

# Source Separation and Beamforming Background

Ian K. Proudler

Dept. of Electronic & Electrical Engineering, University of Strathclyde

UDRC Summer School, 24 - 27 May, 2019, Edinburgh, UK.



University Defence Research Collaboration (UDRC)  
Signal Processing in the Information Age



[dstl]



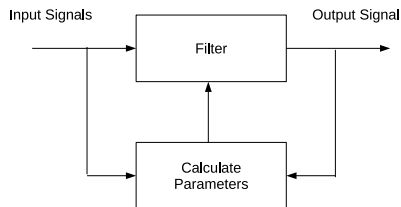
Engineering and Physical Sciences  
Research Council

# Source Separation and Beamforming Background: Overview

1. Overview
2. Signal Separation
3. Non-adaptive beamforming
4. Adaptive signal processing for beamforming
5. Application of linear algebra to array problems
6. More adaptive signal processing for beamforming
7. Blind source separation
8. Summary

# Signal Separation

- ▶ Signal separation requires two components:



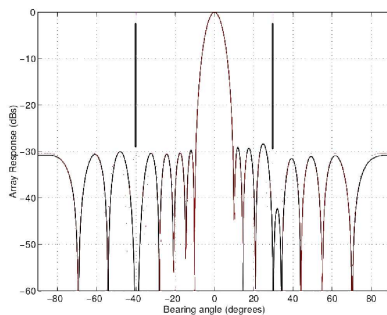
- ▶ A parametrised mechanism to separate the signals (a “filter”)
  - ▶ A means to select the parameters
  - ▶ Performance limited by ‘optimum’ filter
- 
- ▶ Conventionally we have two “filter” mechanisms:
    - ▶ Temporal filter – separate by frequency
    - ▶ Spatial filter (aka beamformer) – separate by AOA
  - ▶ We will focus on narrowband beamforming in this talk
  - ▶ Broadband beamforming requires a space-time filter

# Signal Separation

- ▶ Parameter selection – the interesting part
- ▶ Three cases:
  - ▶ Non-adaptive – we know everything about the scenario
  - ▶ “Adaptive” – we don’t know everything
  - ▶ “Blind” – we don’t know anything (sort of)
- ▶ Important parameters:
  - ▶ AOA of signals
  - ▶ Array calibration
  - ▶ Noise statistics

# Non-Adaptive Source Separation

- ▶ Covered in first talk



- ▶ Lots of good optimisation algorithms (DSP text books e.g. Rabiner & Gold - Temporal filters but basically the same for beamforming)
- ▶ Only  $(N - 1)$  nulls
- ▶ Spatially distributed noise can't be removed

- ▶ Beamformer weights via constrained optimisation (offline)
- ▶ Gain towards wanted signal = 1
- ▶ Gain towards other signals = 0
- ▶ Noise gain as small as possible

# Adaptive Source Separation

- ▶ Aka adaptive beamforming
- ▶ Assume the known parameters are:
  - ▶ AOA of the wanted signal(s)
  - ▶ Array calibration
- ▶ Beamformer weights via constrained optimisation but online this time
- ▶ Gain towards wanted signal = 1
- ▶ Minimise energy of output
- ▶ NB. Could use an AOA algorithm here and fixed beamforming but computationally costly

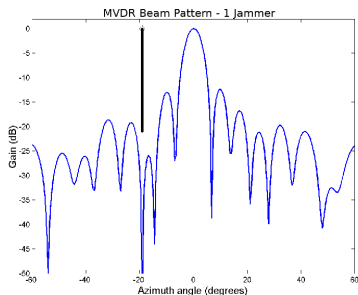
# Adaptive Source Separation

- ▶ Beamformer weights:  $\mathbf{w}$
- ▶ Sensor data at time  $n$ :  $\mathbf{x}(n)$
- ▶ Output at time  $n$ :  $y(n) = \mathbf{w}^H \mathbf{x}(n)$
- ▶ Energy in output:  $J = \sum_{n=0}^{N-1} |y(n)|^2 = \|\mathbf{w}^H \mathbf{X} \mathbf{X}^H \mathbf{w}\|_2^2$
- ▶ Data matrix:  $\mathbf{X} = [\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(N-1)]$
- ▶ Constraint:  $\mathbf{w}^H \mathbf{a}(\theta) = 1$
- ▶ Sample covariance matrix:  $\mathbf{R} = \mathbf{X} \mathbf{X}^H$

