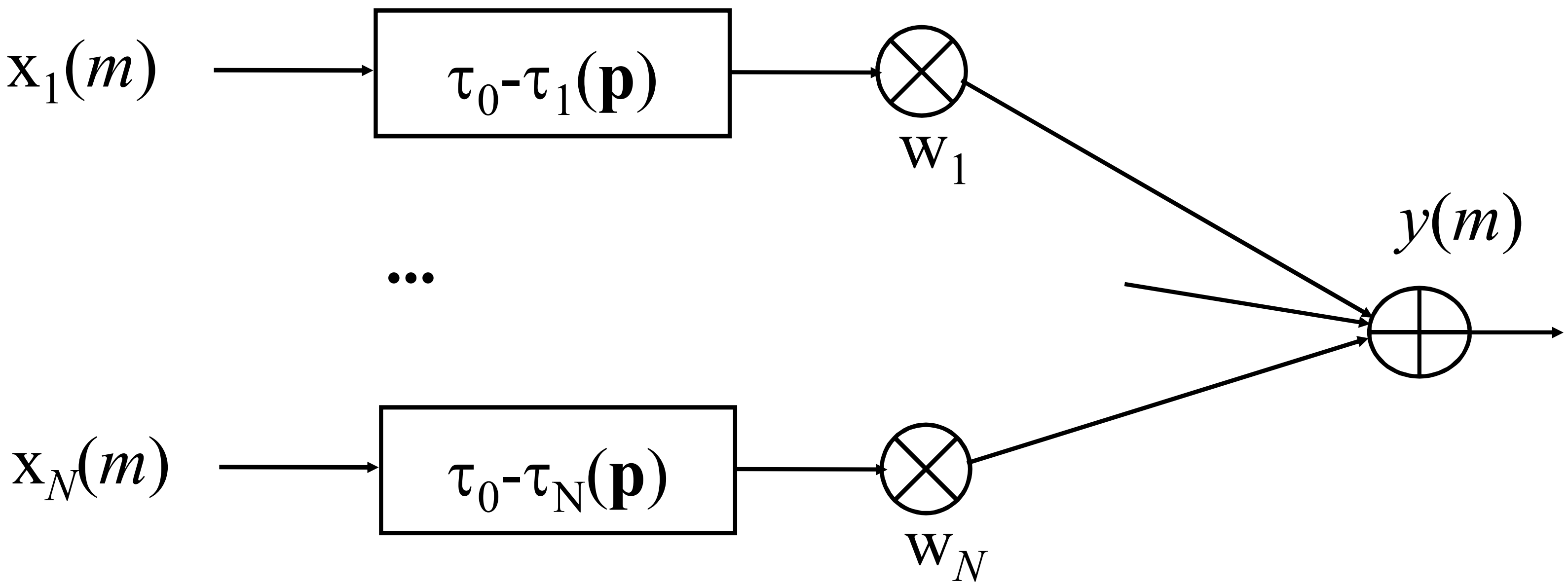
$$\mathbf{x}(m) = \sum_{p=0}^{Q} \mathbf{h}_p s(m-p) + \mathbf{v}(m); \quad y(m) = \sum_{q=0}^{P \gg Q} \mathbf{w}_q^H \mathbf{x}(m-q)$$

$$\underset{(NP\times1)}{\mathbf{w}} = \begin{bmatrix} \mathbf{w}_0 \\ \vdots \\ \mathbf{w}_P \end{bmatrix}; \quad \underset{(NQ\times1)}{\mathbf{h}} = \begin{bmatrix} \mathbf{h}_0 \\ \vdots \\ \mathbf{h}_Q \end{bmatrix}; \quad \mathbf{w}^H \begin{bmatrix} \underset{(NQ\times1)}{\mathbf{h}} & \underset{(N\times1)}{\mathbf{0}} & \cdots & \underset{(N\times1)}{\mathbf{0}} \\ \underset{(N\times1)}{\mathbf{0}} & \underset{(NQ\times1)}{\mathbf{h}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \underset{(N\times1)}{\mathbf{0}} \\ \underset{(N\times1)}{\mathbf{0}} & \cdots & \underset{(N\times1)}{\mathbf{0}} & \underset{(NQ\times1)}{\mathbf{h}} \end{bmatrix} = \mathbf{w}^H \mathbf{H} = \underset{(P-Q+1)\times1}{\mathbf{h}_d}$$

$$\hat{\mathbf{w}}_{MVDR} = \arg\min_{\mathbf{w}} \left\{ \mathbf{w}^H \mathbf{R}_{ST} \mathbf{w} \right\} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{H} = \mathbf{h}_d$$

Steered delay and sum beamformer

- Fractional delays align in time all the components of the signal coming from the direction of interest.
- Short (narrow-band type) weight vector cancels interferences while preserving desired signal.