# Minimum Variance Distortionless Response (MVDR)

- ▶ Minimum Variance := Minimise energy of output
- ▶ Distortionless Response := Gain towards wanted signal = 1
- ▶

$$\mathbf{w} = \frac{\mathbf{R}^{-1}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{R}^{-1}\mathbf{a}(\theta)}$$



- ▶ Gain towards wanted signal = 1
- ▶ Small gain (null) towards other signal
- ▶ Noise gain not controlled  
In fact adapted to that particular noise realization



# Minimum Variance Distortionless Response (MVDR)

- ▶ Multiple noise realizations (blocks of data)

# Minimum Variance Distortionless Response (MVDR)

- ▶ Stabilisation procedures: there are many different ways of reducing the effects of adapting to the noise realizations.
- ▶ All effectively try to 'remove' influence of noise
- ▶ Diagonal loading

$$\mathbf{w} = \text{Arg Min} \left( \|\mathbf{w}^H (\mathbf{R} + \mu I) \mathbf{w}\|_2^2 \right) \text{ st. } \mathbf{w}^H \mathbf{a}(\theta) = 1$$

- ▶ Noise subspace manipulation  
Average noise subspace eigenvalues
- ▶ Penalty Function Method

$$\mathbf{w} = \text{Arg Min} \left( \|\mathbf{w}^H \mathbf{R} \mathbf{w}\|_2^2 + \kappa \|\mathbf{w} - \mathbf{w}_0\|_2^2 \right)$$

“Soft” constraint make adapted beam pattern lie close to the desired pattern.

# Linear Algebra

- ▶ MVDR weight vector depends on covariance matrix  $\mathbf{R}$
- ▶ This matrix has structure
- ▶ Hermitian (symmetric)

$$\mathbf{R}^H = (\mathbf{X}\mathbf{X}^H)^H = \mathbf{X}\mathbf{X}^H = \mathbf{R}$$

- ▶ We can use linear algebra to study / manipulate the covariance matrix
- ▶ Eigenvalue decomposition of Hermitian matrix

$$\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$$

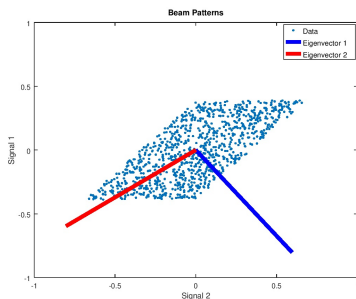
- ▶ Eigenvectors:  $\mathbf{U}$  is a unitary matrix

$$\mathbf{U}^H\mathbf{U} = \mathbf{I}$$

- ▶ Eigenvalues:  $\mathbf{\Lambda}$  is diagonal, all elements are  $\geq 0$
- ▶ Rank of  $M$  is number of non-zero eigenvalues

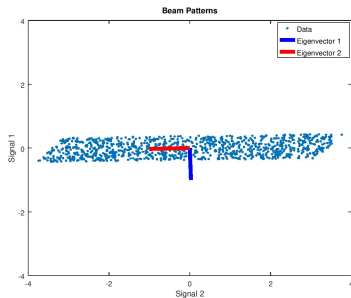
# Eigenvalue Decomposition

- ▶ Eigenvectors are not steering vectors  $\mathbf{X} = \mathbf{A}\mathbf{S}$



2 equal power signals

- ▶ Scatter plot
- ▶ Covariance matrix EVD
- ▶ Eigenvectors approximately steering vectors when powers are dissimilar



2 signals with power ratio  
10:1

# Eigenvalue Decomposition

- ▶ Consider two signals

$$\mathbf{X} = \mathbf{a}(\theta_1)\mathbf{s}_1^T + \mathbf{a}(\theta_2)\mathbf{s}_2^T + \mathcal{N}$$

- ▶ Covariance matrix

$$\mathbf{R} = \mathbf{X}\mathbf{X}^H = \mathbf{A}\mathbf{D}\mathbf{A}^H + \sigma^2\mathbf{I}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}(\theta_1) & \mathbf{a}(\theta_2) \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}$$

- ▶  $\mathbf{A}\mathbf{D}\mathbf{A}^H$  is rank two. EVD:

$$\mathbf{A}\mathbf{D}\mathbf{A}^H = \mathbf{U} \begin{bmatrix} \Lambda_{\mathbf{A}} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{U}^H$$

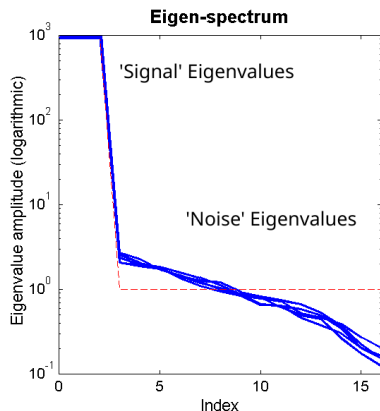
- ▶ Covariance matrix EVD

$$\mathbf{R} = \mathbf{U} \begin{bmatrix} \Lambda_{\mathbf{A}} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{U}^H + \sigma^2\mathbf{I} = \mathbf{U} \begin{bmatrix} \Lambda_{\mathbf{A}} + \sigma^2\mathbf{I} & 0 \\ 0 & \sigma^2\mathbf{I} \end{bmatrix} \mathbf{U}^H$$

# Eigenvalue Spectrum

## ► Eigenvalue spectrum

$$\begin{bmatrix} \Lambda_{\mathbf{A}} + \sigma^2 \\ \sigma^2 I \end{bmatrix}$$



- Two large eigenvalues
- five noise realizations
- Noise eigenvalues not equal – finite data

# Signal and Noise Subspaces

- ▶ Covariance matrix EVD

$$\mathbf{R} = \mathbf{U} \begin{bmatrix} \mathbf{\Lambda}_A + \sigma^2 I & 0 \\ 0 & \sigma^2 I \end{bmatrix} \mathbf{U}^H$$

- ▶ Partition eigenvectors

$$\mathbf{U} = [ \mathbf{U}_1 \quad \mathbf{U}_2 ]$$

- ▶ Orthogonal subspaces

$$\mathbf{U}_1^H \mathbf{U}_1 = I \quad \mathbf{U}_1^H \mathbf{U}_2 = 0$$

- ▶ Covariance matrix EVD

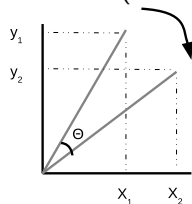
$$\mathbf{R} = \mathbf{U}_1 (\mathbf{\Lambda}_A + \sigma^2) \mathbf{U}_1^H + \mathbf{U}_2 (\sigma^2 I) \mathbf{U}_2^H$$

## Rotation Matrices

- ▶ Eigenvectors:  $\mathbf{U}$  is a unitary matrix

$$\mathbf{U}^H \mathbf{U} = \mathbf{I}$$

- ▶ Can be considered as a rotation in N-dimensional space
- ▶ 2-D case (Givens rotations)



$$\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta)^* & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

- ▶ Can build N-D rotation from 2-D ones

$$\mathbf{U} = [\bullet] \dots \begin{bmatrix} I & 0 & 0 & 0 & 0 \\ 0 & \cos(\theta) & 0 & \sin(\theta) & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & -\sin(\theta)^* & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & I & 0 \end{bmatrix} \dots [\bullet]$$