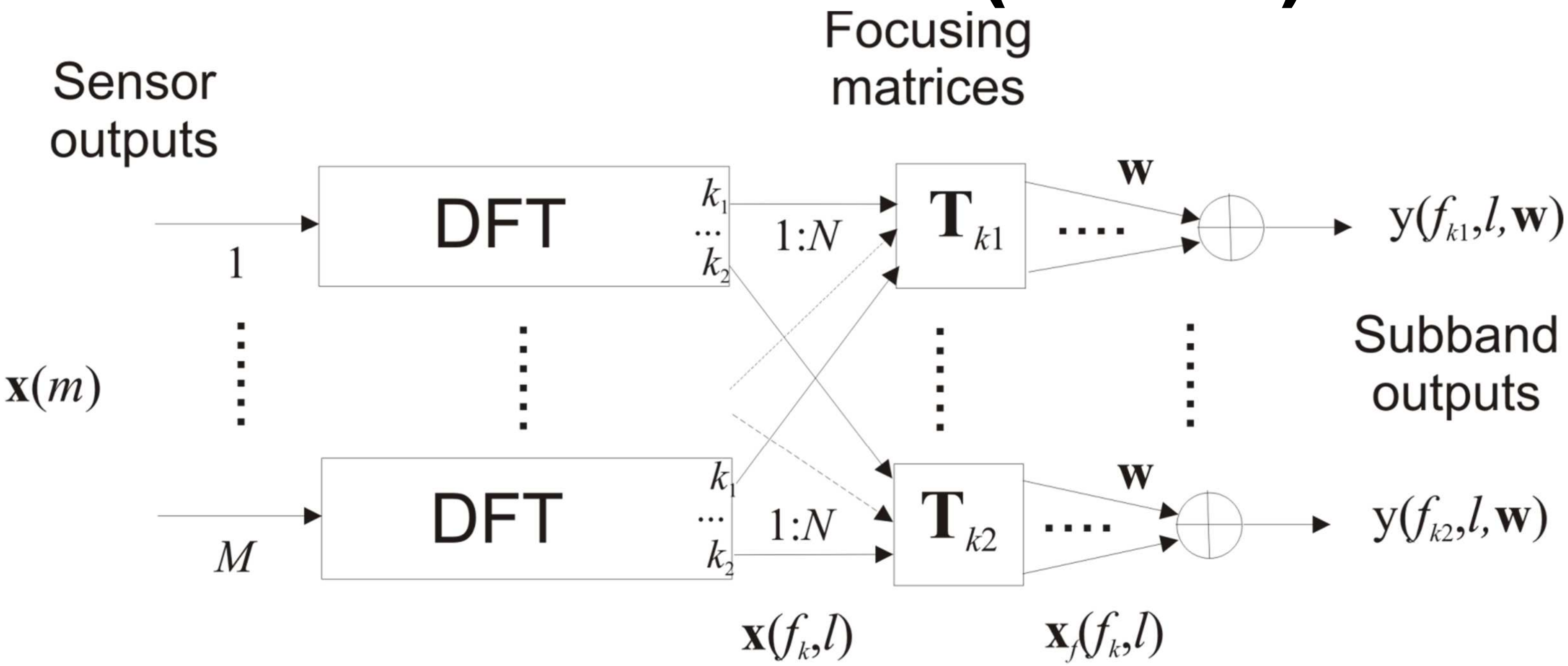
Focused beamformers

- Steered beamformers, preferred in radio-astronomy, sonar and ultrasound, have the *minimum number* of free parameters among wide-band beamformers.
- Interfering sources not well compressed in rank with *loss of degrees of freedom*.
- Alternate formulation of steered beamforming in the frequency domain after binning.
  - Realign all the source steering vectors at bin frequencies onto the corresponding steering vectors of a virtual array at least *within a sufficiently wide angular sector* centered on the direction of interest.
  - Linear transformation (*focusing matrices*) of bin outputs.
  - Unitary transformations preserve *orthogonality between constraint and blocking subspaces* for adaptive beamforming.

Focusing matrices

- Orthogonal (Procrustes) focusing matrices
  - work only on the union of small angular sectors, one or two beamwidths wide;
  - rather high focusing errors;
  - careful scaling of amplitudes and phases of steering vectors is needed to avoid signal distortion;
  - no off-sector spatial filtering capability.
- LS and equiripple interpolation:
  - spatial filtering of off-sector sources;
  - *near singular* focusing matrices for narrow sectors;
  - wide angular sectors possible but with rather lage errors;
  - *virtual narrow-band array manifold must be within the span of the (harmonic) modal decomposition of the manifolds of all frequencies*;
  - reduced virtual array size often required to full-fill this requirement;
  - no directional ambiguity is allowed!

Wide-band steered (focused) beamformer (STBF)



$$\mathbf{T}_k\mathbf{a}\left(f_k,\mathbf{p},\mathbf{c}\right)\cong\mathbf{a}_0\left(\mathbf{p},\mathbf{c}\right);$$

$$y(k,l,\mathbf{w})=\mathbf{w}^H\mathbf{x}_f\left(f_k,l\right);\quad \mathbf{x}_f\left(f_k,l\right)=\mathbf{T}_k\mathbf{x}\left(f_k,l\right)\cong\mathbf{a}_0\left(\mathbf{p},\mathbf{c}\right)s\left(f_k,l\right);$$

$$l=1,2,\ldots,L;\quad k=k_1,k_1+1,\ldots,k_2$$

Procrustes unitary focusing

- Based on the SVD and the orthogonal polar factorization of matrices.
- i.i.d. steering vectors errors would induce bias and can be consistently corrected in a LS sense by a proper PCA.

$$\underset{(N\times Q\geq N)}{\mathbf{A}_0} = \left[\mathbf{a}_o(\mathbf{p}_1) \quad \cdots \quad \mathbf{a}_o(\mathbf{p}_Q) \right]; \quad \underset{(N\times Q)}{\mathbf{A}_j} = \left[\mathbf{a}_j(\mathbf{p}_1, f_j) \quad \cdots \quad \mathbf{a}_j(\mathbf{p}_Q, f_j) \right];$$

$$\hat{\mathbf{A}}_j = \mathbf{A}_j + \mathbf{E}_j; \quad E\left[\mathbf{E}_j \mathbf{E}_j^H \right] = \lambda_v \mathbf{I}_N; \quad \mathbf{T}_j = \underset{\mathbf{T}^H\mathbf{T}=\mathbf{I}_N}{\arg\min} \left| \mathbf{A}_0 - \mathbf{T}\hat{\mathbf{A}}_j \right|_F$$

$$\hat{\mathbf{A}}_j \overset{SVD}{=} \mathbf{U}_j \boldsymbol{\Sigma}_j \mathbf{V}_j^H; \quad \mathbf{A}_0 \mathbf{V}_j \left(\boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_j^T - \lambda_v \mathbf{I}_N \right)^{\frac{1}{2}} \mathbf{U}_j^H \overset{SVD}{=} \mathbf{Q}_{1j} \boldsymbol{\Gamma}_j \mathbf{Q}_{2,j}^H$$

$$\mathbf{T}_j = \mathbf{Q}_{1j} \mathbf{Q}_{2,j}^H$$

Remarks on unitary focusing

- Possible only within the union of small angular sectors (< one beamwidth for each sector across its central direction).
- Good academic simulation results based on heuristic (magic) angle selection and equation weighting schemes.
- Non-convergence of iterated focusing.
- It is essential to match normalized and phase centered steering vectors:
  - Unitary matrices cannot change vector L2 norm. Steering vector normalization prevents irreducible fitting errors.
  - Phase centering minimizes phase rotations within the sector and the overall fitting error.

SAPIENZA
UNIVERSITÀ DI ROMA

Adaptive STBF

- Approximately common signal model for all frequencies after preliminary *focusing.*
- *Single weight vector of length* N for all frequencies.
- Refined *quiescent vector in LCMV structure* against multipath.
- Capability of *isolating*, *combining* or *suppressing* reverberation modes on the basis of their relative group delays.
- ML-STBF is almost *independent of the signal spectrum* and intrinsically robust to model and finite sample errors.
- Targeted to seismic, underwater, ultrasound and audio processing.