# Singular Value Decomposition

- ▶ Not all matrices of interest are Hermitian
- ▶ Singular value decomposition of a matrix  $\mathbf{M}$ :  $N$  rows &  $M$  columns

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$$

$\mathbf{U}$  is  $N \times N$ ,  $\mathbf{\Sigma}$  is  $N \times M$ , and  $\mathbf{V}$  is  $M \times M$

- ▶ Singular vectors:  $\mathbf{U}$  and  $\mathbf{V}$  are unitary matrices
- ▶ Singular values:  $\mathbf{\Sigma}$  is diagonal, all elements are  $\geq 0$
- ▶ Rank of  $M$  is number of non-zero singular values
- ▶ Relation to EVD

$$\mathbf{R} = \mathbf{M}\mathbf{M}^H = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H\mathbf{V}\mathbf{\Sigma}\mathbf{U}^H = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^H$$

Eigenvalues are the square of the singular values

## Stabilized MVDR Beamformer

- ▶ Recall basic MVDR beamformer suffers from weight jitter
- ▶ Average noise eigenvalues

# Array Calibration Errors

- ▶ MVDR minimises power in output signal.
- ▶  $\mathbf{w} = 0$  would do this
- ▶ 'Look direction' constraint protect wanted signal

$$\mathbf{w}^H \mathbf{a}(\theta) = 1$$

- ▶ What if  $\mathbf{a}(\theta)$  is incorrect?
- ▶ Wanted signal looks like an unwanted one!
- ▶ Add extra constraints
  - ▶ More than one 'Look direction' constraint
  - ▶ Flatten main lobe – gradient constraint

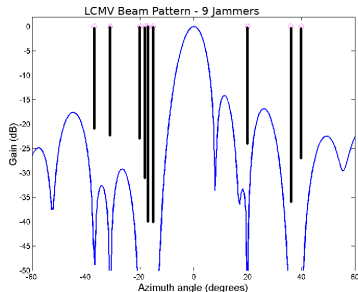
# Linearly Constrained Minimum Variance (LCMV)

- ▶ Minimum Variance = Minimise energy of output
- ▶ Linearly Constrained = More than one constraint

$$\mathbf{w}^H \mathbf{C} = \mathbf{g}^T$$

- ▶ Solution

$$\mathbf{w} = \mathbf{R}^{-1} \mathbf{C} (\mathbf{C}^H \mathbf{R}^{-1} \mathbf{C})^{-1} \mathbf{g}$$



- ▶ Gain in wanted direction = 1
- ▶ Gain towards other directions = 0

# Linearly Constrained Minimum Variance (LCMV)

- ▶ LCMV is a constrained minimisation problem

$$\mathbf{w} = \text{Arg Min} \left( \|\mathbf{w}^H \mathbf{R} \mathbf{w}\|_2^2 \right) \text{ st. } \mathbf{w}^H \mathbf{C} = \mathbf{g}^T$$

- ▶ If there are  $M$  constraints,  $M$  components of  $\mathbf{w}$  are effectively fixed
- ▶ Thus only  $N - M$  'degrees of freedom' in the choice of  $\mathbf{w}$  i.e. can only null out  $N - M$  signals
- ▶ Thus have to have  $N - M > 0$
- ▶ Sometimes the constraints can be linearly dependent or nearly so

# Linearly Constrained Minimum Variance (LCMV)

Consider

$$\mathbf{w}^H \mathbf{C} = \mathbf{g}^T$$

or

$$[\mathbf{w}^H \mathbf{C} - \mathbf{g}^T] = [\mathbf{w}^H, -1] \begin{bmatrix} \mathbf{C} \\ \mathbf{g}^T \end{bmatrix} = 0$$

Take SVD

$$[\mathbf{w}^H, -1] U \Sigma V^H = 0$$

$V$  is full rank so

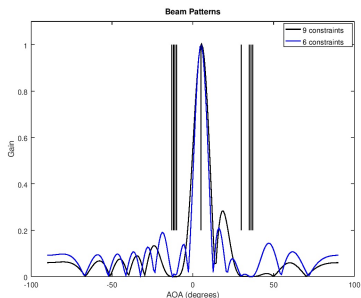
$$[\mathbf{w}^H, -1] U \Sigma = 0$$

If  $N - R$  singular values are small

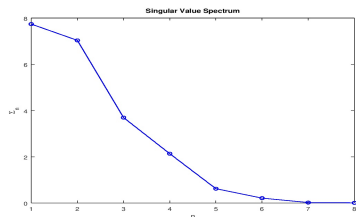
$$[\mathbf{w}^H, -1] U_1 \Sigma_1 = 0$$

Let  $U_1 \Sigma_1 = \begin{bmatrix} \tilde{\mathbf{C}} \\ \tilde{\mathbf{g}}^T \end{bmatrix}$  then  $\mathbf{w}^H \tilde{\mathbf{C}} = \tilde{\mathbf{g}}^T$  and  $\tilde{\mathbf{C}}$  only has  $R$  columns

# Linearly Constrained Minimum Variance (LCMV)

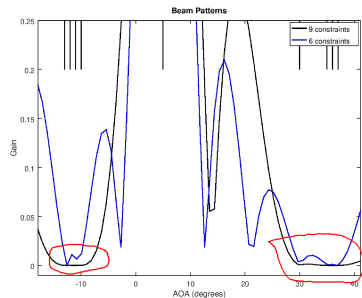


- ▶ Beam patterns
- ▶ 9 Constraints
- ▶ 6 Constraints



- ▶ Singular value spectrum
- ▶ 3 small singular values

# Linearly Constrained Minimum Variance (LCMV)



- ▶ Beam patterns
- ▶ 9 Constraints
- ▶ 6 Constraints

- ▶ Constraints not achieved due to non-zero singular values  
Mostly used for beam shape



# Blind Source Separation

- ▶ What if we don't know: AOA of one signal and array calibration
- ▶ Recall that

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathcal{N}$$

- ▶ Covariance matrix

$$\mathbf{R} = \mathbf{X}\mathbf{X}^H = \mathbf{A}\mathbf{D}\mathbf{A}^H + \sigma^2\mathbf{I}$$

Assume that  $\mathbf{D} = \mathbf{I}$ . If not redefine array manifold  $\mathbf{A}$  so that

$$\mathbf{A} \leftarrow \mathbf{A}\mathbf{D}^{\frac{1}{2}}$$

- ▶ SVD of  $\mathbf{A}$

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$$

- ▶ Covariance matrix EVD

$$\mathbf{X}\mathbf{X}^H = \mathbf{U} [\mathbf{\Sigma}\mathbf{V}^H\mathbf{V}\mathbf{\Sigma} + \sigma^2\mathbf{I}] \mathbf{U}^H = \mathbf{U} [\mathbf{\Sigma}^2 + \sigma^2\mathbf{I}] \mathbf{U}^H$$

# Blind Source Separation

- ▶ Let  $\tilde{\Sigma} = \Sigma + \sigma I$

$$\mathbf{X}\mathbf{X}^H = \mathbf{U}\tilde{\Sigma}^2\mathbf{U}^H$$

- ▶ But

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathcal{N} = \mathbf{U}\Sigma\mathbf{V}^H\mathbf{S} + \mathcal{N}$$

- ▶ Thus

$$\mathbf{U}^H\mathbf{X} = \Sigma\mathbf{V}^H\mathbf{S} + \mathbf{U}^H\mathcal{N}$$

so

$$\Sigma^{-1}\mathbf{U}^H\mathbf{X} = \mathbf{V}^H\mathbf{S} + \Sigma^{-1}\mathbf{U}^H\mathcal{N}$$

assuming  $\Sigma^{-1}$  exists!