MV-STBF

- Suppresses multipath delayed more than the *source only* correlation time.
- Averages sample focused subband matrices.

$$\mathbf{x}_{(F)}\left(f_k,l\right)=\mathbf{T}_k\mathbf{x}\left(f_k,l\right)$$

$$\mathbf{R}_{xx}^{(F)}\left(k\right)=E\left[\mathbf{x}_{(F)}\left(f_k,l\right)\mathbf{x}_{(F)}^H\left(f_k,l\right)\right]=\mathbf{T}_k\mathbf{R}_{xx}\left(k\right)\mathbf{T}_k^H$$

$$\hat{\mathbf{R}}_{xx}^{(F)}\left(k\right)=\frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_{(F)}\left(f_k,l\right)\mathbf{x}_{(F)}^H\left(f_k,l\right)$$

$$\tilde{\mathbf{w}}_{MV}=\arg\min_{\mathbf{w}}\left\{\mathbf{w}^H\left[\frac{1}{K}\sum_k\hat{\mathbf{R}}_{xx}^{(F)}\left(k\right)\right]\mathbf{w}\right\}\quad s.t.\quad\begin{cases}\mathbf{C}^H\mathbf{w}=\mathbf{d}\\ \mathbf{C}^H\mathbf{w}_0=\mathbf{d}\\ \left|\mathbf{Q}^H\left(\mathbf{w}-\mathbf{w}_0\right)\right|^2\le\gamma^2\end{cases}$$

Stochastic ML-STBF

- Output *subband signals* are assumed as zero mean, circular, Gaussian and mutually independent w.r.t. time and frequency.
- Fast and safe modified Newton algorithm for training.
- Built-in linear and quadratic robustness constraints.
- Implicit time domain de-correlation of multipath and focusing errors delayed beyond *one sampling period*!

$$\tilde{\mathbf{w}}_{ML} = \arg\min_{\mathbf{w}} \left\{ \sum_k \ln\left[\mathbf{w}^H \hat{\mathbf{R}}_{xx}^{(F)}(k)\mathbf{w} + C_k \mathbf{w}^H \mathbf{w} \right] \right\} \quad s.t. \quad \begin{cases} \mathbf{C}^H \mathbf{w} = \mathbf{d} \\ \mathbf{C}^H \mathbf{w}_0 = \mathbf{d} \\ \left| \mathbf{Q}^H (\mathbf{w} - \mathbf{w}_0) \right|^2 \le \gamma^2 \\ C_k \ge 0 \end{cases}$$

Choice of the adaptive STBF quiescent vector

- *Focused subband steering vectors are not generally aligned on the same direction* and do not have the same magnitude.
- The optimal quiescent vector (i.e., the effective focused array response) departs from the (scaled) steering vector of the virtual array at the pointing direction.
- An optimal quiescent vector should be within the ball of focused steering vectors.
- Different optimality criteria can be established!
- All work much better than the standard one…

Wide-band quiescent vectors

- The ML-STBF theory allows the development of a set of powerful, spectrum insensitive wideband *matched-field* quiescent vectors

$$\mathbf{w}_{0\,PB-LS}(\mathbf{p},\mathbf{c}) \propto \arg\min_{\mathbf{x},\{\theta_k\}} \left\{ \sum_k \left| \mathbf{Q}\left[\mathbf{T}_k \mathbf{a}(f_k,\mathbf{p},\mathbf{c}) - \mathbf{x}e^{j\theta_k} \right] \right|^2 \right\}$$

$$\mathbf{w}_{0\,WAVES}(\mathbf{p},\mathbf{c}) \propto \arg\max_{\mathbf{x}} \left\{ \sum_k \frac{p_k \left| \mathbf{x}^H \mathbf{T}_k \mathbf{a}(f_k,\mathbf{p},\mathbf{c}) \right|^2}{\left| \mathbf{x} \right|^2} \right\}$$

$$\mathbf{w}_{0\,ML}(\mathbf{p},\mathbf{c}) \propto \arg\max_{\mathbf{x}} \left\{ \sum_k \frac{\left| \mathbf{x}^H \mathbf{T}_k \mathbf{a}(f_k,\mathbf{p},\mathbf{c}) \right|^2}{\left| \mathbf{x}^H \mathbf{T}_k \mathbf{a}(f_k,\mathbf{p},\mathbf{c}) \right|^2 + \sigma_k^2 \left| \mathbf{x} \right|^2} \right\}$$

Sapienza Università di Roma

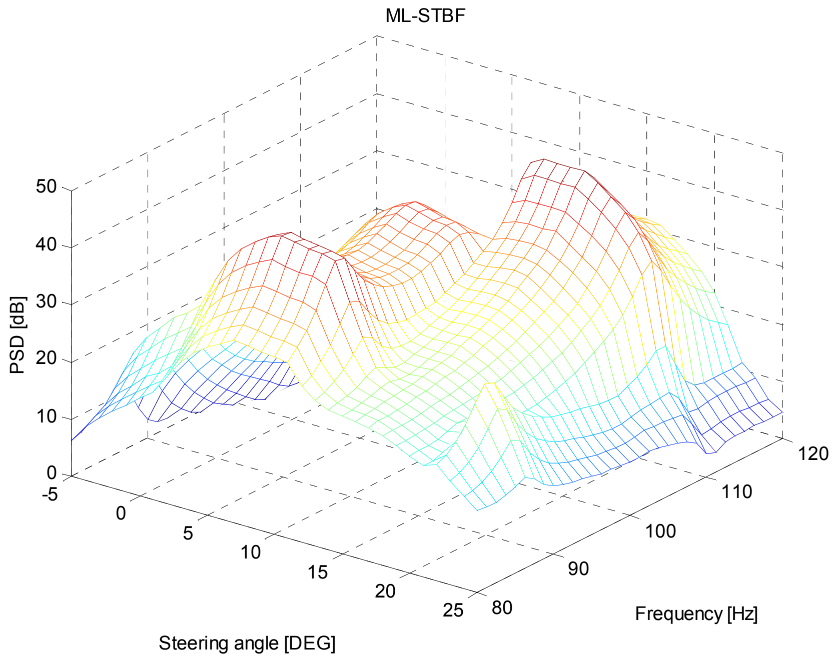
Quiescent vector choice

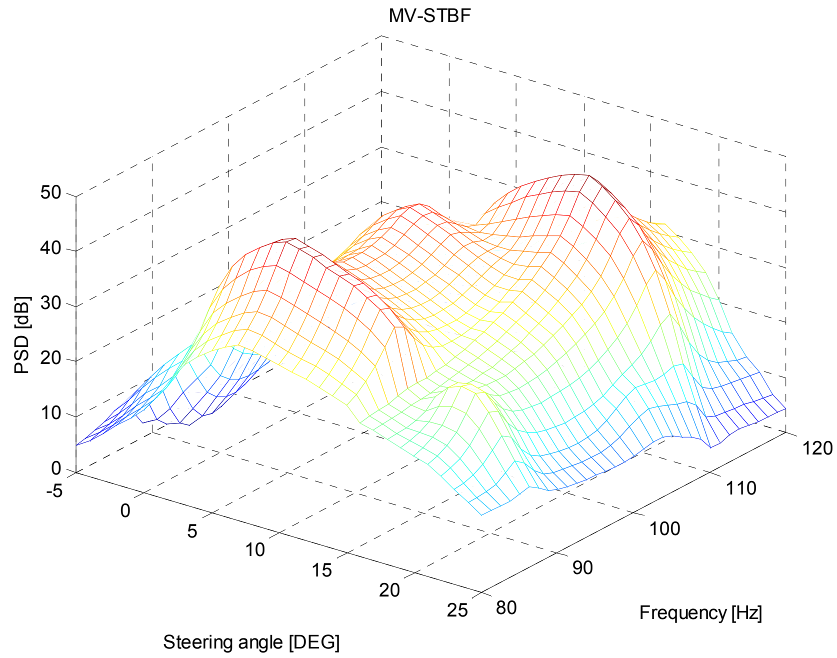
- ML vector preferable for applications to reverberant fields (spectrum estimation, imaging) since optimizes log-spectrum (real cepstrum) response.
- WAVES vector useful only for moderate reverberation level.
- PB-LS vector very robust under multipath.
- Further 1-D phase and magnitude control of the frequency response is required for correct signal extraction.

SAPIENZA
UNIVERSITÀ DI ROMA

Comparison between the ML-STBF and the robust MV-STBF

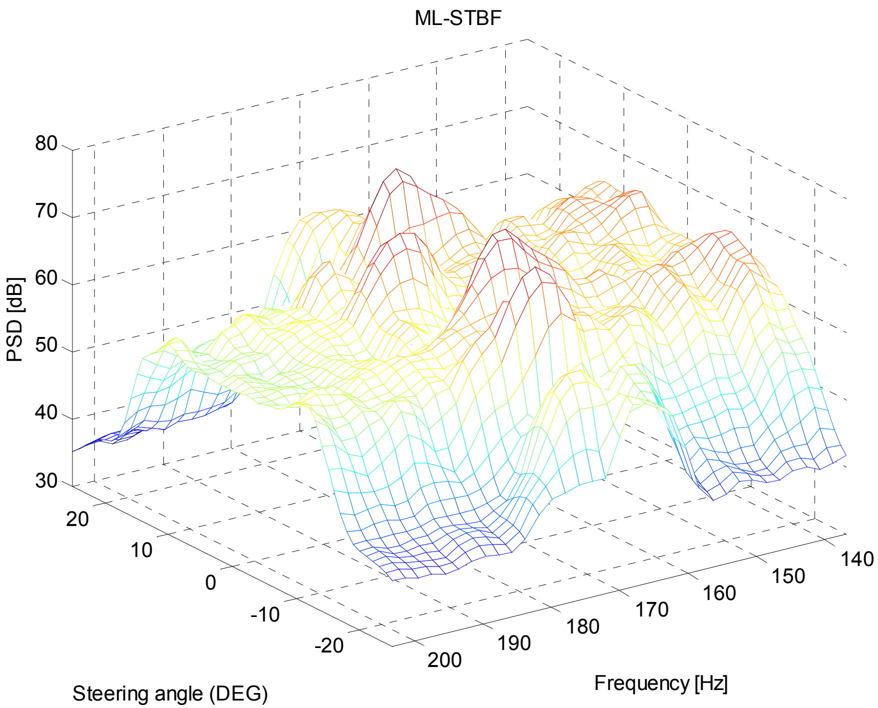
- Two coloured, uncorrelated far-field sources at 7 and 15 degrees from broadside
- ULA 10 sensors, SNR 20 dB, 80-120 Hz, L=100

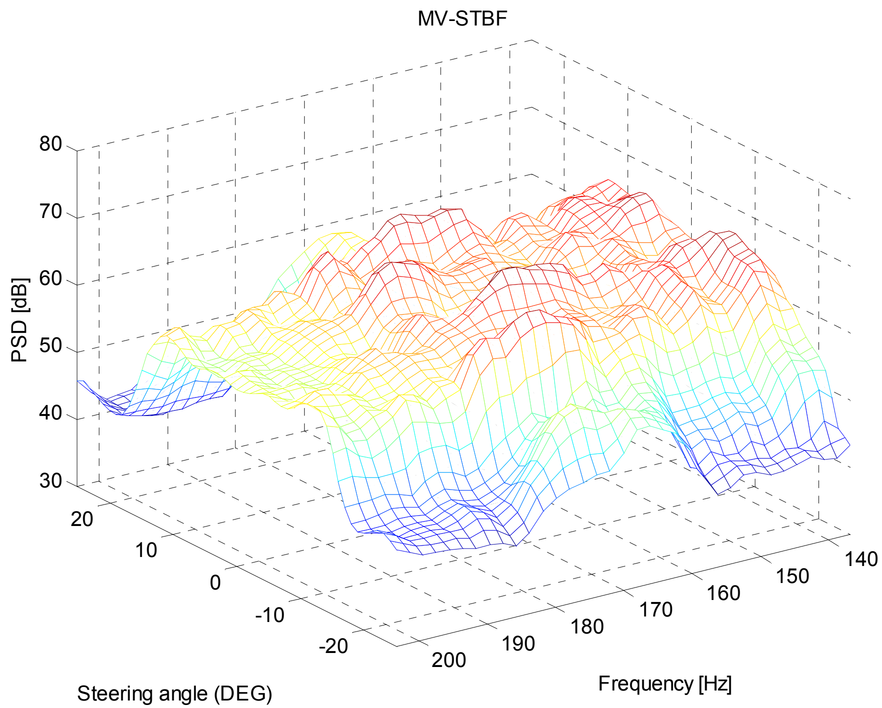




Mediterranean Vertical Array Data - wavenumber scan

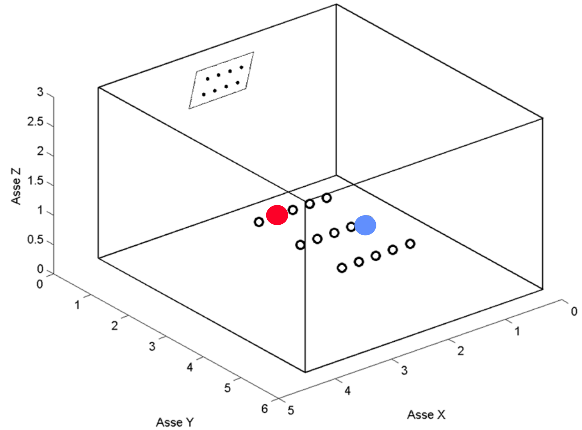
- Vertical ULA, 48 sensors moored in shallow water near the isle of Elba (SACLANT), WAVES vector, narrowband source at 160 Hz