i.e.

$$\mathbf{Y} \equiv \Sigma^{-1}\mathbf{U}^H\mathbf{X} = \mathbf{V}^H\mathbf{S} + \tilde{\mathcal{N}}$$

where  $\tilde{\mathcal{N}} = \Sigma^{-1}\mathbf{U}^H\mathcal{N}$  is a noise term

# Blind Source Separation

- ▶ We have

$$\mathbf{Y} = \mathbf{V}^H \mathbf{S} + \tilde{\mathcal{N}}$$

- ▶ so  $\mathbf{S}$  could be extracted from  $\mathbf{Y}$  if we knew  $\mathbf{V}^H$
- ▶ Then

$$\hat{\mathbf{S}} = (\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^H) \mathbf{X}$$

- ▶ cf. bank of beamformers

$$\hat{\mathbf{S}} = \begin{bmatrix} \mathbf{w}_1^H \\ \vdots \\ \mathbf{w}_N^H \end{bmatrix} \mathbf{X}$$

- ▶ Blind signal separation is limited by what a bank of beamformers can do e.g.  $N$  sensors  $\rightarrow N - 1$  nulls

# Blind Source Separation

- ▶ How to estimate  $\mathbf{V}^H$ ?  
NB

$$\mathbf{Y}\mathbf{Y}^H == \mathbf{V}^H\mathbf{S}\mathbf{S}^H\mathbf{V} + \sigma^2\mathbf{\Sigma}^{-2}$$

but  $\mathbf{S}\mathbf{S}^H = \mathbf{I}$  so

$$\mathbf{Y}\mathbf{Y}^H == \mathbf{I} + \sigma^2\mathbf{\Sigma}^{-2}$$

i.e. the second order statistics of  $\mathbf{Y}$  will not help us estimate  $\mathbf{V}^H$

- ▶ Can however use higher order statistics
- ▶ Can also use nonlinear cost function

# Blind Source Separation

- ▶ E.g. 'FastICA' - iteration to minimise 'negentropy'

$$J(Y) = H(\mathbf{Y}_{\text{Gauss}}) - H(\mathbf{Y})$$

- ▶  $\mathbf{Y}_{\text{Gauss}}$  is Gaussian data with same covariance matrix as  $\mathbf{Y}$ ,  
 $H(\mathbf{Y})$  is the entropy of  $\mathbf{Y}$

$$H(Y) = - \int p_Y(y) \log(p_Y(y)) dy$$

- ▶ Iteration

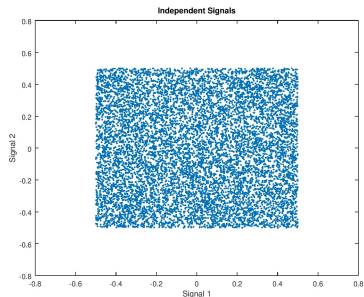
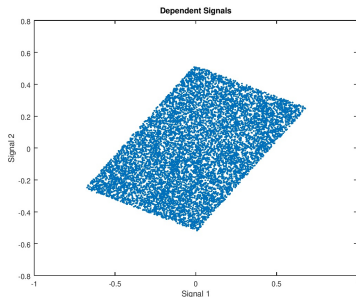
$$\mathbf{V}_i(k+1) = G(\mathbf{V}_i(k)^H \mathbf{X}) \mathbf{Y} - G'(\mathbf{V}_i(k)^H \mathbf{X}) \mathbf{V}_i(k)$$

$$G(\mathbf{v}) = \tanh(\alpha \mathbf{v}), \mathbf{v} \exp(-\mathbf{v}^2/2), \text{ or } \mathbf{v}^3$$

where  $1 \leq \alpha \leq 2$

# Blind Source Separation

- ▶ Higher order statistics
- ▶ Statistical independence  $P(x, y) = P(x)P(y)$
- ▶ Scatter diagram