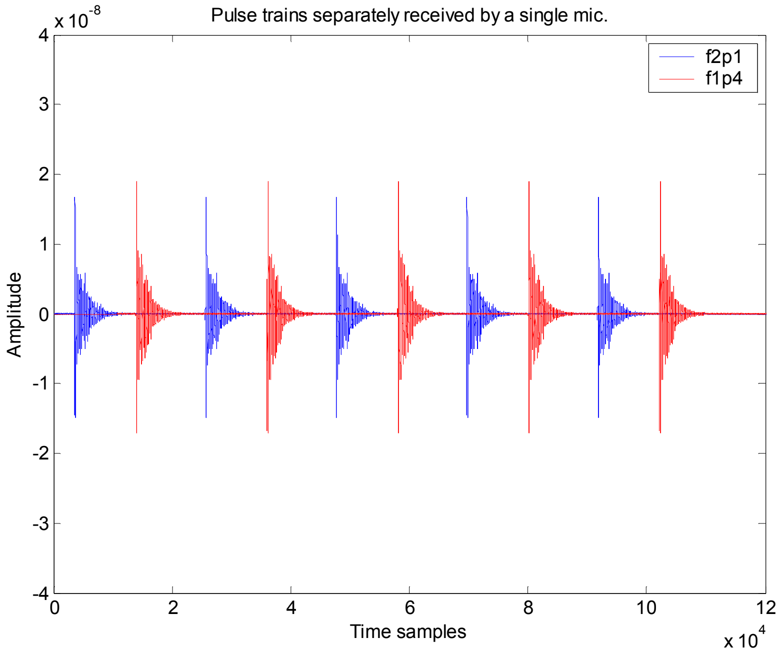


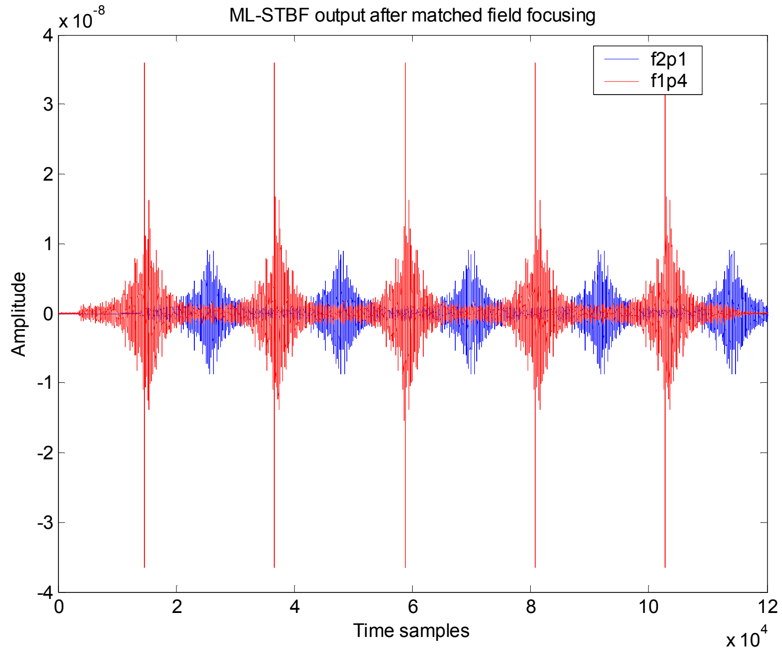


Acoustic source separation

- INFOCOM array (8 microphones)
- Bandwidth 400-1150 Hz
- ML-STBF + ML quiescent vector
- Unitary matched-field focusing

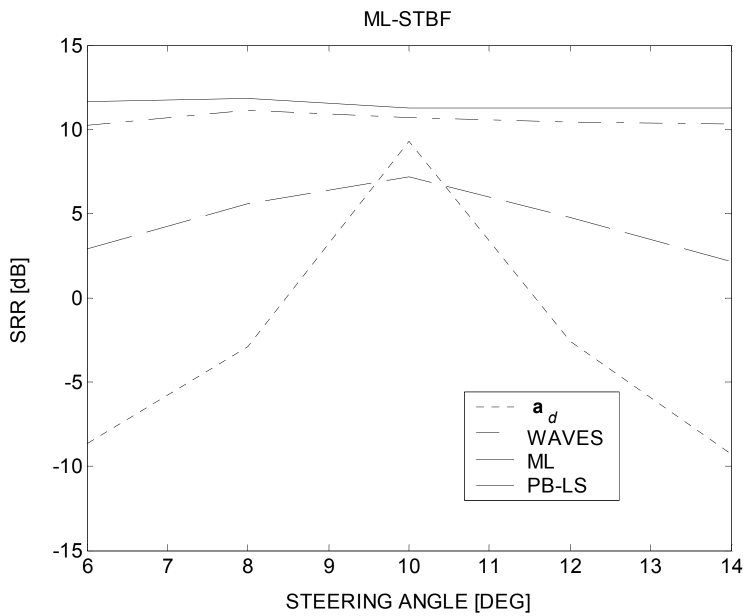


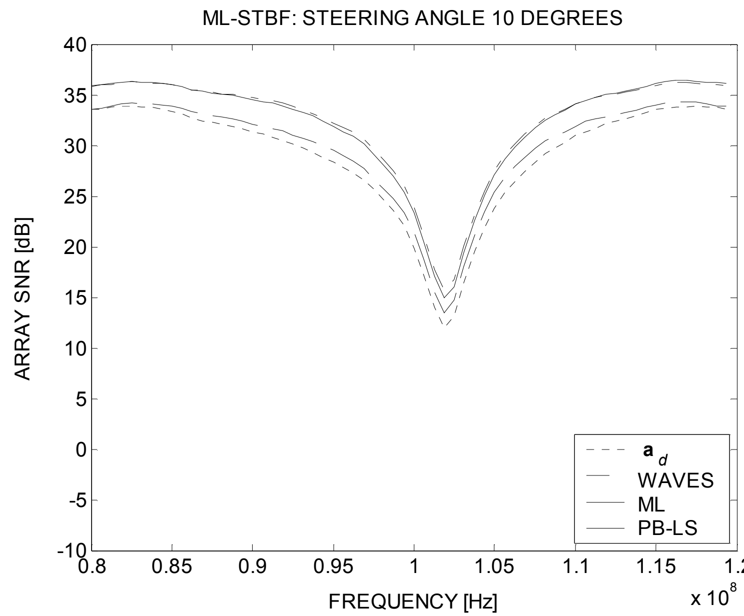




Random scattering

- Optimal quiescent vectors can largely improve output SNR and multipath suppression/combination with ML-STBF and MV-STBF.
- ULA 25 sensors, 0.8-1.2 GHz, L=100, 5% RMS array errors, 10 randomly located scatterers + far-field AR interference.





Wide-band parametric localization techniques

- Incoherent techniques
  - Separate localization within each subband with beamforming, MUSIC, ML, WSF, etc.…
  - Clustering of location estimates in frequency.
  - Low SNR clustering breakdown.
  - Low independent sample number.
  - No focusing issues.
  - High number of narrow-band targets.
- Coherent techniques
  - Single basic statistic for the entire bandwidth (STCM, subband SCM set, etc…).
  - Overall wide-band functional *not completely consistent* (focusing, spectral leakage, pathological signals with few modes).
  - Bias and excess estimation variance at high SNR.
  - Low SNR thresholds.
  - Number of detectable sources lower than the sensor number.

Coherent wide-band estimators

- Often assume independent bin information.
- Approximate ML or WSF estimates
  - *High statistical efficiency* for all scenarios.
  - *Non-robust* to mis-modeling at high SNR.
  - *Uncertain convergence* to the true parameters.
  - *Need bootstrapping* by suboptimal estimators.
  - Lower SNR threshold limited by bootstrapping.
- Focusing techniques
  - Start from *focused SCMs* or STCM.
  - Performance *limited by focusing errors* at high SNR.
  - *Near-optimal performance at low SNR* with final WSF/MODE type estimators and full beamspace width.
  - Virtual array choice very important for results (ULA or canonical harmonic decomposition are the preferred manifolds for focusing).
  - Reduced Fisher information for reduced size beamspaces.

Wide-band coherent WSF and Approximated ML estimators

- Assume the availability of *independent subband SCM estimates* and the proper Gaussian (elliptical) statistical model for data.
- Independency means *non-overlapping bin filter* responses (insufficient for signal copy).
- Approximation mainly comes from:
  - ignoring correlations among bin;
  - finite bin bandwith;
  - *non-closed subspace fitting with subband steering vectors* measured at the central bin frequency;
  - spectral leakage and aliasing (non-white, non-stationary, statistically dependent noise background).

Wide-band WSF and AML

- Given in advance:
  - (whitened) SCM estimates;
  - number of wide-band sources;
  - array manifold for all directions and frequency of interest;
  - good initial location guesses within a fraction of beamwidth.

$$\left[\hat{\mathbf{p}}_{AML},\left\{\hat{\lambda}_{v,k}\right\}\right]=\underset{\mathbf{p},\left\{\lambda_{v,k}\right\}}{\arg\min}\,\mathrm{trace}\left\{\sum_k \lambda_{v,k}^{-1}\left[\mathbf{I}_N-\mathbf{A}\left(\mathbf{p},f_k\right)\mathbf{A}^{\dagger}\left(\mathbf{p},f_k\right)\right]\hat{\mathbf{R}}_{xx}\left(k\right)\right\}$$

$$\left[\hat{\mathbf{p}}_{UML},\left\{\hat{\mathbf{P}}_k\right\},\left\{\hat{\lambda}_{v,k}\right\}\right]=\underset{\mathbf{p},\left\{\lambda_{v,k}\right\}}{\arg\min}\,\mathrm{trace}\left\{\sum_k \hat{\mathbf{R}}_{xx}\left(k\right)\left[\mathbf{A}\left(\mathbf{p},f_k\right)\mathbf{P}_k\mathbf{A}^H\left(\mathbf{p},f_k\right)+\lambda_{v,k}\mathbf{I}_N\right]^{-1}\right\}$$

$$\hat{\mathbf{R}}_{xx}\left(k\right)\cong\hat{\mathbf{E}}_S\left(k\right)\hat{\mathbf{\Lambda}}_S\left(k\right)\hat{\mathbf{E}}_S^H\left(k\right)+\hat{\lambda}_{v,k}\mathbf{I}_N$$

$$\left[\hat{\mathbf{p}}_{WSF},\left\{\hat{\mathbf{C}}_k\right\},\left\{\hat{\lambda}_{v,k}\right\}\right]=\underset{\mathbf{p},\left\{\mathbf{C}_k\right\}}{\arg\min}\left\{\sum_k\left|\mathbf{A}\left(\mathbf{p},f_k\right)\mathbf{C}_k-\hat{\mathbf{E}}_S\left(k\right)\mathbf{W}\left(k\right)\right|_F^2\right\}$$

SAPIENZA
UNIVERSITÀ DI ROMA

Wide-band coherent focusing, CSSM (Wang-Kaveh 1985)

- Focused SCMs are averaged together to form an *universal covariance matrix* USCM).
- A narrow-band location estimator is applied to the USCM.
- Approximates a conditional (A)ML estimator.

$$\mathbf{T}_k \mathbf{A}\left(\mathbf{p}, f_k\right) \cong \mathbf{A}_0\left(\mathbf{p}\right);$$

$$\mathbf{R}_{xx}^{F}\left(k\right) = \mathbf{T}_k \mathbf{R}_{xx}\left(k\right) \mathbf{T}_k^H \cong$$

$$\cong \mathbf{A}_0\left(\mathbf{p}\right) \mathbf{P}_k \mathbf{A}_0^H\left(\mathbf{p}\right) + \lambda_{v,k} \mathbf{R}_{vv}^{F}(k);$$

$$\mathbf{R}_{xx}^{0} = \left(\sum_k \alpha_k\right)^{-1} \left[\sum_k \alpha_k \mathbf{R}_{xx}^{F}\left(k\right)\right] \cong \mathbf{A}_0\left(\mathbf{p}\right) \mathbf{P}_0 \mathbf{A}_0^H\left(\mathbf{p}\right) + \lambda_0 \mathbf{R}_{vv}^{0}$$

WAVES subspace focusing (Di Claudio, Parisi, 2001)

- CSSM focusing errors impair consistency of estimates;
- Focusing *only weighted signal subspaces* allows to control the statistical impact of sample (corrupted) subspaces.
- Subband sample signal eigenvectors play the role of *independent observations* within a *pseudo-data matrix (Weighted AVErage of Signal Subspaces, WAVES).*

$$\hat{\mathbf{Z}} \overset{def}{=} \left[\alpha_{k_1} \mathbf{T}(k_1) \hat{\mathbf{E}}_s(k_1) \mathbf{W}(k_1) \quad \cdots \quad \alpha_{k_J} \mathbf{T}(k_J) \hat{\mathbf{E}}_s(k_J) \mathbf{W}(k_J) \right]$$

$$\hat{\mathbf{Z}} \overset{w.p.1}{\rightarrow} \mathbf{A}_o(\mathbf{p}) \mathbf{C}_0$$

WAVES location estimator

- A USCM is generated from WAVES and used for narrow-band parametric estimation.
- The optimal, universal noise SCM is the same as in CSSM at least in white noise.
- The optimal subspace weighting is the same as the WSF one.
- WAVES *approximates a coherent wide-band WSF* or UML.

$$\left(\sum_{k=k_1}^{k_J}\alpha_k\right)^{-1}\hat{\mathbf{Z}}\hat{\mathbf{Z}}^H \simeq \mathbf{A}_o(\mathbf{p})\mathbf{P}_0\mathbf{A}_0^H(\mathbf{p})+\lambda_o\mathbf{R}_{vv}^0;\quad \lambda_o=0+O\left(N^{-1}\right)$$