- ▶ Calculate rotation to align scatter plot with axes

# Blind Source Separation

- ▶ Estimating the 'hidden' rotation matrix

$$\mathbf{Y} = \mathbf{V}^H \mathbf{S} + \tilde{\mathcal{N}}$$

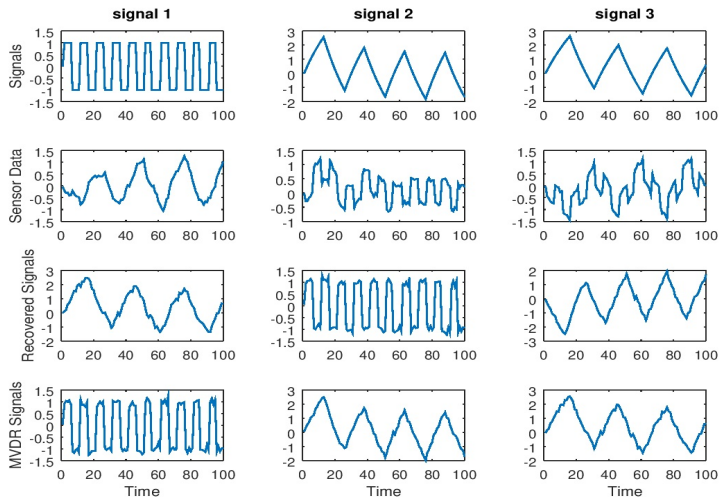
- ▶ Loop through all pairs of signals
- ▶ Rotate to align with axes
- ▶ Repeat until rotation angle is below a threshold

$$\mathbf{Q}_n \mathbf{Q}_{n-1} \dots \mathbf{Q}_1 \mathbf{Y} = \hat{\mathbf{S}}$$

i.e.  $J(\hat{\mathbf{S}}) < \epsilon$  but  $J(\mathbf{S}) = 0$  so  $\hat{\mathbf{S}} \approx \mathbf{S}$

- ▶ Can show that  $\hat{\mathbf{S}}$  is  $\mathbf{S}$  up to scaling and permutation of the signal order

# Blind Source Separation

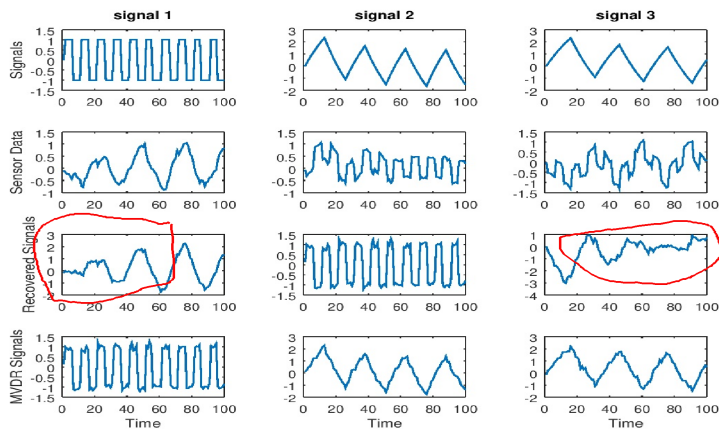


► 3 signals, 3 sensors, SNR = 20dB, MVDR as benchmark



# Blind Source Separation

- ▶ Need more data to calculate higher-order statistics



- ▶ Previous plot: 1000 data samples, This plot: 100 data samples

# Summary

- ▶ Signal Separation: filter and parameters  
Performance limited by 'optimum' filter
- ▶ Non-adaptive beamforming  
Good optimisation algorithms
- ▶ Adaptive signal processing for beamforming  
Constrain direction of main beam, reduce everything else  
Weight jitter, calibration errors  
Lots of linear algebra
- ▶ Blind source separation Higher-order statistics or nonlinear optimisation  
Lots of data needed
- ▶ Acknowledgment: John Mather (QinetiQ).