WAVES optional robust weighting

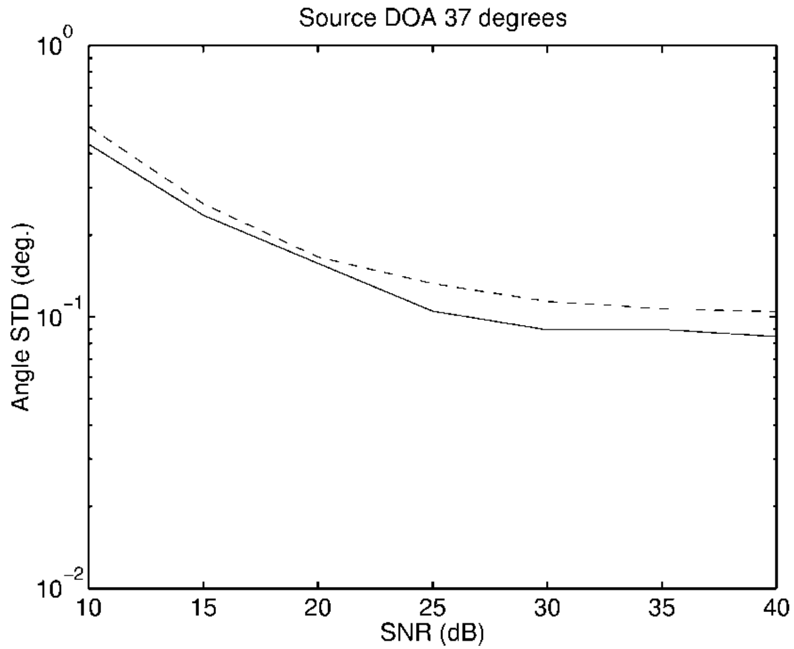
- The WAVES is assumed as generated by samples of an elliptical, zeo-mean, non-Gaussian multivariate distribution.
- The USCM signal subspace is estimated by a *robust pseudocovariance* trimmed on a *Least Informative Distribution* (Huber 1981).
- Fine for coherent sources and strong interferers.
- Slight performance loss in non-critical scenarios.

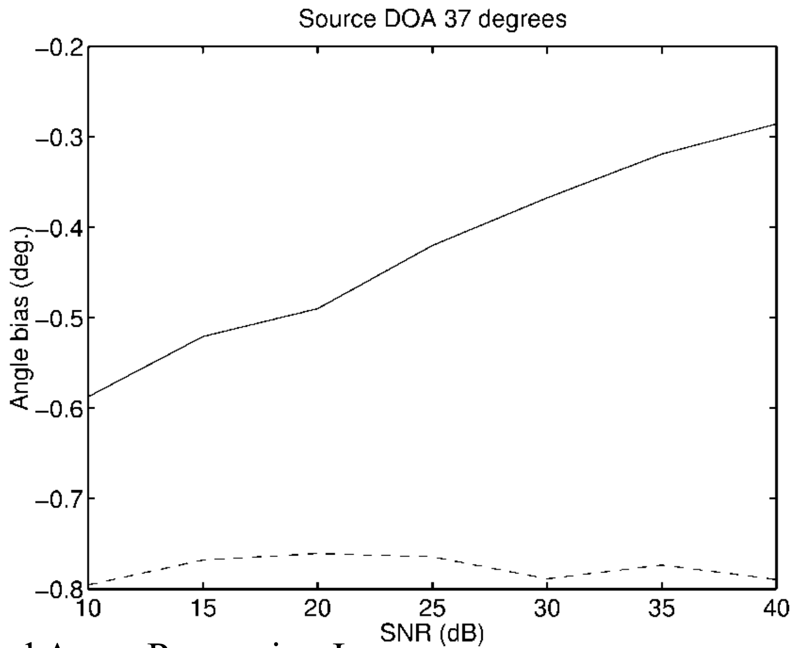
$$\tilde{\mathbf{y}}_k = \tilde{\mathbf{X}}^{-1}\mathbf{T}_k\tilde{\mathbf{e}}_k\rho_k ; s(x) = \left(x+a\right)^{-1} ; \sum_{k=1}^{\eta} s\left(\left|\tilde{\mathbf{y}}_k\right|\right)^2 \tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k^H = \sum_{k=1}^{\eta} s\left(\left|\tilde{\mathbf{y}}_k\right|\right)^2 \mathbf{I}$$

$$\tilde{\mathbf{X}} \overset{SVD}{=} \begin{bmatrix} \tilde{\mathbf{U}}_s & \tilde{\mathbf{U}}_n \end{bmatrix} \begin{bmatrix} \tilde{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \tilde{\Sigma}_n \end{bmatrix} \tilde{\mathbf{V}}^H$$

CSSM vs. WAVES

- ULA focused in beamspace (N=8, M=100, J=33, D=4), two white sources and two coloured AR interference sources.
- Continuous line: WAVES; Dashed line: CSSM: WAVES has smaller bias.





CSSM vs. WAVES comparison

- CSSM may have a slightly higher performance than WAVES at low SNR and with nearly white sources (it is a conditional estimator).
- However WAVES can better afford non-white spectra, strong and weak sources, discard empty bins and asymptotically converges to the correct subspace even if noise SCM is mis-specified.
- Under certain conditions (unitary focusing, no focusing errors), WAVES is asymptotically equivalent to the wide-band WSF.
- Both CSSM and WAVES can employ the same focusing techniques.

SAPIENZA
UNIVERSITÀ DI ROMA

Notes on coherent focusing

- Focusing of any kind can decorrelate coherent sources with a delay larger than one sampling period.
- USCM signal subspace rank shrinks for really coherent wave-fronts and requires the use of MODE or WSF for final location estimator.
- Focusing is effective for fractional bandwidths up to 50% and reasonably wide sectors.
- Unitary coherent focusing requires preliminary beamforming to find source clusters.
- Unitary focusing can be performed only within the *union of source clusters, one-two beamwidths wide each* , to avoid excessive errors.

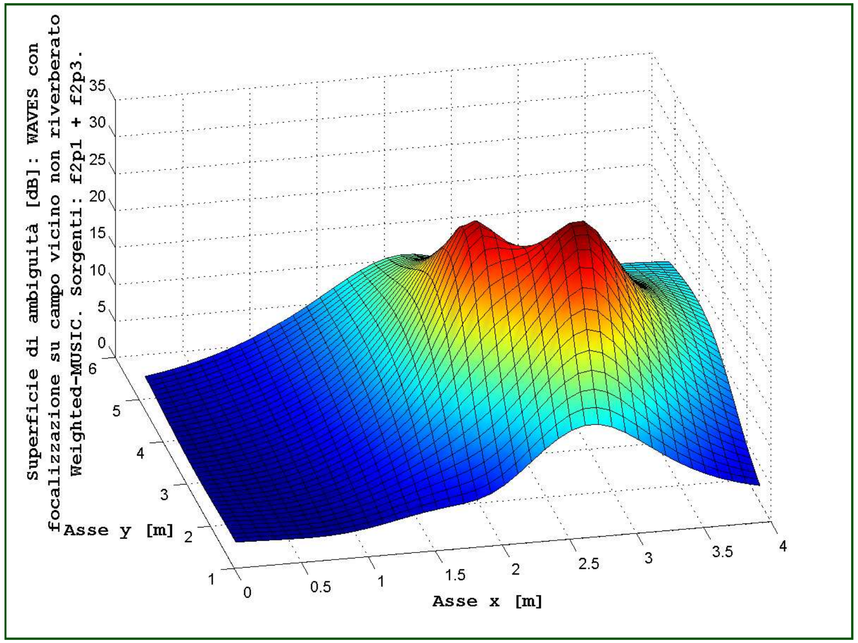
SAPIENZA
UNIVERSITÀ DI ROMA

Matched field WAVES

- WAVES puts in evidence and smooths out random array perturbations, seen as *excess noise in the USCM*.
- If steering vectors are derived by a matched-field model, robust localization exploiting even 3-D reflections is possible.
- WAVES noise subspace weighted with the inverse sample noise eigenvalues (EW-MUSIC) better compensates for reverberation effects even in case of wrong estimation of the source number.

3-D localization experiment in reverberant field by MF-WAVES

- INFOCOM Dpt. microphone array, 600-1200 Hz
- Two acoustic sources in reverberant room at (2.10, 3.11, 0.82) and (2.70, 3.11, 0.82) m
- WAVES + EW-MUSIC
- Matched-field focusing
- 15 calibration points
- 6 dB SRR
- Sep. 1/5 beamwidth
- Low bias



Beamforming invariance focusing

- It is possible to focus the wide-band array on a virtual array of smaller size, generally derived by linearly (beamspace) transforming an ULA manifold or an harmonic basis in azimuth.
- First the *minimum LS error subspace* is found for a given angular sector and defines the final virtual array manifold.
- The *steering vectors at each frequency of interest are fitted* to the virtual target.
- Weighted LS error functional with angular gradient error control (important for asymptotical performance) gives best results.

Sapienza Università di Roma

Weighted LS beamforming invariance

Tentative virtual array response

$$\mathbf{A}_k = \left[\mathbf{a}(\mathbf{p}_1, f_k) \quad \cdots \quad \mathbf{a}(\mathbf{p}_M, f_k)\right]; \quad \mathbf{A}_0 = \left[\mathbf{a}_0(\mathbf{p}_1) \quad \cdots \quad \mathbf{a}_0(\mathbf{p}_M)\right]$$

Diagonal weight matrix

$$\mathbf{W}\mathbf{A}_k^H \overset{QRD}{=} \mathbf{Q}_k\mathbf{R}_k; \quad \mathbf{W}\mathbf{A}_0^H \overset{QRD}{=} \mathbf{Q}_0\mathbf{R}_0; \quad \mathbf{C} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{Q}_0^H\mathbf{Q}_k\mathbf{Q}_k^H\mathbf{Q}_0$$

$$\mathbf{C} \overset{EVD}{=} \left[\underset{(N_0\times\eta)}{\mathbf{U}_1} \quad \underset{[N_0\times(N_0-\eta)]}{\mathbf{U}_2}\right]\begin{bmatrix}\boldsymbol{\Lambda}_1 \cong \mathbf{I}_\eta & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Lambda}_2 < \mathbf{I}_{N_0-\eta}\end{bmatrix}\left[\underset{(N_0\times\eta)}{\mathbf{U}_1} \quad \underset{[N_0\times(N_0-\eta)]}{\mathbf{U}_2}\right]^H$$

$$\mathbf{W}\mathbf{A}_k^H\mathbf{T}_k^H \overset{LS}{\cong} \mathbf{Q}_0\mathbf{U}_1\underset{(\eta\times\eta)}{\mathbf{C}_0}; \quad \mathbf{B}_0 = \mathbf{C}_0^H\mathbf{U}_1^H\mathbf{Q}_0^H\mathbf{W}^{-H} = \left[\mathbf{b}_0(\mathbf{p}_1) \quad \cdots \quad \mathbf{b}_0(\mathbf{p}_M)\right]$$

Low error subspace

Arbitrary matrix

Refined virtual array response

Focusing filters

- A bank of D&S wide-band beamformers is applied to the ST snapshot and *realigns all frequencies as in a single narrow-band snapshot*, albeit with correlated signals and noise.
- Any *set of focusing matrices for a finely spaced frequency set* can be rendered as a focusing filter bank by frequency interpolation.
- Sophisticated *numerical Inverse FT approximation* is required (wildly oscillating functions!).

$$\mathbf{x}_f(m) = \underset{(NP\times\eta)}{\mathbf{W}}^H \mathbf{x}_{ST}(m); \quad \mathbf{W}^H \mathbf{E}(\upsilon_k) \cong \mathbf{T}_k; \quad k = 1,2,\ldots,K$$

$$\mathbf{T}_k \mathbf{a}(\mathbf{p}, f_k) \cong \mathbf{a}_0(\mathbf{p}) \Rightarrow \mathbf{x}_f(m) \cong \begin{bmatrix} \mathbf{a}_0(\mathbf{p}_1) & \cdots & \mathbf{a}_0(\mathbf{p}_D) \end{bmatrix} \begin{bmatrix} s_1(m) \\ \vdots \\ s_D(m) \end{bmatrix} + \mathbf{v}_f(m)$$

STCM based MUSIC (BASS-ALE)

- Exploits the STCM noise subspace.
- Not capable of coping with coherent sources.
- Small source power impact on functional.

$$\mathbf{R}_{ST} = \underset{(NP\times\eta)}{\mathbf{E}_S} \boldsymbol{\Lambda}_S \mathbf{E}_S^H + \lambda_v \mathbf{E}_v \mathbf{E}_v^H$$

$$\underset{[NP\times(NP-\eta)]}{\mathbf{E}_v^H} \mathbf{E}(\upsilon_k) \overset{R-SVD}{=} \begin{bmatrix} \mathbf{U}_{1,k} & \mathbf{U}_{2,k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 \cong \mathbf{I}_{N-\eta_k} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \cong \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{1,k} & \mathbf{V}_{2,k} \end{bmatrix}^H$$

$$\hat{\mathbf{p}}_{BASS-ALE} = \arg\min_{\mathbf{p}} \left\{ \frac{1}{K} \sum_k \left| \mathbf{a}^H(\mathbf{p}, f_k) \mathbf{V}_{1,k} \right|^2 \right\}$$

Work in progress

- The rich structure of the STCM can be exploited for high resolution MUSIC type estimators for high resolution narrow-band signal subspace *circumventing SCM estimation* (STCM-MUSIC).
- STCM-MUSIC is very robust to the source covariance structure and approaches the wide-band CRB even with coherent and strongly coloured signals.
- Exploits source power information.
- For any wide-band estimator, pure sinusoidal or cyclostationary signal estimates remain the main challenge.